

Mémoire présenté
pour l'obtention du Diplôme Universitaire d'actuariat de l'ISFA
et l'admission à l'Institut des Actuaires

le 05/09/2022

Par : Jean-Paul JOHN MATHEWS

Titre : Modélisation des dépendances au sein d'un portefeuille non-vie et impacts sur le besoin en fonds propres

Confidentialité : Non Oui (Durée : 1 an 2 ans)

Les signataires s'engagent à respecter la confidentialité ci-dessus

*Membres présents du jury de l'Institut
des Actuaires :*

Entreprise :

Nom : Milliman France

Signature :



Membres présents du Jury de l'ISFA :

Directeur de Mémoire en entreprise :

Nom : Pierre MIEHE

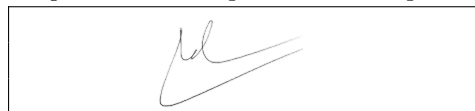
Signature :



Autorisation de publication et de mise en ligne sur un site de diffusion de documents actuariels (après expiration de l'éventuel délai de confidentialité)

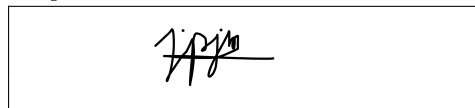
Secrétariat :

Signature du responsable entreprise



Bibliothèque :

Signature du candidat

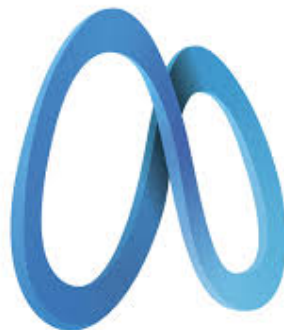


INSTITUT DE SCIENCE FINANCIÈRE ET D'ASSURANCES

Mémoire pour le titre d'Actuaire
Modélisation des dépendances au sein d'un
portefeuille non-vie et impacts sur le besoin en
fonds propres

Auteur :
Jean-Paul JOHN MATHEWS

Tuteur pédagogique :
Stéphane LOISEL
Tuteur d'entreprise :
Pierre MIEHE



7 septembre 2022

Mots-Clés

Dépendances, Dépendances de queue, Théorie des copules, Simulation de copules, Indépendance, Ajustement des lois marginales, VaR, TVaR, Besoin en fonds propres, Théorème de Sklar, Tau de Kendall, Rho de Spearman, Corrélations, Sinistres "as if", Copules archimédiennes, Copules elliptiques, Méthode IFM et CML, Copule empirique, Test d'ajustement du χ^2 , fonction $K(z)$.

Résumé

Pour satisfaire les demandes du régulateur exigeant une marge sur sa solvabilité, mais également les actionnaires requérant une minimisation du besoin en fonds propres pour une meilleure rentabilité, l'assureur se doit de garantir une gestion optimale de sa solvabilité et donc d'estimer un besoin en fonds propres le plus optimal possible.

Dans l'établissement du besoin en fonds propres d'une compagnie d'assurance non-vie, plusieurs hypothèses fortes sont généralement admises. Parmi ces hypothèses, l'hypothèse d'indépendance entre les risques au sein d'un portefeuille, ou encore, si dépendance des risques admise, la considération de risques suivant la même famille de lois, sont des hypothèses fortes pouvant conduire à une erreur dans l'estimation du besoin en fonds propres optimal. Dans ce cadre, une modélisation réaliste et juste des dépendances entre les risques à l'aide de copules peut avoir un impact sur le calcul du besoin en fonds propres. En particulier, le mémoire s'attachera à étudier et modéliser les dépendances aussi bien positives que négatives pour chaque couple de risques, en estimant les bons paramètres de copule et en travaillant, par l'intermédiaire de tests d'ajustement, sur un choix optimal de copule paramétrique pour chaque couple de risques. Une méthode de modélisation de l'ensemble des risques du portefeuille, par le biais de copules multivariées, sera également proposée afin de représenter fidèlement les dépendances du portefeuille.

Il sera également mesuré, à partir de simulations de copules paramétriques décrivant de façon réaliste la structure de dépendance de chaque couple de sinistres, l'erreur omise par l'hypothèse d'indépendance entre les risques sur l'estimation du besoin en fonds propres. Sur la base des conclusions de l'étude, il sera enfin discuté de la pertinence de l'hypothèse d'indépendance des risques établie généralement en assurance non-vie.

Keywords

Dependencies, Tail dependencies, Copula theory, Copula simulation, Independence, Marginal adjustment, VaR, TVaR, Capital requirement, Sklar's theorem, Kendall's Tau, Spearman's Rho, Correlations, "as if" claims, Archimedean copulas, Elliptical copulas, IFM and CML method, Empirical copula, χ^2 fit test, $K(z)$ function.

Abstract

In order to satisfy the demands from the regulator for a sufficient margin on its solvency, but also the shareholders who require a minimization of the need for equity capital for better return, the insurer must guarantee optimal management of its solvency and therefore estimate the most optimal need for equity capital.

In the process of capital requirements estimation of a non-life insurance company, several strong assumptions are generally accepted. Amongst these assumptions, the assumption of independence between risks within a portfolio, or even, if risk dependence is admitted, the consideration of risks following the same family of distributions, are strong assumptions that can lead to an error in the estimation of the optimal capital requirement. In this context, a realistic and accurate modeling of the dependencies between risks using copulas can have an impact on the calculation of the capital requirement. In particular, the dissertation will study and model both positive and negative dependencies for each pair of risks, by estimating the right copula parameters and by working, through fit tests, on an optimal choice of parametric copula for each pair of risks. A method of modeling all the risks of the portfolio, through multivariate copulas, will also be proposed in order to faithfully represent the dependencies of the portfolio.

It will also be measured, using parametric copula simulations which realistically describe the dependence structure of each pair of claims, the error omitted by the assumption of independence between risks on the estimation of the capital requirement. Based on the conclusions of the study, the relevance of the assumption of independence of risks generally established in non-life insurance will finally be discussed.

Note de Synthèse

Dans l'élaboration du besoin en fonds propres pour un assureur non-vie, il est généralement admis plusieurs hypothèses dans la modélisation des risques d'un portefeuille. Il est notamment établi une hypothèse d'indépendance entre les multiples risques d'un portefeuille, ou, si la dépendance entre les risques est prise en compte, il est convenu de faire suivre les lois marginales suivant une même famille de lois. Ces hypothèses ne permettent pas de représenter les risques de façon réaliste au sein d'un portefeuille non-vie et peuvent influencer le niveau de solvabilité d'un assureur dans ce cadre. Les enjeux d'une modélisation des dépendances entre les risques réalistes sont entre autres une modélisation plus fine du besoin en fonds propres pour l'assureur. La juste estimation du besoin en fonds propres est nécessaire pour un assureur afin de répondre aux intérêts, à la fois du régulateur, et aussi des actionnaires. En effet, le régulateur exige une vision prudentielle avec une marge de solvabilité à disposer dans ses fonds propres, tandis que les actionnaires désirent une minimisation du besoin en fonds propres afin d'optimiser leur rentabilité.

Les copules sont un outil permettant de distinguer le comportement de la fonction de répartition marginale dans la structure de dépendance. Elles permettent de déterminer les structures de dépendance de tout type, en particulier les dépendances non linéaires entre sinistres et les dépendances de queue de distribution. Tout au long de ce mémoire, il sera sujet de modéliser au mieux la structure de dépendance à partir des couples de risques possédant une corrélation avérée dans les données extraites.

Théorie des copules

Une copule C de dimension n est une fonction de répartition de dimension n définie de $[0, 1]^n$ vers $[0, 1]$. Elle peut s'écrire de la forme suivante, $\forall (u_1, \dots, u_n) \in [0, 1]^n$:

$$C(u_1, \dots, u_n) = P(U_1 \leq u_1, U_2 \leq u_2, \dots, U_n \leq u_n), U_1, U_2, \dots, U_n \sim U([0; 1]) \quad (1)$$

Le théorème fondamental de la théorie des copules est le théorème de Sklar. Ce théorème permet en effet d'établir le lien entre une fonction multivariée et ses lois marginales. Il permet également de constituer une condition suffisante à l'existence des copules.

Soit F une fonction de répartition d'un vecteur $(X_1, X_2, X_3, \dots, X_n)$ dont les fonctions de répartition marginales sont de la forme $F_i, i = 1, \dots, n$. Il existe alors une unique copule C telle que

$$F(x_1, x_2, \dots, x_n) = C(F_1(x_1), F_2(x_2), \dots, F_n(x_n)) \quad (2)$$

Inversement, étant donnée la copule C et ses marginales $F_i, i=1, 2, \dots, n$, $C(F_1(x_1), F_2(x_2), \dots, F_n(x_n))$ définit une fonction de répartition multivariée de dimension n .

Par ailleurs, ce théorème est également utile pour construire des fonctions copules à partir de distributions multivariées. Si les lois marginales sont continues, alors :

$$C(u_1, \dots, u_n) = F(F_{X_1}^{-1}(u_1), \dots, F_{X_n}^{-1}(u_n))$$

Tout au long de notre étude, plusieurs copules usuelles seront utilisées. Certaines disposent de propriétés particulières que l'on exposera ci-dessous.

- La copule d'indépendance : $\prod(u) = u_1 \times u_2 \times \dots \times u_d$.
- Les copules elliptiques telles que la copule gaussienne ou la copule de Student.

La copule gaussienne est définie à partir de familles de lois gaussiennes. La copule gaussienne multivariée est définie telle que :

$$C_{\Sigma}^{Ga} = \phi_{\Sigma}(\phi^{-1}(u_1), \phi^{-1}(u_2), \dots, \phi^{-1}(u_n)) \quad (3)$$

où ϕ_{Σ} est la fonction de répartition de la loi multivariée $\mathcal{N}(0_{\mathbb{R}^n}, \Sigma)$ et ϕ la fonction de répartition des lois normales suivant $\mathcal{N}(0, 1)$.

La copule de Student se définit par des lois marginales de loi de Student liées à une copule de Student. Soient $T_{\Sigma, \nu}$ et T_{ν} respectivement les fonctions de répartition des lois $T(\Sigma, \nu)$ et $T(\nu)$. La copule de Student se définit alors par :

$$C(u_1, \dots, u_n) = T_{\Sigma, \nu}(T^{-1}(u_1), \dots, T^{-1}(u_n))$$

- Les copules archimédiennes ont la particularité d'obtenir des dépendances asymétriques avec des dépendances de queue différentes. Cette notion de copule archimédienne a été définie par GENEST et MACKAY [1986]. Dans le cas bivarié, on peut la définir ainsi :

$$C(u, v) = \phi^{-1}(\phi(u) + \phi(v)), u, v \in [0, 1]$$

Avec ϕ un générateur strict, une fonction décroissante convexe sur $(0, 1] \rightarrow [0, \infty]$ telle que $\phi(1) = 0$ et $\phi(0) = \infty$. Le générateur ϕ est dit non-strict si $\phi(0) < \infty$.

Les copules de Frank, Clayton, Gumbel et AMH sont définies ci-dessous.

Soit le générateur ϕ défini tel que $\phi(x) = -\log\left(\frac{e^{-ax}-1}{e^{-a}-1}\right)$, avec $a \neq 0$. Alors, la copule de Frank peut s'écrire ainsi :

$$C_a(u, v) = -\frac{1}{a} \log\left(1 + \frac{(e^{-au} - 1)(e^{-av} - 1)}{e^{-a} - 1}\right)$$

Le générateur de la copule de Clayton s'écrit quant à lui :

$$\phi(t) = \frac{t^{-a} - 1}{a}, a \in]0, +\infty[$$

La forme multivariée de la copule de Clayton s'écrit de la façon suivante :

$$C_a(u_1, \dots, u_n) = (u_1^{-a} + \dots + u_n^{-a} - n + 1)^{-\frac{1}{a}}$$

La copule de Gumbel admet comme générateur $\phi(x) = (-\log(x))^a$ avec $a \geq 1$.

La copule correspondante s'écrit ainsi :

$$C_a(u, v) = \exp(-(-\log(u))^a + (-\log(v))^a)^{\frac{1}{a}}$$

La copule de Gumbel est symétrique, ce qui n'est pas le cas en général pour une copule archimédienne.

Ali, Mikhail et Haq (1978) ont créé une copule archimédienne dont la particularité est son paramètre couvrant un intervalle réduit $[-1, 1]$. Le générateur de la copule s'écrit $\phi(t) = \ln\left(\frac{1-\theta(1-t)}{t}\right)$, donc la copule AMH s'écrit de la façon suivante :

$$C_\theta(u, v) = \frac{uv}{1 - \theta(1-u)(1-v)}$$

où le paramètre de copule est $\theta \in [0, 1]$.

Il est également possible de construire une copule empirique à partir d'observations données. Pour cela, il nous faut calculer notamment la statistique de rang de la loi jointe. Les copules empiriques peuvent constituer un outil de comparaison afin de calibrer une copule paramétrique adéquate à la base de données que l'on possède.

Deuhevles[1979] est à l'origine de la notion de copule empirique. La fonction s'appuie sur le rang des observations afin d'extraire la structure de dépendance.

Soit (X_1, \dots, X_d) une suite de variables aléatoires et $(x_1^n, \dots, x_d^n)_{1 \leq n \leq N}$ un échantillon d'observations de taille N . Fixons également $(r_1^n, \dots, r_d^n)_{1 \leq n \leq N}$ la statistique de rang (ou rang des observations) associée à cet échantillon multivarié. Donc pour tout $1 \leq i \leq d$, r_i^n est le rang de x_i^n dans $(x_i^n)_{1 \leq n \leq N}$, c'est à dire $\sum_{n=1}^N \mathbb{1}_{X_i^n \leq x_i^n}$ pour tout $1 \leq i \leq d$.

On peut donc définir la copule empirique sur l'ensemble $\mathbf{E} = \left\{ \left(\frac{n_1}{N}, \dots, \frac{n_d}{N} \right), 1 \leq i \leq d, n_i = 0, \dots, N \right\}$ de la façon suivante :

$$\hat{C}_N \left(\frac{n_1}{N}, \dots, \frac{n_d}{N} \right) = \frac{1}{N} \sum_{n=1}^N \prod_{i=1}^d \mathbb{1}_{\{r_i^n \leq n_i\}}$$

Présentation du jeu de données et recherche de dépendances sur notre jeu de données

Le jeu de données sur lequel se base l'étude provient de l'agence TSA (Transportation Security Administration), agence américaine gérant la sûreté aéroportuaire dans la plupart des aéroports des Etats-Unis. Cette agence a été créée à la suite des attentats du 11 Septembre 2001 afin de mieux protéger la population lors des trajets aériens. Une des fonctions principales de la TSA consiste à organiser des contrôles de passagers et de leurs objets personnels avant des embarquements dans les différents aéroports du pays. Lors de ces différents checkpoints, il peut arriver qu'un passager subisse un sinistre de type responsabilité civile. Le passager a la possibilité de déclarer tout type de sinistre à l'encontre de la TSA. Ainsi, des sinistres tels que des blessures, des pertes ou détérioration d'objets, vols peuvent être déclarés à la suite des contrôles.

Le jeu de données comporte l'ensemble des sinistres déclarés par des passagers pendant la période 2002-2015. 145012 sinistres ont été répertoriés tout au long de cette période. Le but de l'étude sera de déterminer en premier lieu des dépendances entre chaque type de sinistre, puis de modéliser les dépendances de types de sinistre déclarés à l'aide de copules. Enfin, il s'agira d'estimer l'impact de notre modélisation de dépendance par les copules sur le besoin en fonds propres de la TSA pour couvrir les sinistres.

Le traitement de la base consiste à calculer en premier lieu les montants de sinistres clos sur la base 2014. On estimera ainsi chacun des montants de sinistre selon l'indice de consommation 2014, afin d'obtenir des montants de sinistres comparables sur une même année.

Ensuite, il s'agira d'agrèger les montants de sinistres clos non-nuls selon le type de sinistres et la date d'incidence du sinistre. Le choix d'agrégation selon la date d'incidence résulte du fait que les sinistres sont tous de courte portée. En effet, les sinistres sont tous des événements soudains

sans répercussion sur le jour d'après.

L'agrégation des sinistres permet de distinguer les corrélations existantes entre chaque type de sinistre. Afin de justifier l'étude de corrélations à partir de cette base agrégée, l'hypothèse est ici posée d'estimer que les sinistres d'un même type de sinistres soient indépendants entre eux d'un jour à l'autre. Cette hypothèse semble réaliste car il s'agit d'un environnement où les individus sont de passage. Aussi, les sinistres concernent avant tout les passagers, l'évènement est ainsi centré sur le passager.

L'étude de corrélations des types de sinistre de la base de données se base essentiellement sur les mesures de dépendance que sont le coefficient de corrélation de Pearson, le tau de Kendall et le rho de Spearman. Le coefficient de corrélation de Pearson détermine avant tout les corrélations de type linéaire tandis que le tau de Kendall et le rho de Spearman se basent sur les statistiques de rang des observations. Ci-dessous se trouve les résultats de corrélation pour chacun des couples de sinistre de la base de données :

Tau Kendall	Perte biens	Dommages biens	Dom. corporels	Vehicule	Employé
Perte biens	1	0,307	-0,0787	-0,012	0,020
Dommages biens	0,307	1	-0,019	-0,057	0,041
Dom. corporels	-0,0787	-0,019	1	No data	No data
Véhicule	-0,012	-0,057	No data	1	No data
Employé	0,020	0,041	No data	No data	1

TABLE 1 – Tau de Kendall

Nous remarquons une corrélation positive nette du couple "Perte de biens/Dommages biens". Le couple "Perte de biens/Dommages corporels" présente quant à lui une corrélation faiblement négative (-0,08 environ) tout comme le couple "Dommages aux biens/Véhicule motorisé". Le couple "Dommages aux biens/Perte employé" présente une corrélation faiblement positive. Ces tau de Kendall faibles s'expliquent en partie par le nombre d'observations faible pour certains couples de sinistre.

A travers ces mesures, on peut conclure que ces quatre couples présentent une dépendance faible. La modélisation de ces dépendances par les copules est possible pour ces couples car l'établissement de la dépendance au sein de ces couples est vérifié. Nous pouvons à présent modéliser les dépendances de ces couples de sinistre par les copules.

Estimation du paramètre de copule paramétrique et choix de la bonne copule

Les copules empiriques modélisent les dépendances de nos couples de sinistre de façon la plus fidèle possible. Le but de cette section est de déterminer les paramètres de copules paramétriques afin d'ajuster correctement les copules paramétriques à la copule empirique de chaque couple de risques.

Tout d'abord, un travail sur l'ajustement des lois marginales de chaque couple de sinistres a été réalisé. L'ajustement consiste à estimer des lois de la famille exponentielle se rapprochant des lois marginales. Pour ce faire, une méthode de maximum de vraisemblance a été réalisée. Le test d'ajustement de nos lois marginales est un test de Kolmogorov Smirnov. Ci-dessous se trouvent les résultats du test, avec notamment des p-value bien inférieures à 1%.

Ajustement	D	P-Value
Perte de biens - Log-normale	1 571 323	$< 2.2 \times 10^{-16}$
Dommages aux biens - Gamma	13 379	$< 2.2 \times 10^{-16}$
Dommages corporels - Log-normale	226 171	$< 2.2 \times 10^{-16}$
Véhicules motorisées - Gamma	6197,8	$< 2.2 \times 10^{-16}$
Perte employé - Gamma	2330,6	$< 2.2 \times 10^{-16}$

TABLE 2 – Résultats du test de Kolmogorov Smirnov

Au vu des p-values, l'hypothèse H_0 selon laquelle les lois marginales suivent bien leurs lois théoriques respectives est rejetée pour chaque loi (p-value bien inférieur au seuil de 1% fixé). L'approche semi-paramétrique s'impose donc ici. Dans l'approche semi-paramétrique, on ne réalise pas d'hypothèse sur les marges car on l'estime par une méthode non-paramétrique. Ainsi, la loi marginale est remplacée ici par un estimateur non-paramétrique d'une distribution marginale défini ainsi :

$$\hat{F}_e(v) = \frac{\sum_{i=1}^m \mathbf{1}_{(y_i \leq v)}}{m} \quad (4)$$

avec y_i la charge d'un type de sinistre à une date i donnée et m la taille de l'échantillon lié à chaque type de sinistres.

La méthode du maximum de pseudo-vraisemblance (dit méthode CML) est retenue afin d'estimer les paramètres de copule. En effet, elle a l'avantage d'utiliser seulement les estimateurs non-paramétrique de lois marginales au lieu de lois de familles connues, et d'ainsi déterminer de façon plus précise le paramètre adéquat pour chaque copule paramétrique.

Copules	Dommages/ Perte de biens	Perte de biens/ Dommages corporels	Dommages biens/ Véh motorisé	Dommages biens/ Perte employé
Gaussienne	0.4486 (0.464)	-0.1103 (-0.123)	-0.0984 (-0.0898)	0.0665 (0.0418)
Student(ddl=4)	0.4154 (0.464)	-0.1233 (-0.123)	-0.0820 (-0.0898)	0.0548 (0.0418)
Frank	3.0038 (3.0006)	0 (-0.7079)	0 (-0.5164)	0.1925 (0.2399)
Clayton	0.4851 (0.887)	0 (-0.1452)	0 (-0.108)	0.1476 (0.0547)
Gumbel	1.3883 (1.4436)	1 (1)	1 (1)	1.0262 (1.027)
AMH	0.7875 (0.956)	0 (-0.3843)	0 (-0.274)	0.1083 (0.116)

TABLE 3 – Paramètre de la copule selon la méthode CML , paramètre selon la méthode des moments entre parenthèses

Notons que les copules archimédiennes ne peuvent modéliser seulement les dépendances positives. Donc les paramètres minimaux trouvés pour les copules archimédiennes dans le cas de couples présentant des dépendances négatives paraissent cohérents dans ce contexte.

Ci-dessous se trouvent les mesures de dépendances de queue des copules paramétriques déterminées par maximum de pseudo-vraisemblance :

Copules	Dommages/ Perte de biens	Perte de biens/ Dommages corporels	Dommages biens/ Véh motorisé	Dommages biens/ Perte employé
Gaussienne	0 0	0 0	0 0	0 0
Student(ddl=4)	0.2102 0.2102	0.0525 0.0525	0.0596 0.0596	0.0879 0.0879
Frank	0 0	0 0	0 0	0 0
Clayton	0.2396 0	0 0	0 0	0.0091 0
Gumbel	0 0.3525	0 0	0 0	0 0.0350
AMH	0 0	0 0	0 0	0.5 0

TABLE 4 – Coefficients de dépendance de queues inférieur et supérieur pour chaque copule paramétrique calibrée au jeu de données

On observe globalement des dépendances de queue assez nettes pour le couple "Dommages aux biens/Perte de biens".

Dans l'élaboration d'un choix de copule paramétrique décrivant de façon optimale la structure de dépendance au sein d'un couple de risques, il nous faut réaliser des tests d'ajustements des copules paramétriques, obtenus par maximum de pseudo-vraisemblance, avec les copules empiriques. Le test d'ajustement adapté est ici le test d'ajustement du χ^2 . Ce test permet de vérifier l'hypothèse selon laquelle un effectif observé regroupé en classes est conforme à l'effectif théorique regroupé en classes identiques à ceux de l'effectif observé. L'effectif théorique est calculé dans notre étude à partir de la probabilité théorique d'appartenance de la copule paramétrique dans un carré donné grâce à la fonction prob sur R. En multipliant la probabilité théorique par le nombre total d'observation pour le couple de sinistres, nous pouvons obtenir l'effectif théorique pour chaque classe appliqué à une copule paramétrique donnée. L'hypothèse H_0 du test stipule une non-conformité entre les effectifs, donc une p-value élevée signifie que l'on ne peut pas dire que les effectifs soient non-conformes. Une p-valeur élevée sera donc recherchée dans le choix de la bonne copule paramétrique. Ci-dessous se trouvent les choix de copule pour chaque couple de sinistre :

Couples de sinistre	Copule	\hat{a}_{CML}
Dommages/Perte de biens	Gumbel	1.3883
Perte de biens/Dommages corp	Student	-0.1233
Dommages aux biens/Veh motorisé	Student	-0.0820
Dommages aux biens/Perte employé	Student	0.0548

TABLE 5 – Meilleurs copules selon test du χ^2 pour chaque couple de sinistre

La copule de Gumbel possède en particulier une dépendance de queue supérieure mais pas inférieure, ce qui signifierait que les sinistres à charge élevée ont un réel impact dans la structure de dépendance du couple "Dommages aux biens/Perte de biens". Pour les autres couples, il existerait une dépendance de queue symétrique au vu des copules de Student présentant une structure de dépendances symétrique.

Approche multirisque par produit des copules bivariées

Les copules sont également des outils très utiles dans la modélisation de risques multiples. Beaucoup a été réalisé concernant les copules bivariées mais peu d'études sont présentes à ce jour sur les copules à grandes dimensions. Dans cette section, nous cherchons à modéliser les risques associées aux 4 copules bivariées paramétriques modélisés dans la précédente section. Plus précisément, il sera sujet de construire une copule multivariée de dimension 5 (nombre de marginales associées aux 4 couples étudiées) par le biais de produits de copules bivariées bien

choisies, ainsi que d'étudier les dépendances et dépendances de queues des couples au sein de la copule multivariée.

Il est admis que multiplier deux copules bivariées ne permet pas d'obtenir en général une copule. Cependant, par une transformation astucieuse des marginales, il est possible d'obtenir une copule bien définie par un produit de copules bivariées. Le théorème ci-dessous introduit par E. Liebscher [2008] définit cette copule :

Theoreme 1 Soient les copules $\tilde{C}_1, \dots, \tilde{C}_K : [0, 1]^d \rightarrow [0, 1]$. Soient les fonctions $g_{ei} : [0, 1] \rightarrow [0, 1]$ avec $e = 1, \dots, K, i = 1, \dots, d$. Chaque fonction est strictement croissante ou est identiquement égale à 1. Supposons que $\prod_{e=1}^K g_{ei}(v) = v, \forall v \in [0, 1], i = 1, \dots, d$ et $\lim_{v \rightarrow 0} g_{ei}(v) = 0, \forall e = 1, \dots, K, i = 1, \dots, d$. Alors :

$$C(u_1, \dots, u_d) = \prod_{e=1}^K \tilde{C}_e(g_{e1}(u_1), \dots, g_{ed}(u_d)) \quad (5)$$

est aussi une copule.

Une version revisitée, introduite par G. Mazo et al. [2014], est mise en place dans le but de construire notre copule multivariée.

Produit de copules transformées revisitée

Le produit selon $e \in \{1, \dots, K\}$ peut être établie selon $e \in E$, avec E un ensemble de données finies, tel que :

$$C(u_1, \dots, u_d) = \prod_{e \in E} \tilde{C}_e(g_{e1}(u_1), \dots, g_{ed}(u_d)). \quad (6)$$

E représente en particulier un couple de variables en jeu. Il s'agit dans notre cas d'un couples de marginales uniformes de la copule multivariée. Soit U_1, U_2, \dots, U_d les variables aléatoires uniformes. Le couple $\{i, j\}$ de la paire (U_i, U_j) est inclus dans l'ensemble E tel que $E \subset \{\{ij\} : i, j = 1, \dots, d, j > i\}$. Il s'agit d'un sous ensemble de l'ensemble des couples d'indices. Le cardinal de E , noté $|E|$, est égale au plus à $d(d-1)/2$. En suivant les préceptes du théorème du produit des copules bivariées exposées dans la section précédente, la copule multivariée se définit dans ce cas-ci ainsi :

$$C(u_1, \dots, u_d) = \prod_{\{ij\} \in E} \tilde{C}_{\{ij\}}(u_i^{1/n_i}, u_j^{1/n_j}). \quad (7)$$

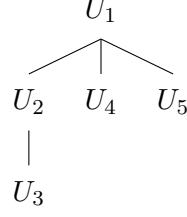
où $\tilde{C}_{\{ij\}}$ sont des copules bivariées, n_i le nombre de marginales liées à la variable U_i et n_j le nombre de marginales liées à la variable U_j .

Une représentation graphique par arbre permet de mieux se représenter les différentes liaisons entre les lois marginales uniformes et ainsi de déterminer E et les variables $n_i, i = 1, \dots, d$.

L'application qui suit reprend les couples de risque étudiées dans le cas bivarié. A partir des 4 copules bivariées modélisées précédemment, nous pouvons obtenir 5 lois marginales uniformes, c'est à dire une pour chaque risque. Dans le détail :

- U_1 : variable aléatoire uniforme du risque "Dommages aux biens".
- U_2 : variable aléatoire uniforme du risque "Perte de biens".
- U_3 : variable aléatoire uniforme du risque "Dommages corporels".
- U_4 : variable aléatoire uniforme du risque "Perte employé".
- U_5 : variable aléatoire uniforme du risque "Véhicule".

D'après les couples établis dans la modélisation des copules bivariées, nous pouvons représenter les liaisons entre nos risques comme ceci :



Dans ce cadre, $E = \{\{12\}, \{14\}, \{15\}, \{23\}\}$, $n_1 = 3, n_2 = 2, n_3 = n_4 = n_5 = 1$. Donc la copule multivariée décrivant nos 5 risques se définit par :

$$C(u_1, u_2, u_3, u_4, u_5) = \tilde{C}_{\{12\}}(u_1^{1/3}, u_2^{1/2}) \tilde{C}_{\{14\}}(u_1^{1/3}, u_4) \tilde{C}_{\{15\}}(u_1^{1/3}, u_5) \tilde{C}_{\{23\}}(u_2^{1/2}, u_3) \quad (8)$$

Il est supposé dans ce modèle PBC (Product of Bivariate Copula) revisité que les couples $\{ij\}$ n'appartenant pas à E disposent de lois uniformes indépendantes tels que $C_{\{ij\}}(u_i, u_j) = u_i u_j$, C représentant la copule bivariée issue du modèle PBC multivarié. D'autre part, on remarque que l'agrandissement de l'arbre, c'est à dire une connexion élevée entre les lois uniformes augmentent les variables n_i , ce qui d'après l'équation du modèle PBC revisité, affaiblit les dépendances de la copule multivariée.

La section poursuit l'étude avec une implémentation de la copule multivariée modélisant nos cinq risques ainsi que l'étude des dépendances de la copule obtenue. Un comparatif avec les copules bivariées paramétriques sera établi.

Application numérique et étude des dépendances

La simulation numérique de la copule multivariée s'articule selon le modèle PBC exposé préalablement ainsi que ses propriétés probabilistes associées. Soient $\tilde{C}_{\{ij\}}$, copules bivariées dépendant du paramètre θ_{ij} . La procédure de simulation est la suivante :

- Pour tout $\{ij\} \in E$, générer $(U_i^{(ij)}, U_j^{(ij)}) \sim \tilde{C}_{\{ij\}}$
- Pour tout $i=1, \dots, d$, calculer $U_i = \max_{k \in \{1, \dots, d\}: \{ki\} \in E} \left\{ \left(U_i^{(ki)} \right)^{n_i} \right\}$

En implémentant l'algorithme pour nos couples de risques, les variables U_1, U_2, U_3, U_4 et U_5 composent la copule multivariée de dimension 5 modélisant les dépendances des risques du portefeuille.

Par la simulations de 10000 éléments représentant la copule multivariée, il est possible de mesurer les dépendances deux à deux des risques de la copule multivariée. Ci-dessous se trouvent les mesures de dépendance de type tau de Kendall pour les risques 2 à 2 dans le cadre de la copule multivariée. Un comparatif avec les dépendances des copules bivariées est également réalisée.

Tau de Kendall	Dommages/ Perte de biens	Dommages aux biens/ Perte employé	Dommages aux biens Vehicule	Perte biens/ Dmg corporels
Copule PBC	0,1124	0,02	-0,0172	-0,0487
Copule bivarié	0,2797	0,0349	-0,052	-0,0787

TABLE 6 – Mesure des dépendances 2 à 2 dans le cadre multivarié et comparaison avec le cadre bivarié

Comme constaté dans les remarques sur le modèle, par rapport aux copules bivariées paramétriques obtenus préalablement dans l'étude, on observe une dépendance globalement plus faible des risques 2 à 2 au sein de la copule PBC multivariée. Il est à noter que les dépendances

négatives sont également dépréciées. On peut expliquer cette diminution de la dépendance par l'établissement seulement de 4 couples parmi 10 possibles ($|E|= 10$ au maximum lorsque $d=5$), ce qui suppose que les autres couples possibles sont indépendantes. Également, n_1 est ici égal à 3, ce qui réduit la dépendance pour les couples associées à la variable "Dommages aux biens" (notamment pour le couple "Dommages aux biens/Perte de biens" qui présente la dépendance la plus élevée parmi tous les risques).

Pour compléter le comparatif des dépendances, la dépendance de queues pour les risques 2 à 2 est exposée ci-dessous :

Tail	Dommages/ Perte de biens	Dommages aux biens/ Perte employé	Dommages aux biens Vehicule	Perte biens Dmg corporels
λ_L	0,02	0	0,03	0,03
λ_U	0,11	0,03	-0,07	0,07

TABLE 7 – Mesure des dépendances de queue de chaque couple au sein de la copule PBC multivariée

En comparant avec les dépendances de queue des copules bivariées, on observe une diminution de la dépendance de queue inférieure ainsi qu'une faible diminution de la dépendance de queue supérieure. L'asymétrie présumée des dépendances dans le cas bivarié (avec une copule de Gumbel dont la dépendance de queue à droite était significative) subsiste dans ce cadre multivariée.

La méthode PBC présente des propriétés avantageuses. En effet, elle ne nécessite que la connaissance des copules bivariées au sein des risques à modéliser. De plus, elle préserve (comme vu ci-dessus) l'asymétrie des queues de distribution et est un modèle flexible. Cependant, certaines limites sont à noter. D'une part les couples dont la copule bivarié n'est pas définie sont établies comme indépendantes, ce qui génère de facto une dépendance à la baisse. D'autre part, un graphe trop connecté, avec des nombreuses liaisons entre chaque risque, affaiblit également la dépendance.

L'utilisation de la méthode PBC pour cette étude semble néanmoins adéquate dans ce contexte, étant donnée la relative faiblesse des dépendances modélisés dans le cas bivarié ainsi que le faible nombre de couples parmi les risques du portefeuille. Il sera intéressant d'établir à partir des copules multivariées obtenus l'impact des dépendances de l'ensemble des risques sur le besoin en fonds propres.

Modélisation de l'impact des dépendances sur le besoin en fonds propres

Pour satisfaire les engagements auprès des assurés, un assureur doit maintenir un besoin en fonds propres optimal pour son portefeuille. En supposant que l'agence TSA indemnise de son propre chef les sinistrés, il nous faut établir un besoin en fonds propres selon chacun des couples de sinistres retenus pour mesurer l'impact des dépendances modélisées sur le besoin en fonds propres. Une comparaison du besoin en fonds propres selon une situation d'indépendance ou de dépendance entre les risques d'un même couple permet de mesurer l'impact des dépendances sur le besoin en fonds propres.

Les différentes mesures de risque utilisées pour calculer le besoin en fonds propres sont la VaR et la TVaR, mesures très utilisées en assurance et en finance. La VaR est une mesure de risque reflétant justement la distribution globale de la variable aléatoire, tandis que la TVaR sera plus sensible aux distributions de queue de la variable aléatoire.

Nous procédons en premier lieu à une simulation d'un grand nombre de charges de sinistres (10000 sinistres) selon la copule paramétrique choisie dans le cas dépendant, et selon la copule d'indépendance dans le cas indépendant. Il s'agira d'estimer la charge de sinistre agrégée pour un couple de sinistres afin de définir la distribution du niveau de fonds propres. Enfin, en appliquant la VaR et TVaR, les résultats du besoin en fonds propres seront comparables selon le cas dépendant ou indépendant.

Dommages/Perte de biens	$VaR_{75\%}$	$VaR_{99.5\%}$	$TVaR_{75\%}$	$TVaR_{99.5\%}$
Cas dépendant (Gumbel)	4882,37	23454,03	17983,24	4006024
Cas indépendant	5 041,51	19 251,08	17 490,83	3 144 863
Ecart	-3,26%	21,83%	2,82%	27,38%

TABLE 8 – Comparaison des mesures de risque en euros pour le couple Dommages aux biens/Perte de biens dans le cas dépendant et indépendant

Pour le couple "Dommages aux biens/Perte de biens", on constate un écart positif avec un besoin en fonds propres supérieur lorsque l'on suppose la dépendance entre les sinistres par rapport au besoin en fonds propres dans le cas indépendant. La mesure à adopter est plutôt la TVaR, car le couple est sensible à la distribution de queues (nous avons vu préalablement une dépendance de queue à droite nette pour ce couple).

Perte biens/Dom. corporels	$VaR_{75\%}$	$VaR_{99.5\%}$	$TVaR_{75\%}$	$TVaR_{99.5\%}$
Cas dépendant (tCopula)	8356,021	31985,77	28589,22	5138989
Cas indépendant	8650,82	31836,09	30614,77	5084060
Ecart	-3,53%	0,47%	-7,09%	1,08%

TABLE 9 – Comparaison des mesures de risque en euros pour le couple Perte de biens/Dommages corporels dans le cas dépendant et indépendant

Le couple "Perte de biens/Dommages corporels" présente une structure de dépendance négative. Au regard de la VaR à 75%, il apparaît que le besoin en fonds propres dans le cas dépendant est inférieur au besoin en fonds propres si l'on suppose l'indépendance entre les risques.

Dommages aux biens/Veh.motorisé	$VaR_{75\%}$	$VaR_{99.5\%}$	$TVaR_{75\%}$	$TVaR_{99.5\%}$
Cas dépendant (tCopula)	4041,611	10948,71	10965,66	1 556704
Cas indépendant	4184,03	10131,45	11697,59	1 390 401
Ecart	-3,45%	8,06%	-6,67%	6,53%

TABLE 10 – Comparaison des mesures de risque en euros pour le couple Dommages aux biens/Véhicule motorisé dans le cas dépendant et indépendant

De même, le couple "Dommages aux biens/Véhicule motorisé", de structure de dépendance négative, présente un besoin en fonds propres inférieur dans le cas dépendant par rapport au cas indépendant.

Dommages aux biens/Perte employé	$VaR_{75\%}$	$VaR_{99.5\%}$	$TVaR_{75\%}$	$TVaR_{99.5\%}$
Cas dépendant (tCopula)	4600,816	11230,97	12647,62	1543714
Cas indépendant	4582,55	10714,16	12552,96	1447175
Ecart	0,4%	4,8%	0,75%	6,67%

TABLE 11 – Comparaison des mesures de risque en euros pour le couple Dommages aux biens/Perte employé dans le cas dépendant et indépendant

Enfin, au vu de la VaR à 75%, le besoin en fonds propres est légèrement supérieur dans le cas dépendant par rapport au cas indépendant pour le couple "Dommages aux biens/Perte employé".

Un approfondissement de la modélisation du besoin en fonds propres dans un cadre multirisques peut aussi être réalisé. A partir de la copule multivariée obtenue dans la section précédente par la méthode PBC, il est en effet possible de générer les charges de sinistres agrégées des 5 risques du portefeuille selon la même méthode de simulation que dans le cas bivarié (charge obtenue par la fonction inverse de répartition empirique pour chaque marge appliquée à chaque uniforme U_1, U_2, U_3, U_4 et U_5).

Au regard de la répartition des quantiles des charges dans le cas dépendant et indépendant, il apparaît qu'un besoin de fonds propres à la baisse se fait sentir, sauf en $VaR_{99,5\%}$.

$VaR_{75\%}$ dépendant	11 156,3378
$VaR_{75\%}$ indépendant	11 358,92
Ecart	-1,78%

TABLE 12 – $VaR_{75\%}$ dans le cadre multirisque

Au vu de la dépendance de queue présente pour les charges les plus élevées, la VaR semble moins approprié dans la mesure du besoin en fonds propres de ces risques. La mesure par le biais de la TVaR, plus sensible aux queues de distribution semble ici plus adaptée.

Mesure par TVaR	$TVaR_{75\%}$	$TVaR_{99,5\%}$
Multirisques dépendant	33 426,89	5 505 137
Multirisques indépendant	33 389,16	5 106 543
Ecart	0,11%	7,81%

TABLE 13 – Mesures par TVaR dans le cadre multirisque

Les mesures par la TVaR établissent un besoin en fonds propres supplémentaire dans le cas dépendant qui se révèle quasiment négligeable (0,11%). En comparaison de l'impact des dépendances dans le cas bivarié, il apparaît dans ce cadre multirisques que les dépendances ne renvoient que légèrement à la hausse le besoin en fonds propres (0,11% vs 2,82% dans le cas bivarié de la copule de Gumbel).

Nous avons pu à travers cette étude déterminer l'impact de structures de dépendance de toutes natures, positives comme négatives, sur le besoin en fonds propres. Il en résulte qu'un besoin en fonds propres supérieur se fait ressentir lorsque l'on suppose la dépendance entre les charges de sinistre. Toutefois, si la structure de dépendance du couple est négative, il apparaît à l'inverse que le besoin en fonds propres dans le cas dépendant est inférieur au cas indépendant. Ce phénomène participe au phénomène de compensation du besoin en fonds propres entre de multiples risques de dépendances diverses (le fameux effet "diversification").

L'étude de risques multiples du portefeuille, par la méthode de produits de copules bivariées (PBC), a permis d'analyser les dépendances des couples de risques dans un cadre multivarié. Il apparaît alors, du fait principalement de la présence de paires de marginales indépendantes, une dépendance plus faible au sein des différents couples de risques de la copule multivariée par rapport aux dépendances mesurées par les copules paramétriques bivariées. Néanmoins, la présente d'asymétrie dans les queues de distribution subsiste dans le cadre multirisques, ce qui doit retenir l'attention de l'assureur dans son choix final de modélisation des risques (une modélisation par copule multivariée gaussienne ou par copule indépendante supposerait l'absence de dépendances de queues). L'impact des dépendances sur besoin en fonds propres est moins significatif dans le cadre multirisques, on note cependant un léger besoin en capital supplémentaire dans le cas dépendant, due en grande partie à la présence de dépendances de queue.

L'hypothèse d'indépendance entre les risques est donc discutable ici vu les natures diverses de dépendance entre les risques et le nombre important de risques différents dans notre base (7 types de sinistre). Toutefois, l'impact des dépendances sur le besoin en fonds propres pour le couple "Dommages aux biens/Perte de biens" peut être similaire à l'impact des dépendances sur le besoin en fonds propres pour l'ensemble de la base de la TSA, étant donné le nombre conséquent de charges de sinistres pour le couple "Dommages aux biens/Perte de biens" par rapport au nombre de sinistres des autres couples possibles.

Summary note

In developing the capital requirement for a non-life insurer, several assumptions are generally accepted when modelling the risks in a portfolio. Especially, an independence assumption between multiple risks in a portfolio is made, or, if dependence between risks is taken into account, it is accepted that marginal laws follow the same family of laws. These assumptions do not enable to represent in a realistic way risks within a non-life portfolio and may influence the solvency level of an insurer in this context. The challenges of modelling realistic risk dependencies include a more accurate modelling of the insurer's capital requirements. The right estimation of capital requirement is necessary for an insurer in order to meet interests of both regulator and shareholders. Indeed, the regulator requires a prudential vision with a solvency margin to be available in its equity, while the shareholders want to minimise the need for equity in order to optimise their return. Copulas are a tool to distinguish the behaviour of the marginal distribution function in the dependency structure. It enable to define dependency structures of any type, especially non-linear dependencies between claims and tail dependencies. Throughout this paper, it will be about modelling dependency structure as best as possible from the risk pairs including proven correlation in the extracted data.

Copula theory

An n-dimensional copula C is an n-dimensional distribution function defined from $[0, 1]^n$ to $[0, 1]$. It can be written in the following form, $\forall(u_1, \dots, u_n) \in [0, 1]^n$:

$$C(u_1, \dots, u_n) = P(U_1 \leq u_1, U_2 \leq u_2, \dots, U_n \leq u_n), U_1, U_2, \dots, U_n \sim U([0; 1]) \quad (9)$$

The fundamental theorem of copula theory is Sklar's theorem. This theorem enables to establish the link between a multivariate function and its marginal laws. It also provides a sufficient condition for the existence of copulas. Let F be a distribution function of a vector $(X_1, X_2, X_3, \dots, X_n)$ whose marginal distribution functions are $F_i, i = 1, \dots, n$. Then there is a unique copula C such that

$$F(x_1, x_2, \dots, x_n) = C(F_1(x_1), F_2(x_2), \dots, F_n(x_n)) \quad (10)$$

Reversely, given the copula C and its marginals $F_i, i = 1, 2, \dots, n, C(F_1(x_1), F_2(x_2), \dots, F_n(x_n))$ defines a multivariate distribution function of dimension n. This theorem is also useful to build copula functions from multivariate distributions. If the marginal laws are continuous, then :

$$C(u_1, \dots, u_n) = F(F_{X_1}^{-1}(u_1), \dots, F_{X_n}^{-1}(u_n))$$

Throughout our study, several usual copulas will be used. Some of them have particular properties that will be explained below.

- The independence copula : $\prod(u) = u_1 \times u_2 \times \dots \times u_d$.
- Elliptical copulas such as the Gaussian copula or the Student copula. The Gaussian copula is defined from families of Gaussian distributions. The multivariate Gaussian copula

is defined as :

$$C_{\Sigma}^{Ga} = \phi_{\Sigma}(\phi^{-1}(u_1), \phi^{-1}(u_2), \dots, \phi^{-1}(u_n)) \quad (11)$$

where ϕ_{Σ} is the distribution function of the multivariate distribution $\mathcal{N}(0_{\mathbb{R}^n}, \Sigma)$ and ϕ is the distribution function of the normal distributions following $\mathcal{N}(0, 1)$.

The Student's copula is defined by marginal laws of Student's law related to a Student's copula. Let $T_{\Sigma, \nu}$ and T_{ν} be the distribution functions of $T(\Sigma, \nu)$ and $T(\nu)$ respectively. The Student's copula is then defined by :

$$C(u_1, \dots, u_n) = T_{\Sigma, \nu}(T^{-1}(u_1), \dots, T^{-1}(u_n)) \quad (12)$$

- Archimedean copulas have the special feature to get asymmetric dependencies with different tail dependencies. This notion of archimedean copula was defined by GENEST and MACKAY [1986]. In bivariate cases, it can be defined as follows :

$$C(u, v) = \phi^{-1}(\phi(u) + \phi(v)), u, v \in [0, 1]$$

With ϕ a strict generator, a convex decreasing function on $(0, 1] \rightarrow [0, \infty]$ such that $\phi(1) = 0$ and $\phi(0) = \infty$. The generator ϕ is called non-strict if $\phi(0) < \infty$.

The Frank, Clayton, Gumbel and AMH copulas are defined below.

Let the generator ϕ defined such that $\phi(x) = -\log\left(\frac{e^{-ax}-1}{e^{-a}-1}\right)$ with $a \neq 0$. Then Frank copula can be written as follows :

$$C_a(u, v) = -\frac{1}{a} \log\left(1 + \frac{(e^{-au}-1)(e^{-av}-1)}{e^{-a}-1}\right)$$

The generator of the Clayton copula is written as follows :

$$\phi(t) = \frac{t^{-a}-1}{a}, a \in]0, +\infty[$$

The multivariate form of the Clayton copula is written as follows :

$$C_a(u_1, \dots, u_n) = (u_1^{-a} + \dots + u_n^{-a} - n + 1)^{-\frac{1}{a}}$$

Gumbel copula admits as generator $\phi(x) = (-\log(x))^a$ with $a \geq 1$.

Gumbel copula is written as follows :

$$C_a(u, v) = \exp(-(-\log(u))^a + (-\log(v))^a)^{\frac{1}{a}}$$

The Gumbel copula is symmetric, which is not the case in general for an archimedean copula.

Ali, Mikhail and Haq (1978) created an archimedean copula whose particularity is its parameter covering a reduced interval $[-1, 1]$. The generator of the copula is written $\phi(t) = \ln\left(\frac{1-\theta(1-t)}{t}\right)$, so the AMH copula is written as follows :

$$C_{\theta}(u, v) = \frac{uv}{1 - \theta(1-u)(1-v)}$$

where the copula parameter is $\theta \in [0, 1]$.

It is also possible to build an empirical copula from given observations. To do this, we need to calculate the rank statistic of the joint distribution. Empirical copulas can be used as a comparative tool in order to calibrate a parametric copula to the database we have.

Deuvels [1979] created the notion of the empirical copula. The function needs rank of the observations to extract the dependence structure.

Let (X_1, \dots, X_d) be a sequence of random variables and $(x_1, \dots, x_n)_{1 \leq n \leq N}$ a sample of observations of size N . Let us also fix $(r_1, \dots, r_n)_{1 \leq n \leq N}$ the rank statistic (or rank of observations) associated with this multi-variate sample. Thus for any $1 \leq i \leq d$, r_i^n is the rank of x_i^n in $(x_i^n)_{1 \leq n \leq N}$, that is to say $\sum_{n=1}^N \mathbb{1}_{X_i^n \leq x_i^n} \forall 1 \leq i \leq d$.

We can then define empirical copula on the cluster $\mathbf{E} = \{(\frac{n_1}{N}, \dots, \frac{n_d}{N}), 1 \leq i \leq d, n_i = 0, \dots, N\}$ as follows :

$$\hat{C}_N \left(\frac{n_1}{N}, \dots, \frac{n_d}{N} \right) = \frac{1}{N} \sum_{n=1}^N \prod_{i=1}^d \mathbb{1}_{\{r_i^n \leq n_i\}}$$

Presentation of the dataset and search for dependencies on our dataset

The dataset on which the study is based comes from the Transportation Security Administration (TSA), the US agency that manages airport security at most US airports. This agency was created following September 11, 2001 attacks in order to better protect the public during air travel. One of the main functions of the TSA is to organise checks on passengers and their personal belongings before boarding flights at country's various airports. During these various checkpoints, it may happen that a passenger suffers a civil liability claim. The passenger has the possibility to report any type of claim to the TSA. Thus, claims such as injuries, loss of or damage to objects, theft can be declared following the checks. The dataset includes all passenger claims for the period 2002-2015. 145012 claims were recorded over this period. The aim of the study will be to first determine the dependencies between each type of claim, and then to model the dependencies of the types of claim reported using copulas. Finally, the impact of our dependency modelling by copulas on the TSA's capital requirement to cover claims will be estimated.

The treatment of the base consists of first calculating the closed claim amounts on the 2014 base. Each of the claim amounts will then be estimated according to the 2014 consumption index, in order to obtain comparable claim amounts in the same year. Next, the amounts of non-zero closed claims are aggregated by claim type and claim incidence date. The choice of aggregation by incidence date results from the fact that claims are all short term. Indeed, the claims are all sudden events with no impacts on the following day. The aggregation of claims allows us to distinguish existing correlations between each type of claim. In order to justify the study of correlations from this aggregated base, the assumption is made that claims of the same type of claim are independent from each other from one day to the next. This assumption seems realistic as this is an environment where individuals are passing through. Also, the claims are referring to the passengers, so the event is passenger centric.

The correlation study of the types of claims in the database is mainly based on the dependency measures Pearson's correlation coefficient, Kendall's tau and Spearman's rho. The Pearson correlation coefficient primarily determines linear correlations, while Kendall's tau and Spearman's rho are based on the rank statistics of the observations. Below are the correlation results for each of the claim pairs in the database :

We find a clear positive correlation for the pairing 'Property loss/Property damage', while the pairing 'Property loss/Personal injury' shows a weakly negative correlation (about -0.08) as well as the pairing 'Property damage/Motor vehicle'. The pair "Property damage/Employee loss" shows a weakly positive correlation. These low Kendall's ratios are partly explained by the low number of observations for some claim pairs.

Through these measures, it can be concluded that these four couples have a low dependency. The modelling of these dependencies by copulas is possible for these pairs because the establishment

Kendall's tau	Property loss	Property damage	Personal injury	Vehicle	Employee
Property loss	1	0,307	-0,0787	-0,012	0,020
Property damage	0,307	1	-0,019	-0,057	0,041
Personal injury	-0,0787	-0,019	1	No data	No data
Car	-0,012	-0,057	No data	1	No data
Employee	0,020	0,041	No data	No data	1

TABLE 14 – Kendall's tau

of the dependency within these pairs is verified. We can now model the dependencies of these claim pairs by copulas.

Estimating the parametric copula parameter and choosing the right copula

The empirical copulas model the dependencies of our loss pairs as closely as possible. The purpose of this section is to define the parameters of the parametric copulas in order to fit accurately the parametric copulas to the empirical copula of each risk pair. Firstly, work was done on the fit of the marginal laws of each pair of claims. The fit consists on estimating the laws of the exponential family which are close to the marginal laws. To do this, a maximum likelihood method was used. The fit test of our marginal laws is a Kolmogorov Smirnov test. Below are the results of the test, with p-values well under 1%.

Fit	D	P-Value
Property loss - Log-normal	1 571 323	$< 2.2 \times 10^{-16}$
Property damage - Gamma	13 379	$< 2.2 \times 10^{-16}$
Personal injury - Log-normal	226 171	$< 2.2 \times 10^{-16}$
Vehicle - Gamma	6197,8	$< 2.2 \times 10^{-16}$
Employee loss - Gamma	2330,6	$< 2.2 \times 10^{-16}$

TABLE 15 – Results of the Kolmogorov Smirnov test

With regard to the p-values, the H_0 hypothesis which states that the marginal laws follow their respective theoretical laws is rejected for each law (p-value well under the fixed 1%). The semi-parametric approach is therefore required here. In the semi-parametric approach, the hypothesis on the margins is not done because it is estimated by a non-parametric method. Thus, the marginal law is replaced here by a non-parametric estimator of a marginal distribution defined as follows :

$$\hat{F}_e(v) = \frac{\sum_{i=1}^m \mathbf{1}_{(y_i \leq v)}}{m} \quad (13)$$

with y_i the amount of a type of claim at a given date i and m the sample size for each type of claim.

The maximum pseudo-likelihood method (known as the CML method) is used to estimate the copula parameters. Indeed, it has the benefit of using only non-parametric estimators of marginal laws instead of laws of known families, and thus to define most precisely the correct parameter for each parametric copula.

Copula	Property loss/ damage	Property loss/ Personal injury	Property damage/ Vehicle	Property damage/ Employee loss
Gaussian	0.4486 (0.464)	-0.1103 (-0.123)	-0.0984 (-0.0898)	0.0665 (0.0418)
Student(ddl=4)	0.4154 (0.464)	-0.1233 (-0.123)	-0.0820 (-0.0898)	0.0548 (0.0418)
Frank	3.0038 (3.0006)	0 (-0.7079)	0 (-0.5164)	0.1925 (0.2399)
Clayton	0.4851 (0.887)	0 (-0.1452)	0 (-0.108)	0.1476 (0.0547)
Gumbel	1.3883 (1.4436)	1 (1)	1 (1)	1.0262 (1.027)
AMH	0.7875 (0.956)	0 (-0.3843)	0 (-0.274)	0.1083 (0.116)

TABLE 16 – Copula parameter with CML method, parameter according to the moments method in brackets.

Note that Archimedean copulas can model only positive dependencies. Therefore, the minimal parameters found for Archimedean copulas in the case of pairs with negative dependencies seem consistent in this context.

Below are the measures of tail dependencies of parametric copulas defined by maximum pseudo-likelihood :

Copula	Property loss/ damage	Property loss/ Personal injury	Property damage/ Vehicle	Property damage/ Employee loss
Gaussian	0 0	0 0	0 0	0 0
Student(ddl=4)	0.2102 0.2102	0.0525 0.0525	0.0596 0.0596	0.0879 0.0879
Frank	0 0	0 0	0 0	0 0
Clayton	0.2396 0	0 0	0 0	0.0091 0
Gumbel	0 0.3525	0 0	0 0	0 0.0350
AMH	0 0	0 0	0 0	0.5 0

TABLE 17 – Lower and upper tail dependency coefficients for each parametric copula calibrated to the dataset

Overall, there are fairly clear tail dependencies for the "Property Damage/Loss of Property" pair.

In developing a choice of parametric copula that describes in the best possible way the dependence structure within a risk pair, we need to perform fit tests of the parametric copulas, obtained by maximum pseudo-likelihood, with the empirical copulas. The suitable fitting test here is the χ^2 goodness of fit test. This test checks the hypothesis that an observed population grouped into classes is consistent with the theoretical population grouped into classes identical to those of the observed population. The theoretical number is calculated in our study from the theoretical probability of the parametric copula to belong to a given square using "prob" function on R. By multiplying the theoretical probability by the total number of observations for the pair of claims, we can obtain the theoretical size for each class applied to a given parametric copula. The H_0 hypothesis of the test states a non-conformity between the numbers, so a high p-value means that the numbers cannot be said to be non-conforming. A high p-value will therefore be desired in the choice of the right parametric copula. Below are the best copula choices for each pair of claims :

Claim pairs	Copula	\hat{a}_{CML}
Property loss/damage	Gumbel	1.3883
Property loss/Personal injury	Student	-0.1233
Property damage/Vehicle	Student	-0.0820
Property damage/Employee loss	Student	0.0548

TABLE 18 – Best copulas according to χ^2 goodness of fit test for each pair of claims

Especially, the Gumbel copula has a higher but not a lower tail dependency, which would mean that high amount claims have a real impact on the dependency structure of the property damage/Property loss. For the other pairs, there would be a symmetric tail dependence with regard to the Student copulas with symmetric dependence structure.

Multi-risk approach by product of bivariate copulas

Copulas are also very useful tools for modelling multiple risks. Much has been done on bivariate copulas but few studies are present to date on high dimensional copulas. In this section, we seek to model the risks associated with the 4 parametric bivariate copulas modelled in the previous section. More precisely, it will be a matter of constructing a multivariate copula of dimension 5 (number of marginals associated with the 4 pairs studied) through products of well-chosen bivariate copulas, as well as studying the dependencies and tail dependencies of the pairs within the multivariate copula.

It is accepted that multiplying two bivariate copulas does not usually result in a copula. However, by a clever transformation of the marginals, it is possible to obtain a well-defined copula by a product of bivariate copulas. The theorem below introduced by E. Liebscher [2008] defines this copula :

Theorem 1 Let be the copulas $\tilde{C}_1, \dots, \tilde{C}_K : [0, 1]^d \rightarrow [0, 1]$. Let be the functions $g_{ei} : [0, 1] \rightarrow [0, 1]$ with $e = 1, \dots, K, i = 1, \dots, d$. Each function is strictly increasing or is identically equal to 1. Let suppose that $\prod_{e=1}^K g_{ei}(v) = v, \forall v \in [0, 1], i = 1, \dots, d$ et $\lim_{v \rightarrow 0} g_{ei}(v) = 0, \forall e = 1, \dots, K, i = 1, \dots, d$. Then :

$$C(u_1, \dots, u_d) = \prod_{e=1}^K \tilde{C}_e(g_{e1}(u_1), \dots, g_{ed}(u_d)) \quad (14)$$

is also a copula.

A revised version, introduced by G. Mazo et al [2014], is implemented in order to construct our multivariate copula.

Product of transformed copulas revisited

The product depending on $e \in \{1, \dots, K\}$ can be established with $e \in E$, with E a finite dataset, such as :

$$C(u_1, \dots, u_d) = \prod_{e \in E} \tilde{C}_e(g_{e1}(u_1), \dots, g_{ed}(u_d)). \quad (15)$$

In particular, E represents a pair of variables in play. In our case, it is a pair of uniform marginals of the multivariate copula. Let U_1, U_2, \dots, U_d uniform random variables. The couple $\{i, j\}$ of the pair (U_i, U_j) is included in E such as $E \subset \{\{i, j\} : i, j = 1, \dots, d, j > i\}$. This is a subset of the set of index pairs. The cardinal of E, denoted $|E|$, is equal to at most $d(d-1)/2$. Following the precepts of the product theorem of bivariate copulas outlined in the previous section, the

multivariate copula is defined in this case as follows :

$$C(u_1, \dots, u_d) = \prod_{\{ij\} \in E} \tilde{C}_{\{ij\}}(u_i^{1/n_i}, u_j^{1/n_j}). \quad (16)$$

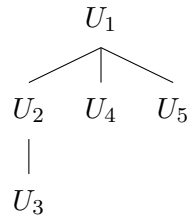
where $\tilde{C}_{\{ij\}}$ are bivariate copula, n_i the number of functions linked to the variable U_i et n_j the number of marginals related to the variable U_j .

A graphical representation by tree makes it possible to better represent the various links between the uniform marginal laws and thus to determine E and the variables $n_i, i = 1, \dots, d$.

The following application uses the risk pairs studied in the bivariate case. From the 4 bivariate copulas modelled previously, we can obtain 5 uniform marginal laws, i.e. one for each risk. In detail :

- U_1 : uniform random variable of risk "Property damage".
- U_2 : uniform random variable of risk "Property loss".
- U_3 : uniform random variable of risk "Personal injury".
- U_4 : uniform random variable of risk "Employee loss".
- U_5 : uniform random variable of risk "vehicle".

Based on the pairs established in the bivariate copula modelling, we can represent the links between our risks as follows :



In this context, $E = \{\{12\}, \{14\}, \{15\}, \{23\}\}$, $n_1 = 3, n_2 = 2, n_3 = n_4 = n_5 = 1$. Thus the multivariate copula describing our 5 risks is defined as :

$$C(u_1, u_2, u_3, u_4, u_5) = \tilde{C}_{\{12\}}(u_1^{1/3}, u_2^{1/2}) \tilde{C}_{\{14\}}(u_1^{1/3}, u_4) \tilde{C}_{\{15\}}(u_1^{1/3}, u_5) \tilde{C}_{\{23\}}(u_2^{1/2}, u_3) \quad (17)$$

It is supposed in this revisited PBC (Product of Bivariate Copula) model that the couples $\{ij\}$ not belonging to E have independent uniform laws such that $C_{ij}(u_i, u_j) = u_i u_j$, C representing the bivariate copula resulting from the multivariate PBC model. On the other hand, we notice that the enlargement of the tree, i.e. a high connection between the uniform laws increases the n_i variables, which according to the equation of the revisited PBC model, weakens the dependencies of the multivariate copula. The section continues the study with an implementation of the multivariate copula modelling our five risks and the study of the dependencies of the resulting copula. A comparison with the parametric bivariate copulas will be established.

Numerical application and dependencies study

The numerical simulation of the multivariate copula is based on the PBC model and its associated probabilistic properties. Let be $\tilde{C}_{\{ij\}}$, bivariat copulas depending on parameter θ_{ij} . Simulation procedure is the following :

- For all $\{ij\} \in E$, generate $(U_i^{(ij)}, U_j^{(ij)}) \sim \tilde{C}_{\{ij\}}$
- For all $i=1, \dots, d$, calculate $U_i = \max_{k \in \{1, \dots, d\} : \{ki\} \in E} \left\{ \left(U_i^{(ki)} \right)^{n_i} \right\}$

By implementing the algorithm for our risk pairs, the variables U_1, U_2, U_3, U_4 and U_5 compose the multivariate copula of dimension 5 modelling the portfolio risk dependencies.

By simulating 10000 elements representing the multivariate copula, it is possible to measure the pairwise dependencies of the risks in the multivariate copula. Below are Kendall's tau dependence measures for the pairwise risks in the multivariate copula. A comparison with the dependencies of the bivariate copulas is also made.

Kendall's tau	Damage/ Property loss	Property damage/ Employee loss	Property damage Vehicle	Property loss/ Personal injury
PBC copula	0,1124	0,02	-0,0172	-0,0487
Bivariate copula	0,2797	0,0349	-0,052	-0,0787

TABLE 19 – Measurement of pairwise dependencies in the multivariate framework and comparison with the bivariate framework

As noted in the remarks on the model, compared to the parametric bivariate copulas obtained earlier in the study, there is an overall weaker dependence of the 2-to-2 risks within the multivariate PBC copula. Note that the negative dependencies are also depreciated. This decrease in dependence can be explained by the establishment of only 4 pairs out of 10 possible pairs ($|E|=10$ at most when $d=5$), which implies that the other possible pairs are independent. Also, n_1 is here equal to 3, which reduces the dependence for the pairs associated with the "Property damage" variable (notably for the pair "Property damage/Loss of property" which presents the highest dependence among all the risks).

To complete the comparison of dependencies, the tail dependency for risks 2 to 2 is shown below :

Tail	Damages/ Property loss	Property damage/ Employee loss	Property damage Vehicle	Property loss Personal injury
λ_L	0,02	0	0,03	0,03
λ_U	0,11	0,03	-0,07	0,07

TABLE 20 – Measuring the tail dependencies of each pair within the multivariate PBC copulae

Comparing with the tail dependencies of the bivariate copulas, there is a decrease in the lower tail dependence and a small decrease in the upper tail dependence. The presumed asymmetry of the dependencies in the bivariate case (with a Gumbel copula whose right tail dependence was significant) remains in this multivariate setting.

The PBC method has advantageous properties. Indeed, it only requires knowledge of the bivariate copulas within the risks to be modelled. Moreover, it preserves (as seen above) the asymmetry of the distribution tails and is a flexible model. However, some limitations should be noted. On the one hand, the pairs whose bivariate copula is not defined are established as independent, which de facto generates a downward dependence. On the other hand, an overly connected graph, with many links between each risk, also weakens the dependence.

The use of the PBC method for this study nevertheless seems adequate in this context, given the relative weakness of the dependencies modelled in the bivariate case as well as the small number of pairs among the risks in the portfolio. It will be interesting to establish from the multivariate copulas obtained the impact of the dependencies of all the risks on the capital requirement.

Modelling the impact of dependencies on capital requirements

In order to meet policyholder liabilities, an insurer must keep an optimal capital requirement for its portfolio. Assuming that the TSA compensates victims on its own, we need to establish a capital requirement for each of the claim pairs selected to measure the impact of the modelled dependencies on the capital requirement. A comparison of the capital requirement according to whether we suppose independence or dependence between the risks of the same couple makes it possible to measure the impact of dependencies on the capital requirement.

The different risk measures used to calculate the capital requirement are VaR and TVaR, which are widely used in insurance and finance. VaR is a risk measure that accurately represents the overall distribution of the random variable, while TVaR will be more sensitive to the tail distributions of the random variable.

We first simulate a large number of claim expenses (10,000 claims) according to the chosen parametric copula in the dependent case, and according to the independence copula in the independent case. It will be necessary to estimate the aggregate claim amounts for a couple of claims in order to define the distribution of the level of equity. Finally, by applying the VaR and TVaR, the results of the capital requirement will be similar for the dependent and independent cases.

Property loss/damage	$VaR_{75\%}$	$VaR_{99.5\%}$	$TVaR_{75\%}$	$TVaR_{99.5\%}$
Dependent case (Gumbel)	4882,37	23454,03	17983,24	4006024
Independent case	5 041,51	19 251,08	17 490,83	3 144 863
Gap	-3,26%	21,83%	2,82%	27,38%

TABLE 21 – Comparison of risk measures in euros for the Property Damage/Loss pair in the dependent and independent case

For the "Property Damage/Loss" pair, we find a positive difference with a higher capital requirement when we assume the dependence between the claims with respect to the capital requirement in the independent case. The measure to adopt is rather the TVaR, as the pair is sensitive to the tail distribution (we have previously seen a net right-tail dependence for this pair).

Property loss/Personal injury	$VaR_{75\%}$	$VaR_{99.5\%}$	$TVaR_{75\%}$	$TVaR_{99.5\%}$
Dependent case (tCopula)	8356,021	31985,77	28589,22	5138989
Independent case	8650,82	31836,09	30614,77	5084060
Gap	-3,53%	0,47%	-7,09%	1,08%

TABLE 22 – Comparison of risk measures in euros for the property loss/personal injury pair in the dependent and independent case

The property loss/personal injury pair has a negative dependency structure. Based on the VaR at 75%, it appears that the capital requirement in the dependent case is lower than the capital requirement if we assume independence between risks.

Property damage/Vehicle	$VaR_{75\%}$	$VaR_{99.5\%}$	$TVaR_{75\%}$	$TVaR_{99.5\%}$
Dependent case (tCopula)	4041,611	10948,71	10965,66	1 556704
Independent case	4184,03	10131,45	11697,59	1 390 401
Gap	-3,45%	8,06%	-6,67%	6,53%

TABLE 23 – Comparison of risk measures in euros for the Property Damage/Vehicle pair in the dependent and independent case

Similarly, the property damage/motor vehicle pair, with a negative dependency structure, has a lower capital requirement in the dependent case than in the independent case.

Property damage/Employee loss	$VaR_{75\%}$	$VaR_{99.5\%}$	$TVaR_{75\%}$	$TVaR_{99.5\%}$
Dependent case (tCopula)	4600,816	11230,97	12647,62	1543714
Independent case	4582,55	10714,16	12552,96	1447175
Gap	0,4%	4,8%	0,75%	6,67%

TABLE 24 – Comparaison des mesures de risque en euros pour le couple Dommages aux biens/Perte employé dans le cas dépendant et indépendant

Finally, given the 75% VaR, the capital requirement is slightly higher in the dependent case than in the independent case for the "Property Damage/Employee loss" combination.

This study has enabled us to identify impact of all types of dependency structures, both positive and negative, on the equity requirement. The result is that a higher equity requirement is felt when we assume dependence between claim expenses. However, if the dependency structure of the couple is negative, it appears that the capital requirement in the dependent case is lower than in the independent case. This phenomenon is part of the phenomenon of compensation of the capital requirement between multiple risks of various dependencies (the famous "diversification" effect).

The study of multiple risks in the portfolio, using the product of bivariate copulas (PBC) method, made it possible to analyse the dependencies of the risk pairs in a multivariate framework. It appears that, mainly due to the presence of independent marginal pairs, there is less dependence within the different risk pairs of the multivariate copula compared to the dependencies measured by the parametric bivariate copulas. Nevertheless, the presence of asymmetry in the tails of the distribution remains in the multirisk framework, which should be of interest to the insurer in his final choice of risk modelling (modelling by Gaussian multivariate copula or by independent copula would imply the absence of tail dependencies). The impact of dependencies on capital requirements is less significant in the multi-risk framework, but there is a slight additional capital requirement in the dependent case, largely due to the presence of tail dependencies.

The assumption of independence between risks is therefore questionable here given the diverse nature of dependencies between risks and the large number of different risks in our database (7 types of claims). However, the impact of the dependencies on the capital requirement for the "Property Damage/Property loss" pair may be similar to the impact of the dependencies on the capital requirement for the TSA base as a whole, given the large number of claim expenses for the "Property Damage/Property loss" pair compared to the number of claims of the other possible pairs.

Remerciements

Je souhaite avant tout remercier mon tuteur d'entreprise Pierre MIEHE pour sa disponibilité et ses conseils avisés. Il a su m'accorder du temps que je sais très précieux pour lui.

Je remercie également le corps enseignant de l'ISFA qui m'a permis d'avoir la possibilité de rédiger un mémoire d'actuaire, en particulier mon tuteur pédagogique Stéphane LOISEL pour le suivi du mémoire.

Je souhaite remercier chaleureusement l'ensemble de l'équipe Milliman Mind pour son soutien lors de la rédaction de mon mémoire, en particulier Son Tuyen NGUYEN et Markiyan BOMK pour leurs aides précieuses. Leur accompagnement m'a permis d'avancer plus sereinement dans la confection de ce mémoire.

Un grand merci également à Frédéric PLANCHET pour sa relecture critique.

Enfin, un grand merci à mes amis proches, et surtout ma famille pour leur soutien inconditionnel tout au long de la confection de ce mémoire, merci pour les nombreuses relectures, merci de m'avoir encouragé et motivé durant cette longue épreuve qu'est la rédaction d'un mémoire.

Table des matières

Résumé	2
Note de synthèse	5
Summary note	17
Remerciements	27
Introduction générale	37
1 Théorie des copules	39
1.1 Définitions et concepts fondamentaux	39
1.1.1 Définition d'une copule	39
1.1.2 Théorème de Sklar	40
1.1.3 Propriété d'invariance de la copule	40
1.2 Construction d'une copule	41
1.3 Les copules usuelles	41
1.3.1 Copule d'indépendance	41
1.3.2 Copule gaussienne	41
1.3.3 Copule de Student	42
1.3.4 Copules archimédiennes	44
1.3.4.1 Copule de Frank	46
1.3.4.2 Copule de Clayton	48
1.3.4.3 Copule de Clayton inverse	49
1.3.4.4 Copule de Gumbel	51
1.3.4.5 Copule AMH (Ali, Mikhail et Haq)	52
1.3.5 Copules empiriques	54
2 Présentation des différentes mesures de dépendance	57
2.1 Les mesures de concordance	57
2.1.1 Ordre et mesure de concordance	57
2.1.2 Tau de Kendall	58
2.1.3 Rho de Spearman	59
2.1.4 Remarques	60
2.1.5 Récapitulatif des tau de Kendall aux copules archimédiennes	60
2.2 Dépendance de queues	60
2.2.1 Probabilités avec la forme analytique des fonctions copules	61
2.2.2 Caractérisation de la dépendance des queues	61
2.2.3 Exemples de calculs de dépendances de queues	63
2.3 Estimation de la dépendance de queues	65
2.3.1 Approche paramétrique	65
2.3.2 Approche non-paramétrique	66

3	Méthodes de simulation et inférence statistique des copules	67
3.1	Simulations de copule	67
3.1.1	Copules archimédiennes	67
3.1.2	Méthode de distribution	68
3.2	Inférence statistique des copules	69
3.2.1	Méthode des moments	69
3.2.2	Méthode du maximum de vraisemblance	70
3.2.3	Méthode Inference Functions for Margins (IFM)	70
3.3	Choix de la bonne copule	71
3.3.1	La fonction $K(z)$	71
3.3.2	Test du χ^2	72
4	Présentation, gestion et analyse de la base de données	75
4.1	Présentation de la base et mise en contexte	75
4.1.1	Description des variables disponibles	75
4.1.2	Présentation des types de sinistre	76
4.2	Analyse des données et transformation de la base	77
4.2.1	Analyse du nombre de sinistres refusés	77
4.2.2	Analyse du nombre de sinistres déclarés	78
4.2.3	Traitement des montants de sinistres clos "as if" 2014	81
4.2.4	Traitement des dates d'incidence, agrégation des sinistres clos et transformation de la base à étudier	81
4.3	Étude de corrélations entre types de sinistre et recherche de dépendance	82
4.3.1	Analyse des données finales	82
4.3.2	Recherche de dépendance	83
4.4	Conclusions sur les dépendances observées	95
5	Estimation de la copule empirique et modélisation de dépendances	97
5.1	Représentation de la copule empirique	97
5.1.1	Densité empirique	97
5.1.2	Copule empirique	100
5.2	Ajustement lois marginales	102
5.2.1	Test d'ajustement des lois marginales	109
5.2.2	Calcul de la marge estimée non-paramétrique	109
5.3	Estimation du paramètre de copule par maximum de pseudo-vraisemblance	110
5.4	Coefficients de dépendance des queues et estimation empirique de dépendance des queues	111
5.5	Test d'ajustement statistique et graphique menant au choix de la bonne copule	112
5.5.1	Test d'ajustement du χ^2	112
5.5.2	Ajustement graphique par la fonction $K(z)$	113
5.6	Approche multirisque par produit des copules bivariées	117
5.6.1	Produit de copules transformées - cas théorique	117
5.6.2	Produit de copules transformées revisité	117
5.6.3	Application numérique et étude des dépendances	118
5.7	Modélisation du besoin en fonds propres	120
5.7.1	Mise en contexte et hypothèses	120
5.7.2	Les différentes mesures de risque appliquées au modèle : VaR et TVaR	120
5.7.3	Simulations suivant l'indépendance ou la dépendance entre les risques	121
5.7.4	Impact des dépendances sur le besoin en fonds propres	122
5.8	Conclusion sur l'impact des dépendances modélisées	124
	Conclusion	127

Annexes	131
Annexe 1	131
Annexe 2	133
Bibliographie	134

Tables des figures

1.1	Tirage aléatoire de 2 lois uniformes liées par une copule Gaussienne, de coefficient de corrélation 0,75	43
1.2	Tirage aléatoire de 2 lois uniformes liées par une copule de Student, de coefficient de corrélation $\rho=0,75$ et de degrés de liberté $\nu=4$	44
1.3	Tirage aléatoire de 2 lois uniformes liées par une copule de Frank, de paramètre $\theta=4$	46
1.4	Graphique de la fonction de répartition de la copule de Frank	47
1.5	Contour copule de Frank	47
1.6	Tirage aléatoire de 2 lois uniformes liés à une copule de Clayton de paramètre $a=3$	48
1.7	Fonction de répartition de la copule de Clayton	49
1.8	Contour de la copule de Clayton	49
1.9	Tirage aléatoire de 2 lois uniformes liées à une copule de Clayton inverse, de paramètre $a=3$	50
1.10	Tirage aléatoire de 2 lois uniformes liées à une copule de Gumbel de paramètre $a=3$	51
1.11	Fonction de répartition de la copule de Gumbel	52
1.12	Contour de la copule de Gumbel	52
1.13	Tirage aléatoire de 2 lois uniformes liées à une copule AMH, de paramètre $\theta = 0,7$	53
1.14	Fonction de répartition de la copule AMH	54
1.15	Contour de la copule AMH	54
4.1	Évolution du nombre de sinistres selon l'année d'incidence	78
4.2	Évolution du nombre de sinistres clos non-nuls dans la catégorie "perte de biens".	79
4.3	Évolution du nombre de sinistres clos non-nuls dans la catégorie "dommages matériels"	79
4.4	Évolution du nombre de sinistres clos non-nuls dans la catégorie "dommages corporels"	80
4.5	Évolution du nombre de sinistres clos non-nuls dans la catégorie "véhicule motorisé"	80
4.6	Dépendogramme du couple perte de biens /dommages aux biens	86
4.7	Dépendogramme du couple perte de biens/dommages corporels	87
4.8	Dépendogramme du couple dommages aux biens/dommages corporels	88
4.9	Dépendogramme du couple perte employé/perte de biens	88
4.10	Dépendogramme du couple perte de biens/véhicule motorisé	89
4.11	Dépendogramme du couple dommages aux biens/perte employé	89
4.12	Dépendogramme du couple dommages aux biens/véhicule motorisé	90
4.13	K-Plot du couple perte de biens/dommages aux biens	91
4.14	K-Plot du couple perte de biens/dommages corporels	91
4.15	K-Plot du couple dommages aux biens/dommages corporels	92
4.16	K-Plot du couple perte employé/perte de biens	92
4.17	K-Plot du couple perte de biens/véhicule motorisé	93
4.18	K-Plot du couple dommages aux biens/perte employé	93
4.19	K-Plot du couple dommages aux biens/véhicule motorisé	94

5.1	Histogramme 3D du couple perte de biens/dommages aux biens	98
5.2	Histogramme 3D du couple perte de biens/dommages corporels	98
5.3	Histogramme 3D du couple dommages aux biens/véhicule motorisé	99
5.4	Histogramme 3D du couple dommages aux biens/perte employé	99
5.5	Copule empirique du couple Dommages aux biens/Perte de biens	100
5.6	Copule empirique du couple Perte de biens/Dommages corporels	101
5.7	Copule empirique du couple Dommages aux biens/Véhicule motorisé	101
5.8	Copule empirique du couple Dommages aux biens/Perte employé	102
5.9	Densité estimée type "Dommages aux biens"	103
5.10	Densité estimée type "Perte de biens"	103
5.11	Densité estimée type "Véhicule motorisé"	104
5.12	Densité estimée type "Dommages corporels"	104
5.13	Densité estimée type "Perte employé"	105
5.14	Densité ajustée type "Perte de biens"	106
5.15	Densité ajustée type "Dommages aux biens"	107
5.16	Densité ajustée type "Dommages corporels"	107
5.17	Densité ajustée type "Véhicule motorisé"	108
5.18	Densité ajustée type "Perte employé"	108
5.19	Fonction $K(z)$ empirique pour le couple "Dommages/Perte de biens	114
5.20	Fonction $K(z)$ représentée pour chaque copule calibrée en fonction de $K(z)$ empirique (en rouge) pour le couple "Dommage/Perte de biens"	114
5.21	Fonction $K(z)$ représentée pour chaque copule calibrée en fonction de $K(z)$ empirique (en rouge) pour le couple "Perte de biens/Dommages corporels"	115
5.22	Fonction $K(z)$ représentée pour chaque copule calibrée en fonction de $K(z)$ empirique (en rouge) pour le couple "Dommages aux biens/Véhicule motorisé"	115
5.23	Fonction $K(z)$ représentée pour chaque copule calibrée en fonction de $K(z)$ empirique (en rouge) pour le couple "Dommages aux biens/Perte employé"	116

Liste des tableaux

1	Tau de Kendall	8
2	Résultats du test de Kolmogorov Smirnov	9
3	Paramètre de la copule selon la méthode CML , paramètre selon la méthode des moments entre parenthèses	9
4	Coefficients de dépendance de queues inférieur et supérieur pour chaque copule paramétrique calibrée au jeu de données	10
5	Meilleures copules selon test du χ^2 pour chaque couple de sinistre	10
6	Mesure des dépendances 2 à 2 dans le cadre multivarié et comparaison avec le cadre bivarié	12
7	Mesure des dépendances de queue de chaque couple au sein de la copule PBC multivariée	13
8	Comparaison des mesures de risque en euros pour le couple Dommages aux biens/Perte de biens dans le cas dépendant et indépendant	14
9	Comparaison des mesures de risque en euros pour le couple Perte de biens/Dommages corporels dans le cas dépendant et indépendant	14
10	Comparaison des mesures de risque en euros pour le couple Dommages aux biens/Véhicule motorisé dans le cas dépendant et indépendant	14
11	Comparaison des mesures de risque en euros pour le couple Dommages aux biens/Perte employé dans le cas dépendant et indépendant	14
12	$VaR_{75\%}$ dans le cadre multirisque	15
13	Mesures par TVaR dans le cadre multirisque	15
14	Kendall's tau	20
15	Results of the Kolmogorov Smirnov test	20
16	Copula parameter with CML method, parameter according to the moments method in brackets.	21
17	Lower and upper tail dependency coefficients for each parametric copula calibrated to the dataset	21
18	Best copulas according to χ^2 goodness of fit test for each pair of claims	22
19	Measurement of pairwise dependencies in the multivariate framework and comparison with the bivariate framework	24
20	Measuring the tail dependencies of each pair within the multivariate PBC copulae	24
21	Comparison of risk measures in euros for the Property Damage/Loss pair in the dependent and independent case	25
22	Comparison of risk measures in euros for the property loss/personal injury pair in the dependent and independent case	25
23	Comparison of risk measures in euros for the Property Damage/Vehicle pair in the dependent and independent case	25
24	Comparaison des mesures de risque en euros pour le couple Dommages aux biens/Perte employé dans le cas dépendant et indépendant	26
2.1	Tableau récapitulatif des générateurs de copules archimédiennes et de leurs tau de Kendall	60

2.2	Dépendance de queues pour les différentes copules usuelles	65
3.1	Copules archimédiennes avec leurs paramètres, leurs générateurs ainsi que la transformée de Laplace-Stieltjes inversée	68
3.2	Copules archimédiennes avec leurs paramètres, leurs générateurs ainsi que la fonction $K(z)$ associée	71
4.1	Nombre de sinistres par type de sinistre	77
4.2	Nombre de sinistres refusés par type de sinistre	78
4.3	Coefficient de corrélation de Pearson	83
4.4	Tau de Kendall	84
4.5	Rho de Spearman	84
4.6	P-value du test de corrélation de Pearson	85
4.7	P-value du test de corrélation de Kendall	85
4.8	P-value du test de corrélation de Spearman	85
5.1	Maximum de log-vraisemblance selon les lois testées	105
5.2	Résultats du test de Kolmogorov Smirnov	109
5.3	Paramètre de la copule selon la méthode CML, paramètre selon la méthode des moments entre parenthèses	110
5.4	Coefficients de dépendance de queues inférieur et supérieur pour chaque copule paramétrique calibrée au jeu de données	111
5.5	Estimation empirique des dépendances de queue inférieur et supérieur par la méthode de la copule de queue empirique	111
5.6	Effectifs empiriques du couple "Dommages aux biens/Perte employé"	113
5.7	Effectifs théoriques de la copule de Student de paramètre 0,0548 et ddl=4 pour le couple "Dommages aux biens/Perte employé"	113
5.8	Résultats du test pour le couple "Dommages aux biens/Perte employé"	113
5.9	Meilleures copules selon test du χ^2 pour chaque couple de sinistres	113
5.10	Mesure des dépendances 2 à 2 dans le cadre multivarié et comparaison avec le cadre bivarié	119
5.11	Mesure des dépendances de queue de chaque couple au sein de la copule PBC multivariée	119
5.12	Comparaison des mesures de risque en euros pour le couple Dommages aux biens/Perte de biens dans le cas dépendant et indépendant	122
5.13	Comparaison des mesures de risque en euros pour le couple Perte de biens/Dommages corporels dans le cas dépendant et indépendant	122
5.14	Comparaison des mesures de risque en euros pour le couple Dommages aux biens/Véhicule motorisé dans le cas dépendant et indépendant	123
5.15	Comparaison des mesures de risque en euros pour le couple Dommages aux biens/Perte employé dans le cas dépendant et indépendant	123
5.16	$VaR_{75\%}$ dans le cadre multirisque	124
5.17	Mesures par TVaR dans le cadre multirisque	124

Introduction générale

Afin de calculer les besoins en fonds propres d'un assureur sur un portefeuille multirisques, l'assureur pose souvent comme hypothèse que les risques constituant son portefeuille sont indépendants ou que, s'il y a des dépendances, les risques appartiennent à la même distribution. Dans un contexte où la sinistralité s'est accrue, où les catastrophes liées à l'environnement se réalisent de plus en plus fréquemment, les dépendances entre branches de risque sont amenées à augmenter. Le régulateur, les agences de notation ainsi que l'assuré réclament des fonds propres couvrant avec assez de marge la sinistralité, tandis que les intérêts des investisseurs seraient plutôt enclins à investir le moins possible dans le capital de l'assurance pour une rentabilité maximale. Il faut donc pouvoir modéliser correctement les dépendances entre risques afin d'optimiser ce besoin en fonds propres. Nous verrons tout au long de ce mémoire qu'une modélisation plus réaliste des liaisons entre risques est possible grâce à la théorie des copules.

L'objectif de ce mémoire est d'analyser et de modéliser les dépendances de plusieurs risques d'un portefeuille non-vie à l'aide de copules pour pouvoir estimer le niveau en fonds propres requis pour le portefeuille. Les dépendances aussi bien positives que négatives seront modélisées de façon réaliste à l'aide des copules paramétriques calibrées en fonction. Le but de l'étude sera de mesurer l'impact des dépendances entre les risques d'un portefeuille sur le besoin en fonds propres, et ainsi discuter de la pertinence de l'hypothèse d'indépendance entre les risques généralement admise en assurance non-vie.

Le mot "copula", introduit par le mathématicien Abe Sklar en 1959, désigne le lien ou une attache en latin. Il s'agit d'une fonction pouvant lier des fonctions marginales avec leur distribution multivariée. Une des propriétés fortes de la copule est qu'elle permet de distinguer les dépendances de la distribution jointe du comportement de lois marginales, ce qui évite bien des hypothèses fortes sur le comportement des marges dans l'étude des dépendances.

Les données sur lesquelles sera basée l'application de notre étude représentent un portefeuille de sinistres survenus entre 2002 et 2015 et recensés par la TSA (Transportation Security Administration), agence américaine chargée de la sûreté aéroportuaire dans la plupart des aéroports des Etats-Unis.

Les bases théoriques exposées dans ce mémoire nous permettront d'appliquer un processus bien précis nous menant vers des études de dépendances et plus particulièrement de dépendances extrêmes les plus précises possibles. En effet, il s'agira dans un premier temps d'établir une copule empirique servant de référence car se rapprochant le plus possible des données. Puis de calibrer des copules paramétriques en estimant leur paramètre de copule. Une première étude de la dépendance globale comme de queue peut alors se faire à travers des mesures que l'on exposera. Enfin, il sera nécessaire de tester l'ajustement de ces différentes copules paramétriques par rapport à la copule empirique. Ces tests permettront à terme un choix optimal de copule paramétrique avec laquelle les mesures de dépendance représenteront idéalement les dépendances entre nos branches de risque. L'évaluation du besoin en fonds propres sera réalisée avec les mesures de risque utilisées en assurance et en finance (VaR et TVaR) sur la base des charges de sinistre simulées par copule paramétrique préalablement choisie.

En premier lieu, nous aborderons dans ce mémoire l'ensemble de la théorie des copules, qui

comprend notamment un théorème fondamental pour l'application de copules, le théorème de Sklar. Nous poserons les bases théoriques de construction d'une copule. Les copules usuelles implémentées sous Milliman Mind, plateforme web de Milliman permettant d'améliorer la performance de modèles actuariels, seront présentées avec un bref descriptif des dépendances de queue sur chacune de ces copules.

Dans un second temps, nous nous pencherons plus en détail sur les différentes mesures de dépendance, que ce soit des mesures globales de dépendance telles que le tau de Kendall, le coefficient de Pearson ou le rho de Spearman, que des mesures de dépendance de queue (Tail Dependence Coefficient en anglais). Toutes ces mesures ont la particularité de pouvoir s'exprimer en fonction de la copule qui leur est associée. Un choix de copule adéquat et bien assimilé à notre jeu de données est crucial dans la pertinence de ces mesures de dépendance et dans la bonne représentation des dépendances d'un portefeuille de risques donné.

Nous poursuivrons les principes fondamentaux de l'étude par la présentation des différentes méthodes de simulation de copules qui sont utilisées sous Milliman Mind. Ces méthodes permettent de construire des copules à partir d'un vecteur aléatoire donné et d'un paramètre de copule. Certaines telles que la méthode des distributions sont plus appropriées pour les copules dites elliptiques. Nous analyserons également les différentes méthodes d'estimation du paramètre de copule telles que la méthode du maximum de vraisemblance, la méthode IFM ou encore la méthode des moments. Une fois l'estimateur du paramètre de la copule obtenu, nous avons une copule paramétrique théorique. Pour pouvoir valider cette copule théorique comme celle adéquate au jeu de données, il faut pouvoir établir des tests permettant de comparer ces copules théoriques à la copule empirique de référence. Nous proposerons donc deux méthodes permettant de choisir la copule paramétrique idoine aux données d'entrée : la méthode utilisant la fonction $K(z)$ qui permet de comparer graphiquement la fonction K empirique de la théorique, et la méthode par test du X^2 qui nécessite au préalable une répartition en classes des valeurs empiriques.

Le second volet de ce mémoire sera consacré à l'application de ces principes théoriques exposés lors du premier volet. Dans un premier temps, il s'agira d'exposer le traitement des données initiales extraites de la base de la TSA, et notamment le travail d'agrégation des charges de sinistre selon les dates et les types de sinistre. Puis une analyse des corrélations entre les types de sinistre sera réalisée afin de se donner un aperçu des couples de risque pouvant présenter des dépendances.

La modélisation des dépendances sera étudiée en premier lieu par le calcul de la copule empirique et l'étude des lois marginales. Le travail se poursuivra ensuite par l'estimation des paramètres de copule par approche semi-paramétrique puis de tests d'ajustement afin d'établir un bon choix de copule.

Une modélisation des copules en grande dimension sera proposée pour des risques associés aux couples modélisés par des copules bivariées. Cette modélisation permettra de se faire une représentation plus fidèle des dépendances dans l'ensemble du portefeuille de données.

Enfin, la modélisation du besoin en fonds propres sera entreprise par l'intermédiaire de simulations basées sur les copules choisies. Il s'agira notamment de mesurer l'erreur commise sur le niveau en fonds propres lorsque l'hypothèse d'indépendance entre les risques est établie.

Chapitre 1

Théorie des copules

Les fonctions copules sont des outils mathématiques qui permettent de pallier les insuffisances du cadre d'analyse gaussien des structures de dépendances. Elles permettent notamment d'isoler la structure de dépendance d'une loi jointe vis-à-vis des lois marginales. Elles sont un outil très puissant pour la modélisation de dépendances extrêmes et peuvent être plus adaptées à la gestion de risques financiers. Les copules constituent un outil très intéressant en vue d'une étude de dépendance entre plusieurs branches de risques assurantiels qui semblent dépendantes, et contribuent à l'établissement de provisions plus justes pour couvrir ces risques.

1.1 Définitions et concepts fondamentaux

1.1.1 Définition d'une copule

Une copule C de dimension n est une fonction de répartition de dimension n définie de $[0, 1]^n$ vers $[0, 1]$. Elle peut s'écrire de la forme suivante, $\forall (u_1, \dots, u_n) \in [0, 1]^n$:

$$C(u_1, \dots, u_n) = P(U_1 \leq u_1, U_2 \leq u_2, \dots, U_n \leq u_n), U_1, U_2, \dots, U_n \sim U([0; 1]) \quad (1.1)$$

A travers cette définition, on peut s'apercevoir que la fonction de répartition jointe caractérise la dépendance entre les variables aléatoires. Cependant, la copule C va permettre également de distinguer le comportement de la fonction de répartition marginale dans cette structure de dépendance (voir théorème de Sklar).

Plusieurs propriétés caractérisent en outre les copules.

Une copule de dimension n est une fonction $C : [0, 1]^n \rightarrow [0, 1]$ telle que :

1. $C(1, \dots, 1, u_i, 1, \dots, 1) = u_i, \forall i \in 1, \dots, n, u_i \in [0, 1]$;
2. $C(u_1, \dots, u_{i-1}, 0, u_{i+1}, \dots, u_n) = 0, \forall i \in 1, \dots, n, u_i \in [0, 1]$;
3. C est n -croissante :

$$\sum_{u_k \in a_k, b_k, 1 \leq k \leq n} (-1)^{\text{Card}\{k; u_k = a_k\}} C(u_1, \dots, u_n) \geq 0$$

$\forall a = (a_1, \dots, a_n), b = (b_1, \dots, b_n) \in [0, 1]^n$ tels que $a \leq b$.

Les conditions 1 et 2 impliquent que les marges sont uniformément réparties sur $[0, 1]$. La condition 3 est équivalente au fait que C soit croissante par marge et qu'elle vérifie une propriété de Lipschitz qui garantit sa continuité.

1.1.2 Théorème de Sklar

Le théorème de Sklar est fondamental dans la théorie des copules. En effet, il permet d'établir le lien entre une fonction multivariée et ses lois marginales. Il constitue également une condition suffisante à l'existence des copules.

Soit F une fonction de répartition d'un vecteur $(X_1, X_2, X_3, \dots, X_n)$ dont les fonctions de répartition marginales sont de la forme $F_i, i = 1, \dots, n$. Il existe alors une unique copule C telle que

$$F(x_1, x_2, \dots, x_n) = C(F_1(x_1), F_2(x_2), \dots, F_n(x_n)) \quad (1.2)$$

Inversement, étant donnée la copule C et ses marginales $F_i, i=1, 2, \dots, n$, $C(F_1(x_1), F_2(x_2), \dots, F_n(x_n))$ définit une fonction de répartition multivariée de dimension n .

Si les fonctions de répartition marginales sont inversibles, on peut démontrer le théorème de la façon suivante :

$$\begin{aligned} F(x_1, \dots, x_n) &= P[X_1 \leq x_1, \dots, X_n \leq x_n] \\ &= P[F_1(X_1) \leq F_1(x_1), \dots, F_n(X_n) \leq F_n(x_n)] \\ &= P[U_1 \leq F_1(x_1), \dots, U_n \leq F_n(x_n)], U_1, \dots, U_n \sim U([0, 1]) \\ &= C(F_1(x_1), \dots, F_n(x_n)) \end{aligned}$$

avec $F_1(X_1), \dots, F_n(X_n) \sim U([0, 1])$, car, pour F_i inversible :

$$P[F_i(X_i) \leq u] = P[X_i \leq F_i^{-1}(u)] = F_i(F_i^{-1}(u)) = u, \text{ pour } u \in [0, 1].$$

Ce théorème est par ailleurs particulièrement utile pour construire des fonctions copules à partir de distributions multivariées. Si les lois marginales sont continues, alors :

$$C(u_1, \dots, u_n) = F(F_{X_1}^{-1}(u_1), \dots, F_{X_n}^{-1}(u_n))$$

En outre, si la distribution multivariée est absolument continue, alors elle admet une densité :

$$f(x_1, \dots, x_n) = f_{X_1}(x_1) \times \dots \times f_{X_n}(x_n) \times c(F_{X_1}(x_1), \dots, F_{X_n}(x_n))$$

où c désigne la densité de la copule C définie de la façon suivante :

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

On peut caractériser également la densité de la copule à partir de la densité multivariée et des densités jointes du vecteur aléatoire considéré :

$$c(u_1, \dots, u_n) = \frac{f(F_{X_1}^{-1}(u_1), \dots, F_{X_n}^{-1}(u_n))}{f_{X_1}(F_{X_1}^{-1}(u_1)) \times \dots \times f_{X_n}(F_{X_n}^{-1}(u_n))}$$

1.1.3 Propriété d'invariance de la copule

Les copules sont invariantes, c'est-à-dire que si ϕ_1 et ϕ_2 sont strictement croissantes alors la copule du couple (X_1, X_2) est identique à la copule du couple $(\phi_1(X_1), \phi_2(X_2))$.

Démonstration :

Soient F_i la fonction de répartition X_i et ϕ_i^{-1} la fonction inverse généralisée de ϕ_i . On a

$P[\phi_i(X_i) \leq x_i] = P[X_i \leq \phi_i^{-1}(x_i)] = F_i \circ \phi_i^{-1}(x_i)$, donc $G_i = F_i \circ \phi_i^{-1}$ est la fonction de répartition de $\phi_i(X_i)$ et par définition $C(u, v) = F_{X_1, X_2}(F_1^{-1}(u), F_2^{-1}(v)) = P[X_1 \leq F_1^{-1}(u), X_2 \leq F_2^{-1}(v)] = P[\phi_1(X_1) \leq \phi_1(F_1^{-1}(u)), \phi_2(X_2) \leq \phi_2(F_2^{-1}(v))] = P[\phi_1(X_1) \leq G_1^{-1}(u), \phi_2(X_2) \leq G_2^{-1}(v)]$.

Dans le cadre de simulation de copules, cette propriété est utilisée avec ϕ_i correspondant à des fonctions de répartition.

1.2 Construction d'une copule

Comme vu précédemment, le théorème de Sklar suppose la continuité des marges. L'unicité liée à la restriction des marginales continues limite l'utilisation des copules à ce cas et exclut de ce fait le cas des variables aléatoires discrètes.

Prenons par exemple la copule logistique de Gumbel. Soit la fonction bivariée suivante :

$$F(x_1, x_2) = (1 + e^{-x_1} + e^{-x_2})^{-1}.$$

On remarquera que cette copule est symétrique. Cette propriété est, comme vu par la suite, caractéristique des copules archimédiennes.

On a alors : $F_{X_1}(x_1) = F_X(x_1, \infty) = (1 + e^{x_1})^{-1}$ et $F_{X_2}(x_2) = F_X(\infty, x_2) = (1 + e^{x_2})^{-1}$, donc on peut en déduire que $F_{X_1}^{-1}(u) = \ln \frac{u}{1-u}$ et $F_{X_2}^{-1}(v) = \ln \frac{v}{1-v}$.

Donc on peut finalement écrire la copule de Gumbel :

$$C(u, v) = F_X(F_{X_1}^{-1}(u), F_{X_2}^{-1}(v)) = [1 + \frac{1-u}{u} + \frac{1-v}{v}]^{-1} = \frac{uv}{u+v-uv}.$$

1.3 Les copules usuelles

Le théorème de Sklar nous permet de construire des fonctions de répartition multivariées à partir de fonction de répartition marginales données.

Dans cette partie, nous allons présenter les principaux types de copules fréquemment utilisées dans la pratique. Ces copules sont notamment présentes à la modélisation sur la plateforme web qu'est Milliman Mind. Certaines possèdent des propriétés particulières que nous allons pouvoir exposer.

1.3.1 Copule d'indépendance

Soit $X = (X_1, \dots, X_d)$ un vecteur aléatoire. Les composantes de ce vecteur sont indépendantes si et seulement si $\mathbb{F}(x_1, \dots, x_d) = F_1(x_1) \times \dots \times F_d(x_d)$. La copule d'indépendance est donc définie par :

$$\Pi(u) = u_1 \times \dots \times u_d.$$

1.3.2 Copule gaussienne

La copule gaussienne est définie à partir de familles de lois gaussiennes. La copule gaussienne multivariée est définie telle que :

$$C_{\Sigma}^{Ga} = \phi_{\Sigma}(\phi^{-1}(u_1), \phi^{-1}(u_2), \dots, \phi^{-1}(u_n)) \quad (1.3)$$

où ϕ_Σ est la fonction de répartition de la loi multivariée $\mathcal{N}(0_{\mathbb{R}^n}, \Sigma)$ et ϕ la fonction de répartition des lois normales suivant $\mathcal{N}(0, 1)$.

Dans le cas bivarié, si le coefficient de corrélation ρ est associé à la matrice de covariance Σ , ϕ_ρ s'écrit également :

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

(1.4)

Ce résultat nous permet de déduire l'expression de la copule gaussienne bivariée, en nous appuyant du théorème de Sklar :

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

Avec Z variable univariée suivant une loi normale standard.

La copule gaussienne est la plus populaire des copules ainsi que la plus facile à implémenter, puisque le seul paramètre à estimer est la matrice de corrélation linéaire. La dépendance gaussienne est entièrement déterminée par cette matrice. Il est également possible d'estimer le coefficient de corrélation de rang ou encore le coefficient de corrélation linéaire.

Cependant, un inconvénient majeur à proposer des copules gaussiennes est qu'elle sous-estime la dépendance de ses queues. En effet, le coefficient de corrélation des queues est de 0 (ce qui, comme vu en profondeur dans le chapitre précédent, désigne une absence de dépendance des queues) sauf si les variables sont parfaitement corrélées ($\rho = 1$). Une très petite corrélation est donc attendue au niveau des queues (de l'ordre d'1 événement tous les 100 voire 200 ans). Par conséquent, l'utilisation des copules gaussiennes pour mesurer des risques assurantiels au niveau des queues de distribution (des risques extrêmes) est limitée.

1.3.3 Copule de Student

Tout comme pour la copule normale, on peut définir indirectement la copule de Student (ou t-Copula) par l'intermédiaire du théorème de Sklar. Un vecteur aléatoire suivant la loi de Student se définit par des lois marginales de loi de Student liées à une copule de Student. Ainsi, on peut définir la copule de Student de la façon suivante :

Soient $T_{\Sigma, \nu}$ et T_ν respectivement les fonctions de répartition des lois $T(\Sigma, \nu)$ et $T(\nu)$. La copule de Student se définit alors par :

$$C(u_1, \dots, u_n) = T_{\Sigma, \nu}(T^{-1}(u_1), \dots, T^{-1}(u_n))$$

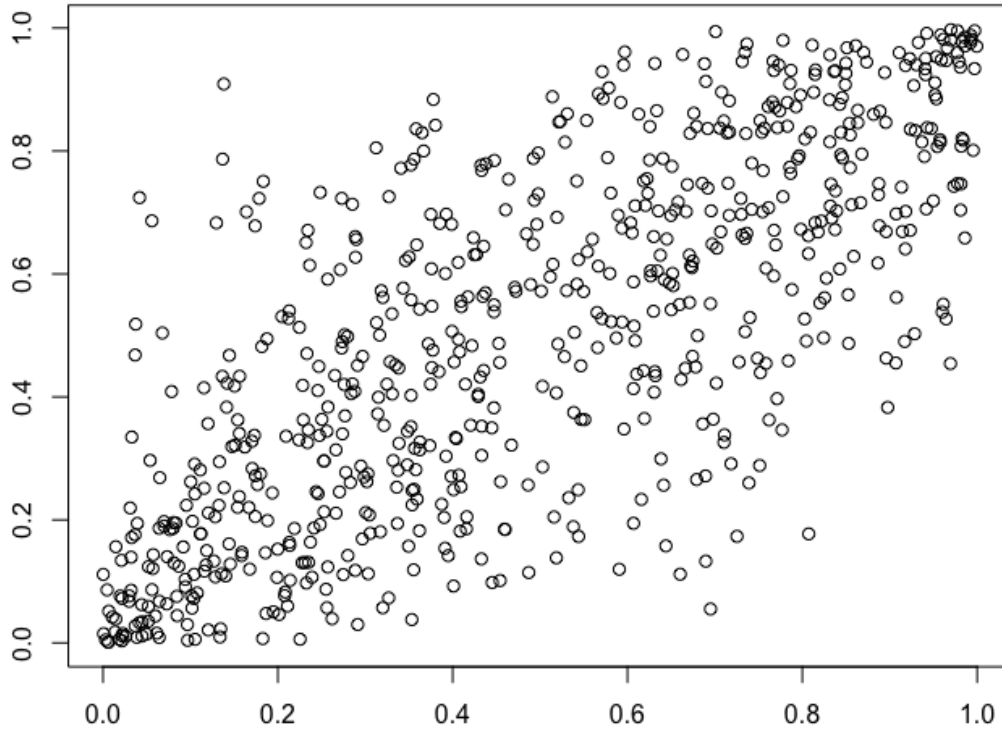


FIGURE 1.1 – Tirage aléatoire de 2 lois uniformes liées par une copule Gaussienne, de coefficient de corrélation 0,75

Dans le cas bivarié, la copule de Student, de coefficient de corrélation ρ et de degré de liberté ν est donnée par :

$$C_{\rho,\nu}(u, v) = \int_{-\infty}^{t_{\nu}^{-1}(u)} \int_{-\infty}^{t_{\nu}^{-1}(v)} \frac{\Gamma(\frac{\nu+2}{2})}{2\pi\Gamma(\frac{\nu}{2})\sqrt{1-\rho^2}} \times \left(1 + \frac{s^2 + t^2 - 2\rho st}{\nu(1-\rho^2)}\right)^{-\frac{\nu+2}{2}} ds dt$$

Où Γ est une loi Gamma et t_{ν} est la fonction de répartition de la loi de Student à ν degrés de liberté :

$$t_{\nu}(x) = \int_{-\infty}^x \frac{\Gamma(\frac{\nu+2}{2})}{\sqrt{\rho\pi}\Gamma(\frac{\nu}{2})} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} dt$$

Comme nous le verrons en détail dans le prochain chapitre, la copule de Student capte correctement les dépendances de queues à gauche et à droite. Si le degré de liberté $\nu \rightarrow \infty$, alors la copule de Student converge vers la copule gaussienne.

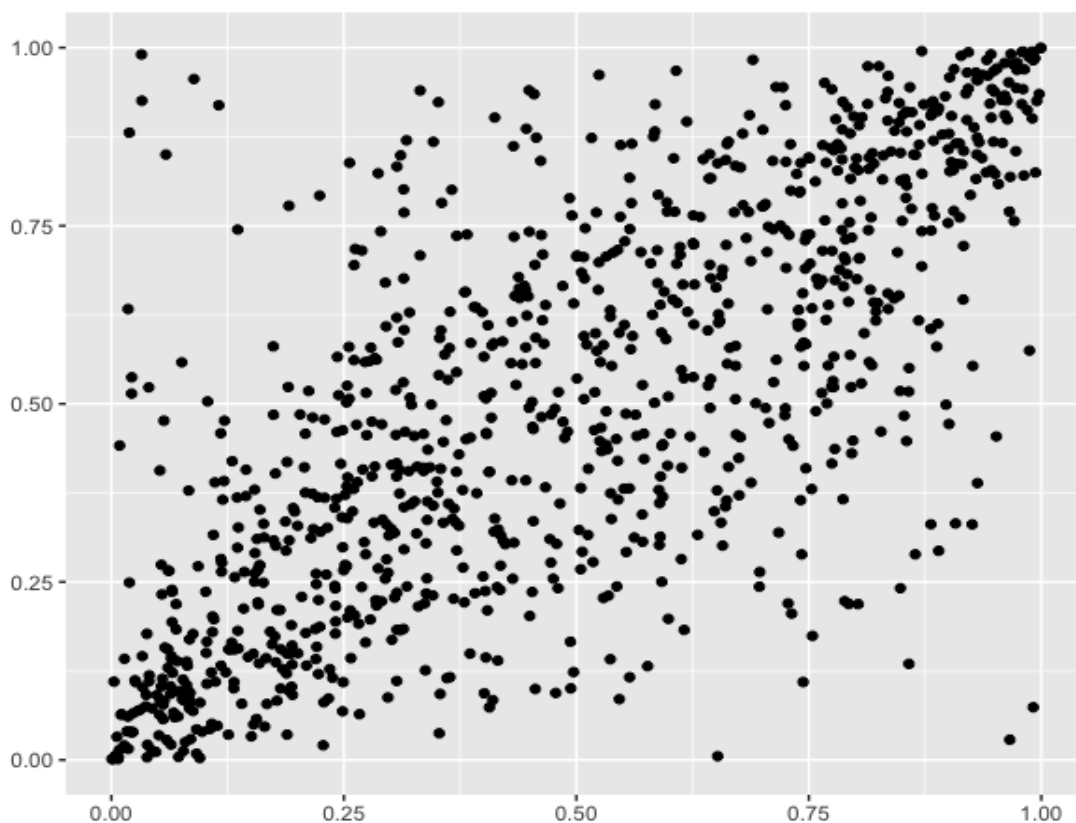


FIGURE 1.2 – Tirage aléatoire de 2 lois uniformes liées par une copule de Student, de coefficient de corrélation $\rho=0,75$ et de degrés de liberté $\nu=4$

1.3.4 Copules archimédiennes

Les copules archimédiennes ont la particularité de pouvoir décrire des structures de dépendances diverses. En effet, il est possible d'obtenir des dépendances asymétriques, où les dépendances de queues à gauche et à droite diffèrent. Les principales copules archimédiennes sont les copules de Frank, de Gumbel et de Clayton.

Cette notion de copule archimédienne a été définie par GENEST et MACKAY [1986]. Dans le cas bivarié, on peut la définir ainsi :

$$C(u, v) = \phi^{-1}(\phi(u) + \phi(v)), u, v \in [0, 1]$$

Avec ϕ un générateur strict, une fonction décroissante convexe sur $(0, 1] \rightarrow [0, \infty]$ telle que $\phi(1) = 0$ et $\phi(0) = \infty$. Le générateur ϕ est dit non-strict si $\phi(0) < \infty$.

On peut généraliser cette définition à l'ordre n comme suit :

$$C(u_1, \dots, u_n) = \phi^{-1}(\phi(u_1) + \dots + \phi(u_n))$$

C est une copule si et seulement si le générateur ϕ est une fonction strictement monotone, c'est-à-dire si :

$$(-1)^k \frac{d^k}{du^k} \phi^{-1}(u) \leq 0, \forall k \in \mathbb{N}$$

Une source importante de générateurs de copules archimédiennes est l'inversion de la transformée de Laplace des fonctions de répartition, c'est à dire de la forme : $\phi^{-1}(x) = \mathbb{E}(e^{-xS})$, $x \in [0, \infty[$ avec S une variable aléatoire non-négative. En effet, le fait que ces inverses soient des transformées de Laplace assure que C_ϕ soit bien une copule. De plus, cela facilite grandement les simulations.

Propriétés des copules archimédiennes

Voici des propriétés utiles à la construction de copules multivariées.

Si C est une copule archimédienne de générateur ϕ alors :

- C est symétrique , c'est à dire $C(u_1, u_2) = C(u_2, u_1), \forall (u_1, u_2) \in [0, 1]^2$;
- C est associative, c'est à dire $C(C(u_1, u_2), u_3) = C(u_1, C(u_2, u_3)), \forall (u_1, u_2, u_3) \in [0, 1]^3$;
- si $c > 0$ est une constante, alors $c\phi$ est aussi un générateur de C ;
- $\delta(u) = C(u, u)$ vérifie $\delta(u) < u, \forall u \in [0, 1]$

1.3.4.1 Copule de Frank

Soit le générateur ϕ défini tel que $\phi(x) = -\log\left(\frac{e^{-ax}-1}{e^{-a}-1}\right)$, avec $a \neq 0$. Alors la copule de Frank peut s'écrire ainsi :

$$C_a(u, v) = -\frac{1}{a} \log\left(1 + \frac{(e^{-au} - 1)(e^{-av} - 1)}{e^{-a} - 1}\right)$$

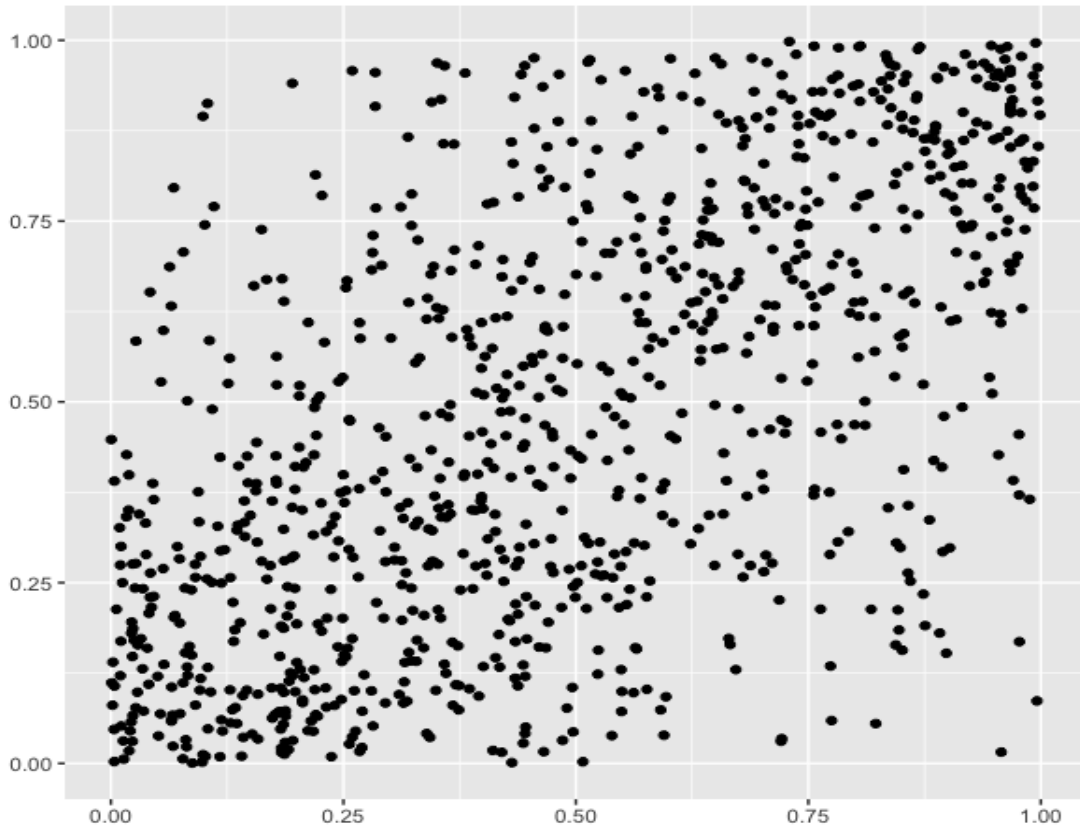


FIGURE 1.3 – Tirage aléatoire de 2 lois uniformes liées par une copule de Frank, de paramètre $\theta=4$

On peut apercevoir ici que la copule, qui est par ailleurs symétrique, présente une accumulation de points en $(0,0)$ et $(1,1)$. ce qui traduit une corrélation entre les faibles montants de sinistre et une corrélation entre les sinistres de montants importants.

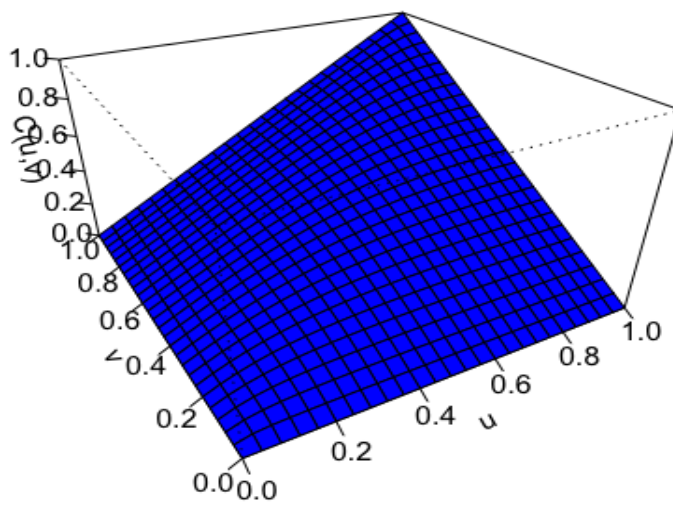


FIGURE 1.4 – Graphique de la fonction de répartition de la copule de Frank

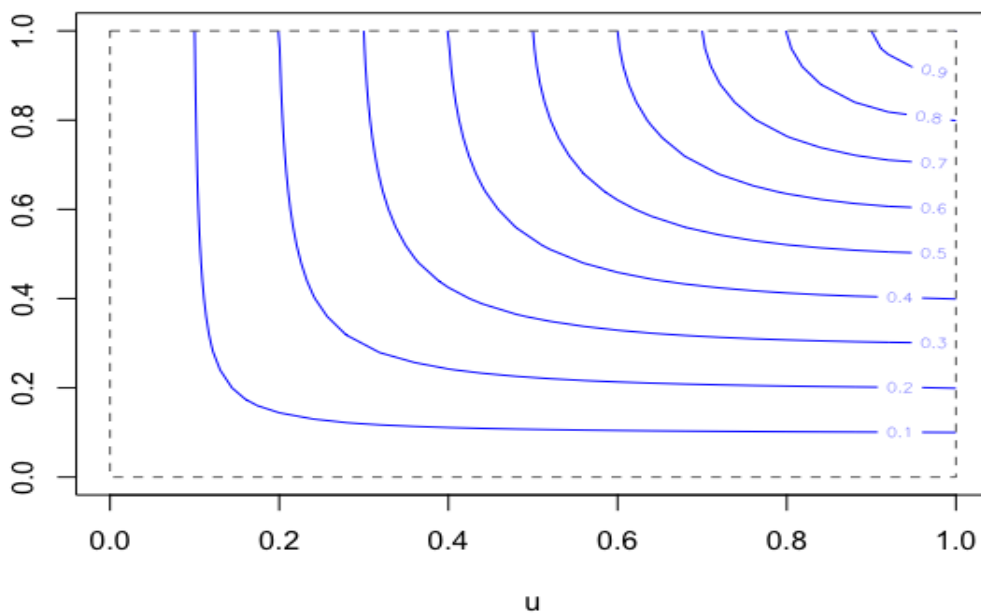


FIGURE 1.5 – Contour copule de Frank

1.3.4.2 Copule de Clayton

Le générateur de la copule de Clayton s'écrit :

$$\phi(t) = \frac{t^{-a} - 1}{a}, a \in]0, +\infty[$$

On peut donc écrire la copule de Clayton sous la forme multivariée de la façon suivante :

$$C_a(u_1, \dots, u_n) = (u_1^{-a} + \dots + u_n^{-a} - n + 1)^{-\frac{1}{a}}$$

Il existe une dépendance de queue à gauche pour la copule de Clayton. Cependant, la copule n'admet pas de dépendance de queue à droite. Nous verrons ce cas de dépendance plus en détail dans le prochain chapitre.

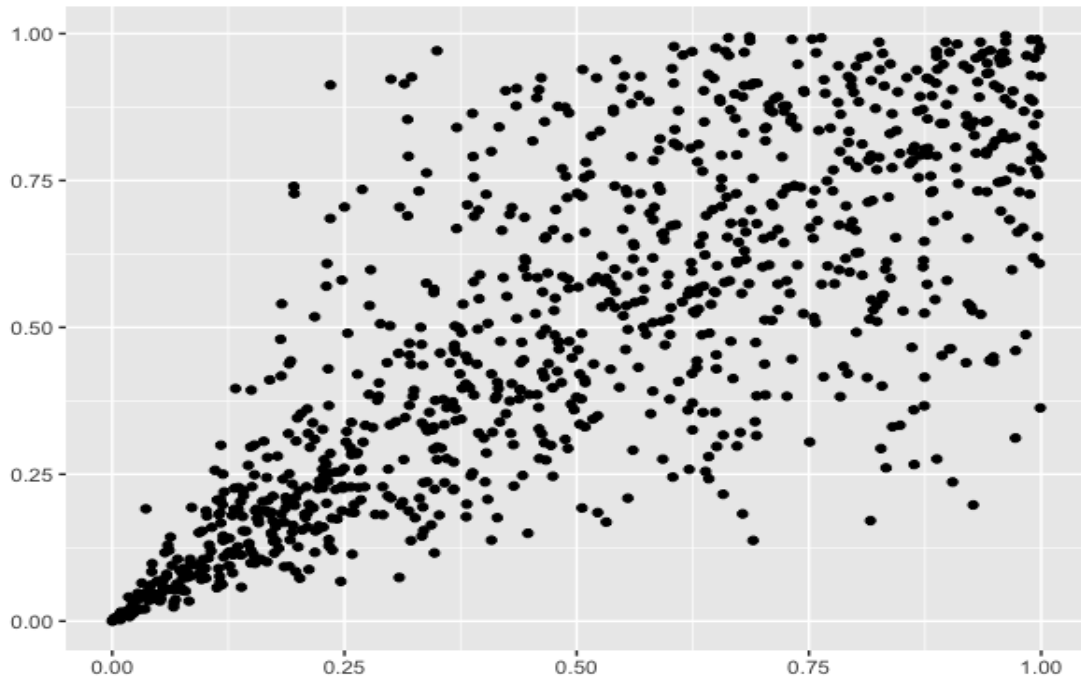


FIGURE 1.6 – Tirage aléatoire de 2 lois uniformes liés à une copule de Clayton de paramètre $a=3$

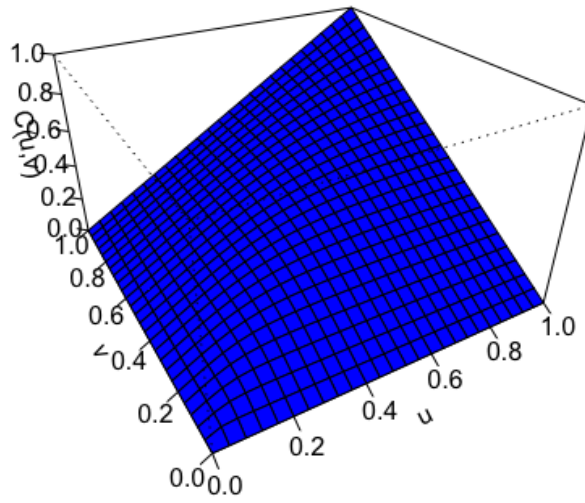


FIGURE 1.7 – Fonction de répartition de la copule de Clayton

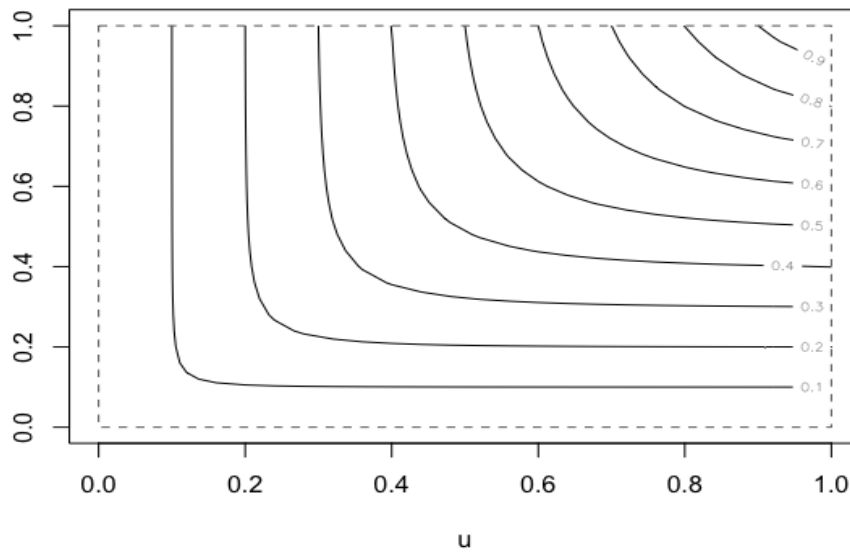


FIGURE 1.8 – Contour de la copule de Clayton

1.3.4.3 Copule de Clayton inverse

Il existe également une possibilité de représenter les dépendances de queue à droite, c'est à dire des réalisations ayant de forte valeur, en choisissant de modéliser une copule de Clayton inverse. Cette copule est notamment simulable sur Milliman Mind. Elle s'écrit de la façon suivante :

$$C'(u_1, \dots, u_n) = 1 - C(u_1, \dots, u_n) = 1 - (u_1^{-a} + \dots + u_n^{-a} - n + 1)^{-\frac{1}{a}}$$

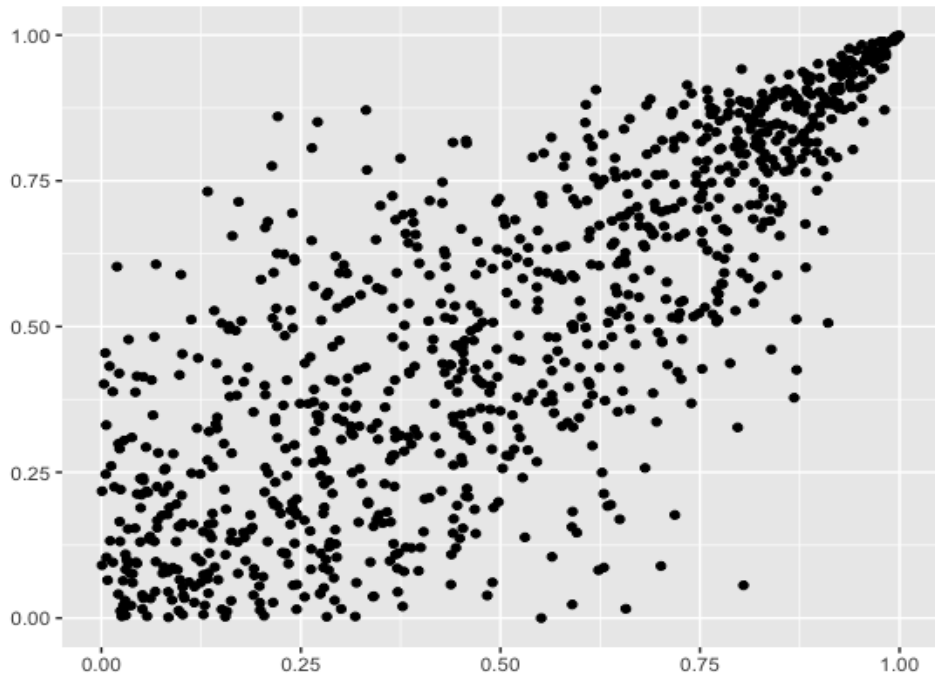


FIGURE 1.9 – Tirage aléatoire de 2 lois uniformes liées à une copule de Clayton inverse, de paramètre $a=3$

1.3.4.4 Copule de Gumbel

La copule de Gumbel admet comme générateur $\phi(x) = (-\log(x))^a$ avec $a \geq 1$. La copule correspondante s'écrit ainsi :

$$C_a(u, v) = \exp(-(-\log(u))^a + (-\log(v))^a)^{\frac{1}{a}}$$

On peut remarquer ici que la copule est symétrique, ce qui n'est pas le cas en général pour une copule archimédienne.

La copule de Gumbel n'a pas de dépendance de queue à gauche mais admet une dépendance de queue à droite. Nous verrons plus en détail les calculs menant à cette interprétation dans le prochain chapitre. 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'assurances ou actifs financiers.

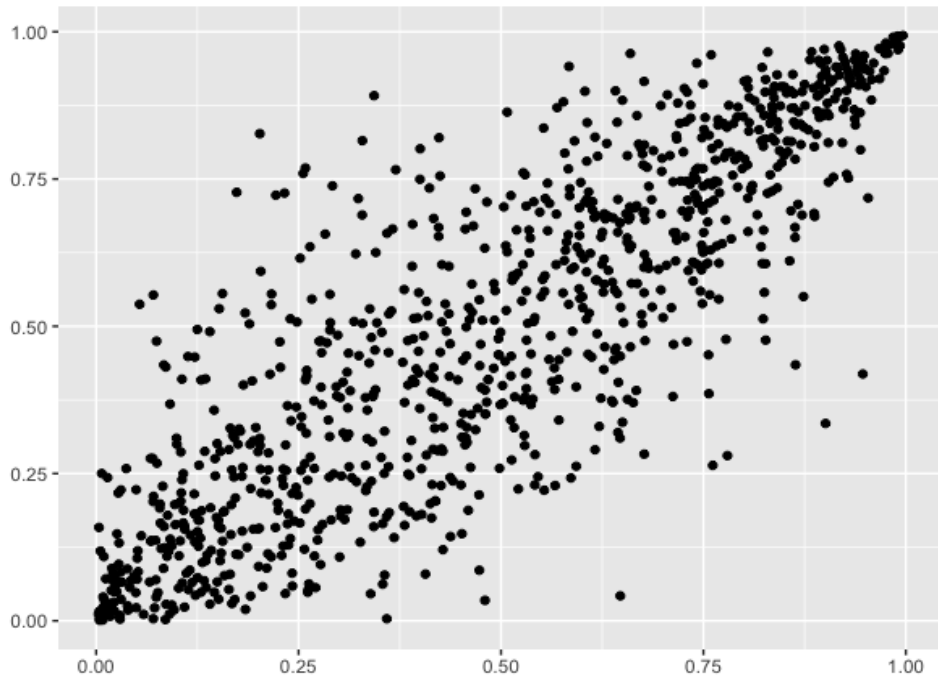


FIGURE 1.10 – Tirage aléatoire de 2 lois uniformes liées à une copule de Gumbel de paramètre $a=3$

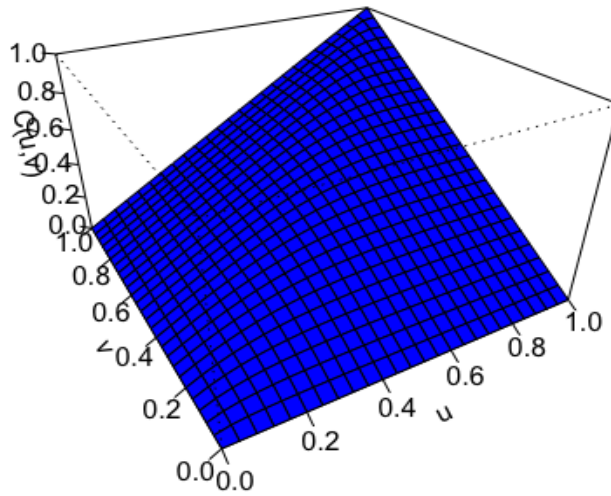


FIGURE 1.11 – Fonction de répartition de la copule de Gumbel

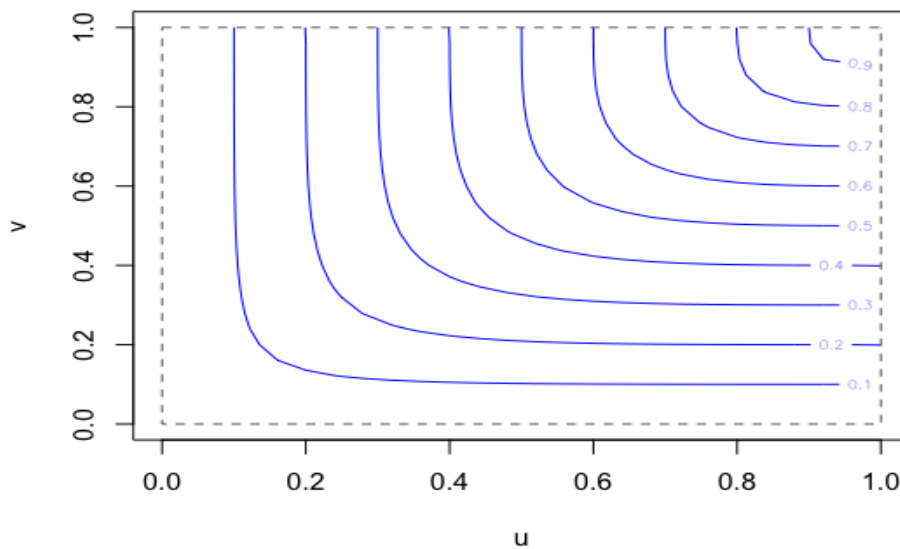


FIGURE 1.12 – Contour de la copule de Gumbel

1.3.4.5 Copule AMH (Ali, Mikhail et Haq)

Ali, Mikhail et Haq (1978) ont créé une copule archimédienne dont la particularité est son paramètre couvrant un intervalle réduit $[-1, 1]$. Le générateur de la copule s'écrit $\phi(t) =$

$\ln\left(\frac{1-\theta(1-t)}{t}\right)$, donc la copule AMH s'écrit de la façon suivante :

$$C_{\theta}(u, v) = \frac{uv}{1 - \theta(1-u)(1-v)}$$

où le paramètre de copule $\theta \in [-1, 1]$. La copule mesure aussi bien les dépendances positives que négatives. Comme nous le verrons plus en détail dans le prochain chapitre, il y a dépendance de queue à gauche si $\theta = 1$.



FIGURE 1.13 – Tirage aléatoire de 2 lois uniformes liées à une copule AMH, de paramètre $\theta = 0,7$

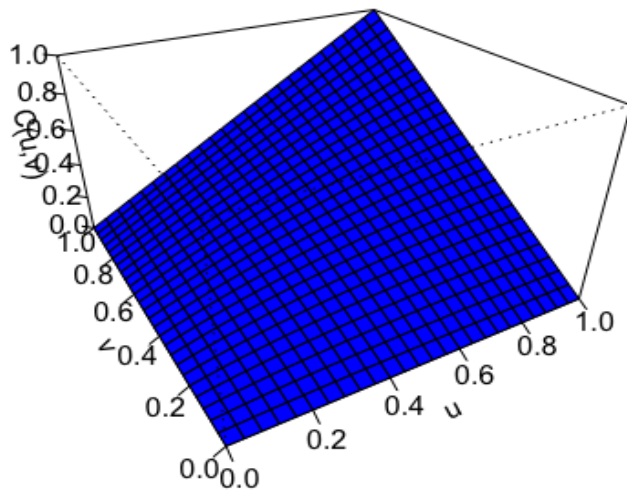


FIGURE 1.14 – Fonction de répartition de la copule AMH

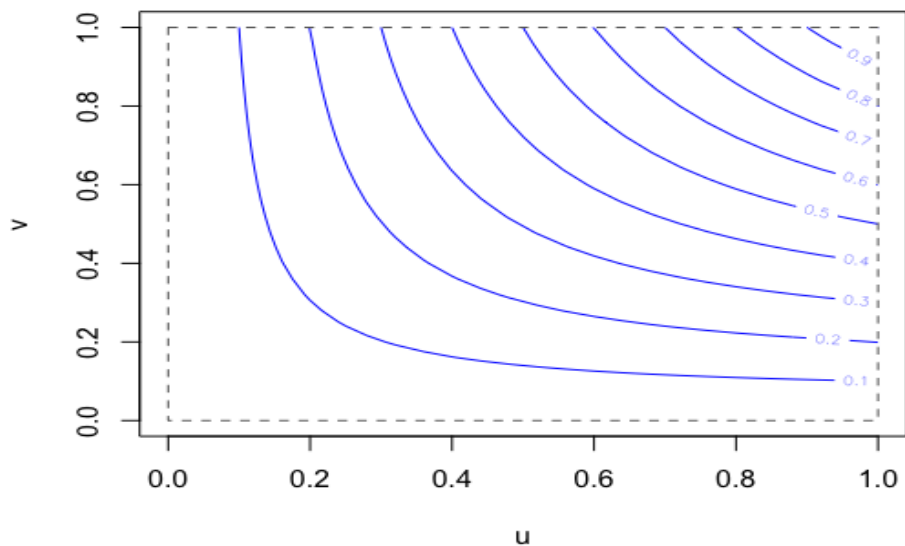


FIGURE 1.15 – Contour de la copule AMH

1.3.5 Copules empiriques

Nous pouvons également construire une copule empirique d'un ensemble de variables aléatoires à partir de leurs observations respectives. Pour ce faire, nous allons nous appuyer sur la statistique de rang de la loi jointe. Les copules empiriques trouvent leur utilité dans l'estimation

de copule paramétrique à partir d'observations données. En effet, elles peuvent constituer un outil de comparaison afin de calibrer une copule adéquate à la base de données que l'on possède.

Deuhevles[1979] est à l'origine de la notion de copule empirique. La fonction s'appuie sur le rang des observations afin d'extraire la structure de dépendance.

Soit (X_1, \dots, X_d) une suite de variables aléatoires et $(x_1^n, \dots, x_d^n)_{1 \leq n \leq N}$ un échantillon d'observations de taille N . Fixons également $(r_1^n, \dots, r_d^n)_{1 \leq n \leq N}$ la statistique de rang (ou rang des observations) associée à cet échantillon multivarié. Donc pour tout $1 \leq i \leq d$, r_i^n est le rang de x_i^n dans $(x_i^n)_{1 \leq n \leq N}$, c'est à dire $\sum_{n=1}^N \mathbb{1}_{X_i^n \leq x_i^n}$ pour tout $1 \leq i \leq d$.

On peut donc définir la copule empirique sur l'ensemble $\mathbf{E} = \{(\frac{n_1}{N}, \dots, \frac{n_d}{N}), 1 \leq i \leq d, n_i = 0, \dots, N\}$ de la façon suivante :

$$\hat{C}_N \left(\frac{n_1}{N}, \dots, \frac{n_d}{N} \right) = \frac{1}{N} \sum_{n=1}^N \prod_{i=1}^d \mathbb{1}_{\{r_i^n \leq n_i\}}$$

Cette copule présente certains inconvénients comme le fait de représenter plusieurs points de discontinuité, ce qui limite l'utilisation de la densité lorsqu'elle est requise. Dehevles a tout même établi une relation entre \hat{C}_N et c_N :

$$\hat{C}_N \left(\frac{i}{N}, \frac{j}{N} \right) = \sum_{p=1}^i \sum_{q=1}^j c_N \left(\frac{p}{N}, \frac{q}{N} \right)$$

La copule empirique \hat{C}_N offre l'avantage de s'affranchir des fonctions de répartition marginales. Cependant, l'estimateur empirique naturel de la copule utilise les lois marginales et s'exprime ainsi :

$$\mathbb{C}_N(u) = \mathbb{F}_N(F_{1,N}^{-1}(u_1), \dots, F_{d,N}^{-1}(u_d)), \forall u = (u_1, \dots, u_d) \in [0, 1]^d.$$

où $\{F_{i,N}(\cdot)\}_{i=1}^d$ sont les marges empiriques univariées. Les deux copules coïncident sur le même ensemble \mathbf{E} et \hat{C}_N converge asymptotiquement vers \mathbb{C}_N tel que :

$$\sup_{u \in [0,1]^d} \left| \hat{C}_N(u) - \mathbb{C}_N(u) \right| = \frac{1}{N}.$$

Chapitre 2

Présentation des différentes mesures de dépendance

Nous avons abordé dans le chapitre précédent les principales notions caractérisant la copule ainsi que les différentes copules usuelles que l'on sera amené à utiliser dans le cadre de notre application. Nous pouvons interpréter la copule $C(\cdot)$ d'un vecteur aléatoire $X = (X_1, X_2)$ comme une normalisation de la répartition jointe $F(\cdot)$ après avoir éliminé les effets de marges des fonctions $F_1(\cdot)$ et $F_2(\cdot)$. La copule caractérise ainsi la dépendance des composantes de X . Nous verrons dans ce chapitre que plusieurs mesures de dépendance peuvent être exprimées à partir de la copule $C(\cdot)$. Nous considérerons dans un premier temps les mesures de concordance en évoquant notamment le tau de Kendall et le rho de Spearman. Puis nous aborderons les dépendances de queues et l'application de ces nouvelles mesures sur les copules usuelles.

2.1 Les mesures de concordance

Avant de définir ce que sont les mesures de concordance, attachons-nous tout d'abord à la définition de l'ordre de concordance.

2.1.1 Ordre et mesure de concordance

Definition 1 Soient \mathbb{C}_1 et \mathbb{C}_2 deux copules bivariées.

On dit que \mathbb{C}_1 est plus petite que \mathbb{C}_2 et on note $\mathbb{C}_1 \prec \mathbb{C}_2$, si et seulement si :

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

L'ordre \prec est appelé l'ordre de concordance.

On peut remarquer que cette notion d'ordre de concordance n'est que partielle car on ne peut comparer toutes les copules entre elles.

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

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

Avec $W^d(u) = \max\left(\sum_{i=1}^d u_i - d + 1; 0\right)$ et $M^d(u) = \min(u_1, \dots, u_d) \forall u \in [0, 1]^d$.

Une démonstration de cette inégalité se trouve en annexe 1.

On peut désormais définir la mesure de concordance.

Definition 2 Une mesure numérique κ d'association entre deux variables aléatoires continues X_1 et X_2 , dont la copule est $\mathbb{C}(\cdot)$, est une mesure de concordance si elle satisfait les conditions suivantes :

1. κ est définie pour toute paire (X_1, X_2) de variables aléatoires continues ;
2. $-1 = \kappa_{\langle X, -X \rangle} \leq \kappa_{\langle \mathbb{C} \rangle} \leq \kappa_{\langle X, X \rangle} = 1$;
3. $\kappa_{\langle X_1, X_2 \rangle} = \kappa_{\langle X_2, X_1 \rangle}$;
4. Si X_1 et X_2 sont indépendantes, alors $\kappa_{\langle X_1, X_2 \rangle} = 0$;
5. $\kappa_{\langle -X_1, X_2 \rangle} = \kappa_{\langle X_1, -X_2 \rangle} = -\kappa_{\langle X_1, X_2 \rangle}$;
6. Si $\mathbb{C}_1 \prec \mathbb{C}_2$ alors $\kappa_{\langle \mathbb{C}_1 \rangle} \leq \kappa_{\langle \mathbb{C}_2 \rangle}$;
7. Si $(X_{1,m}, X_{2,m})$ est une suite de variables aléatoires continues dont la copule est $\mathbb{C}_m(\cdot)$ et si $\mathbb{C}_m(\cdot)$ converge vers $\mathbb{C}(\cdot)$ alors $\lim_{m \rightarrow \infty} \kappa_{\langle \mathbb{C}_m \rangle} = \kappa_{\langle \mathbb{C} \rangle}$.

A partir de la propriété (6), on peut en déduire que l'ordre de concordance implique l'ordre sur la mesure κ .

Une propriété importante de la mesure de concordance $\kappa_{\langle X, Y \rangle}$ est que si $\alpha(\cdot)$ et $\beta(\cdot)$ sont des fonctions strictement monotones, alors $\kappa_{\langle \alpha(X), \beta(Y) \rangle} = \kappa_{\langle X, Y \rangle}$. Par exemple, si $Y = f(X)$ et $f(\cdot)$ est décroissante, alors $\kappa_{\langle X, Y \rangle} = \kappa_{\langle X, f(X) \rangle} = \kappa_{\langle X, -X \rangle} = -1$.

Nous allons désormais introduire la notion de mesure de concordance et de discordance à travers le théorème suivant, cette mesure sera notamment utilisée pour définir le tau de Kendall.

Théorème 1 Soient (X_1, Y_1) et (X_2, Y_2) deux vecteurs indépendants de variables aléatoires continues de fonctions de répartition jointes, respectivement $\mathbb{H}_1(\cdot, \cdot)$ et $\mathbb{H}_2(\cdot, \cdot)$. Soient $F(\cdot)$ et $G(\cdot)$ les marges associées respectivement à X_1, X_2 et Y_1, Y_2 . Soient $\mathbb{C}_1(\cdot, \cdot)$ et $\mathbb{C}_2(\cdot, \cdot)$ les copules associées respectivement à $\mathbb{H}_1(\cdot, \cdot)$ et à $\mathbb{H}_2(\cdot, \cdot)$ telles que $\mathbb{H}_1(x, y) = \mathbb{C}_1(F(x), G(y))$ et $\mathbb{H}_2(x, y) = \mathbb{C}_2(F(x), G(y))$. Si $\mathcal{Q}(\cdot, \cdot)$ est la mesure de concordance et de discordance de (X_1, Y_1) et (X_2, Y_2) ; c'est à dire,

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

alors

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

La démonstration de ce théorème sera abordée en Annexe 2. Définissons désormais le tau de Kendall, très utilisé pour définir des dépendances globales de deux marginales à partir de copules.

2.1.2 Tau de Kendall

Definition 3 Tau de Kendall.

Soient (X_1, X_2) et (Y_1, Y_2) deux paires indépendantes de variables aléatoires, ayant pour fonctions de répartition marginales communes F (pour X_1 et X_2) et G (pour Y_1 et Y_2). Le tau de Kendall se définit comme suit :

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

Il s'agit donc de la différence entre la probabilité de concordance et la probabilité de non concordance.

On peut également lier le tau de Kendall en la définissant ainsi :

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

où la paire (U, V) est de loi \mathbb{C} .

Cette expression est toutefois difficilement manipulable numériquement car elle implique de calculer la différentielle $d\mathbb{C}(u, v)$. Il est ainsi plus facile de calculer le tau de Kendall avec l'expression équivalente suivante (Nelsen(1999) page 131) :

$$\tau = 1 - 4 \int_{[0,1]^2} \partial_{u_1} \mathbb{C}(u_1, u_2) \partial_{u_2} \mathbb{C}(u_1, u_2) du_1 du_2. \quad (2.2)$$

On peut également définir le tau de Kendall d'une manière empirique. Soit un échantillon de taille n de données bivariées $\{(x_k, y_k)\}_{k=1}^n$. Soit c et d respectivement le nombre de paires concordantes et discordantes dans cet échantillon. Le tau de Kendall en version empirique est ainsi défini par :

$$\tau_n = \frac{c - d}{c + d} = \frac{c - d}{C_2^n}, C_2^n = \frac{n!}{2(n-2)!} = \frac{n(n-1)}{2}.$$

On peut aussi définir cette version empirique en fonction de la densité de copule empirique

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

On peut noter également que, dans le cas bivarié, si $\mathbb{C}(\cdot)$ est une copule archimédienne de générateur $\phi(\cdot)$ alors

$$\tau = 1 + 4 \int_0^1 \frac{\phi(u)}{\phi'(u)} du.$$

A partir de cette définition, nous pouvons aisément calculer la plupart des tau de Kendall pour les copules archimédiennes.

Exemple de calcul de tau de Kendall

Essayons de calculer le tau de Kendall pour la copule de Gumbel. Pour rappel, le générateur ϕ de la copule de Gumbel s'écrit $\phi(x) = (-\log(x))^a$ avec $a \geq 1$. Donc :

$$\tau = 1 + 4 \int_0^1 \frac{(-\log(x))^a}{-a \times \frac{1}{x} \times (-\log(x))^{a-1}} dx = 1 + \frac{4}{a} \int_0^1 x \log(x) dx$$

En utilisant ensuite une intégration par parties, on arrive au résultat $\tau = 1 - \frac{1}{a}$. On peut remarquer ici que le tau de Kendall croit en fonction de a , ce qui est évident car a est un paramètre de mesure de dépendance entre les risques. La copule de Gumbel indique donc uniquement des dépendances positives.

Tout comme le tau de Kendall, le rho de Spearman est basé sur la concordance et la discordance de couples de variables aléatoires.

2.1.3 Rho de Spearman

Definition 4 *Rho de Spearman*

Soient (X_1, Y_1) , (X_2, Y_2) et (X_3, Y_3) trois paires indépendantes de variables aléatoires de même fonction de répartition jointe H . Le coefficient de corrélation de Spearman se définit comme suit :

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

Exprimons désormais le rho de Spearman en fonction de la copule \mathbb{C} .

Théorème 2 Soit (X, Y) deux variables continues de copule \mathbb{C} . Le rho de Spearman de (X, Y) est défini par :

$$\rho_S(X, Y) = 3\mathcal{Q}(\mathbb{C}, \Pi) = 12 \int_{I^2} uv d\mathbb{C}(u, v) - 3 = 12 \int_{I^2} \mathbb{C}(u, v) dudv - 3.$$

2.1.4 Remarques

1. On peut considérer que ρ_S est une mesure de distance moyenne entre la copule $\mathbb{C}(\cdot)$ et la copule d'indépendance car

$$\rho_S(\mathbb{C}) = 12 \int_{I^2} (\mathbb{C}(u, v) - uv) dudv.$$

2. Le tau de Kendall et le rho de Spearman ne dépendent pas du comportement des lois marginales, mais uniquement de la structure de dépendance.
3. Soit un échantillon de taille n de données bivariées $\{(x_k, y_k)\}_{k=1}^n$. On définit la version empirique du rho de Spearman en fonction de la densité de la copule ainsi :

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

On note généralement que les valeurs τ et ρ pour une copule $\mathbb{C}(\cdot)$ sont différentes.

2.1.5 Récapitulatif des tau de Kendall aux copules archimédiennes

Le tableau récapitulatif suivant exprime les tau de Kendall pour chaque copule usuelle archimédienne étudiée précédemment, en fonction de leur paramètre de copule.

Copule	$\phi_a(u)$	τ_a
Gumbel	$(-\log(u))^a, a \geq 1$	$1 - \frac{1}{a}$
Franck	$-\log\left(\frac{e^{-au}-1}{e^{-a}-1}\right), a \neq 0$	$1 - \frac{4(1-D_1(a))}{a}$ avec $D_1(a) = \frac{1}{a} \int_0^a \frac{t}{e^t-1} dt$
Clayton	$\frac{(u^{-a}-1)}{a}, a > 0$	$\frac{a}{a+2}$

TABLE 2.1 – Tableau récapitulatif des générateurs de copules archimédiennes et de leurs tau de Kendall

2.2 Dépendance de queues

Il y a de nombreuses manières de mesurer la dépendance parmi les risques, par le tau de Kendall ou le rho de Spearman par exemple qui se concentrent sur la dépendance globale. La dépendance de queues est une mesure plus spécialisée dans les queues d'évènements extrêmes, ce qui en fait un concept clé dans de nombreuses applications financières et assurantielles. La mesure de dépendance de queue est un indicateur de dépendance de queue (Tail Dependence Coefficient en anglais).

Afin de mieux comprendre l'expression de l'indicateur de dépendance de queues, attachons-nous tout d'abord à décrire les expressions analytiques des fonction copules.

2.2.1 Probabilités avec la forme analytique des fonctions copules

Les expressions analytiques ci-dessous permettront de déduire facilement les calculs de dépendances de queue ou encore les algorithmes de simulation des copules. Nous exposons ici un cas bidimensionnel mais les résultats peuvent être facilement généralisables pour les dimensions supérieures.

$$\mathbb{P}(U_1 \leq u_1, U_2 \geq u_2) = \mathbb{P}(U_1 \leq u_1) - \mathbb{P}(U_1 \leq u_1, U_2 \leq u_2) = u_1 - C(u_1, u_2)$$

$$\mathbb{P}(U_1 \leq u_1 | U_2 \leq u_2) = \frac{\mathbb{P}(U_1 \leq u_1, U_2 \leq u_2)}{\mathbb{P}(U_2 \leq u_2)} = \frac{C(u_1, u_2)}{u_2}$$

$$\mathbb{P}(U_1 \leq u_1 | U_2 \geq u_2) = \frac{\mathbb{P}(U_1 \leq u_1, U_2 \geq u_2)}{\mathbb{P}(U_2 \geq u_2)} = \frac{u_1 - C(u_1, u_2)}{1 - u_2}$$

$$\mathbb{P}(U_1 \geq u_1 | U_2 \geq u_2) = \mathbb{P}(U_1 \geq u_1) - \mathbb{P}(U_1 \geq u_1, U_2 \leq u_2) = 1 - u_1 - u_2 + C(u_1, u_2)$$

$$\mathbb{P}(U_1 \geq u_1 | U_2 \leq u_2) = \frac{\mathbb{P}(U_1 \geq u_1, U_2 \leq u_2)}{\mathbb{P}(U_2 \leq u_2)} = \frac{1 - u_1 - u_2 + C(u_1, u_2)}{1 - u_2}$$

On peut également définir une fonction copule conditionnelle :

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

$$= \lim_{\Delta u \rightarrow 0^+} \frac{\mathbb{P}(U_1 \leq u_1, u_2 \leq U_2 \leq u_2 + \Delta u)}{\mathbb{P}(u_2 \leq U_2 \leq u_2 + \Delta u)}$$

$$= \lim_{\Delta u \rightarrow 0^+} \frac{C(u_1, u_2 + \Delta u) - C(u_1, u_2)}{\Delta u}$$

$$= \partial_2 C(u_1, u_2)$$

2.2.2 Caractérisation de la dépendance des queues

L'indicateur de dépendance des queues (TDC, Tail dependence coefficient en anglais) entre deux variables aléatoires X et Y , de fonction de répartition jointe H et de fonctions de répartitions marginales respectivement F et G , a été introduit dans le but de mesurer la dépendance entre des événements extrêmes.

La dépendance des queues à gauche est une mesure de dépendance dans le coin bas-gauche d'une représentation de copule et est défini comme suit :

$$\lambda_l(X|Y) = \lim_{u \rightarrow 0^+} \mathbb{P}[X \leq F^{-1}(u) | Y \leq G^{-1}(u)] = \lim_{u \rightarrow 0^+} \frac{C(u, u)}{u}. \quad (2.3)$$

Il existe également la mesure de dépendance à droite qui mesure la dépendance dans le coin haut-droit :

$$\lambda_u(X|Y) = \lim_{u \rightarrow 1^-} \mathbb{P}[X > F^{-1}(u) | Y > G^{-1}(u)] = \lim_{u \rightarrow 1^-} \frac{1 - u - u + C(u, u)}{1 - u}. \quad (2.4)$$

Démonstration

Soit deux variables aléatoires X et Y de copule C , on a :

$$\begin{aligned}\lambda_t(X|Y) &= \lim_{u \rightarrow 0^+} \mathbb{P}[X \leq F^{-1}(u) | Y \leq G^{-1}(u)] \\ &= \lim_{u \rightarrow 0^+} \frac{\mathbb{P}(X \leq F^{-1}(u), Y \leq G^{-1}(u))}{\mathbb{P}(Y \leq G^{-1}(u))} \\ &= \lim_{u \rightarrow 0^+} \frac{C(u, u)}{u}.\end{aligned}$$

De même, on peut écrire :

$$\begin{aligned}\lambda_u(X|Y) &= \lim_{u \rightarrow 1^-} \mathbb{P}(X > F^{-1}(u) | Y > G^{-1}(u)) \\ &= \lim_{u \rightarrow 1^-} \frac{\mathbb{P}(X > F^{-1}(u), Y > G^{-1}(u))}{\mathbb{P}(Y > G^{-1}(u))} \\ &= \lim_{u \rightarrow 1^-} \frac{1 - \mathbb{P}(X \leq F^{-1}(u) \cup Y \leq G^{-1}(u))}{1 - \mathbb{P}(Y \leq G^{-1}(u))} \\ &= \lim_{u \rightarrow 1^-} \frac{1 - (\mathbb{P}(X \leq F^{-1}(u)) + \mathbb{P}(Y \leq G^{-1}(u)) - \mathbb{P}(X \leq F^{-1}(u), Y \leq G^{-1}(u)))}{1 - \mathbb{P}(Y \leq G^{-1}(u))} \\ &= \lim_{u \rightarrow 1^-} \frac{1 - (u + u - C(u, u))}{1 - u}\end{aligned}$$

Il est à noter que le théorème de Sklar déduit immédiatement que les mesures λ_t et λ_u sont déterminées par la copule C reliant F , G et H . Les deux mesures indiquent l'existence d'une dépendance seulement si elles sont comprises dans l'intervalle $[0,1]$. Si la valeur est égale à zéro, il n'y a pas de dépendance de queue.

On peut s'apercevoir toutefois qu'il peut y avoir deux principales limites à l'utilisation de cette mesure de dépendance :

- l'indicateur n'établit pas de taux de convergence pour les limites en 1 et 0.
- La dépendance est seulement mesurée sur l'axe de la diagonale $x=y$.

Cependant, ces mesures de dépendance restent adaptées pour l'étude de dépendances asymptotiques.

On peut facilement déduire les dépendances de queue à gauche et à droite d'un couple de variables à partir de l'expression analytique de sa copule.

Nous étudierons et comparerons par la suite, à l'aide de cette mesure de dépendance, les dépendances de queue de plusieurs copules modélisées sur la plateforme Milliman Mind.

2.2.3 Exemples de calculs de dépendances de queues

Copule de Gumbel

Rappelons l'expression de la copule de Gumbel $C_a(u, v) = \exp(-(-\log(u))^a + (-\log(v))^a)^{\frac{1}{a}}$. On peut ainsi calculer le coefficient de dépendance de queue supérieur :

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

et le coefficient de dépendance inférieur :

$$\begin{aligned}\lambda_L &= \lim_{u \rightarrow 0} \frac{u^{2\frac{1}{a}}}{u} \\ &= 0\end{aligned}$$

On note donc que la copule de Gumbel a uniquement une dépendance de queue supérieure, c'est à dire qu'il y a uniquement une dépendance entre les fortes valeurs.

Copule de Clayton

Rappelons que la copule de Clayton est définie par :

$$C_a(u, v) = (u^{-a} + v^{-a} - 1)^{-\frac{1}{a}}$$

Donc le coefficient de dépendance de queue supérieur s'écrit :

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

Et le coefficient de dépendance de queue inférieur s'écrit :

$$\begin{aligned}
\lambda_L &= \lim_{u \rightarrow 0} \frac{C(u, u)}{u} \\
&= \lim_{u \rightarrow 0} \frac{(2u^{-a} - 1)^{-\frac{1}{a}}}{u} \\
&= \lim_{u \rightarrow 0} (2 - u^a)^{-\frac{1}{a}} \\
&= 2^{-\frac{1}{a}}.
\end{aligned}$$

Il y a donc ici une dépendance à gauche mais pas de dépendance supérieure, au niveau des grandes valeurs.

Copule de Student

Soient X_1, X_2 deux variables aléatoires de lois de Student bivariée à ν degrés de liberté et de corrélation ρ . Comme $X_2|X_1 \sim t_{\nu=1}$, alors $\mathbb{E}(X_2|X_1 = x) = \rho x$ et $Var(X_2|X_1 = x) = \frac{\nu+x^2}{1+\nu}(1-\rho)$. Donc

$$\begin{aligned}
\lambda_U &= \lim_{x \rightarrow \infty} \mathbb{P}(X_2 > x | X_1 > x) \\
&= \lim_{x \rightarrow \infty} 2\bar{t}_{\nu+1} \left(\left(\frac{1+\nu}{\nu+x^2} \right)^{1/2} \frac{x+\rho x}{\sqrt{1-\rho^2}} \right) \\
&= \lim_{x \rightarrow \infty} 2\bar{t}_{\nu+1} \left(\left(\frac{1+\nu}{\frac{\nu}{x^2}+1} \right)^{1/2} \frac{\sqrt{1-\rho}}{\sqrt{1+\rho}} \right) \\
&= 2\bar{t}_{\nu+1} \left(\left(\frac{(1+\nu)(1-\rho)}{1+\rho} \right)^{1/2} \right)
\end{aligned}$$

où $\bar{t}_{\nu+1}$ est la fonction de survie de la loi de Student à $\nu+1$ degrés de liberté. A noter que pour cette copule, $\lambda_L = \lambda_U$.

Copule AMH

Rappelons que la copule AMH est définie par :

$$C_\theta(u, v) = \frac{uv}{1 - \theta(1-u)(1-v)}$$

Calculons désormais la dépendance de queue inférieure λ_L :

$$\begin{aligned}
\lambda_L &= \lim_{u \rightarrow 0} \frac{C(u, u)}{u} = \lim_{u \rightarrow 0} \frac{u}{1 - \theta(1-u)^2} \\
&= \begin{cases} \lim_{u \rightarrow 0} \frac{u}{1 - \theta(1-u)^2} = 0.5; \theta = 1; \\ \lim_{u \rightarrow 0} \frac{u}{1 - \theta(1-u)^2} = 0; \theta < 1. \end{cases}
\end{aligned}$$

λ_U s'écrit :

$$\lambda_U = \lim_{u \rightarrow 1} \frac{[1 - 2u + C(u, u)]}{1 - u} = \lim_{u \rightarrow 1} \left[1 - \frac{u(1 - \theta(1-u))}{1 - \theta(1-u)^2} \right] = 0.$$

Donc tant que $\lambda_L \neq 0$ quand $\theta = 1$, la dépendance de queue à gauche existe asymptotiquement pour $\theta = 1$. Pour tout $\theta < 1$, les valeurs extrêmes (à gauche ou à droite) sont asymptotiquement indépendantes.

Copule	λ_L	λ_U
Gumbel	0	$2 - 2^{\frac{1}{a}}$
Clayton	$2^{-\frac{1}{a}}$	0
Frank	0	0
AMH	0.5 si a=1 ,sinon 0	0
Student	$2_{\nu+1} \left(\left(\frac{(1+\nu)(1-\rho)}{1+\rho} \right)^{1/2} \right)$	$2_{\nu+1} \left(\left(\frac{(1+\nu)(1-\rho)}{1+\rho} \right)^{1/2} \right)$
Gaussienne	0	0

TABLE 2.2 – Dépendance de queues pour les différentes copules usuelles

2.3 Estimation de la dépendance de queues

Les mesures de dépendance de queues calculées auparavant sont exprimées en fonction de copules, ce qui peuvent les rendre inutiles dans les cas où la copule à étudier n'est pas connue. La mesure de dépendance de queue classique ne serait dans ce cas pas applicable. Il existe plusieurs méthodes pour estimer un coefficient de dépendance de queue, dont celle de l'approche paramétrique qui se base sur l'hypothèse que d'une part la structure de dépendance entre les variables est représentée par une copule archimédienne et d'autre part que la limite du théorème de Juri et Wüthrich (2002) permet d'estimer la dépendance de queue.

Il existe également plusieurs méthodes dans la littérature proposant d'estimer la mesure de dépendance des queues empiriquement, de façon non-paramétrique, en utilisant entre autres les copules empiriques, le tau de Kendall ou encore l'approximation de copules (par les copules maximum et d'indépendance). Il est à noter que toutes ces méthodes ne sont pas toutes fiables selon la simulation de copules étudiée. Ces méthodes de mesures de dépendance de queues sont indicatives et sont à utiliser dans une approche comparative.

2.3.1 Approche paramétrique

Frahm et al. (2005) ont publié des approches paramétriques basées sur le degré d'information sur les distributions de variables. Ces auteurs ont présenté une estimation en utilisant une distribution spécifique, une copule spécifique et une classe de copule. Parmi ces trois méthodes, la dernière relative aux copules archimédiennes est basée sur le résultat de Juri et Wüthrich(2002). Elle est la plus générale.

Nous allons présenter ici l'approche la plus générale parce-que cette approche utilise toute une classe de copule, à savoir la classe de copule archimédienne. La limite vers laquelle tend la copule archimédienne établie par Juri et Wüthrich(2002) permet d'estimer la dépendance de queue.

Soit la fonction suivante :

$$F_t = \frac{C(\min(x,t))}{C(t,t)}, 0 \leq x \leq 1, t \in [0, 1]$$

Alors la copule associée à cette fonction s'écrit :

$$C_t(u,v) = \frac{C(F_t^{-1}(u), F_t^{-1}(v))}{C(t,t)}$$

Quand t tend vers 0, cette copule tend vers une copule de Clayton. Ainsi, on peut exprimer le coefficient de dépendance de queue inférieure à partir du paramètre de la copule :

$$\lim_{t \rightarrow 0} C_t(u, v) = C_{Cl}(u, v, \theta), \theta > 0 \lambda_L = 2^{-\frac{1}{\theta}}$$

2.3.2 Approche non-paramétrique

La normalité asymptotique et la consistance forte de l'estimateur que l'on va présenter ont été étudiées dans Schmidt et Stadtmüller (2006) :

$$\begin{aligned} \hat{\lambda}_U &= \hat{\Lambda}_U(1, 1) \text{ ou } \hat{\Lambda}_U^{EVT}(1, 1) \\ \hat{\lambda}_L &= \hat{\Lambda}_L(1, 1) \text{ ou } \hat{\Lambda}_L^{EVT}(1, 1) \end{aligned}$$

où $\hat{\Lambda}_U(x, y)$ est la copule de queue supérieure empirique et $\hat{\Lambda}_U^{EVT}(x, y)$ est la copule de queue supérieure estimée par la fonction stable de queue de dépendance. Les définitions de copule de queue empirique et fonctions de dépendance de queue stable sont présentes dans Schmidt et Stadtmüller (2006). Généralement, les deux méthodes retournent des résultats similaires, les différences dépendent de la taille de l'échantillon et du choix des paramètres utilisés pour l'estimation.

Il existe également d'autres méthodes d'estimations non-paramétriques (DobricSchmidt 2005, CaillaultGuégan 2005 entre autres) que l'on ne présentera pas ici. Ces estimations ont différentes précisions qui dépendent des données utilisées. Pour les auteurs, le meilleur estimateur n'est pas consistant et varie selon les cas. De plus, pour des coefficients de dépendance de queue faibles voir nuls, l'estimation non paramétrique n'est pas fiable, sur-estimant ou ne distinguant pas le cas où la dépendance de queue n'existe pas.

Chapitre 3

Méthodes de simulation et inférence statistique des copules

Nous avons abordé précédemment les coefficients de dépendance de queue, outils très utiles à la mesure des dépendances extrêmes. Dans le cadre de notre étude, il sera très utile d'estimer notamment les dépendances de queues pour pouvoir établir des dépendances extrêmes entre branches de risque (dans notre étude, il s'agira des branches auto et habitation d'un même portefeuille). Cette estimation de dépendance de queues sera aussi un outil permettant de calibrer la copule paramétrique adéquate à notre jeu de données. Pour ce faire, il faut tout d'abord pouvoir simuler des copules paramétriques à partir d'un jeu de données, puis estimer le paramètre de copule adéquat aux données. Nous aborderons donc lors de ce chapitre les différentes méthodes de simulation des copules elliptiques et archimédiennes, ainsi que l'inférence statistique des copules. Nous verrons par la suite qu'à partir de ces outils, nous pourrons établir une copule paramétrique adéquate à notre jeu de données par l'intermédiaire du test du χ^2 .

3.1 Simulations de copule

Dans la plateforme web qu'est Milliman Mind il a été développé une fonction permettant de simuler des copules selon un ensemble de données et un paramètre de copule donnée. Nous allons aborder dans cette partie les différentes méthodes utilisées dans Milliman Mind pour construire une copule à partir d'un vecteur aléatoire dont on ne connaît pas les lois marginales.

3.1.1 Copules archimédiennes

Nous nous concentrerons ici sur une méthode de simulation des copules archimédiennes s'appuyant sur la transformée de Laplace-Stieltjes inversée (méthode implémentée dans Milliman Mind), outil très efficace pour des observations de grandes dimensions. Tout d'abord, nous pouvons établir un lien entre la transformée de Laplace-Stieltjes et la fonction génératrice d'une copule archimédienne par le biais de ce théorème :

Théorème 3 *Une fonction ϕ sur $[0, \infty]$ est une transformée de Laplace-Stieltjes d'une fonction marginale si et seulement si ϕ est complètement monotone et $\phi(0) = 1$.*

L'algorithme suivant simulera une copule archimédienne de dimension d avec un générateur ϕ , où $\mathcal{LS}^{-1}(\phi)$ est la transformée de Laplace-Stieltjes inversée de ϕ .

Algorithme 1 (Marshall, Olkin)

1. Générer aléatoirement $V \sim F = \mathcal{LS}^{-1}(\phi)$

2. Générer aléatoirement une suite de variables aléatoires indépendantes et identiquement distribuées $X_i \sim U[0, 1], i \in \{1, \dots, d\}$

3. Renvoyer (U_1, \dots, U_d) , où $U_i = \phi\left(-\frac{\log(X_i)}{V}\right), i \in \{1, \dots, d\}$.

Cet algorithme est notamment efficace pour de grande dimensions, car nous avons besoin que de $d + 1$ nombres aléatoires pour générer des observations de dimension d . Seulement un échantillon $V \sim F$ est requis, cet échantillon étant indépendant de la dimension.

Dans le tableau suivant, nous allons décrire les fonctions F associées à chaque copule archimédienne. La fonction de la loi Stable que l'on aperçoit dans le tableau se note $S(\alpha, \beta, \gamma, \delta, 1), \alpha \in (0, 2], \beta \in [-1, 1], \gamma \in [0, \infty), \delta \in \mathbb{R}$, de fonction caractéristique $\exp(i\delta t - \gamma^\alpha |t|^\alpha (1 - i\beta \text{signe}(t)w(t, \alpha)))$, $t \in \mathbb{R}$. $\text{signe}(t) = \mathbf{1}_{[0, \infty)}(t) - \mathbf{1}_{(-\infty, 0]}(t)$, $w(t, \alpha) = \tan((\alpha\pi)/2)$ pour $\alpha \neq 1$, et $w(t, \alpha) = -2 \log(|t|)/\pi$ pour $\alpha = 1$. Pour les copules AMH et Frank, F est discrète avec une fonction de masse $(y_k)_{k \in \mathbb{N}}, k \in \mathbb{N}$.

Copule	Paramètre a	$\phi(x)$	F
Gumbel	$[1, \infty)$	$(-\log(x))^a$	$S(\frac{1}{a}, 1, (\cos(\frac{\pi}{2a}))^a, 0, 1)$
Clayton	$(0, \infty)$	$\frac{t^{-a}-1}{a}$	$\Gamma(1/a, 1)$
Frank	\mathbb{R}^*	$-\log\left(\frac{e^{-ax}-1}{e^{-a}-1}\right)$	$y_k = \frac{(1-e^{-a})^k}{ka}, k \in \mathbb{N}$
AMH	$[-1, 1)$	$\frac{1-a}{e^t-a}$	$y_k = (1-a)a^{k-1}, k \in \mathbb{N}$

TABLE 3.1 – Copules archimédiennes avec leurs paramètres, leurs générateurs ainsi que la transformée de Laplace-Stieltjes inversée

Il est malheureusement difficile de trouver la fonction F explicitement. En effet, pour une transformée de Laplace-Stieltjes inversée ϕ , il faut vérifier les tables de transformées de Laplace pour trouver F .

Le cas où F est une fonction pas à pas peut être traité par le théorème suivant :

Théorème 4 Soit $F = \mathcal{L}\mathcal{S}^{-1}(\phi)$ et $G(x) = \sum_{k=0}^{\infty} y_k \mathbf{1}_{[x_k, \infty)}(x)$ avec $(x_k)_{k \in \mathbb{N}}$ strictement croissante

et $y_k > 0, k \in \mathbb{N}$ et $\sum_{k=0}^{\infty} y_k = 1$. Alors :

$$F \equiv G \iff \phi(t) = \sum_{k=0}^{\infty} y_k e^{-x_k t}, t \in [0, \infty)$$

L'unicité du théorème de la transformée de Laplace-Stieltjes implique que $F \equiv G$ pour une mesure de Lebesgue nulle, donc $F \equiv G$.

3.1.2 Méthode de distribution

Cette méthode de simulation repose entièrement sur le théorème de Sklar et son corollaire. En effet, rappelons que la copule C associée à un vecteur aléatoire $X = (X_1, \dots, X_N)$ de loi F dont les marginales sont F_{X_1}, \dots, F_{X_n} s'écrit :

$$F(x_1, \dots, x_n) = C(F_{X_1}(x_1), \dots, F_{X_n}(x_n)), \forall (x_1, \dots, x_n) \in \mathbb{R}^n$$

$$C(u_1, \dots, u_n) = F\left(F_{X_1}^{-1}(u_1), \dots, F_{X_n}^{-1}(u_n)\right), \forall (u_1, \dots, u_n) \in [0, 1]^n.$$

Pour simuler un vecteur $U = (U_1, \dots, U_n)$ de copule C , il suffit donc d'appliquer la transformation $U = (F_{X_1}(X_1), \dots, F_{X_n}(X_n))$.

On note que cette méthode est utilisée dans le cas où la distribution F associée à la copule C est plus facile à simuler que la copule C elle-même. Cette méthode est donc très utilisée pour les simuler les copules elliptiques. En effet, on peut simuler par exemple facilement une distribution normale en utilisant la décomposition inférieure de Cholesky ($\mathcal{N}(0, \Sigma) = P\mathcal{N}(0, I)$) avec P vérifiant $PP^t = \Sigma$.

3.2 Inférence statistique des copules

Après avoir abordé les différentes méthodes de simulation implémentées sous Milliman Mind, nous évoquerons dans cette partie les méthodes permettant d'estimer le paramètre d'une copule, le paramètre de la copule restant l'unique inconnue à toute construction de copule. Cette étape est primordiale dans le but d'un bon calibrage de la copule paramétrique avec notre jeu de données. Nous allons étudier ici trois méthodes reconnus dans l'estimation de paramètre de copules : la méthode des moments, la méthode du maximum de vraisemblance et la méthode IFM. On va distinguer ici l'approche paramétrique de l'approche semi-paramétrique. L'approche paramétrique suppose que les lois marginales appartiennent à une famille de lois indexée par un paramètre ; tandis que l'approche semi-paramétrique fait l'hypothèse que les lois marginales n'appartiennent à aucune famille. En approche paramétrique, on estime les paramètres des marges pour pouvoir estimer les fonctions marginales ; en approche semi-paramétrique les marges sont estimées par l'estimateur empirique d'une fonction marginale. Nous allons nous concentrer dans cette partie essentiellement sur l'approche paramétrique, car c'est celle que nous allons adopter pour notre étude de données.

3.2.1 Méthode des moments

La méthode des moments est en fait basée sur l'inversion du rho de Spearman ou du tau de Kendall. En effet, il existe une relation, plus ou moins explicite, entre le rho de Spearman et le tau de Kendall, et le paramètre de la copule θ . Rappelons la version empirique de ces mesures, pour la paire (X_i, X_j) :

$$\hat{\rho}_{i,j} = \frac{\sum_{k=1}^n (\hat{U}_i^{(k)} - \bar{U}_i) (\hat{U}_j^{(k)} - \bar{U}_j)}{\left[\sum_{k=1}^n (\hat{U}_i^{(k)} - \bar{U}_i)^2 \sum_{k=1}^n (\hat{U}_j^{(k)} - \bar{U}_j)^2 \right]^{\frac{1}{2}}}$$

$$\hat{\tau}_{i,j} = \binom{n}{2}^{-1} \sum_{k < l} \text{sign} \left((X_i^{(k)} - X_i^{(l)})(X_j^{(k)} - X_j^{(l)}) \right)$$

où $\hat{U}_i^{(k)} = \hat{F}_i(X_i^{(k)})$, $\bar{U}_i = \sum_{k=1}^n \frac{\hat{U}_i^{(k)}}{n}$, $i = 1, \dots, d$ et $\text{sign}(x) = 1$ si $x > 0$, -1 si $x < 0$ et 0 si $x = 0$.

On sait que ces estimateurs sont non biaisés asymptotiquement et normaux. Dans le cas bivarié,

il n'y a qu'un seul paramètre à estimer, on s'applique à faire correspondre l'estimation sous le modèle du tau de Kendall avec son estimation empirique. Autrement dit, l'estimateur $\hat{\theta}$ vérifie :

$$\tau(\hat{\theta}) = \hat{\tau}_{1,2}$$

Si $\theta \rightarrow \tau(\theta)$ est inversible alors :

$$\hat{\theta} = \tau^{-1}(\hat{\tau}_{1,2})$$

Cette méthode fonctionne de la manière en utilisant le rho de Spearman, ou d'autres coefficients de dépendance. La méthode d'inversion du tau de Kendall ou du rho de Spearman est semi-paramétrique, puisque les marges F_1, \dots, F_d que l'on aperçoit dans l'expression des coefficients empiriques sont implicitement estimées empiriquement.

Cette méthode des moments par inversion du tau de Kendall est très utilisée pour estimer les paramètres de la matrice de corrélation P des copules elliptiques. En effet, pour chaque marge bivariée de la copule, il y a une correspondance entre la ligne et la colonne correspondante de P et le tau de Kendall.

Cas multivarié

Dans un cas multivarié où il n'y a qu'un seul paramètre à estimer (copules archimédiennes par exemple), il y a cette fois plusieurs paires à prendre en compte. L'estimateur vérifie alors :

$$\tau(\hat{\theta}) = \frac{1}{d(d-1)/2} \sum_{i < j} \hat{\tau}_{i,j}$$

où les $\hat{\tau}_{i,j}$ sont les tau de Kendall empiriques des paires (X_i, X_j) . Les propriétés asymptotiques de $\hat{\theta}$ peuvent toujours être établies.

3.2.2 Méthode du maximum de vraisemblance

En approche paramétrique, on suppose que chaque marge F_i appartient à une famille de lois indexée par un paramètre α_i . La densité de notre échantillon s'écrit donc :

$$f(x_1, \dots, x_d; \alpha_1, \dots, \alpha_d; \theta) = c(F_1(x_1; \alpha_1), \dots, F_d(x_d; \alpha_d); \theta) \prod_{i=1}^d f_i(x_i; \alpha_i)$$

Pour estimer le vecteur des paramètres $(\alpha_1, \dots, \alpha_d, \theta)$, il faut maximiser la fonction de vraisemblance suivante :

$$L(\alpha_1, \dots, \alpha_d; \theta) = \prod_{k=1}^n f(X_1^{(k)}, \dots, X_d^{(k)}; \alpha_1, \dots, \alpha_d, \theta).$$

On peut remarquer que cette vraisemblance peut être très compliquée à calculer, ou que l'optimisation numérique peut être lente ou complexe.

Nous allons préférer pour notre étude la méthode IFM se basant sur la vraisemblance.

3.2.3 Méthode Inference Functions for Margins (IFM)

La méthode IFM consiste à estimer les paramètres $\alpha_1, \dots, \alpha_n$ des marginales par la méthode du maximum de vraisemblance, puis d'incorporer ces estimateurs dans l'expression de la log-vraisemblance de la copule. Enfin, il s'agira de maximiser cette expression du log-vraisemblance de la copule pour obtenir un estimateur du paramètre de la copule $\hat{\theta}$. Autrement dit il faut

maximiser la log-vraisemblance suivante :

$$l(\alpha) = \sum_{j=1}^n \log(c(F_1(x_{j,1}; \alpha_1), \dots, F_n(x_{j,n}; \alpha_n); \theta)) + \sum_{j=1}^m \sum_{i=1}^n \log(f_i(x_{j,i}; \alpha_i))$$

D'abord il faudra estimer les paramètres $\alpha_1, \dots, \alpha_n$ par maximisation des log-vraisemblances marginales l_i :

$$\hat{\alpha}_i = \arg \max_{\alpha_i} \sum_{j=1}^n \log(f_i(x_{j,i}; \alpha_i))$$

Ensuite, il s'agira d'estimer le paramètre θ de la façon suivante :

$$\hat{\theta} = \arg \max_{\theta} \sum_{j=1}^m \log(c(F_1(x_{j,1}; \hat{\alpha}_1), \dots, F_n(x_{j,n}; \hat{\alpha}_n); \theta))$$

Les estimateurs obtenus par la méthode IFM, que l'on note $\theta_{IFM} = (\alpha_1, \dots, \alpha_n, \theta)$, sont sans biais, convergent et asymptotiquement normaux. A noter que la méthode IFM, tout comme la méthode du maximum de vraisemblance, est sensible aux éventuelles erreurs dans l'estimation des lois marginales.

3.3 Choix de la bonne copule

Après avoir estimé le paramètre de la copule, il semble important de trouver la copule paramétrique la plus compatible avec notre jeu de données. Nous allons ici présenter deux méthodes permettant de comparer la copule empirique calculée à partir du jeu de données (copule de référence) avec la copule d'intérêt (copule paramétrique théorique).

3.3.1 La fonction $K(z)$

Cette fonction se définit comme étant la fonction de la répartition de la variable aléatoire $C(U, V)$:

$$K(z) = F(C(U, V) < z)$$

Pour une copule archimédienne, cette fonction se définit comme suit :

$$K(z) = z - \frac{\phi(z)}{\phi'(z)}$$

Vous trouverez ci-dessous un tableau exprimant chaque fonction K selon leur copule archimédienne :

Copule	Paramètre a	$\phi(x)$	$K(z)$
Gumbel	$[1, \infty)$	$(-\log(x))^a$	$z(1 - \frac{1}{a} \log(z))$
Clayton	$(0, \infty)$	$\frac{t^{-a}-1}{a}$	$z + az(1 - z^{1/a})$
Frank	\mathbb{R}^*	$-\log\left(\frac{e^{-ax}-1}{e^{-a}-1}\right)$	$z + \frac{1}{a} \log\left(\frac{1-e^{-az}}{1-e^{-a}}\right)$
AMH	$[-1, 1)$	$\frac{1-a}{e^t-a}$	$z + 1 - ae^{-z}$

TABLE 3.2 – Copules archimédiennes avec leurs paramètres, leurs générateurs ainsi que la fonction $K(z)$ associée

A partir d'un échantillon d'observation $(x_1, y_1), \dots, (x_n, y_n)$ issu d'un vecteur aléatoire (X, Y) , nous pouvons désormais estimer la fonction K à partir de cet échantillon. Pour ce faire, il faut suivre la procédure suivante :

1. Définir la pseudo observation z_i pour chaque $i=1, \dots, n$ tel que

$$z_i = \frac{\text{nombre de paires } \{x_j, y_j\} \text{ telles que } x_j < x_i \text{ et } y_j < y_i}{n - 1}$$

2. Définir l'estimateur non-paramétrique de K comme suit :

$$K_n(z) = \frac{\{\text{nombre de } z_i \leq z\}}{n}$$

Nous pourrions ainsi être en mesure d'effectuer une comparaison graphique entre K_a et l'estimateur non paramétrique de K_n calculé à partir de l'échantillon.

En règle générale cette simple comparaison graphique mettra assez nettement en évidence la copule archimédienne à utiliser. Cependant, cette méthode ne peut pas être assez précise en cas de différence non-visible par comparaison graphique. Il faut pouvoir quantifier plus précisément l'écart entre la copule empirique de référence et la copule théorique.

3.3.2 Test du χ^2

La sélection de la meilleure copule doit s'appuyer sur un test statistique. Un test du χ^2 peut être un outil intéressant pour pouvoir tester l'ajustement d'une distribution paramétrique à une distribution empirique et ainsi choisir la bonne copule paramétrique. Il a l'avantage de s'adapter aux distributions multivariées. Il nécessite cependant un découpage en classes, ce qui peut réduire sa puissance puisque le choix des classes peut toujours être critiqué. Il est néanmoins facile d'utilisation.

Le processus d'une sélection de copule se décrit différemment si l'on est amené à effectuer un test du χ^2 .

1. Estimer le paramètre de la copule(par maximum de vraisemblance, IFM)

2. Calcul de la copule empirique

Il s'agit ici de calculer la copule empirique à partir d'un treillis défini.

3. Découper les intervalles de la distribution empirique

Il définit ici un nombre k d'intervalles découpant chacune des marginales uniformes. Les intervalles sont identiques entre chaque marginale uniforme, sans a priori sur la distribution empirique bivariable.

4. Calcul des effectifs et construction d'une table de contingence bivarié

Dans cette étape, il faut calculer les effectifs empiriques et théoriques sur chacun des k^2 intervalles bidimensionnels. Pour ce faire, il nous faut calculer les probabilités bivariées d'appartenance à un intervalle. On construira la table de contingence avec $f_{i,j}, \hat{f}_{i,j}$ étant l'effectif calculé en multipliant la probabilité bivariée d'appartenance à un intervalle au nombre total d'observations.

5. Application critère de Cochran et regroupement en classes

Il faut pouvoir regrouper les intervalles en prenant en compte les intervalles d'effectif faible voir nul dans les regroupements. Les regroupements doivent respecter le critère de Cochran qui recommande des effectifs théoriques au moins supérieurs à 1% du nombre total d'observations dans chaque classe, et supérieurs à 5% du nombre total d'observations dans au moins 80% des classes. Ce regroupement est aussi à effectuer sur le tableau bivarié des effectifs empiriques. Selon les zones de dépendance recherchées, on pourra notamment isoler les cellules des queues de distribution.

6. Calcul du χ^2 et sélection de la copule

Après regroupement en n classe, on peut calculer la statistique bivariée du χ^2 observée comme suit :

$$\chi_{obs}^2 = \sum_{j=1}^n \frac{(O_j - E_j)^2}{E_j}$$

O_j et E_j représentent respectivement les effectifs observés et attendus dans chaque classe j, j=1,...,n. Cette statistique suit une loi du χ^2 (n-r-1) degrés de liberté avec r le nombre de paramètres estimés de la copule. Il faut ensuite se donner un seuil α pour savoir si l'on rejette l'hypothèse d'acceptation de la copule théorique ou non. On calcule ainsi la p-valeur $\hat{\alpha} = \mathbb{P}(\chi_{n-r-1}^2 > \chi_{obs}^2)$. On rejette l'adéquation de la copule paramétrique si $\hat{\alpha} < \alpha$. Parmi les copules acceptées (s'il y en a plusieurs), on sélectionne celle qui maximise la p-valeur.

Chapitre 4

Présentation, gestion et analyse de la base de données

L'objet de ce chapitre est de présenter la base de données sur laquelle nous allons nous appuyer tout au long de notre étude. Une brève mise en contexte de la sûreté aéroportuaire aux Etats-Unis sera introduite. Puis nous procéderons à une description plus détaillée des variables à prendre en compte pour nos modèles. Après avoir abordé les différentes étapes de traitement de la base, quelques études seront exposées concernant le taux de sinistres sans suite et le montant de sinistres clos.

4.1 Présentation de la base et mise en contexte

La base de données choisie provient de la TSA (Transportation Security Administration), l'agence nationale américaine se chargeant de la sécurité dans les transports (aériens notamment). Cette agence fût créée à la suite des attentats du 11 Septembre 2001 afin de mieux protéger la population lors de différents trajets.

Une des fonctions principales de la TSA consiste à organiser des processus de contrôle des passagers et de leurs objets personnels dans différents aéroports du pays avant des embarquements.

Lors de ces différents check-points avant un embarquement, il peut malheureusement arriver qu'un sinistre subi par le passager à l'encontre de la TSA se produise. Blessures, pertes d'effets personnels (bagages, vêtements, flacons..etc.), dégradation d'objets ou encore vols d'objets peuvent se produire lors des procédures de contrôle.

La TSA a pu répertorier dans cette base l'ensemble des sinistres déclarés par des individus à l'encontre de la TSA pendant la période 2002 à 2015. Ces 145012 sinistres se sont produits dans les différents aéroports des Etats-Unis où des employés de la TSA réalisent les procédures de contrôle.

Le but de notre étude sera de modéliser les charges de sinistre pour chaque type de sinistre. Puis nous étudierons les dépendances existantes entre ces types de sinistre à l'aide de copules.

4.1.1 Description des variables disponibles

Chaque ligne de notre base correspond à un sinistre déclaré par une personne à l'encontre de la TSA entre 2002 et 2014. La base de sinistres s'articule autour de 12 variables décrivant de façon détaillée le sinistre déclaré :

- Numéro de sinistre ;

- Nom de la compagnie aérienne ;
- Type de sinistre ;
- Site du sinistre ;
- Item ;
- Statut (sinistre reconnu, rejeté,..etc.) ;
- Date de déclaration du sinistre ;
- Date de survenance du sinistre ;
- Code de l'aéroport ;
- Nom de l'aéroport ;
- Charge de sinistre déclaré ;
- Charge de sinistre clos ;

Les évènements du sinistre étant courts, nous avons remarqué lors du traitement de la base que les numéros de sinistre étaient différents pour chaque personne sinistrée, ce qui veut dire que les sinistres impactent qu'à une petite échelle, au niveau du sinistre ou de sa propriété (bagages,objets..etc.). Dans notre volonté de recherche de corrélation menant à une dépendance des types de sinistre, l'agrégation selon le numéro du sinistre n'a ici pas d'utilité. Le numéro du sinistre ne fera donc pas partie des éléments importants à prendre en compte pour la transformation de notre base de données.

Parmi les variables à retenir pour la transformation de notre base, le type de sinistre, la date de survenance du sinistre ainsi que les charges de sinistre déclarés et clos seront cruciaux.

La date de survenance du sinistre est ici préférée à la date de déclaration du sinistre car nous cherchons à modéliser des dépendances sur des sinistres se produisant à la même date. Les sinistres subis sont tous de portée courte : l'évènement subi a une durée courte. C'est pourquoi nous effectuerons une agrégation des charges de sinistre selon les dates de survenance de sinistre. On fait donc l'hypothèse que les sinistres se produisant un jour donné sont indépendants des sinistres subis à une autre date.

Les charges de sinistres clos serviront de base pour l'étude des dépendances, mais l'étude au préalable des charges de sinistre déclarés et du ratio clos/déclarés permettra de mieux se représenter la base de données.

4.1.2 Présentation des types de sinistre

Nous disposons des données relatives aux types de sinistre suivants :

- Dommages aux biens ;
- Perte de biens des passagers ;
- Dommages corporels ;
- Véhicule motorisé ;
- Vols de passagers ;
- Perte ou dommage matériel lié au travail d'un employé de la TSA ;
- Terminal d'autobus ;
- Réclamation ;

Les dommages aux biens ainsi que la perte de biens des passagers sont les types de sinistre les plus récurrents dans la base de données. Voici un tableau répertoriant l'occurrence des sinistres pour chaque type de sinistre :

Types de sinistres	Occurrences dans la base
Dommages aux biens	56 138
Perte de biens	81 838
Dommages corporels	690
Véhicule motorisé	214
Vols des passagers	427
Perte employé	419
Terminal d'autobus	1
Réclamation	3

TABLE 4.1 – Nombre de sinistres par type de sinistre

Nous étudierons prochainement les taux de sinistres clos selon les types de sinistres. Après avoir analysé les corrélations entre types de sinistres, nous modéliserons les charges de sinistres entre types de sinistre présentant une dépendance par le biais des copules.

4.2 Analyse des données et transformation de la base

Pour mener à bien le travail de corrélations entre types de sinistres, nous avons besoin d'analyser puis de transformer en profondeur notre base de données. Pour ce faire, nous avons travaillé depuis le logiciel Excel avec l'appui de diverses fonctions Mind (fonctions d'agrégation, de filtre notamment). Dans le cadre d'études de corrélations, nous travaillerons sur le logiciel R avec l'aide de tables csv exportées depuis Milliman Mind.

4.2.1 Analyse du nombre de sinistres refusés

Les charges de sinistre déclarés dont le montant est égal à 0 n'ont aucune utilité pour la suite de notre étude, car l'étude se base avant tout sur un sinistre déclaré non-négligeable. C'est pourquoi nous avons décidé de retirer de notre base ces nombreux montants nuls de sinistre déclarés.

Nous appelons par sinistre refusé un sinistre déclaré dont la TSA refuse d'indemniser la personne sinistrée après vérification de la part d'experts. Concrètement, nous pouvons détecter ce types de sinistres dans notre base en se référant aux montants nuls ou vides de charges de sinistres clos. Il existe également la variable "Statut" décrivant l'acceptation ou le refus du sinistre prononcé par la TSA. Cependant, des incohérences au niveau de la fiabilité de cette variable ont été repérés (par exemple un sinistre refusé mais un montant de sinistre clos non nul). C'est pourquoi l'étude des sinistres clos semble plus adéquate à l'analyse de ces sinistres refusés.

Nous remarquons tout d'abord que le nombre de sinistres refusés est assez conséquent dans notre base. Il y a 82 048 sinistres refusés sur notre base contenant 145 012 sinistres, soit environ 56,6% de la base. Plus d'un sinistre sur 2 est donc refusé par la TSA sur la période 2002 à 2015. Les déclarations sont refusées la plupart du temps lorsqu'il a été établi que la TSA n'est aucunement responsable d'une perte ou de dommages sur un objet (bagage non-ouvert par un employé de la TSA par exemple). Il peut arriver également que le contrôle du passager et de ses bagages se fasse sans aucun contact physique avec l'employé de la TSA.

Dans le détail, voici la répartition de sinistres refusés selon le type de sinistre :

Types de sinistres	Nb sinistres refusés	Nb sinistres refusés /Nb sinistres du type	Nb sinistres refusés /Nb sinistres global
Dommages aux biens	28935	51,54%	19,95%
Perte de biens	50471	61,67%	34,80%
Dommages corporels	506	73,33%	0,3%
Véhicule motorisé	105	49,06%	0,2%
Vols des passagers	168	39,34%	0,37%
Perte employé	151	36,75%	0,33%
Terminal d'autobus	1	100%	0%
Réclamation	3	100%	0%

TABLE 4.2 – Nombre de sinistres refusés par type de sinistre

La part de sinistres refusés au sein d'un type de sinistre se trouve être particulièrement conséquente. Les montants non-nuls de charges de sinistres clos seront conservés pour l'étude des dépendances.

4.2.2 Analyse du nombre de sinistres déclarés

Bien que l'étude porte avant tout sur la sinistralité, il semble important d'évaluer également le nombre de sinistres déclarés dans le temps. En effet, la période étudiée de 14 ans est longue et il peut sembler probable que la répartition des sinistres soit fluctuante selon les années. Étudions tout d'abord le nombre de sinistres déclarés tous types confondus :

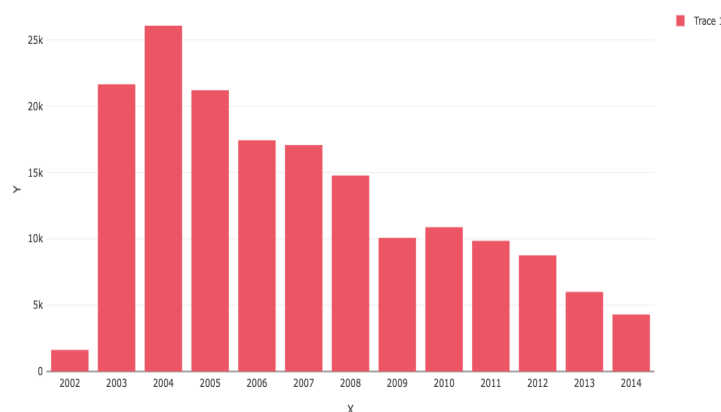


FIGURE 4.1 – Évolution du nombre de sinistres selon l'année d'incidence

On observe ici, après une année 2002 avec un nombre de sinistres faible due à la nouveauté du dispositif de déclaration de sinistre (la TSA ne fut créée que fin 2001), une augmentation soudaine du nombre de sinistres survenus avec un pic à 26070 sinistres survenus en 2004. Une baisse progressive se produit après 2005. En 2009, il est possible que la baisse conséquente observée est due aux effets de la crise de 2008, provoquant notamment une baisse de départ en vacances et de voyages en avion.

Étudions désormais l'évolution du nombre de sinistres survenus non-nuls (sinistre clos non-nul) selon chaque type de sinistre.

Avant même d'obtenir des résultats, on peut s'attendre, au vu du nombre conséquent de sinistres survenus en 2004, à un risque d'obtenir un nombre élevé de sinistres en 2004 dans chacune des catégories.

Comme vu précédemment, les types de sinistre les plus fréquents dans notre base sont les dommages aux biens et la perte de biens matériels.

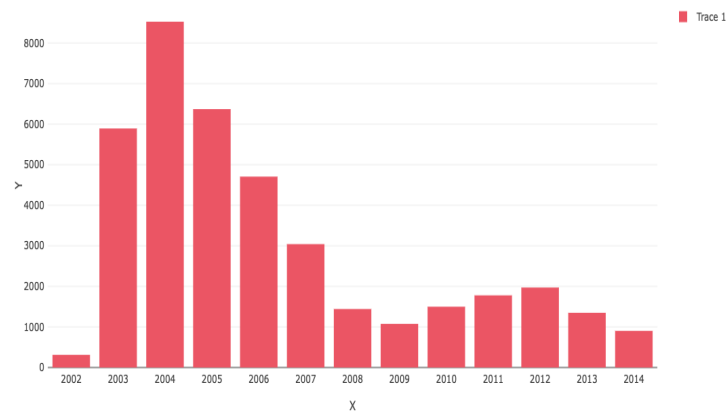


FIGURE 4.2 – Évolution du nombre de sinistres clos non-nuls dans la catégorie "perte de biens".

On peut noter comme prévu un pic du nombre de sinistres en 2004 pour une baisse progressive. La tendance de l'histogramme est semblable à l'évolution du nombre de sinistres global.

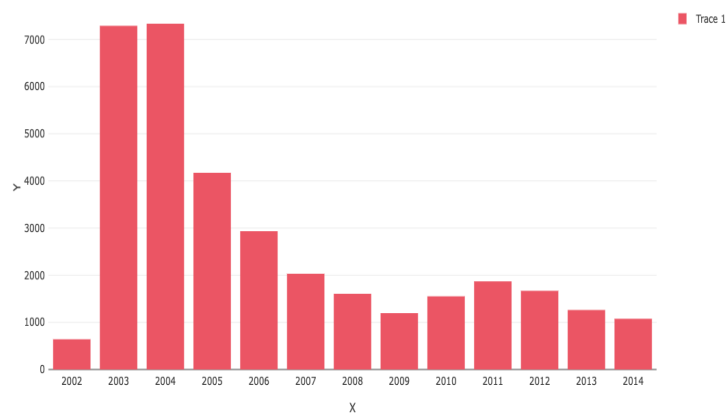


FIGURE 4.3 – Évolution du nombre de sinistres clos non-nuls dans la catégorie "dommages matériels"

Concernant la catégorie des dommages matériels, la tendance est globalement similaire à l'évolution générale. On peut remarquer toutefois un nombre de sinistres plus conséquent en 2003.

La catégorie des dommages corporels, bien que peu présente dans la base de données, épouse aussi globalement la tendance globale du nombre de sinistres survenus :

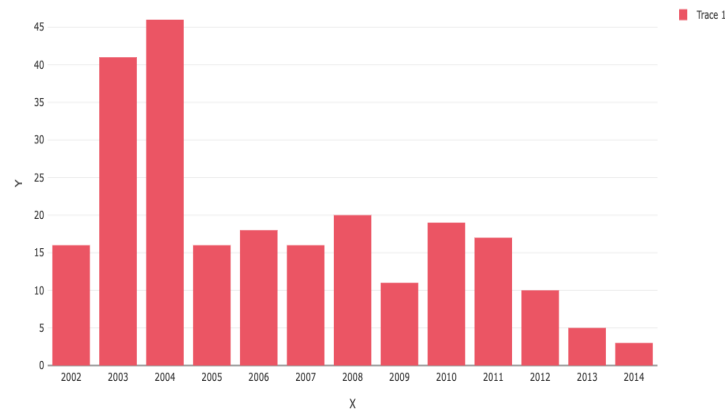


FIGURE 4.4 – Évolution du nombre de sinistres clos non-nuls dans la catégorie "dommages corporels"

La seule nette différence observée par rapport à l'évolution globale du nombre de sinistres concerne l'évolution du nombre de sinistres dans les véhicules motorisés. En effet, on peut remarquer ici une augmentation du nombre de sinistres qui atteint son pic en 2010 avant de diminuer de façon drastique.

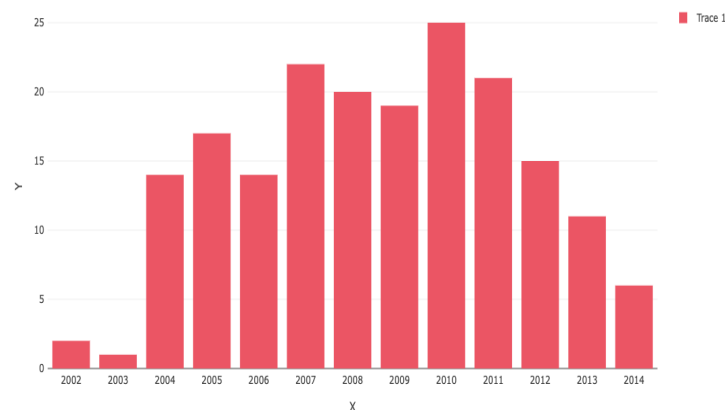


FIGURE 4.5 – Évolution du nombre de sinistres clos non-nuls dans la catégorie "véhicule motorisé"

Les autres types de sinistres tels que la perte de bien d'un employé ou le vol de biens d'un passager ont un nombre de sinistre croissant jusqu'en 2004. La réclamation ainsi que les sinistres liés au terminal d'autobus sont trop peu fréquents pour être raisonnablement représentés ici (de l'ordre de 3 ou 4 occurrences au plus).

4.2.3 Traitement des montants de sinistres clos "as if" 2014

Avant de pouvoir effectuer une agrégation des montants de sinistres clos non-nuls selon le type de sinistre et la date d'incidence, il faut pouvoir examiner et retraiter les montants de sinistre. En effet, bien que les sinistres soient globalement peu onéreux dans cette base de données, un sinistre survenu en 2002 ne coûtera pas le même montant, toutes choses égales par ailleurs (gravité du sinistre, objets et types de sinistres..etc.), que s'il survient en 2014. Les coûts généraux des biens et des produits sont variables en fonction du temps et liés par des indices représentatifs de ces biens.

Les biens sinistrés dans notre cas étant principalement des biens de la consommation courante (bagages, petits objets, véhicules,..etc), nous avons retenu ici l'indice des prix à la consommation pour l'indexation des sinistres. Naturellement, les dommages corporels ne peuvent être livrés à un indice ici car ce type de sinistre porte atteinte au corps humain et non à un bien matériel. Leurs montants ne varieront donc pas selon qu'ils se soient produits en 2002 ou 2014.

Concrètement, nous allons, à l'aide de cet indice des prix à la consommation, estimer tous les montants de sinistres comme s'ils s'étaient produits à une année donnée, 2014 dans notre cas.

L'opération est la suivante. Soit :

- n l'année de cotation
- I_k l'indice de consommation de l'année k
- S_k le montant de sinistre de l'année k
- S_k^n le montant de sinistre "as if" d'un sinistre de l'année k vu en année n .

Alors on a : $S_k^n = S_k \times \frac{I_n}{I_k}$.

Les sinistres "as if" sont ainsi obtenus pour chaque montant de la base des sinistres clos (les montants de sinistre déclarés ne sont pas utiles pour la suite du travail entrepris dans ce mémoire). Cette transformation des montants est une étape indispensable pour une bonne significativité de nos résultats de dépendance à venir. Elle est préalable au traitement de la base détaillée dans la section suivante.

4.2.4 Traitement des dates d'incidence, agrégation des sinistres clos et transformation de la base à étudier

Comme vu auparavant, une ligne dans la base de données initiale représente un sinistre avec ses caractéristiques (numéro de sinistre, item, dates, montants,..etc.). Le but de cette partie est de décrire les procédés de retraitement de la base entrepris pour aboutir à une base finie, prête aux calculs de corrélation entre type de sinistres.

Tout d'abord, les dates n'étant pas formatées de la même manière, un travail de formatage a été réalisé. Ces dates permettront entre autres d'extraire des années pour le calcul de sinistres "as if". Les dates d'incidence sont surtout un élément essentiel dans l'agrégation des montants de sinistre.

Il faut également tâcher à bien avoir des montants de sinistres clos correspondant à des nombres (non-nuls, nuls ou cases vides, suppression du \$).

Dans l'optique d'une étude de la corrélation des charges de sinistres entre plusieurs types de sinistre, il nous faut procéder, pour l'obtention de corrélations consistantes, à une agrégation, un cumul de charges de sinistres sur une période donnée et/ou sur un sinistre donné. Nous remarquons à la lecture de la base qu'aucun numéro de sinistre ne comporte de doublons, ce qui signifie que tous les sinistres sont bien déclarés séparément. Un regroupement de sinistres par période s'impose donc. Nous avons remarqué que les types de sinistre présents dans cette base sont plutôt de courte portée. En effet, une perte ou un dommage d'un matériel est un événement soudain et n'a pas de répercussion apparente sur les jours suivants l'évènement. C'est pourquoi, nous avons décidé de sommer les montants de sinistres clos ayant la même date d'incidence. En réalisant cette transformation, nous supposons l'indépendance des événements entre 2 jours

donnés, ce qui paraît être une hypothèse plausible dans ce contexte (milieu aéroportuaire des USA, évènements centrés sur le passager).

Pour procéder à l'agrégation des sinistres, il a fallu tout d'abord concaténer les chaînes de caractère correspondant à la date d'incidence et le type de sinistre. Puis à l'aide de la fonction d'agrégation Mind MM_GROUPBY (fonction disponible par l'add-in Excel téléchargeable depuis le site web Milliman Mind), nous avons pu sommer les montants de sinistres clos selon le type de sinistre et la date d'incidence. Les montants de sinistres clos vides (=NA) produisent en sortie des sommes de montants de sinistres également. Au final, nous choisirons de supprimer les sommes de montant nuls et vides, pour une meilleure clarté lors de l'estimation des corrélations entre charges de sinistre selon les types de sinistre.

Avec ces nouveaux éléments, nous avons pu établir une nouvelle base comportant des sommes de montant de sinistre clos pour chaque date d'incidence. Une colonne correspond à un type de sinistre. Par des fonctions de filtre selon les valeurs acceptées de montants agrégés (sommes à 0 et valeurs vides sont exclus), nous avons pu exporter à l'aide de la plateforme Milliman Mind les tables servant de base pour l'étude des corrélations à venir.

4.3 Étude de corrélations entre types de sinistre et recherche de dépendance

Dans cette section, nous allons exposer le travail de recherche de dépendances ainsi les différentes corrélations obtenues entre nos types de risque.

Comme vu précédemment, nous avons pris comme base ici des sommes de sinistres clos non-nuls. Les montants de sinistres clos sont des montants que la TSA s'est engagée à reverser à la victime du sinistre en guise d'indemnisation. Il apparaît cohérent d'utiliser les sinistres clos plutôt que les montants de sinistre déclarés, car cela prend en compte la non-responsabilité de la TSA dans certains sinistres (les sinistres clos nuls par exemple). Pour l'étude de l'impact des dépendances entre type de sinistre sur les fonds propres de la TSA, il apparaît également important de ne prendre en compte que ce qu'indemnise la TSA.

Enfin, nous prendrons en compte seulement les sommes de montants strictement supérieures à zéro car il risque d'avoir une masse de points en zéro dans les différents dépendogrammes (diagramme selon le rang des montants), ce qui altère le caractère continu souhaité dans une corrélation. Il y a également un risque d'avoir plusieurs rangs ex-aequo à 0, ce qui altère l'estimation de corrélation pour des mesures de corrélation basées sur le rang tel que le rho de Spearman.

4.3.1 Analyse des données finales

Notre base de données finale comporte 4549 lignes, soit le nombre de dates uniques dans la base initiale de données. Le nombre de types de sinistres étant conséquent (8 types de sinistre) nous n'avons pas trouvé de sinistres clos non-nuls subissant tous les types de sinistre le même jour. Cependant, en procédant par couple de types de sinistre, nous avons obtenu des données plus ou moins volumineuses :

- 4245 dates uniques ont eu un impact sur le couple perte de biens/dommages aux biens
- 178 dates uniques sur le couple perte employé/perte de biens
- 53 dates uniques sur le couple perte employé/vol
- 213 dates uniques sur le couple perte de biens/dommages corporels
- 167 dates uniques sur le couple perte de biens/véhicule motorisé
- 178 dates uniques sur le couple dommages aux biens/perte employé

- 214 dates uniques sur le couple dommages aux biens/dommages corporels
- 169 dates uniques sur le couple dommages aux biens/véhicule motorisé

Les autres couples tels que dommages corporels/perce employé ont trop peu d'observations pour être exploités par une étude de dépendance (moins de 20 observations).

Sans surprise, nous observons un volume d'observations plus conséquent pour le couple perte de biens/dommages aux biens, en cohérence avec le grand nombre de sinistres affilié à ces types de sinistre (cf Graphique de la partie précédente).

4.3.2 Recherche de dépendance

Nous cherchons dans cette section la dépendance qui peut être présente entre chaque couple de risque. Seules les dépendances par couple nommées précédemment seront analysées.

Avec le logiciel R, nous avons pu utiliser trois indicateurs estimant la dépendance entre ces types de sinistre : le coefficient de corrélation linéaire de Pearson, le Tau de Kendall ainsi que le Rho de Spearman. Ces calculs seront accompagnés de tests statistiques d'indépendance permettant d'infirmier ou non la présence de dépendance (calcul de la p-value).

Nous compléterons ensuite notre analyse par une représentation graphique de la dépendance. Un dépendogramme (ou Rank-Rank plot), basé sur les rangs des sommes de montants de sinistre précisera les grandes lignes de dépendance. Un Kendall Plot (K-Plot) conclura cette analyse en précisant et en validant les structures de dépendance supposées par nos indicateurs.

Matrice de corrélation de Pearson

Pearson	Perte biens	Dommages biens	Dom. corporels	Véhicule	Employé
Perte biens	1	0,202	-0,109	0,1217	0,033
Dommages biens	0,202	1	0,14	-0,02	0,051
Dom. corporels	-0,109	0,14	1	No data	No data
Véhicule	0,1217	-0,02	No data	1	No data
Employé	0,033	0,051	No data	No data	1

TABLE 4.3 – Coefficient de corrélation de Pearson

Certains couples n'ayant pas assez d'observations ont la mention "no data" dans les tableaux.

Matrice de corrélation de Kendall

Tau Kendall	Perte biens	Dommages biens	Dom. corporels	Vehicule	Employé
Perte biens	1	0,307	-0,0787	-0,012	0,020
Dommages biens	0,307	1	-0,019	-0,057	0,041
Dom. corporels	-0,0787	-0,019	1	No data	No data
Véhicule	-0,012	-0,057	No data	1	No data
Employé	0,020	0,041	No data	No data	1

TABLE 4.4 – Tau de Kendall

Matrice de corrélation de Spearman

Rho Spearman	Perte biens	Dommages biens	Dom. corporels	Vehicule	Employé
Perte biens	1	0,45	-0,114	-0,026	0,033
Dommages biens	0,45	1	-0,026	-0,093	0,052
Dom. corporels	-0,114	-0,026	1	No data	No data
Véhicule	-0,026	-0,093	No data	1	No data
Employé	0,033	0,052	No data	No data	1

TABLE 4.5 – Rho de Spearman

Avant de pouvoir interpréter ces valeurs, rappelons l'objectif de chaque indicateur. Le tau de Kendall, tout comme le rho de Spearman, est basé sur les rangs des valeurs, tandis que le coefficient de corrélation de Pearson se base uniquement sur les valeurs entrées. Le tau de Kendall a l'avantage de mieux gérer les valeurs ex-aequo (sur la base de paires discordantes et concordantes) par rapport au rho de Spearman. Le coefficient linéaire de Pearson mesure quant à lui la monotonie de la corrélation (tendance affine). On remarque par ailleurs que le coefficient de Pearson peut vite varier selon la présence de valeurs aberrantes.

Chaque indicateur a une valeur comprise entre -1 et 1, 0 désignant une indépendance parfaite, 1 une corrélation positive parfaite, et -1 une corrélation négative parfaite. Il se distingue de l'ensemble de ces résultats une corrélation entre la perte de biens et les dommages aux biens. Un autre couple se distinguant par une possible corrélation est le couple Dommages corporels/Perte de biens.

Plusieurs couples disposent d'une relative corrélation de Pearson (Véhicule/perte de biens ou encore dommages corporels/Dommages de biens) mais ont des rho de Spearman et tau de Kendall proches de zéro. Cela peut s'expliquer par une tendance linéaire sans dépendance claire.

Vérifions ces résultats par des tests statistiques d'indépendance.

Ces derniers devraient permettre de savoir si les coefficients exposés plus haut sont significativement différents de zéro ou non.

Les p-values sont ci-dessous associés à un risque d'erreur de 5%.

p-values	Perte biens	Dommages biens	Dom. corporels	Véhicule	Employé
Perte biens		$< 2,2 \times 10^{-16}$	0,114	0,1194	0,6651
Dommages biens			0,03953	0,7451	0,4968
Dom. corporels				No data	No data
Véhicule					No data
Employé					

TABLE 4.6 – P-value du test de corrélation de Pearson

p-values	Perte biens	Dommages biens	Dom. corporels	Véhicule	Employé
Perte biens		$< 2,2 \times 10^{-16}$	0,09	0,7444	0,69
Dommages biens			0,6812	0,2762	0,4215
Dom. corporels				No data	No data
Véhicule					No data
Employé					

TABLE 4.7 – P-value du test de corrélation de Kendall

p-values	Perte biens	Dommages biens	Dom. corporels	Véhicule	Employé
Perte biens		$< 2,2 \times 10^{-16}$	0,098	0,7435	0,663
Dommages biens			0,706	0,2355	0,489
Dom. corporels				No data	No data
Véhicule					No data
Employé					

TABLE 4.8 – P-value du test de corrélation de Spearman

On observe ici que la plupart des p-value affichées ne rejettent pas l'hypothèse nulle (H_0 = indépendance) au seuil de risque 5%. Seul le couple Dommages aux biens/Perte de biens rejette clairement l'hypothèse nulle. Au seuil 10%, le couple dommages corporels/Perte de biens rejette également l'hypothèse nulle d'indépendance.

On peut noter toutefois une p-value assez faible pour le couple dommages aux biens/Véhicule (0,2762 et 0,2355) pour les tests de corrélation de Spearman et Kendall. Mais le test de Pearson ne suit pas cette tendance (0,7451).

Analysons désormais graphiquement les corrélations pour valider l'hypothèse de dépendance ou non.

Dépendogrammes de couples de risque

Les dépendogrammes permettent d'appréhender de manière graphique la structure de dépendance de deux variables aléatoires, dans notre cas, deux séries de données portant chacune sur des montants agrégés de sinistres pour un type donné. Deux marges uniformes seront créées en

utilisant la classification par rang pour chacun de nos couples.

On parlera de dépendance positive si une majorité de points est sur la diagonale du bas vers le haut du dépendogramme. A l'inverse, on évoquera une dépendance négative si une majorité de points est située sur la diagonale de haut en bas du dépendogramme.

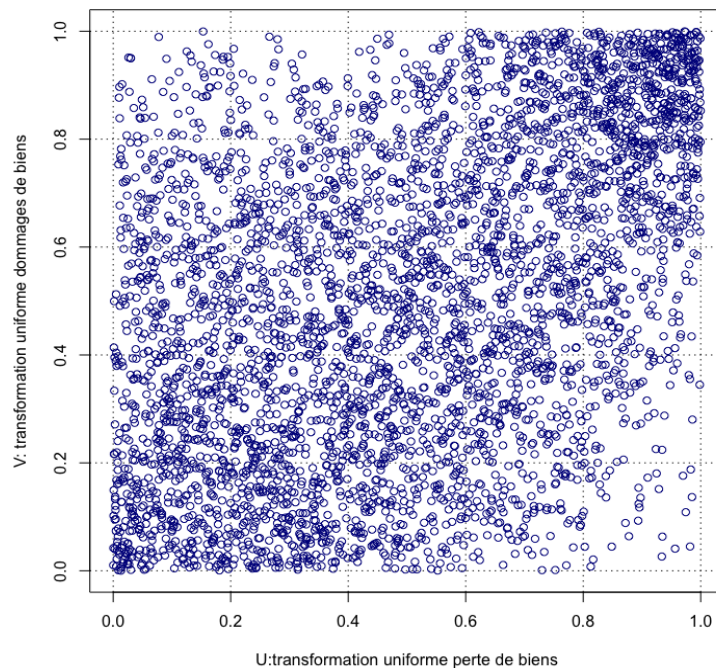


FIGURE 4.6 – Dépendogramme du couple perte de biens /dommages aux biens

On peut voir sur le graphique ci-dessus une majorité de points sur la diagonale du bas vers le haut, déterminant ainsi clairement une dépendance positive entre le type « perte de biens » et « dommages aux biens ». Ainsi, si un montant de sinistre sur la journée est élevé dans la catégorie « perte de biens », il y a une grande probabilité pour que la charge de sinistres liée aux dommages aux biens soit élevée également à la même date. On observe sur ce dépendogramme un amas plus élevé de point sur le coin supérieur droit par rapport au coin inférieur gauche. Cela peut traduire à terme une asymétrie entre les dépendances de queue, sujet sur lequel nous nous pencherons dans le chapitre suivant.

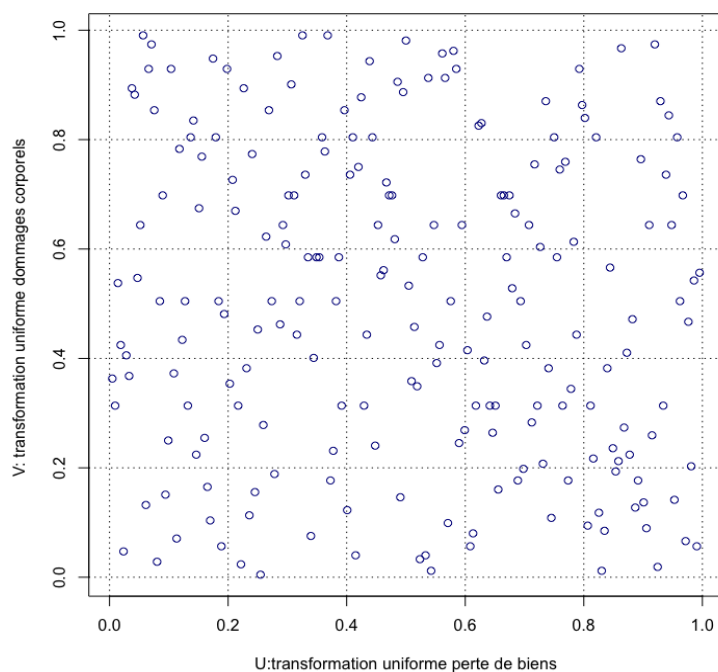


FIGURE 4.7 – Dépendogramme du couple perte de biens/dommages corporels

On distingue ici une plus grande densité de points sur la diagonale de haut vers le bas, traduisant la possibilité d'une dépendance négative entre le couple « perte de biens » et « dommages corporels ». Nous pouvons apercevoir plus spécifiquement un amas de points dans le coin inférieur droit du dépendogramme. Cette dépendance négative n'est cependant pas clairement établie, comme les indicateurs de Spearman et de Kendall l'indiquent. Le faible nombre de valeurs n'aide également pas à la détermination d'une dépendance forte.

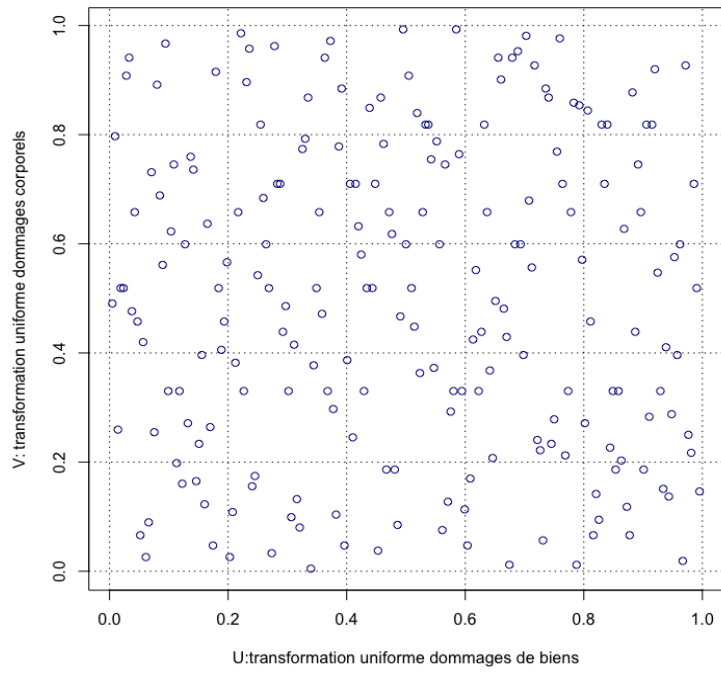


FIGURE 4.8 – Dépendogramme du couple dommages aux biens/dommages corporels

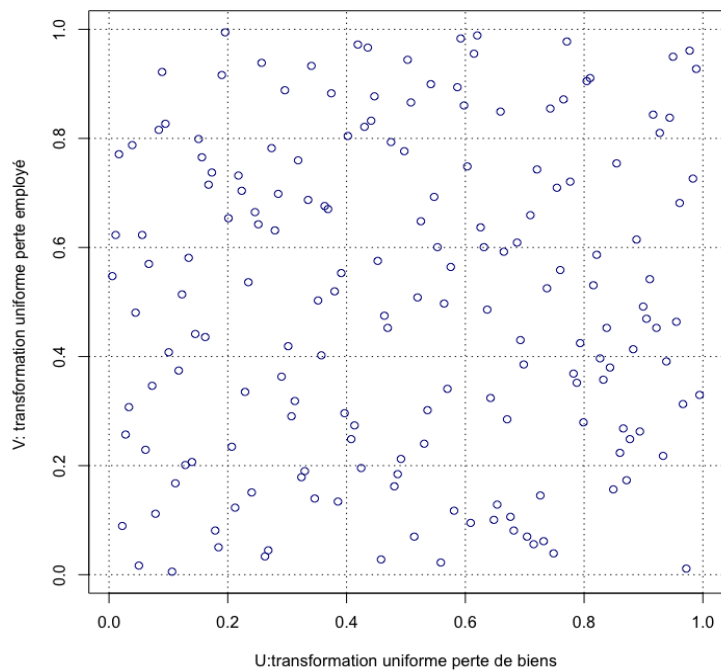


FIGURE 4.9 – Dépendogramme du couple perte employé/perte de biens

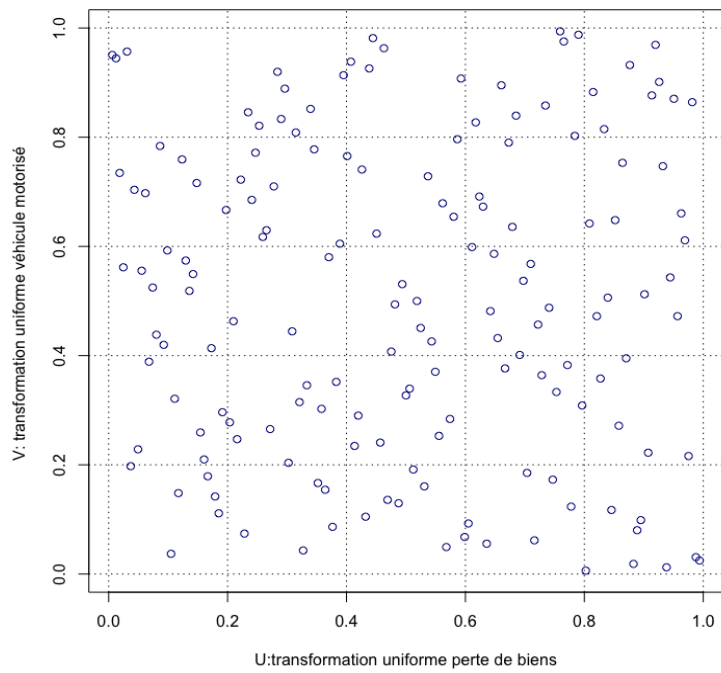


FIGURE 4.10 – Dépendogramme du couple perte de biens/véhicule motorisé

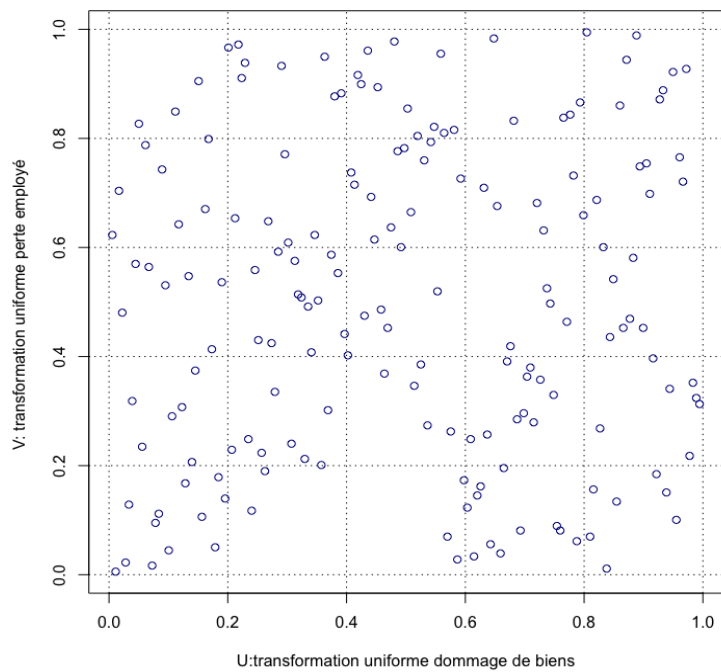


FIGURE 4.11 – Dépendogramme du couple dommages aux biens/perte employé

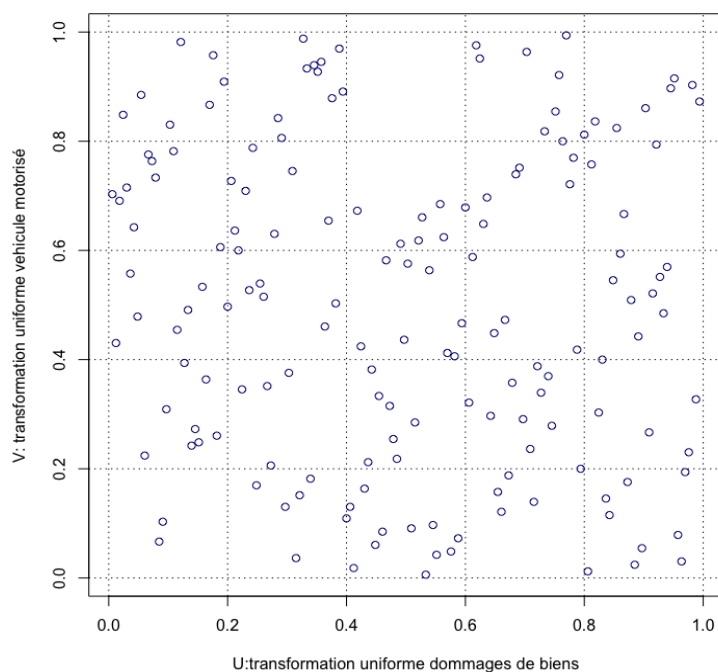


FIGURE 4.12 – Dépendogramme du couple dommages aux biens/véhicule motorisé

On peut remarquer ici que la répartition des points n'est pas parfaite, ce qui permet de rejeter l'hypothèse d'indépendance entre ces deux types de risques. Nous pouvons distinguer sur ce dépendogramme un regroupement légèrement plus prononcé sur la diagonale du haut vers le bas de ce graphique, ce qui traduit une tendance à la dépendance négative dans le couple véhicule motorisé/dommages aux biens. Cependant, le faible nombre de points ainsi que les résultats de corrélations (tau de Kendall, rho de Spearman et leurs p-values) ne permettent pas de confirmer formellement cette hypothèse fondée sur la représentation graphique.

K-Plot de couples de risque

Tout comme les dépendogrammes, les K-Plot décrivent la structure de dépendance pour deux risques donnés. Ce graphique trace l'espérance des statistiques d'ordre sous l'hypothèse nulle d'indépendance par rapport aux valeurs de la copule empirique ordonnée. Les données sont transformées en rangs.

Concrètement, des points sur la diagonale de bas en haut signifient une indépendance entre les deux types de sinistre. Des points au-dessus de la diagonale et en dessous de la courbe incurvée signifient une dépendance positive entre les deux types de risque. À l'inverse, des points en dessous de la diagonale signifient une dépendance négative entre les deux types de risque. Des points sur la courbe incurvée désignent une dépendance positive parfaite entre les deux types de sinistre, tandis que des points situés sur l'axe des abscisses désignent une dépendance négative parfaite entre les deux types de sinistre.

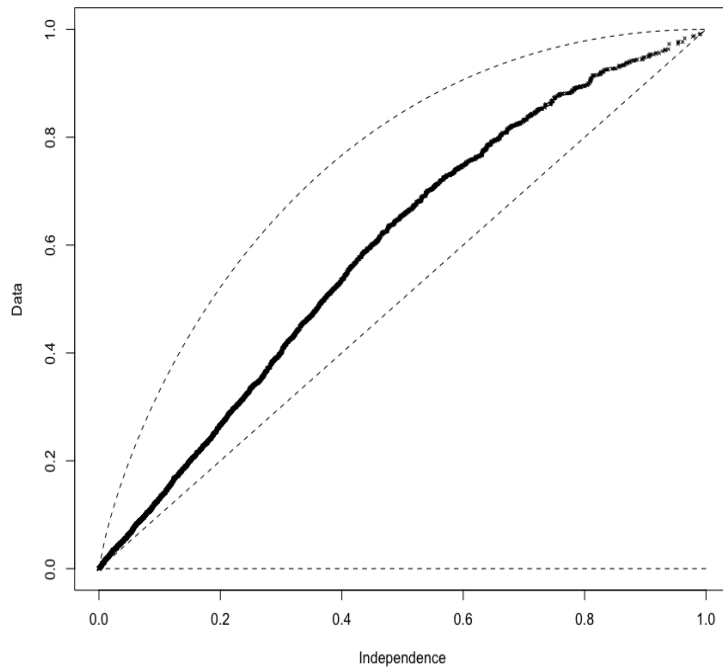


FIGURE 4.13 – K-Plot du couple perte de biens/dommages aux biens

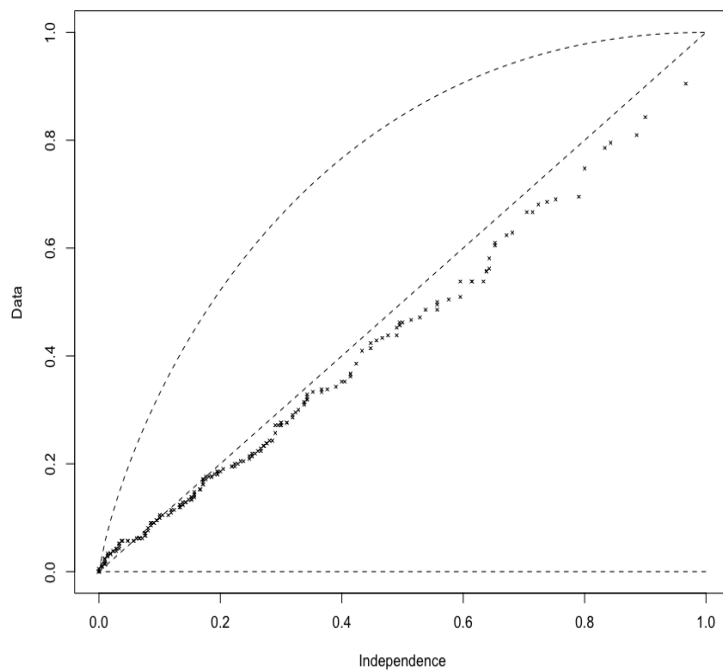


FIGURE 4.14 – K-Plot du couple perte de biens/dommages corporels

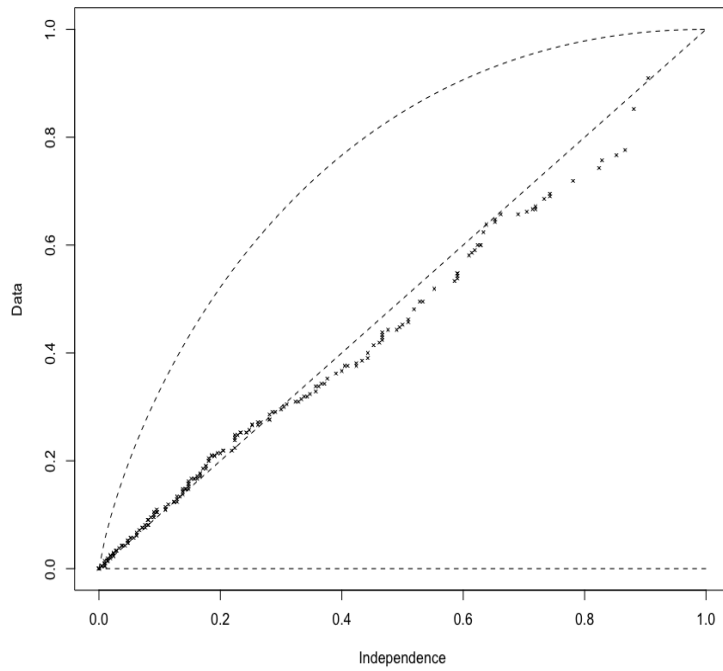


FIGURE 4.15 – K-Plot du couple dommages aux biens/dommages corporels

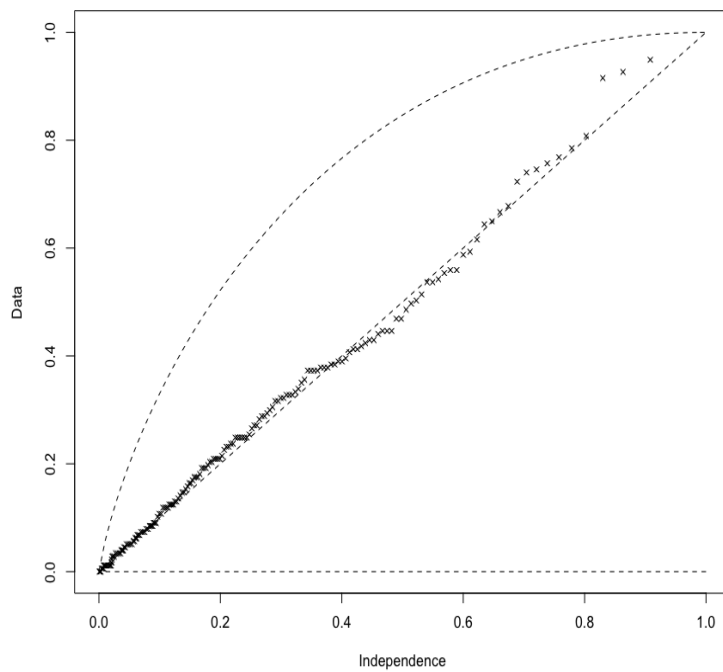


FIGURE 4.16 – K-Plot du couple perte employé/perte de biens

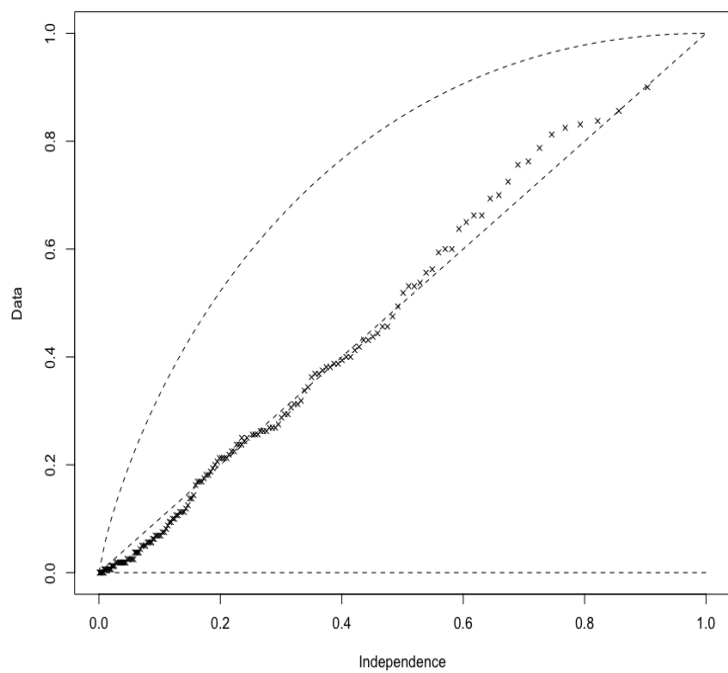


FIGURE 4.17 – K-Plot du couple perte de biens/véhicule motorisé

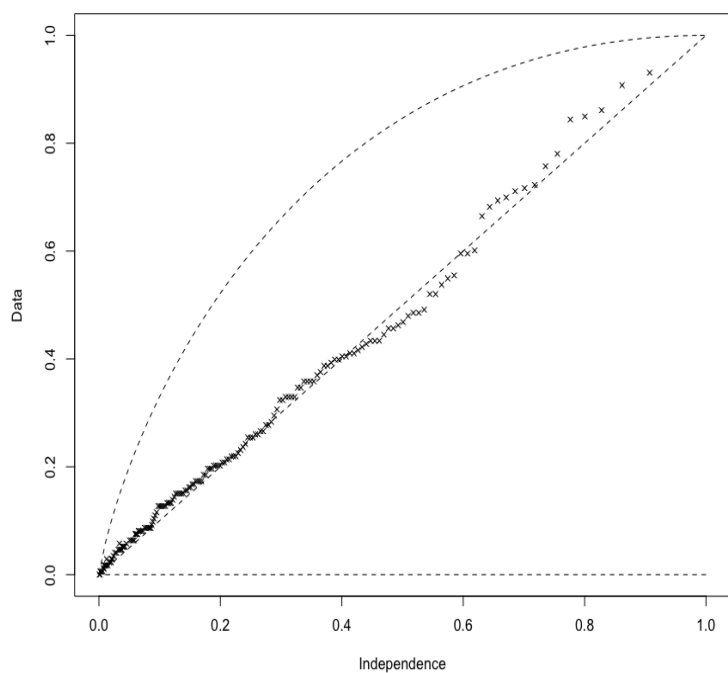


FIGURE 4.18 – K-Plot du couple dommages aux biens/perte employé

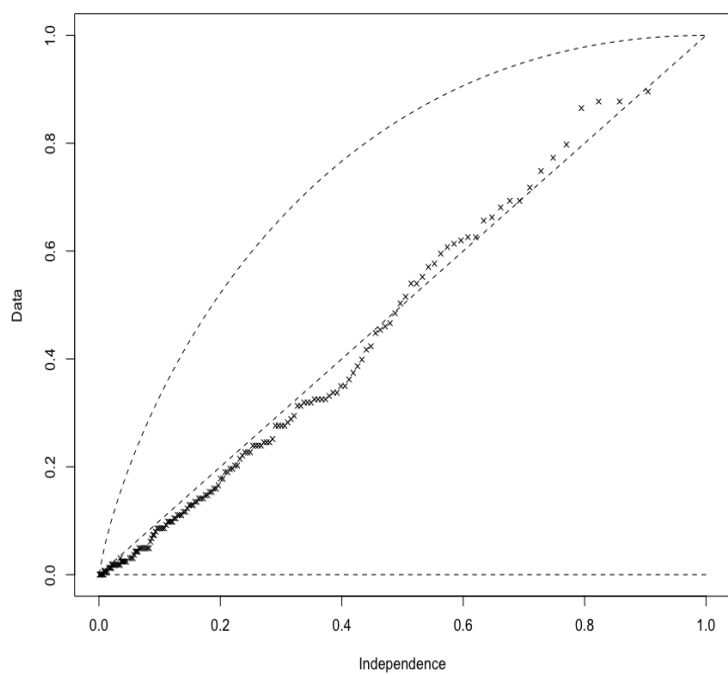


FIGURE 4.19 – K-Plot du couple dommages aux biens/véhicule motorisé

4.4 Conclusions sur les dépendances observées

A partir des indicateurs de corrélation que sont le tau de Kendall, le rho de Spearman ou encore le coefficient de corrélation de Pearson, des différentes représentations graphiques telles que les dépendogrammes et les K-Plot, nous pouvons confirmer l'existence de dépendances entre les charges agrégées de sinistre pour certains couples de sinistres.

Le couple Dommages aux biens/Perte de biens marque suffisamment une dépendance positive. On peut apercevoir une concentration plus marquée sur les grandes valeurs dans le coin supérieur droit du dépendogramme correspondant. Les deux queues possèdent une concentration conséquente de points. Toutefois la queue supérieure dispose d'un amas de points plus conséquent. Nous pouvons donc supposer une asymétrie de la dépendance pour ce couple de types de sinistre, ce qui orientera le choix de copules paramétriques à instaurer lors de notre modélisation à venir.

Pour les autres couples, il existe une dépendance positive ou négative assez faible. Le faible nombre d'observations (entre 100 et 200 observations) n'aide pas à la justesse de l'analyse. Certains couples ont une p-valeur trop élevée pour en tirer une quelconque conclusion lors des tests de corrélation (couples perte de biens/véhicule, perte de biens/employé ou encore dommages corporels/dommages aux biens). Les dépendogrammes ne délimitant aucune tendance, la modélisation de ces dépendances par les copules semble vaine.

Malgré leur effectif faible, il existe néanmoins des couples présentant une dépendance faible consistante voir moyennement consistante. Le couple perte de biens/dommages corporels possède une dépendance négative très faible à exploiter lors de la modélisation. De même pour le couple dommages aux biens/véhicule motorisé qui présente également une dépendance négative peu marquante. Nous allons également essayer de modéliser au mieux la dépendance positive faible du couple dommages aux biens/employé.

Le type de copule choisie à terme permettra de connaître la significativité de l'asymétrie des dépendances de queue. Si une asymétrie se distingue dans nos dépendances, nous pourrions écarter la piste de copules elliptiques, copules présentant une symétrie de dépendances, dans le cadre du choix de la meilleure copule. Cependant, l'asymétrie présente dans les dépendogrammes présentés n'est pas clairement établie, ce qui ne permet pas d'omettre de l'étude les copules elliptiques.

Chapitre 5

Estimation de la copule empirique et modélisation de dépendances

Le but de cette partie est de modéliser au mieux les dépendances de nos couples de types de sinistre par un choix adéquat de la meilleure copule paramétrique. Pour ce faire, nous allons dans un premier temps étudier les densités empiriques ainsi que les copules empiriques associées à nos couples de sinistre. Ces copules empiriques serviront de référence dans le cadre du choix de la meilleure copule paramétrique. Ensuite nous allons travailler sur l'ajustement de nos lois marginales à des lois connues par une méthode de maximum de vraisemblance. Cet ajustement permettra de calibrer une copule paramétrique adéquate. Puis il s'agira de calculer les paramètres de copule pour chaque copule testée. Enfin, nous réaliserons des tests statistiques permettant d'établir un choix de meilleur copule au regard de la copule empirique. Cette dernière analyse sera appuyée par des représentations graphiques des copules choisies avec leurs copules empiriques correspondantes.

5.1 Représentation de la copule empirique

Comme évoqué auparavant, certains couples de sinistre ne disposent que de peu d'observations tout en ayant une dépendance pas assez significative. Nous retiendrons pour la suite de notre étude seulement les couples ayant une dépendance assez significative, c'est-à-dire les couple dommages aux biens/perte de biens, perte de biens/dommages corporels, dommages aux biens/véhicule motorisé et dommages aux biens/perte employé.

5.1.1 Densité empirique

Afin de visualiser la densité de la copule recherchée, il est nécessaire de représenter la densité empirique de chacun des couples de sinistre retenus. Les densités ci-dessous sont représentées sous forme d'histogramme en 3 dimensions avec un quadrillage des valeurs du couples dans un carré 10×10 .

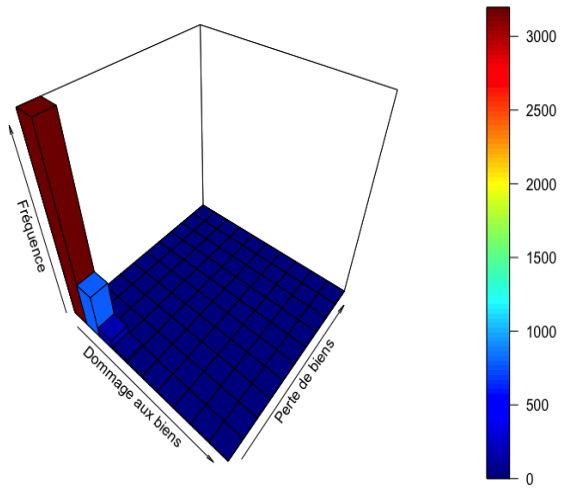


FIGURE 5.1 – Histogramme 3D du couple perte de biens/dommages aux biens

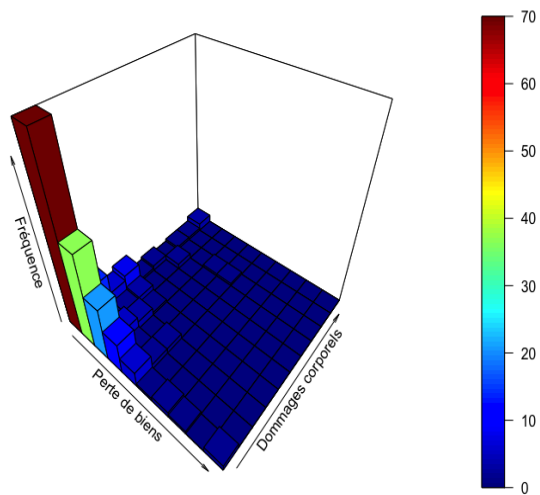


FIGURE 5.2 – Histogramme 3D du couple perte de biens/dommages corporels

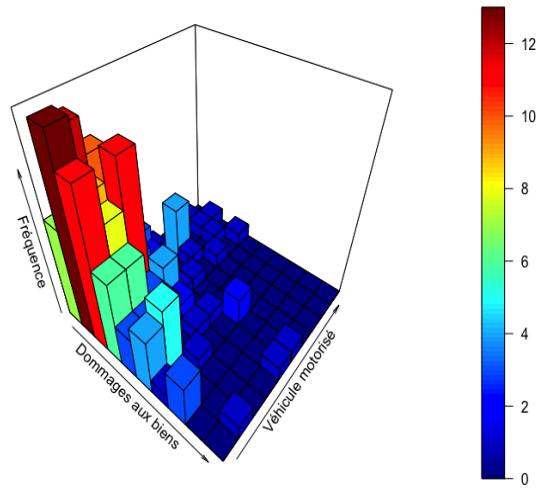


FIGURE 5.3 – Histogramme 3D du couple dommages aux biens/véhicule motorisé

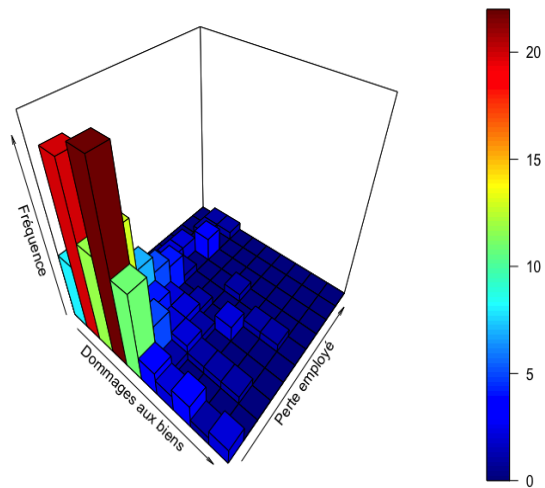


FIGURE 5.4 – Histogramme 3D du couple dommages aux biens/perte employé

On observe globalement dans ces quatre histogrammes 3D des couples d'observation concentrés majoritairement pour de faibles valeurs (distinction moins nette pour le couple dommages

aux biens/véhicule motorisé). Les couples sont donc plus concentrés vers des montants minimum que des montants extrêmes.

5.1.2 Copule empirique

Comme évoqué auparavant, Deuhevles est à l'origine de la notion de la copule empirique, une copule se basant sur le rang des observations et non sur les lois marginales afin d'extraire la structure de dépendance.

La copule empirique bivariée est définie de la façon suivante :

$$\hat{C}_N \left(\frac{n_1}{N}, \frac{n_2}{N} \right) = \frac{1}{N} \sum_{n=1}^N \mathbb{1}_{\{r_1^n \leq n_1\}} \mathbb{1}_{\{r_2^n \leq n_2\}}$$

avec $(r_1^n, r_2^n)_{1 \leq n \leq N}$ le rang des observations d'un échantillon de taille N.

L'expression de cette copule pour chacun de nos couples de sinistre servira de référence dans notre recherche de copule paramétrique représentant au plus près les dépendances entre nos couples de sinistre.

Les copules empiriques ont été représentées graphiquement ci-dessous à l'aide du logiciel R.

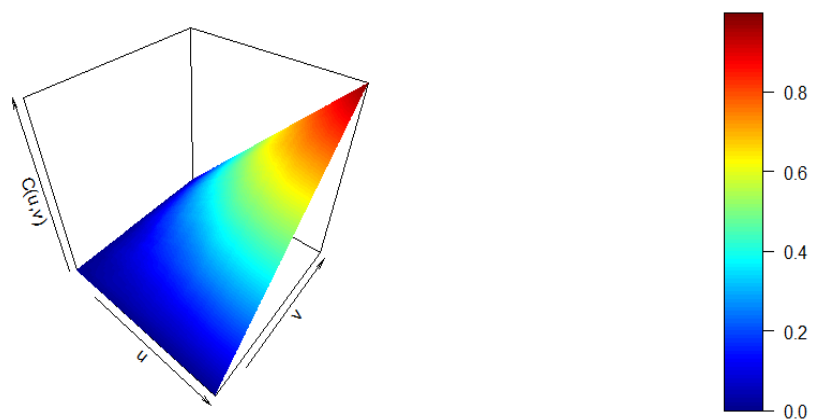


FIGURE 5.5 – Copule empirique du couple Dommages aux biens/Perte de biens

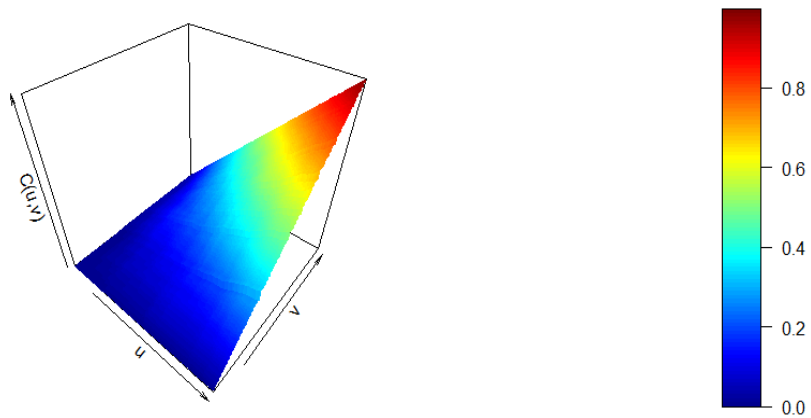


FIGURE 5.6 – Copule empirique du couple Perte de biens/Dommages corporels

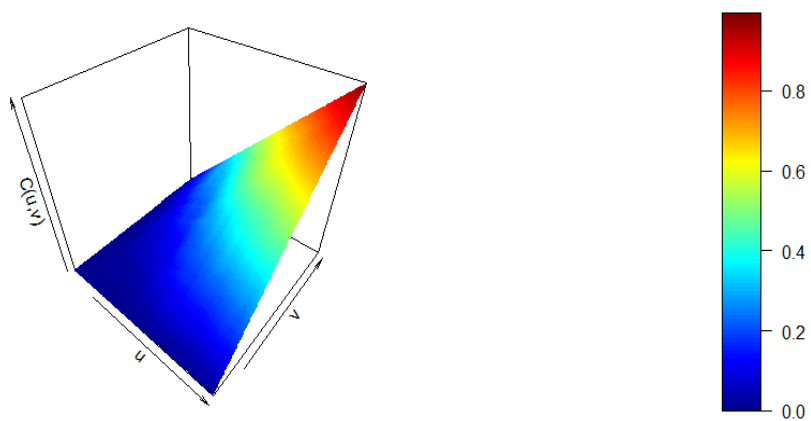


FIGURE 5.7 – Copule empirique du couple Dommages aux biens/Véhicule motorisé

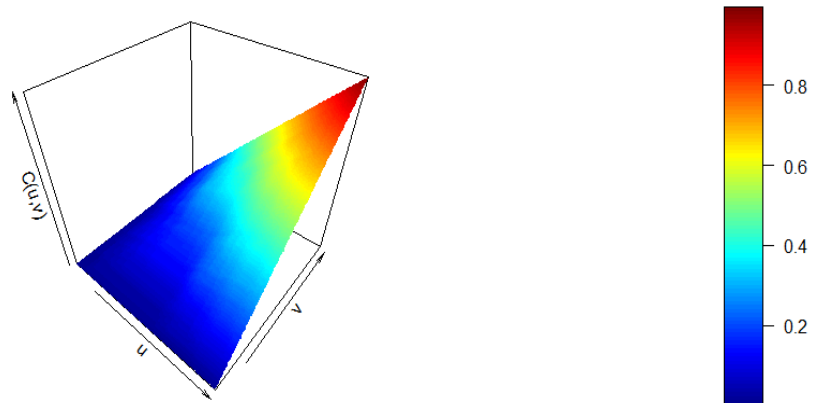


FIGURE 5.8 – Copule empirique du couple Dommages aux biens/Perte employé

On peut remarquer que les quatre graphiques ont la même allure. La surface du premier graphique semble cependant plus homogène du fait d'un nombre d'observations plus grand par rapport aux autres couples (cf. dégradés de couleurs).

5.2 Ajustement lois marginales

Dans l'optique d'un ajustement des paramètres des copules paramétriques, nous avons besoin d'estimer au mieux les marges. Nous essaierons dans cette section d'approcher nos lois marginales par des familles de lois indexées par un paramètre ; il s'agit de l'approche paramétrique. Nous allons utiliser la méthode du maximum de vraisemblance pour estimer le paramètre adéquat à nos lois marginales.

Dans un premier temps, estimons les densités de chacune de nos lois marginales.

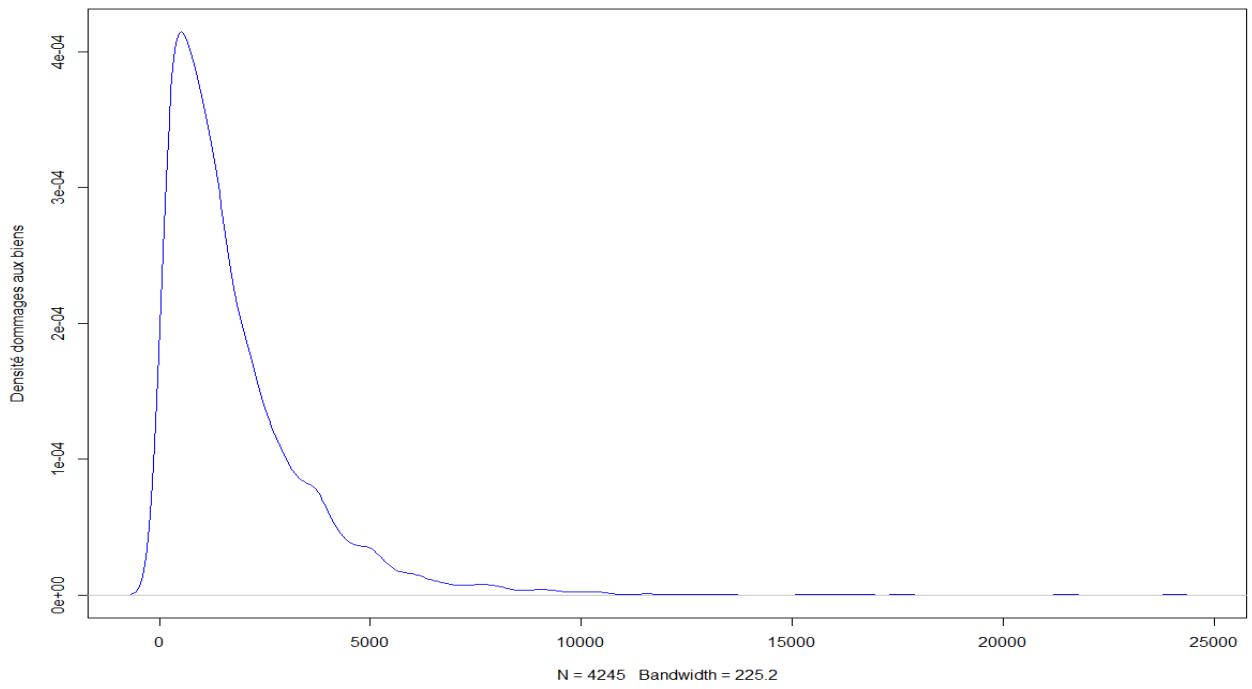


FIGURE 5.9 – Densité estimée type "Dommages aux biens"

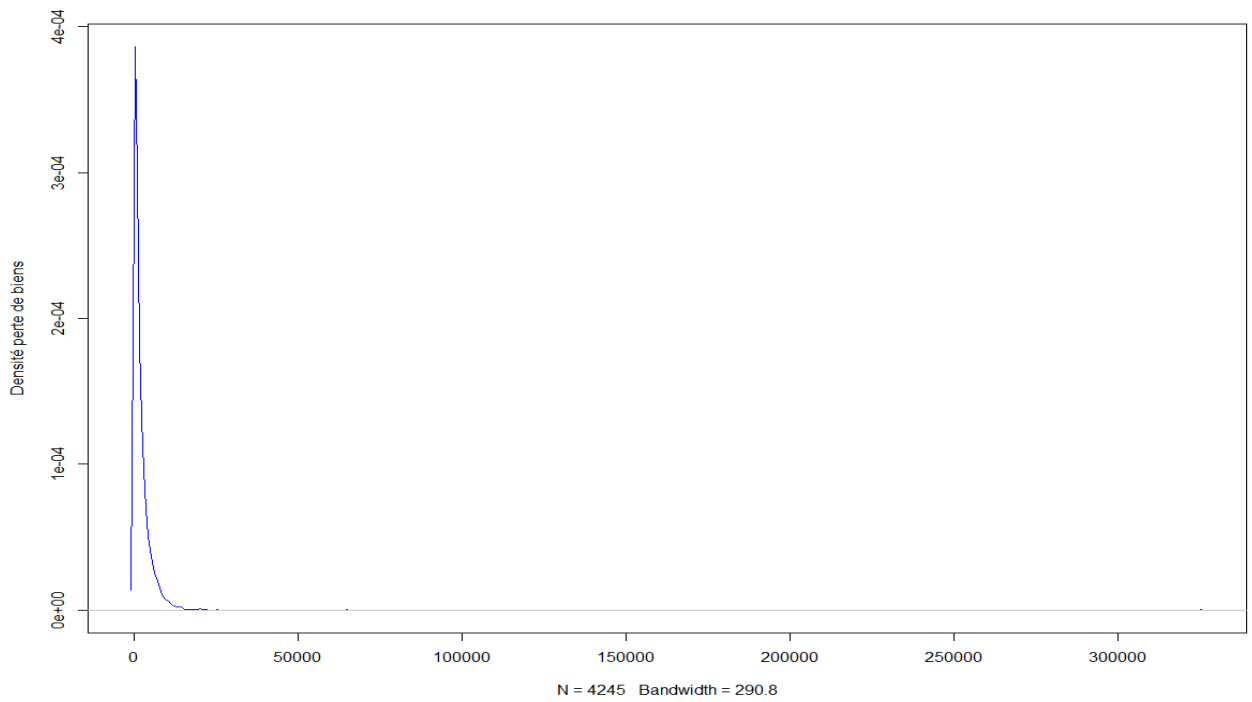


FIGURE 5.10 – Densité estimée type "Perte de biens"

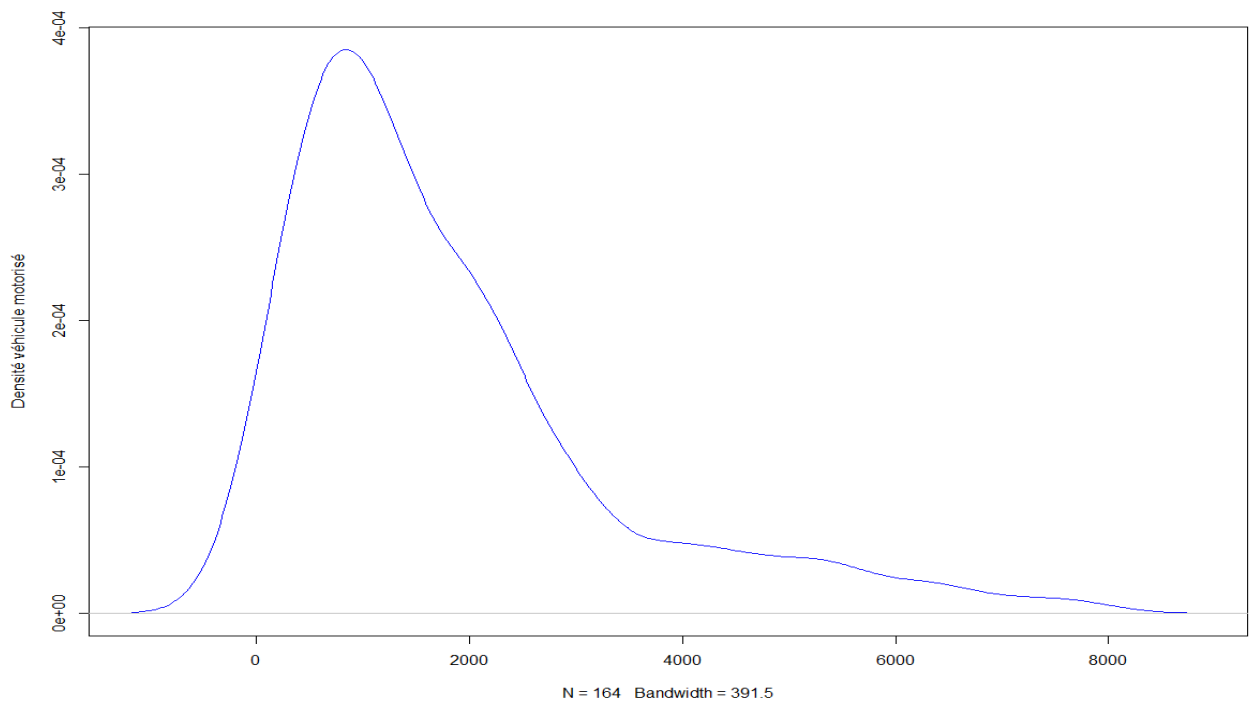


FIGURE 5.11 – Densité estimée type "Véhicule motorisé"

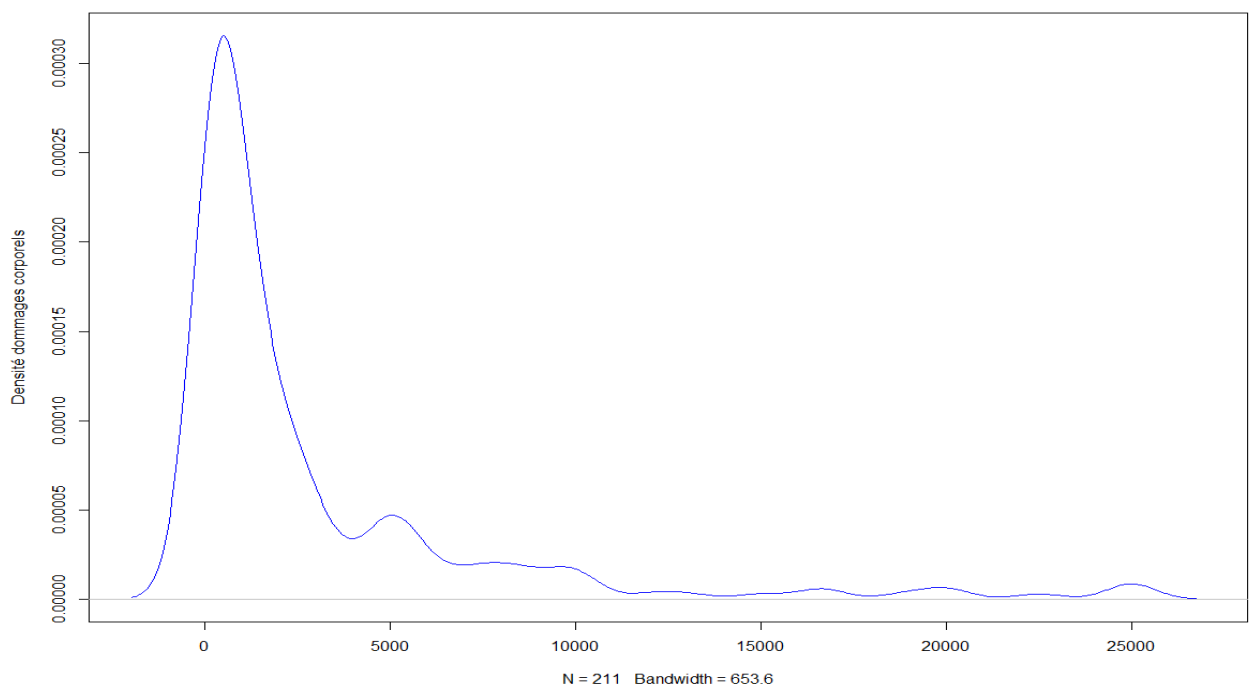


FIGURE 5.12 – Densité estimée type "Dommages corporels"

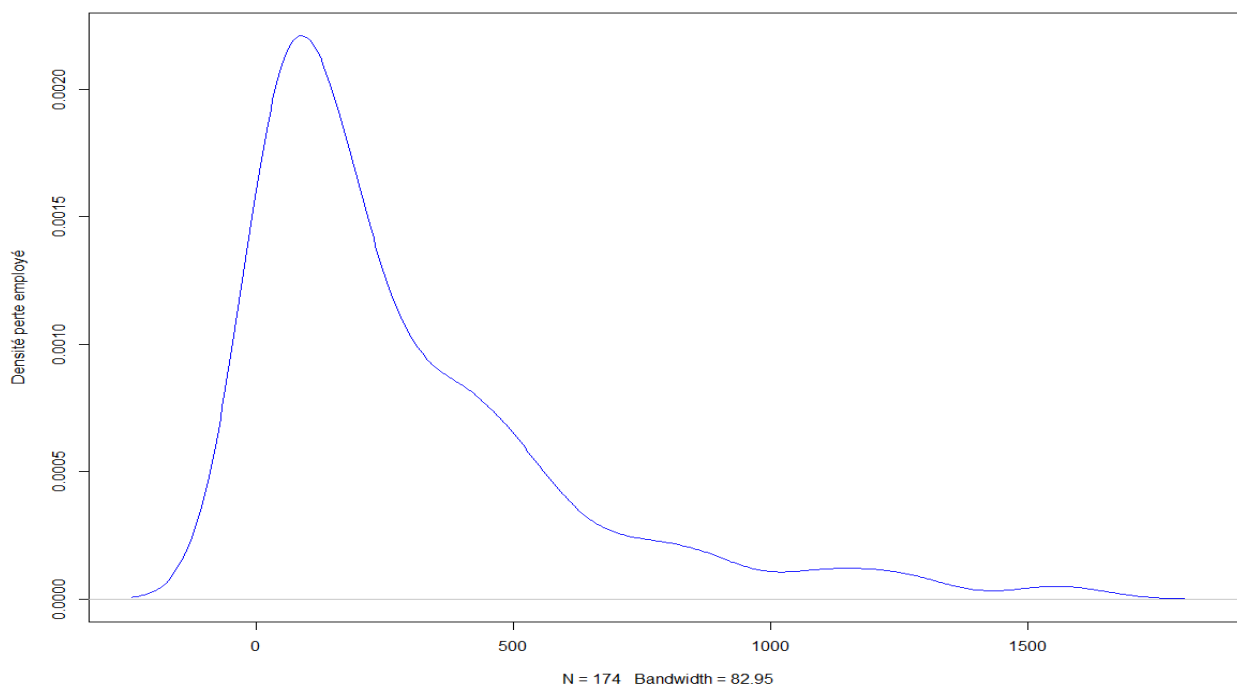


FIGURE 5.13 – Densité estimée type "Perte employé"

Au vu de l'allure générale des courbes, nous pouvons restreindre notre ajustement sur les lois suivantes :

- Loi Normale ;
- Loi Log-Normale ;
- Loi Gamma ;
- Loi Weibull ;
- Loi Exponentielle ;

A l'aide de la fonction `nlm` sur R, nous pouvons obtenir les maximum de log-vraisemblance suivants pour chaque famille de lois :

Loi testée	Perte biens	Dommages biens	Dom. corporels	Véhicule	Employé
Loi normale	-8 683 185	-865 258	-324 236,4	-24 716,43	-1 973,154
Loi log-normale	-36 668,3	-36 183,75	-1 861,162	-1 403,926	-1 162,982
Loi Gamma	-36 691,5	-36 020.61	-1 875,264	-1 389,41	-1 161,062
Loi Weibull					
Loi Exponentielle	-36 825,26	-36 057,03	-1 899,919	-1 393,231	-1 161,591

TABLE 5.1 – Maximum de log-vraisemblance selon les lois testées

L'ajustement avec la loi de Weibull ne converge pas pour chaque loi marginale, et ce quel que soit les paramètres initiaux posés.

Au vu des maximums de log-vraisemblance aperçus dans les tables ci-dessus, on peut en déduire la loi ajustée pour chaque type de sinistres :

- Les charges de sinistre liées à la perte de biens suivent une loi log-normale de paramètre (6,8727 ; 1,9988).
- Les charges de sinistre liées aux dommages aux biens suivent une loi gamma de paramètre (1,185796 ; 0,00066).
- Les charges de sinistre liées aux dommages corporels suivent une loi log-normale de paramètre (6,950758 ; 2,464405).
- Les charges de sinistre liées aux véhicules motorisés suivent une loi gamma de paramètre (1,332159 ; 0,00074).
- Les charges de sinistre liées à la perte employé suivent une loi gamma de paramètre (0,90911 ; 0,003116).

Nous allons désormais représenter les densités ajustées pour chaque type de sinistre avant de tester la significativité de l'ajustement à l'aide du test de Kolmogorov-Smirnov.

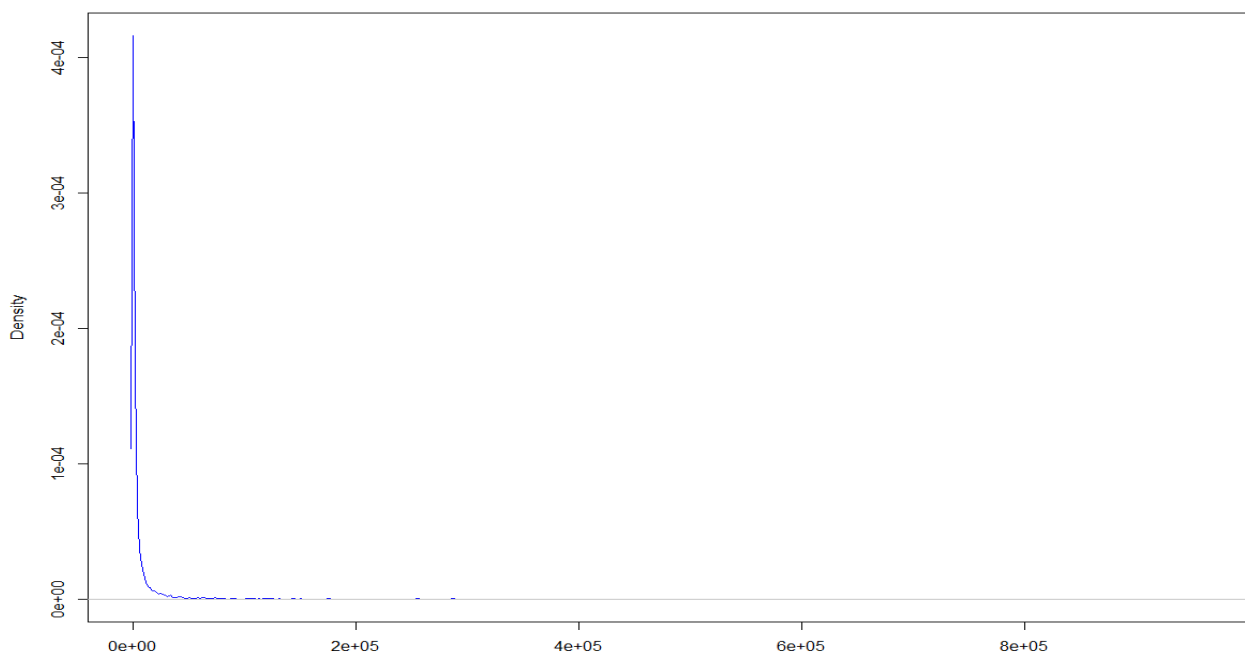


FIGURE 5.14 – Densité ajustée type "Perte de biens"

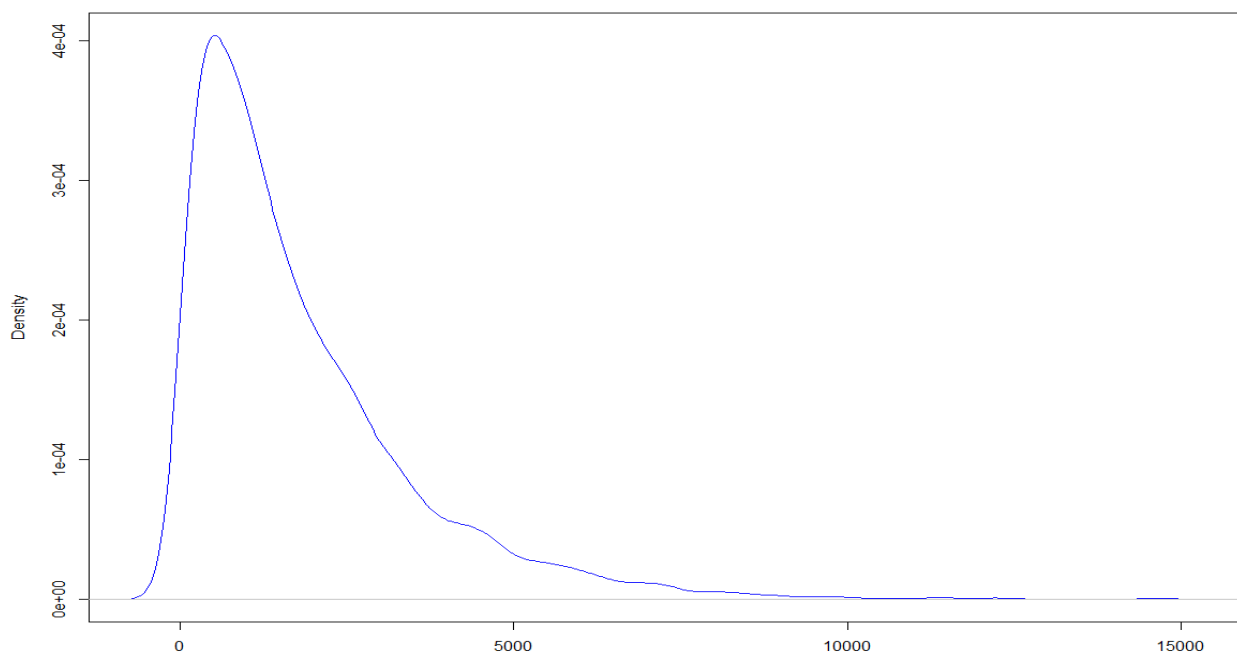


FIGURE 5.15 – Densité ajustée type "Dommages aux biens"

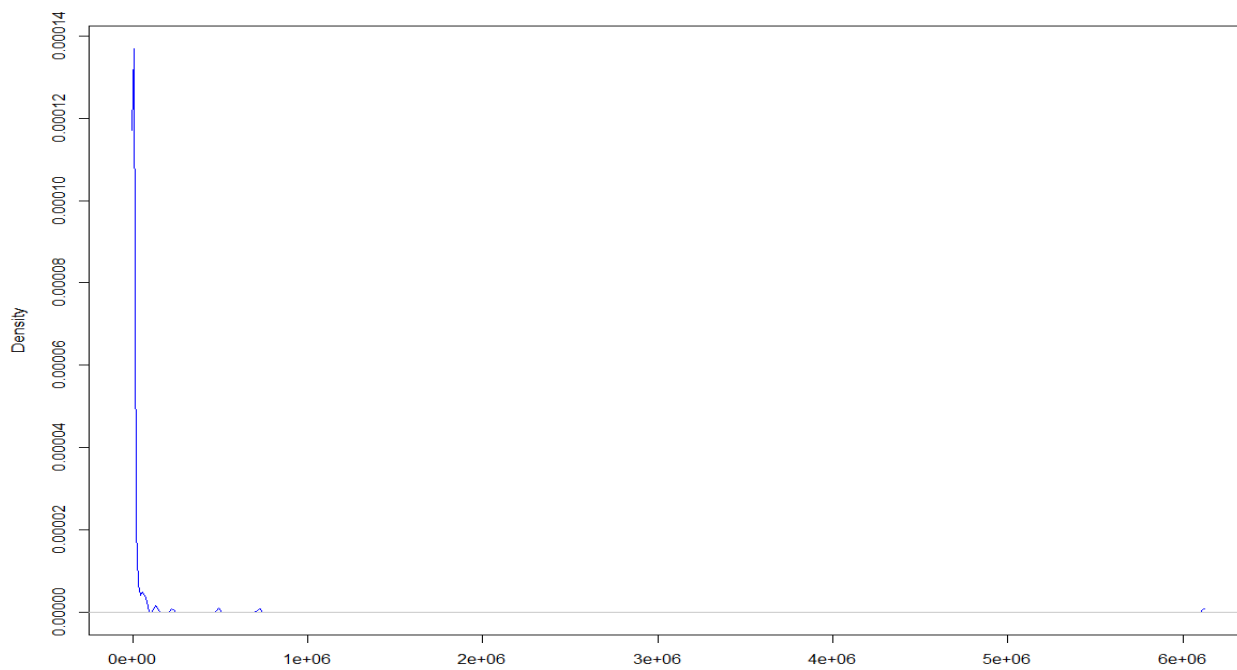


FIGURE 5.16 – Densité ajustée type "Dommages corporels"

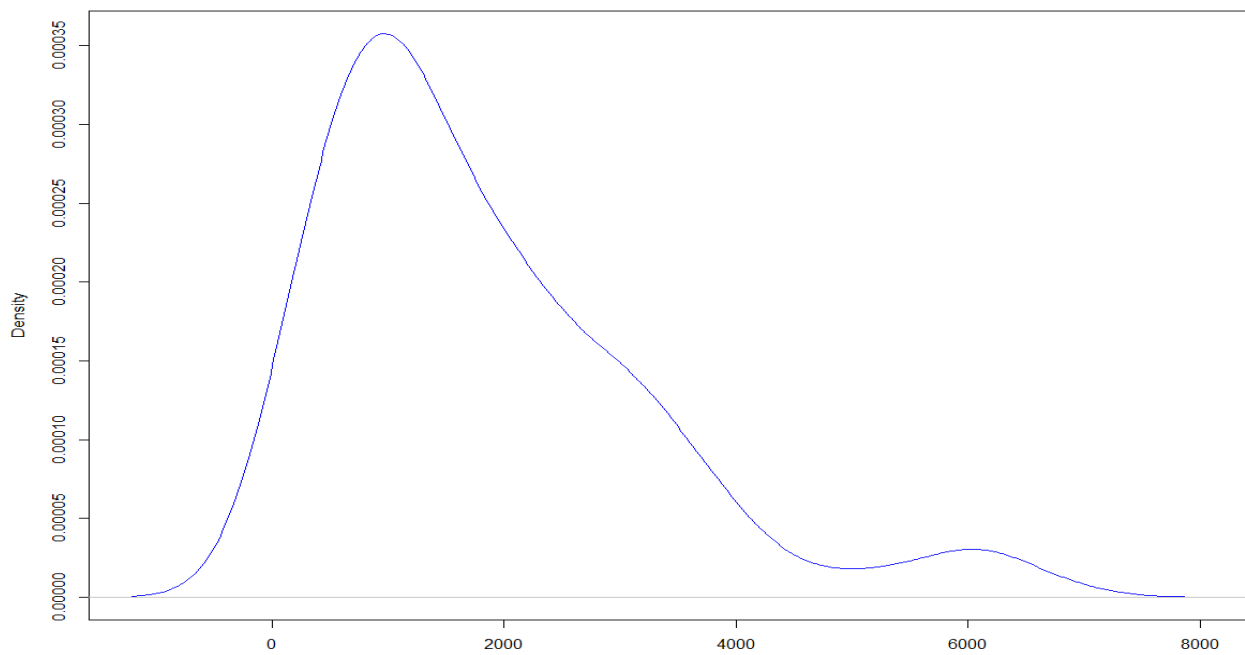


FIGURE 5.17 – Densité ajustée type "Véhicule motorisé"

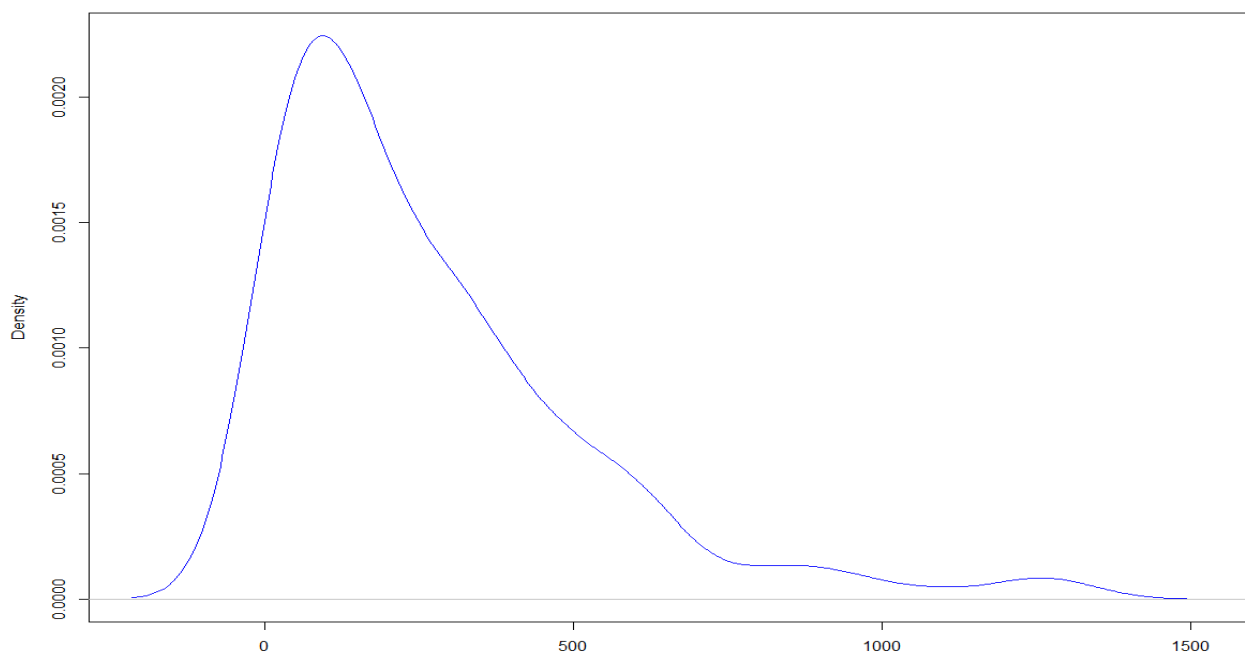


FIGURE 5.18 – Densité ajustée type "Perte employé"

5.2.1 Test d'ajustement des lois marginales

Pour vérifier la qualité d'ajustement de nos lois marginales, nous avons choisi d'établir un test d'ajustement d'une distribution inconnue à une distribution théorique. Il s'agit du test de Kolmogorov Smirnov. Contrairement au test du χ^2 , le test de Kolmogorov Smirnov est exact et est particulièrement bien adapté aux petits échantillons.

Voici comment s'organise le test.

A partir d'un n -échantillon observé de la loi marginale, on construit la bande de confiance au niveau de confiance $1 - \alpha$. Si cette bande ne contient pas entièrement le graphe de la loi théorique, on rejette l'hypothèse H_0 . Dans le cas où la bande contient entièrement le graphe de la loi théorique, on ne rejette pas H_0 au niveau de confiance $1 - \alpha$. Nous allons choisir ici un seuil de confiance à 99%.

Ajustement	D	P-Value
Perte de biens - Log-normale	1 571 323	$< 2.2 \times 10^{-16}$
Dommages aux biens - Gamma	13 379	$< 2.2 \times 10^{-16}$
Dommages corporels - Log-normale	226 171	$< 2.2 \times 10^{-16}$
Véhicules motorisés - Gamma	6197,8	$< 2.2 \times 10^{-16}$
Perte employé - Gamma	2330,6	$< 2.2 \times 10^{-16}$

TABLE 5.2 – Résultats du test de Kolmogorov Smirnov

L'hypothèse nulle H_0 , qui signifie que nos lois marginales suivent bien leurs lois théoriques respectives, est rejetée pour chacun des cas (p-value bien inférieur au seuil de 1% fixé). La qualité d'ajustement n'est donc pas satisfaisante et l'approche paramétrique semble donc peu fiable. Nous estimerons donc les copules par une approche semi-paramétrique, en estimant les lois marginales par leur fonction de répartition empirique.

5.2.2 Calcul de la marge estimée non-paramétrique

Au vu des résultats peu probant concernant l'approche paramétrique, on privilégiera ici l'approche semi-paramétrique. Dans l'approche semi-paramétrique on ne fait pas d'hypothèse sur la marge mais on l'estime par une méthode non paramétrique. Pour estimer le paramètre de dépendance α (paramètre de copule), on construit une pseudo-vraisemblance en remplaçant la marge par son estimé non-paramétrique.

L'estimateur $\hat{F}_e(y)$ d'une distribution marginale $F_e^0(y)$ est la fonction de répartition empirique échelonnée de y_{11}, \dots, y_{mn_m} explicitée par la fonction en escalier suivante :

$$\hat{F}_e(v) = \frac{\sum_{i=1}^m \sum_{j=1}^{n_i} \mathbf{1}_{(y_{ij} \leq v)}}{\sum_{i=1}^m n_i} \quad (5.1)$$

Dans notre cas, n_i est égale à 1, tandis que la variable m est égale à la taille de l'échantillon liée à chaque type de sinistres.

Donc pour chaque type de sinistre, l'estimateur de la distribution marginale se définit comme suit :

$$\hat{F}_e(v) = \frac{\sum_{i=1}^m \mathbf{1}_{(y_i \leq v)}}{m} \quad (5.2)$$

5.3 Estimation du paramètre de copule par maximum de pseudo-vraisemblance

Une fois l'estimation empirique des lois marginales calculée pour chacun de nos types de sinistre, nous pouvons estimer le paramètre de copule par une méthode du maximum de pseudo-vraisemblance (ou méthode CML) se basant sur l'estimation empirique de nos lois marginales. Dans cette section, nous étudierons le paramètre adéquat pour toutes les copules usuelles citées plus tôt dans ce mémoire, à savoir les copules archimédiennes (Clayton, Franck, Gumbel, AMH) et les copules elliptiques (copule gaussienne et de Student). Soit \hat{F}_e une estimation empirique de loi marginale, alors la vraisemblance s'écrira pour chaque couple d'estimations empiriques de lois marginales :

$$L(\alpha) = \prod_{i=1}^m c(\hat{F}_{e_1}(y_i), \hat{F}_{e_2}(y_i), \alpha) \quad (5.3)$$

En maximisant le logarithme de la vraisemblance exprimée ci-dessus, l'estimateur du paramètre de copule s'écrit :

$$\hat{\alpha} = \arg \max_{\alpha \in \mathcal{A}} \sum_{i=1}^m \log \left(c(\hat{F}_{e_1}(y_i), \hat{F}_{e_2}(y_i), \alpha) \right), \mathcal{A} \subset \mathbb{R} \quad (5.4)$$

Cette méthode offre donc l'avantage d'estimer les paramètres de la copule indépendamment de l'estimation des marginales et ainsi atténuer l'erreur commise. De plus, elle est beaucoup moins coûteuse en temps de calcul que les méthodes d'approche paramétrique. Ceci est d'autant plus valable que la dimension de la copule et de ses marginales est importante.

Copules	Dommages/ Perte de biens	Perte de biens/ Dommages corporels	Dommages biens/ Véh motorisé	Dommages biens/ Perte employé
Gaussienne	0.449 (0.464)	-0.110 (-0.123)	-0.098 (-0.090)	0.066 (0.042)
Student(ddl=4)	0.415 (0.464)	-0.123 (-0.123)	-0.082 (-0.090)	0.055(0.042)
Frank	3.004 (3.001)	-0.714 (-0.708)	-0.557 (-0.516)	0.193 (0.240)
Clayton	0.485 (0.887)	-0.044 (-0.145)	0 (-0.108)	0.148 (0.055)
Gumbel	1.388 (1.444)	1 (1)	1 (1)	1.026 (1.027)
AMH	0.787 (0.956)	-0.426 (-0.384)	-0.236 (-0.274)	0.108 (0.116)

TABLE 5.3 – Paramètre de la copule selon la méthode CML, paramètre selon la méthode des moments entre parenthèses

On peut remarquer que les paramètres de copule par la méthode des moments peuvent être négatifs pour les couples "Perte de biens/Dommages corporels" et "Dommages aux biens/Véhicule Motorisé". L'écart entrevu selon l'approche par moment et l'approche CML confirme l'idée que l'approche paramétrique est plus précise pour le calcul du paramètre de copule. Par ailleurs, les couples "Perte de biens/Dommages corporels" et "Dommages aux biens/Véhicule motorisé" semblent dénoter d'une dépendance négative. En appliquant la méthode CML sur de copules de Clayton et de Gumbel pour ces couples, on remarque que les paramètres optimaux se situent sur leur borne inférieure, ce qui peut s'expliquer par le fait que les copules de Gumbel et de Clayton

ne peuvent représenter seulement les dépendances positives entre types de sinistre.

On remarque ainsi que la méthode d'inversion du tau de Kendall (résultant d'une approche paramétrique) affiche des paramètres de copule sensiblement différents de la méthode CML (résultant d'une approche semi-paramétrique), confortant ainsi l'emploi de l'approche semi paramétrique dans l'estimation des paramètres.

Après avoir modélisé les paramètres de copule, il peut être intéressant de calculer le coefficient de dépendance de queue pour chaque copule calibrée à un couple de sinistres, puis de le comparer à un coefficient de dépendance de queue estimée empiriquement.

5.4 Coefficients de dépendance des queues et estimation empirique de dépendance des queues

A partir des différentes copules paramétriques calibrées à nos données, nous pouvons calculer le coefficient de dépendance des queues correspondant. En effet, le coefficient de dépendance de queue à gauche et à droite s'exprime en fonction de la fonction copule correspondante, ce qui découle sur une expression en fonction du paramètre de copule comme vu en première partie de ce mémoire.

Ci-dessous sont disposés les coefficients de dépendance de queue inférieurs et supérieurs pour chaque couple et chaque copule paramétrique calibrée.

Copules	Dommages/ Perte de biens	Perte de biens/ Dommages corporels	Dommages biens/ Véh. motorisé	Dommages biens/ Perte employé
Gaussienne	0 0	0 0	0 0	0 0
Student(ddl=4)	0.2102 0.2102	0.0525 0.0525	0.0596 0.0596	0.0879 0.0879
Frank	0 0	0 0	0 0	0 0
Clayton	0.2396 0	0 0	0 0	0.0091 0
Gumbel	0 0.3525	0 0	0 0	0 0.0350
AMH	0 0	0 0	0 0	0 0

TABLE 5.4 – Coefficients de dépendance de queues inférieur et supérieur pour chaque copule paramétrique calibrée au jeu de données

Afin de se donner un élément de comparaison, nous avons pu estimer empiriquement les coefficients de dépendance de queue pour chaque couple de risques. Pour ce faire, nous avons utilisé la copule de queue empirique, étudiée dans Schmidt et Stadtmüller (2006). Les résultats ci-dessous ont été obtenu à partir de la fonction tdc du package FRAPO sur le logiciel R.

Dommages/ Perte de biens	Perte de biens/ Dommages corporels	Dommages biens/ Véh motorisé	Dommages biens/ Perte employé
0.092 0.077	0.071 0	0 0	0.231 0.154

TABLE 5.5 – Estimation empirique des dépendances de queue inférieur et supérieur par la méthode de la copule de queue empirique

Au regard des dépendances de queue empiriques, on remarque que la copule de Student semble être la plus adéquate pour modéliser les charges de sinistre des couples de sinistre

Dommages/Perte de biens et Dommages aux biens/Perte employé. En observant uniquement les dépendances de queue, on peut également supposer que les couples de sinistre Perte de biens/Dommages corporels et Dommages de biens/Véhicule motorisé puissent être modélisés par des copules de Frank, Gumbel, AMH ou gaussienne.

Pour pouvoir faire un choix approprié de copule paramétrique modélisant au plus près les couples de sinistre, il nous faut tester la structure de dépendance des copules paramétriques obtenus avec leur copule empirique correspondante.

Dans la section suivante, nous nous attacherons à appliquer deux méthodes préalablement introduits dans ce mémoire, la fonction $K(z)$ ainsi que le test par la méthode du χ^2 .

5.5 Test d'ajustement statistique et graphique menant au choix de la bonne copule

5.5.1 Test d'ajustement du χ^2

Le test d'ajustement du χ^2 permet de tester l'hypothèse selon laquelle un effectif observé regroupé en classes est conforme à l'effectif théorique regroupé en classes identiques à ceux de l'effectif observé.

Dans le cadre de l'établissement d'un choix de meilleur copule, nous allons étudier l'effectif empirique selon nos uniformes catégorises en intervalles identiques. De même, l'effectif théorique issu de chacune des copules paramétriques sera catégorisé selon les mêmes intervalles que pour l'effectif empirique.

Pour la catégorisation des effectifs, le critère de Cochran a été rigoureusement appliqué. Le critère de Cochran stipule :

- Toutes les classes ont une espérance non-nulle.
- Au moins 80% des classes ont une espérance supérieure à 5.

Les uniformes (marginales) seront regroupés selon 8 intervalles de taille identique (intervalles de longueur $\frac{1}{8}$ allant de 0 à 1, donc 64 classes pour le couple) pour le couple Dommages/perte de biens. Pour les trois autres couples, le nombre d'observations total étant faible (moins de 250), les effectifs seront regroupés selon 16 classes (donc 4 intervalles de tailles identiques pour chaque uniforme).

Afin de modéliser au mieux l'effectif théorique pour chaque copule paramétrique, nous avons pu calculer la probabilité théorique d'appartenance de la copule dans un carré donné (espace 2D intervalle \times intervalle) grâce à la fonction prob sur R. En multipliant la probabilité théorique par le nombre total d'observations pour le couple de sinistres, nous pouvons obtenir l'effectif théorique pour chaque classe appliquée à une copule paramétrique donnée.

Le test du χ^2 calcule l'écart des effectifs au carré pondéré par leur effectif observé. L'hypothèse H_0 du test stipule la non-conformité entre les effectifs, donc une p-value élevée signifie que l'on ne peut pas dire que les effectifs soient non-conformes. La copule dont la p-valeur est maximale sera sélectionnée comme copule décrivant le mieux la structure de dépendance du couple de sinistres.

Afin d'illustrer le test, vous trouverez ci-dessous les effectifs empiriques du couple "Dommages aux biens/Perte employé" catégorisés selon 16 classes identiques ainsi que l'effectif théorique généré par une copule de Student de paramètre 0,0548 regroupés avec ces 16 classes.

	[0,0.25]]0.25,0.5]]0.5,0.75]]0.75,1]
[0,0.25]	16	7	10	10
]0.25,0.5]	5	12	16	11
]0.5,0.75]	11	14	9	10
]0.75,1]	11	11	9	12

TABLE 5.6 – Effectifs empiriques du couple "Dommages aux biens/Perte employé"

	[0,0.25]]0.25,0.5]]0.5,0.75]]0.75,1]
[0,0.25]	12.82	10.10	9.62	10.95
]0.25,0.5]	10.10	11.99	11.78	9.62
]0.5,0.75]	9.62	11.78	11.99	10.10
]0.75,1]	10.95	9.62	10.10	12.82

TABLE 5.7 – Effectifs théoriques de la copule de Student de paramètre 0,0548 et ddl=4 pour le couple "Dommages aux biens/Perte employé"

La valeur du χ^2 est ici de 10,48 avec un degré de liberté de 15 et une p-value égale à 0,7885. Voici les résultats du test pour chaque copule paramétrée.

	Gauss	Student	Frank	Clayton	Gumbel	AMH
χ^2	14.83	10.48	14.665	13.681	14.518	14.732
P-Value	0.4637	0.7885	0.4758	0.5499	0.4867	0.4709

TABLE 5.8 – Résultats du test pour le couple "Dommages aux biens/Perte employé"

Le meilleur choix de copule pour ce couple est donc bel et bien la copule de Student. Ci-dessous sont disposés les meilleurs choix de copules pour chaque couple de sinistres selon le test du χ^2 .

Couples de sinistre	Copule	\hat{a}_{CML}
Dommages/Perte de biens	Gumbel	1.3883
Perte de biens/Dommages corp	Student	-0.1233
Dommages aux biens/Veh motorisé	Student	-0.0820
Dommages aux biens/Perte employé	Student	0.0548

TABLE 5.9 – Meilleurs copules selon test du χ^2 pour chaque couple de sinistres

5.5.2 Ajustement graphique par la fonction $K(z)$

A l'aide de la fonction $K(z)$, nous cherchons dans cette section à confirmer les résultats du test d'ajustement du χ^2 obtenus précédemment.

Comme évoqué en première partie de ce mémoire, la fonction $K(z)$ se définit comme étant la fonction de répartition de la copule $C(U,V)$ tel que $K(z) = P(C(U, V) < z)$. U et V sont pour cette étude les fonctions de répartition des marginales pour un couple donné. Le but est ici de comparer l'estimation empirique de $K(z)$ de l'expression de $K(z)$ calculée pour chaque copule paramétrée.

Dans un premier temps, à l'aide de la méthode évoquée en première partie de ce mémoire, nous allons exprimer l'estimation empirique de $K(z)$. Ci-dessous est représentée la fonction $K(z)$ estimée empiriquement pour le couple "Dommages/Perte de biens".

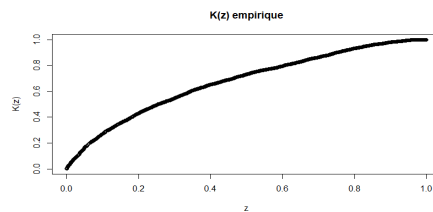


FIGURE 5.19 – Fonction $K(z)$ empirique pour le couple "Dommages/Perte de biens"

Calculons désormais les fonctions $K(z)$ pour chaque copule paramétrique calibrée précédemment dans notre étude. Ci-dessous sont représentées les fonctions $K(z)$ pour chaque copule, comparativement à leur fonction $K(z)$ empirique de référence tracée en rouge ci-dessous.

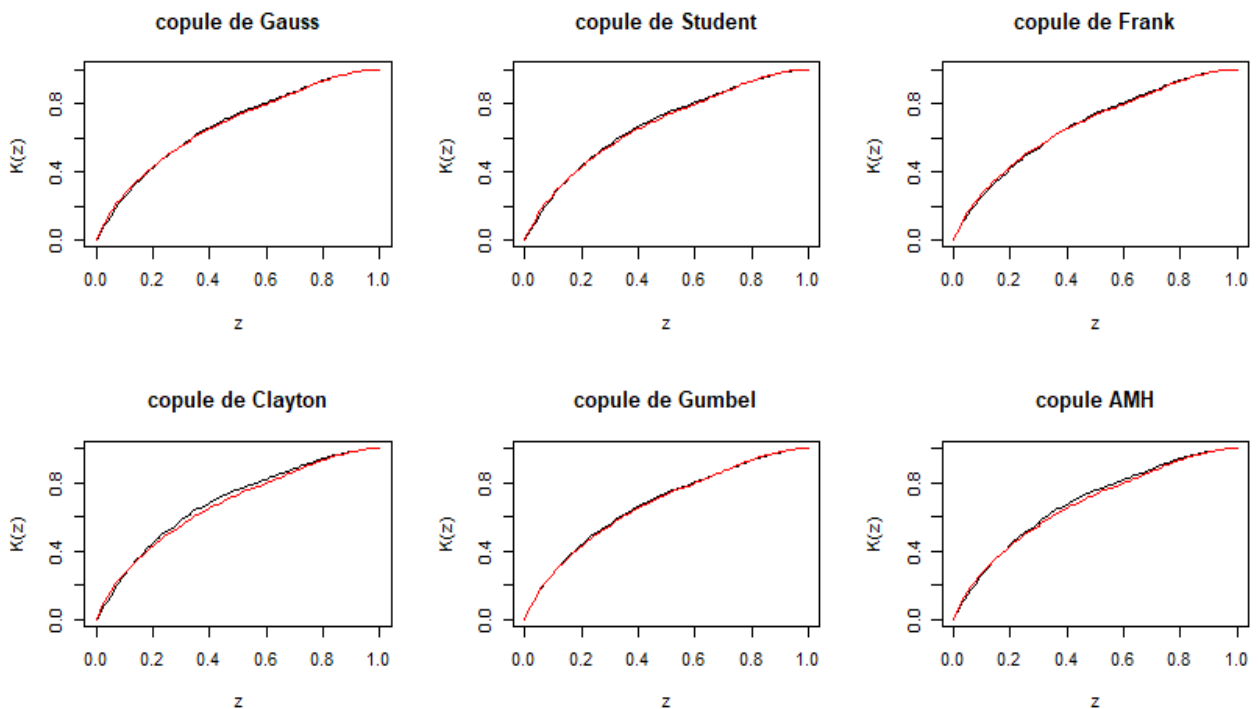


FIGURE 5.20 – Fonction $K(z)$ représentée pour chaque copule calibrée en fonction de $K(z)$ empirique (en rouge) pour le couple "Dommage/Perte de biens"

Nous pouvons observer à partir de la figure ci-dessus que les fonctions $K(z)$ de la copule de Gumbel et Frank se rapprochent le plus de la fonction $K(z)$ empirique. Cela confirme ainsi le

choix de la copule de Gumbel, établi lors du test d'ajustement du χ^2 , comme représentant le mieux la structure de dépendance de ce couple.

Réalisons désormais la même étude pour les trois autres couples de sinistre.

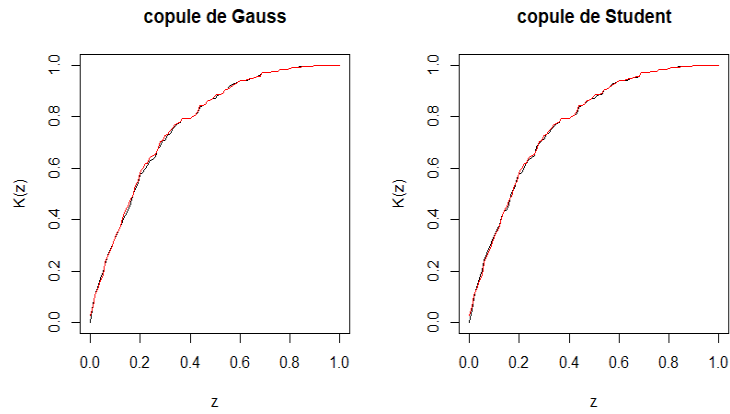


FIGURE 5.21 – Fonction $K(z)$ représentée pour chaque copule calibrée en fonction de $K(z)$ empirique (en rouge) pour le couple "Perte de biens/Dommages corporels"

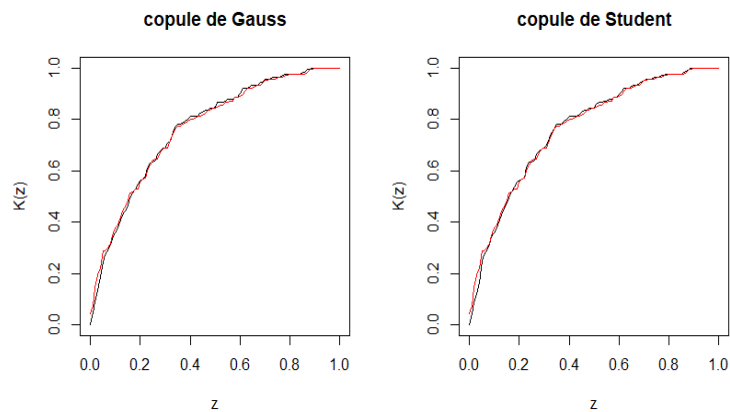


FIGURE 5.22 – Fonction $K(z)$ représentée pour chaque copule calibrée en fonction de $K(z)$ empirique (en rouge) pour le couple "Dommages aux biens/Véhicule motorisé"

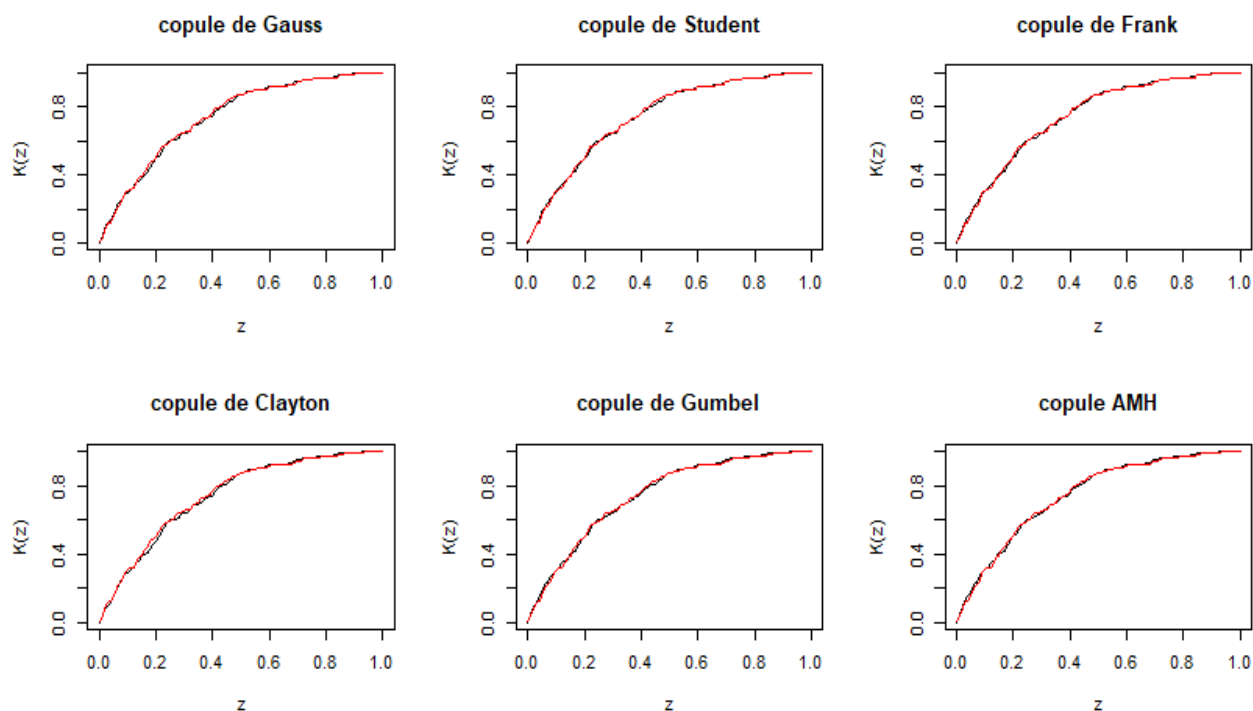


FIGURE 5.23 – Fonction $K(z)$ représentée pour chaque copule calibrée en fonction de $K(z)$ empirique (en rouge) pour le couple "Dommages aux biens/Perte employé"

On peut remarquer au vu des graphiques une proximité globale des fonctions $K(z)$ des copules calibrées avec leur estimation empirique, signe de la pertinence de l'approche semi-paramétrique dans l'estimation des paramètres de copule. Il est difficile à partir de ces représentations graphiques de distinguer nettement une fonction $K(z)$ se rapprochant le plus de son estimation empirique.

Nous pouvons néanmoins apercevoir sur les trois dernières figures une plus grande proximité des copules de Student avec leur estimation non-paramétrique de $K(z)$, ce qui confirme les résultats obtenus par test d'ajustement χ^2 .

On peut noter qu'il résulte de cette modélisation des dépendances la preuve d'une dépendance des queues présente sur chacun des couples de sinistre. En effet, le choix des copules (dépendance de queue supérieure pour la copule de Gumbel, dépendances de queue inférieure et supérieure pour la copule de Student) confirme l'existence des dépendances de queue pour ces couples ; dépendance supposée par l'estimation non-paramétrique des dépendances de queue.

L'utilité de cette modélisation se base principalement sur le calcul de dépendances entre des risques sans avoir à estimer de lois marginales aux risques.

Un autre point qui a été démontré à travers l'étude est la fragilité possible de l'hypothèse d'indépendance entre les risques au sein d'un même portefeuille. Selon que cette hypothèse soit vraie ou fautive, il en résulte un calcul du besoin en fonds propres différent et un impact sur le besoin en fonds propres.

Afin de mesurer l'impact de l'étude des dépendances sur le besoin en fonds propres nécessaire à l'indemnisation des assurés, nous allons calculer dans la seconde partie le besoin en fonds propres selon l'hypothèse de dépendances existantes entre les risques mais aussi selon l'hypothèse d'indépendance.

5.6 Approche multirisque par produit des copules bivariées

Les copules sont également des outils très utiles dans la modélisation de risques multiples. Beaucoup a été réalisé concernant les copules bivariées mais peu d'études sont présentes à ce jour sur les copules à grandes dimensions. Dans cette section, nous cherchons à modéliser les risques associés aux 4 copules bivariées paramétriques modélisées dans la précédente section. Plus précisément, il sera sujet de construire une copule multivariée de dimension 5 (nombre de marginales associées aux 4 couples étudiés) par le biais de produits de copules bivariées bien choisies, ainsi que d'étudier les dépendances et dépendances de queues des couples au sein de la copule multivariée. Un comparatif avec les dépendances obtenues par les copules bivariées modélisées dans la section précédente sera réalisé.

5.6.1 Produit de copules transformées - cas théorique

Il est admis que multiplier deux copules bivariées ne permet pas d'obtenir en général une copule. Cependant, par une transformation astucieuse des marginales, il est possible d'obtenir une copule bien définie par un produit de copules bivariées. Le théorème ci-dessous introduit par E. Liebscher [2008] définit cette copule :

Theoreme 1 Soient les copules $\tilde{C}_1, \dots, \tilde{C}_K : [0, 1]^d \rightarrow [0, 1]$. Soient les fonctions $g_{ei} : [0, 1] \rightarrow [0, 1]$ avec $e = 1, \dots, K, i = 1, \dots, d$. Chaque fonction est strictement croissante ou est identiquement égale à 1. Supposons que $\prod_{e=1}^K g_{ei}(v) = v, \forall v \in [0, 1], i = 1, \dots, d$ et $\lim_{v \rightarrow 0} g_{ei}(v) = 0, \forall e = 1, \dots, K, i = 1, \dots, d$. Alors :

$$C(u_1, \dots, u_d) = \prod_{e=1}^K \tilde{C}_e(g_{e1}(u_1), \dots, g_{ed}(u_d)) \quad (5.5)$$

est aussi une copule.

Il est difficile d'appliquer cette méthode pour $d > 3$ et $K > 2$. En effet, le calcul de la densité de la copule, et de ce fait de la vraisemblance, est difficile, tout comme l'implémentation numérique. Enfin, les contraintes exposées dans le théorème 1, à savoir $\prod_{e=1}^K g_{ei}(v) = v, \forall v \in [0, 1], i = 1, \dots, d$, ne sont pas évidentes dans la mise en pratique de la méthode. C'est pourquoi une version revisitée, introduite par G. Mazo et al. [2014], sera mise en place dans le but de construire notre copule multivariée.

5.6.2 Produit de copules transformées revisitée

Le produit selon $e \in \{1, \dots, K\}$ peut être établie selon $e \in E$, avec E un ensemble de données finies, tel que :

$$C(u_1, \dots, u_d) = \prod_{e \in E} \tilde{C}_e(g_{e1}(u_1), \dots, g_{ed}(u_d)). \quad (5.6)$$

E représente en particulier un couple de variables en jeu. Il s'agit dans notre cas d'un couples de marginales uniformes de la copule multivariée. Soit U_1, U_2, \dots, U_d les variables aléatoires uniformes. Le couple $\{i, j\}$ de la paire (U_i, U_j) est inclus dans l'ensemble E tel que $E \subset \{\{ij\} : i, j = 1, \dots, d, j > i\}$. Il s'agit d'un sous ensemble de l'ensemble des couples d'indices. Le cardinal de E , noté $|E|$, est égale au plus à $d(d-1)/2$. En suivant les préceptes du théorème du produit des copules bivariées exposées dans la section précédente, la copule multivariée se définit dans ce cas-ci ainsi :

$$C(u_1, \dots, u_d) = \prod_{\{ij\} \in E} \tilde{C}_{\{ij\}}(u_i^{1/n_i}, u_j^{1/n_j}). \quad (5.7)$$

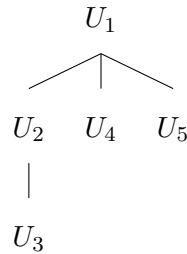
où $\tilde{C}_{\{ij\}}$ sont des copules bivariées, n_i le nombre de marginales liées à la variable U_i et n_j le nombre de marginales liées à la variable U_j .

Une représentation graphique par arbre permet de mieux se représenter les différentes liaisons entre les lois marginales uniformes et ainsi de déterminer E et les variables $n_i, i = 1, \dots, d$.

L'application qui suit reprend les couples de risque étudiés dans le cas bivarié. A partir des 4 copules bivariées modélisées précédemment, nous pouvons obtenir 5 lois marginales uniformes, c'est-à-dire une pour chaque risque. Dans le détail :

- U_1 : variable aléatoire uniforme du risque "Dommages aux biens".
- U_2 : variable aléatoire uniforme du risque "Perte de biens".
- U_3 : variable aléatoire uniforme du risque "Dommages corporels".
- U_4 : variable aléatoire uniforme du risque "Perte employé".
- U_5 : variable aléatoire uniforme du risque "Véhicule".

D'après les couples établis dans la modélisation des copules bivariées, nous pouvons représenter les liaisons entre nos risques comme ceci :



Dans ce cadre, $E = \{\{12\}, \{14\}, \{15\}, \{23\}\}$, $n_1 = 3, n_2 = 2, n_3 = n_4 = n_5 = 1$. Donc la copule multivariée décrivant nos 5 risques se définit par :

$$C(u_1, u_2, u_3, u_4, u_5) = \tilde{C}_{\{12\}}(u_1^{1/3}, u_2^{1/2}) \tilde{C}_{\{14\}}(u_1^{1/3}, u_4) \tilde{C}_{\{15\}}(u_1^{1/3}, u_5) \tilde{C}_{\{23\}}(u_2^{1/2}, u_3) \quad (5.8)$$

Il est supposé dans ce modèle PBC (Product of Bivariate Copula) revisité que les couples $\{ij\}$ n'appartenant pas à E disposent de lois uniformes indépendantes telles que $C_{\{ij\}}(u_i, u_j) = u_i u_j$, C représentant la copule bivariée issue du modèle PBC multivarié. D'autre part, on remarque que l'agrandissement de l'arbre, c'est-à-dire une connexion élevée entre les lois uniformes augmente les variables n_i , ce qui d'après l'équation du modèle PBC revisité, affaiblit les dépendances de la copule multivariée.

La section poursuit l'étude avec une implémentation de la copule multivariée modélisant nos cinq risques ainsi que l'étude des dépendances de la copule obtenue. Un comparatif avec les copules bivariées paramétriques sera établi.

5.6.3 Application numérique et étude des dépendances

La simulation numérique de la copule multivariée s'articule selon le modèle PBC exposé préalablement ainsi que ses propriétés probabilistes associées. Soient $\tilde{C}_{\{ij\}}$, copules bivariées dépendant du paramètre θ_{ij} . La procédure de simulation est la suivante :

- Pour tout $\{ij\} \in E$, générer $(U_i^{(ij)}, U_j^{(ij)}) \sim \tilde{C}_{\{ij\}}$
- Pour tout $i=1, \dots, d$, calculer $U_i = \max_{k \in \{1, \dots, d\}: \{ki\} \in E} \left\{ \left(U_i^{(ki)} \right)^{n_i} \right\}$

En implémentant l'algorithme pour nos couples de risques, les variables U_1, U_2, U_3, U_4 et U_5 composent la copule multivariée de dimension 5 modélisant les dépendances des risques du portefeuille.

Par la simulation de 10000 éléments représentant la copule multivariée, il est possible de mesurer les dépendances deux à deux des risques de la copule multivariée. Ci-dessous se trouvent les mesures de dépendance de type tau de Kendall pour les risques 2 à 2 dans le cadre de la copule multivariée. Un comparatif avec les dépendances des copules bivariées est également réalisé.

Tau de Kendall	Dommages/ Perte de biens	Dommages aux biens/ Perte employé	Dommages aux biens Vehicule	Perte biens/ Dmg corporels
Copule PBC	0,1124	0,02	-0,0172	-0,0487
Copule bivarié	0,2797	0,0349	-0,052	-0,0787

TABLE 5.10 – Mesure des dépendances 2 à 2 dans le cadre multivarié et comparaison avec le cadre bivarié

Comme constaté dans les remarques sur le modèle, par rapport aux copules bivariées paramétriques obtenues préalablement dans l'étude, on observe une dépendance globalement plus faible des risques 2 à 2 au sein de la copule PBC multivariée. Il est à noter que les dépendances négatives sont également dépréciées. On peut expliquer cette diminution de la dépendance par l'établissement seulement de 4 couples parmi 10 possibles ($|E|=10$ au maximum lorsque $d=5$), ce qui suppose que les autres couples possibles sont indépendants. Également, n_1 est ici égal à 3, ce qui réduit la dépendance pour les couples associés à la variable "Dommages aux biens" (notamment pour le couple "Dommages aux biens/Perte de biens" qui présente la dépendance la plus élevée parmi tous les risques).

Pour compléter le comparatif des dépendances, la dépendance de queues pour les risques 2 à 2 est exposée ci-dessous :

Tail	Dommages/ Perte de biens	Dommages aux biens/ Perte employé	Dommages aux biens Vehicule	Perte biens Dmg corporels
λ_L	0,02	0	0,03	0,03
λ_U	0,11	0,03	-0,07	0,07

TABLE 5.11 – Mesure des dépendances de queue de chaque couple au sein de la copule PBC multivariée

En comparant avec les dépendances de queue des copules bivariées, on observe une diminution de la dépendance de queue inférieure ainsi qu'une faible diminution de la dépendance de queue supérieure. L'asymétrie présumée des dépendances dans le cas bivarié (avec une copule de Gumbel dont la dépendance de queue à droite était significative) subsiste dans ce cadre multivarié.

La méthode PBC présente des propriétés avantageuses. En effet, elle ne nécessite que la connaissance des copules bivariées au sein des risques à modéliser. De plus, elle préserve (comme vu ci-dessus) l'asymétrie des queues de distribution et est un modèle flexible. Cependant, certaines limites sont à noter. D'une part les couples dont la copule bivariée n'est pas définie sont établis comme indépendants, ce qui génère de facto une dépendance à la baisse. D'autre part, un graphe trop connecté, avec des nombreuses liaisons entre chaque risque, affaiblit également la dépendance.

L'utilisation de la méthode PBC pour cette étude semble néanmoins adéquate dans ce contexte,

étant donnée la relative faiblesse des dépendances modélisées dans le cas bivarié ainsi que le faible nombre de couples parmi les risques du portefeuille. Il sera intéressant d'établir à partir des copules multivariées obtenues l'impact des dépendances de l'ensemble des risques sur le besoin en fonds propres.

5.7 Modélisation du besoin en fonds propres

5.7.1 Mise en contexte et hypothèses

Afin de respecter ses engagements d'indemnisation de l'assuré, l'assureur doit maintenir un niveau minimum de fonds propres. La gestion optimale du besoin en fonds propres répond à des intérêts externes divers. Le régulateur (par exemple l'ACPR en France) souhaite que l'assureur soit le plus solvable possible avec un niveau de fonds propres le plus élevé. A contrario, l'investisseur souhaite allouer le moins possible de capital dans la compagnie d'assurance afin de gagner en rentabilité, donc minimiser le besoin en fonds propres de l'assureur.

L'objectif de cette section est de mesurer l'impact des dépendances entre les risques sur le besoin en fonds propres pour la société TSA à l'aide de copules. On suppose ici que la TSA prend en charge l'indemnisation entière des victimes.

Pour mener à bien cette étude, nous allons utiliser différentes mesures de risque appliquées aux charges de sinistres agrégés sur un couple de sinistre donné. Plus précisément, le but sera de comparer les mesures de risque en supposant l'indépendance puis la dépendance entre les types de sinistre d'un couple. L'impact des dépendances sur le besoin en fonds propres sera donc caractérisé par la mesure de l'écart entre les mesures de risque dans le cas dépendant et indépendant.

Plusieurs hypothèses ont été émises pour pouvoir mener à bien la modélisation du besoin en fonds propres. Tout d'abord, la charge de sinistre est la seule variable aléatoire du modèle. On suppose que le besoin en fonds propres s'établit sur l'année 2014 (les montants de sinistre ont été calculés "as if" 2014), l'analyse n'est pas dynamique au sens où l'on ne peut estimer le besoin en fonds propres sur une année t quelconque. Enfin, on rappelle que les sinistres à chaque date sont considérés ici comme indépendants et suivant la même loi.

5.7.2 Les différentes mesures de risque appliquées au modèle : VaR et TVaR

Un outil nécessaire à la mesure du besoin en fonds propres en assurance est l'utilisation des mesures de risque. Les mesures de risque permettent de déterminer un niveau de capital à allouer pour un portefeuille donné. Une mesure de risques (selon JES 2010, CHARPENTIER) est une fonction définie sur l'espace des variables aléatoires et prenant ses valeurs dans \mathbb{R} :

- Invariance en loi : X converge en loi vers $Y \implies R(X) = R(Y)$;
- Croissance : $X \geq Y \implies R(X) \geq R(Y)$;
- Invariance par translation : $\forall k \in \mathbb{R}, R(X + k) = R(X) + k$;
- Homogénéité : $\forall \lambda \in \mathbb{R}^+, R(\lambda X) = \lambda R(X)$;
- Sous-additivité : $R(X + Y) \leq R(X) + R(Y)$;
- Convexité : $\forall \beta \in [0, 1], R(\beta X + [1 - \beta]Y) \leq \beta R(X) + [1 - \beta]R(Y)$.

Les mesures de risque utilisées dans le cadre de notre étude du besoin en capital de la société TSA sont la TVaR ainsi que la VaR.

La Value at Risk (VaR) est une mesure très utilisée en assurance comme en finance. La Value at

Risk de niveau α associée à un risque X est donnée par :

$$VaR_\alpha(X) = \inf\{x | F_X(x) \geq \alpha\}. \quad (5.9)$$

Nous pouvons également définir la VaR selon la fonction quantile de la loi X $F_X^{-1}(\alpha)$ tel que $VaR_\alpha(X) = F_X^{-1}(\alpha)$. Cette mesure est une mesure de risque qui n'est pas sous additive. Elle représente simplement la quantité qui va permettre de couvrir un montant de sinistres engendré par le risque X selon un niveau α . Cette mesure est intéressante dans le cadre d'une étude de solvabilité d'une compagnie mais reste peu sensible aux montants extrêmes.

La Tail Value at Risk (TVaR) de niveau α associée au risque X se définit par :

$$TVaR(X; \alpha) = \frac{1}{1 - \alpha} \int_\alpha^1 VaR(X; t) dt. \quad (5.10)$$

La TVaR est donc la moyenne des VaR de niveau supérieur à α . On peut aussi exprimer la TVaR de la façon suivante :

$$TVaR(X; \alpha) = VaR_\alpha(X) + \frac{1}{1 - \alpha} \mathbb{E}(|X - VaR_\alpha(X)|) \quad (5.11)$$

La TVaR est plus sensible aux distributions de queue de la variable aléatoire X du fait de l'intégration d'un chargement de sécurité par rapport à la VaR. C'est une mesure de risque sous-additive.

Ces deux mesures de risque seront calculées lors de l'application, leur interprétation dépendra du couple de sinistres sur lequel se portera la mesure.

5.7.3 Simulations suivant l'indépendance ou la dépendance entre les risques

Nous tâcherons d'exposer dans cette section l'algorithme de simulation des charges de sinistre. Plus précisément, il s'agira de calculer la charge agrégée de sinistres. Cette charge agrégée s'exprime par la somme des charges de sinistres au sein d'un même couple de sinistres. Pour obtenir ces montants de sinistre, nous procéderons à la simulation de couples (u,v) suivant la copule déterminant la dépendance de notre couple de sinistres. Puis par inversion de fonction de répartition, il sera possible d'obtenir les charges de sinistre souhaitées. Voici en détail l'algorithme suivi :

- Effectuer 10000 simulations du vecteur uniforme (u,v) sur $[0;1]$ ayant pour loi jointe $C(u,v)$ (copule paramétrique choisie selon les choix de copule établis pour chaque couple de sinistres, cf 5.5) . Cette simulation se réalise à partir de l'algorithme de Marshall-Olkin utilisé dans la fonction Mind "MM_SIMCOPULA" (détails de l'algorithme exposés préalablement).
- En déduire les charges de sinistres X_1 et X_2 par inversion de la fonction de répartition empirique associée à la loi marginale. X s'exprime donc tel que $X = (X_1, X_2) = (F_1^{-1}(u), F_2^{-1}(v))$. A noter que l'approche paramétrique associant une loi paramétrique à la loi marginale ne semblait pas pertinente comme vu dans la section portant sur l'ajustement des lois marginales. C'est pourquoi nous utiliserons l'approche non-paramétrique avec une fonction de répartition empirique pour chaque loi marginale d'un couple.
- Estimer pour les 10000 charges de sinistres $X^{(i)} = (X_1^{(i)}; X_2^{(i)})$; $i = 1, \dots, 10\ 000$; la charge de sinistre agrégée $S^{(i)} = X_1^{(i)} + X_2^{(i)}$.

Nous avons donc une distribution empirique de la charge de sinistre agrégée grâce aux 10 000 réalisations de S. Nous nous appuierons sur cette distribution pour l'estimation de l'impact des dépendances sur le besoin en fonds propres.

Dans le cas d'une hypothèse d'indépendance entre les sinistres, la méthode suivie est la même,

à l'exception du vecteur uniforme (u,v) simulé selon une loi jointe $C(u,v)$ de type copule indépendante. Cette méthode garantit des charges de sinistre indépendantes entre elles, et donc une distribution empirique S décrivant l'indépendance entre les charges de sinistre.

5.7.4 Impact des dépendances sur le besoin en fonds propres

Cette section présente les résultats du modèle d'évaluation du besoin en fonds propres. Par l'intermédiaire des mesures de risque VaR et TVaR appliquées aux distributions empiriques S des charges agrégées de sinistre d'un couple, nous avons pu établir le besoin en fonds propres pour la société TSA à date fixée (année 2014) dans le cas où l'on suppose l'indépendance ou la dépendance entre les types de sinistre.

Ci-dessous se trouve les résultats selon une VaR et une TVaR à un niveau 75% ou 99,5% pour chacun des quatre couples de sinistre.

Dommages/Perte de biens	$VaR_{75\%}$	$VaR_{99.5\%}$	$TVaR_{75\%}$	$TVaR_{99.5\%}$
Cas dépendant (Gumbel)	4882,37	23454,03	17983,24	4006024
Cas indépendant	5 041,51	19 251,08	17 490,83	3 144 863
Ecart	-3,26%	21,83%	2,82%	27,38%

TABLE 5.12 – Comparaison des mesures de risque en euros pour le couple Dommages aux biens/Perte de biens dans le cas dépendant et indépendant

Rappelons que pour ce couple de risque, la copule de Gumbel représente le mieux la dépendance de risques au sein du couple de sinistres. La copule de Gumbel présente une sensibilité au niveau de la queue de distribution supérieure des lois marginales représentant chaque sinistre du couple. Dans ce contexte, l'interprétation du besoin en fonds propres pour ce couple de sinistres par la mesure de risque TVaR semble pertinente. En effet, la TVaR semble une bien meilleure mesure de risque dans ce cas d'étude car elle est plus sensible aux queues de distribution. Rappelons qu'elle constitue une moyenne des VaR à partir d'un certain seuil.

La VaR semble peu adaptée dans ce cas d'étude car nos charges de sinistre présentent une queue de distribution avec une VaR à croissance non-linéaire selon le seuil.

En observant la TVaR à un niveau de 75%, on constate un besoin en fonds propres supérieur de 2,82% pour couvrir le couple de risques si l'on considère qu'il y a dépendance entre les risques, par rapport à une indépendance entre les risques du couple. En avançant le niveau à 99,5%, on s'aperçoit que l'écart de besoin en fonds propres est plus significatif (27,38%).

Perte biens/Dom. corporels	$VaR_{75\%}$	$VaR_{99.5\%}$	$TVaR_{75\%}$	$TVaR_{99.5\%}$
Cas dépendant (tCopula)	8356,021	31985,77	28589,22	5138989
Cas indépendant	8650,82	31836,09	30614,77	5084060
Ecart	-3,53%	0,47%	-7,09%	1,08%

TABLE 5.13 – Comparaison des mesures de risque en euros pour le couple Perte de biens/Dommages corporels dans le cas dépendant et indépendant

Dommages aux biens/Veh.motorisé	$VaR_{75\%}$	$VaR_{99.5\%}$	$TVaR_{75\%}$	$TVaR_{99.5\%}$
Cas dépendant (tCopula)	4041,611	10948,71	10965,66	1 556704
Cas indépendant	4184,03	10131,45	11697,59	1 390 401
Ecart	-3,45%	8,06%	-6,67%	6,53%

TABLE 5.14 – Comparaison des mesures de risque en euros pour le couple Dommages aux biens/Véhicule motorisé dans le cas dépendant et indépendant

Pour les résultats concernant les deux couples ci-dessus, la dépendance entre les types de sinistre étant négative, nous pouvons observer un écart négatif entre le besoin en fonds propres dans le cas d'une dépendance entre les risques, et une indépendance entre les risques. Il est plus judicieux ici d'interpréter les résultats selon le prisme de la VaR ici, car la t-Copula caractérisant la dépendance entre les risques pour ces deux couples présente une dépendance de queue très faible. Si les queues de distribution sont présentes mais faibles on privilégiera la VaR, moins sensible aux queues de distribution.

Dans le cadre de ces deux couples de sinistre, le besoin en fonds propres est donc moins prononcé si l'on suppose la dépendance entre les risques du couple, par rapport à une indépendance entre les risques d'un couple (écart de respectivement -3,53% et -3,45%).

Dommages aux biens/Perte employé	$VaR_{75\%}$	$VaR_{99.5\%}$	$TVaR_{75\%}$	$TVaR_{99.5\%}$
Cas dépendant (tCopula)	4600,816	11230,97	12647,62	1543714
Cas indépendant	4582,55	10714,16	12552,96	1447175
Ecart	0,4%	4,8%	0,75%	6,67%

TABLE 5.15 – Comparaison des mesures de risque en euros pour le couple Dommages aux biens/Perte employé dans le cas dépendant et indépendant

La dépendance entre les risques 'Dommages aux biens' et 'Perte employé' est caractérisée ici par une t-Copula, avec une présence faible de dépendance de queue à gauche et à droite. La sensibilité aux queues de distribution étant faible, nous privilégierons ici une interprétation des résultats selon la VaR. Rappelons que la dépendance au sein de ce couple de sinistres est positive. Nous pouvons conclure ici un besoin en fonds propres dans le cas dépendant supérieur de 0,4% par rapport au cas indépendant (mesuré par la VaR à 75%). Un besoin en fonds propres plus important semble se dessiner si l'on suppose une dépendance entre les risques par rapport à une indépendance entre les risques.

Un approfondissement de la modélisation du besoin en fonds propres dans un cadre multi-risques peut aussi être réalisé. A partir de la copule multivariée obtenue dans la section précédente par la méthode PBC, il est en effet possible de générer les charges de sinistres agrégées des 5 risques du portefeuille selon la même méthode de simulation que dans le cas bivarié (charge obtenue par la fonction inverse de répartition empirique pour chaque marge appliquée à chaque uniforme U_1, U_2, U_3, U_4 et U_5).

Au regard de la répartition des quantiles des charges dans le cas dépendant et indépendant, il apparaît qu'un besoin de fonds propres à la baisse se fait sentir, sauf en $VaR_{99,5\%}$.

$VaR_{75\%}$ dépendant	11 156,3378
$VaR_{75\%}$ indépendant	11 358,92
Ecart	-1,78%

TABLE 5.16 – $VaR_{75\%}$ dans le cadre multirisque

Au vu de la dépendance de queue présente pour les charges les plus élevées, la VaR semble moins appropriée dans la mesure du besoin en fonds propres de ces risques. La mesure par le biais de la TVaR, plus sensible aux queues de distribution semble ici plus adaptée.

Mesure par TVaR	$TVaR_{75\%}$	$TVaR_{99,5\%}$
Multirisques dépendant	33 426,89	5 505 137
Multirisques indépendant	33 389,16	5 106 543
Ecart	0,11%	7,81%

TABLE 5.17 – Mesures par TVaR dans le cadre multirisque

Les mesures par la TVaR établissent un besoin en fonds propres supplémentaire dans le cas dépendant qui se révèle quasiment négligeable (0,11%). En comparaison de l'impact des dépendances dans le cas bivarié, il apparaît dans ce cadre multirisques que les dépendances ne renvoient que légèrement à la hausse le besoin en fonds propres (0,11% vs 2,82% dans le cas bivarié de la copule de Gumbel).

5.8 Conclusion sur l'impact des dépendances modélisées

L'étude suivie au cours de ce chapitre nous permet de modéliser avec précision les dépendances au sein de plusieurs couples de sinistre par le biais de copules paramétriques. Dans un premier temps, nous avons essayé de calibrer les lois marginales selon des lois de la famille exponentielle. Cependant, il résulte des différents tests effectués (test de Kolmogorov-Smirnov notamment) un mauvais ajustement des lois marginales aux lois paramétriques. Une approche semi-paramétrique a donc été considérée avec l'estimation de fonctions de répartition empirique pour chaque loi marginale. Ces fonctions ont permis le calcul des paramètres de copules par la méthode du pseudo maximum de vraisemblance.

Le choix de la bonne copule paramétrique pour un couple de sinistre donné est une étape cruciale dans notre étude. Pour mener à bien cette étape, des tests d'ajustement par χ^2 et par représentation graphique de la fonction $K(z)$ ont été entrepris.

L'objectif de ce chapitre réside dans la mesure d'un impact de la dépendance entre risques au sein d'un portefeuille. Pour ce faire, nous avons pu mesurer le besoin en fonds propres, par l'intermédiaire de mesures de risques telles que la VaR ou la TVaR, selon que l'on suppose la dépendance (par le biais des copules préalablement choisies) ou l'indépendance des risques pour un couple de risques donné. Il résulte de cette étude que le besoin en fonds propres semble supérieur lorsque l'on considère une dépendance entre les risques par rapport à une indépendance entre les risques. A l'inverse, dans le cas de dépendances négatives caractérisant un couple donné (couples Perte de biens/Dommages corporels et Dommages aux biens/Véhicule motorisé), le besoin en fonds propres lorsque l'on considère la dépendance entre les risques est cette fois inférieur au besoin en fonds propres dans le cas indépendant.

L'étude des copules multivariées nous a permis de modéliser les dépendances de la quasi-totalité des risques du portefeuille. Par la méthode PBC, il en ressort des mesures de dépendances affaiblies entre deux risques donnés, dus notamment à l'indépendance supposée des couples non modélisés (comme vu au début de l'étude, certains couples ne présentent pas ou peu de corrélations, d'où un abandon de modélisation des dépendances pour ces couples). Dans ce cadre multivarié, l'impact des dépendances sur le besoin en fonds propres est moins significatif, l'assureur aurait une faible part de capital à ajouter par rapport à une modélisation par risques indépendants. Il est à souligner dans cette étude l'impact des queues de distribution sur le besoin en capital (utilisation de la VaR au lieu de TVaR, asymétrie vers des dépendances pour des charges extrêmes, ce qui génère un besoin supplémentaire en capital).

Par ailleurs, le modèle appliqué dans ce chapitre projette les sinistres selon l'année civile 2014. Le modèle est donc discret puisqu'il modélise la situation financière à un instant donné de l'année. Une analyse dynamique n'est pas réalisée ici mais serait un axe à étudier. Le modèle décrit est donc une approche simplifiée et ne prétend aucunement couvrir la solvabilité d'une société d'assurance non-vie sur la base du portefeuille de la TSA.

Conclusion

La théorie des copules est un outil flexible qui capte de façon réaliste tous types de dépendances (dépendance non-linéaire) et notamment celles présentes dans les valeurs extrêmes. Le travail proposé dans ce mémoire s'appuie d'un contexte multi-risques d'assurance non-vie, mais peut s'assimiler également dans un cadre Solvabilité 2 (calculs de SCR dans les différents modules de risque) ou un contexte financier (ALM, épargnes).

L'objectif tout au long de ce mémoire a consisté à mesurer l'impact de l'introduction de la dépendance (et notamment des dépendances extrêmes) entre les types de sinistre sur la solvabilité d'une compagnie d'assurance non-vie. Dans ce cadre, l'utilisation des copules s'est révélée déterminante pour construire une structure de dépendance représentant au mieux les différents sinistres.

Nous avons appliqué notre étude sur une base de données représentant un portefeuille de sinistres survenus sur la période 2002 à 2015 enregistrés par la TSA (Transportation Security Administration), agence publique américaine se chargeant de la sûreté aéroportuaire dans la plupart des aéroports des États-Unis.

Le retraitement des données, effectué en grande partie sur la plateforme Milliman Mind, a été réalisé sur la base d'une hypothèse d'indépendance des sinistres selon les jours. Ainsi les charges de sinistres ont été agrégées selon les dates de survenance et les types de sinistre.

Après étude de corrélations entre les couples de sinistre, nous avons retenu quatre couples de sinistres pouvant être sujet à des dépendances. Afin de définir une structure de dépendance représentant de façon fidèle les charges de sinistre, deux étapes sont prépondérantes : estimer un paramètre de copule par une approche paramétrique ou semi-paramétrique, puis choisir judicieusement la bonne copule paramétrique représentant la structure de dépendance en s'appuyant sur le test du χ^2 et test graphique par la fonction $K(z)$. L'ajustement des lois marginales par des lois de la famille exponentielle n'étant pas très concluant, nous avons choisi de procéder à une approche semi-paramétrique pour l'estimation du paramètre de copule. En effet, l'estimation par la méthode du pseudo maximum de vraisemblance (ou CML) semble conduire à des copules plus fidèles aux structures de dépendance de nos couples de risques.

Nous pouvons remarquer que les copules choisies (copule de Gumbel, copules de Student) possèdent toutes soit une dépendance de queue supérieure, soit une dépendance de queue inférieure et supérieure. Les sinistres ont donc une portée volatile et une tendance évoluant vers des charges extrêmes qui n'est pas à négliger dans l'impact sur le besoin en fonds propres.

Parmi les quatre couples de risques, deux couples sont caractérisés par une dépendance positive tandis que les deux restants décrivent une dépendance négative. L'impact du besoin en fonds propres sur les couples de risque est différent selon la nature des dépendances au sein du couple. En effet, une dépendance positive entre les risques décrit un besoin en fonds propres plus élevé par rapport à une hypothèse d'indépendance entre les risques. À l'inverse, une dépendance négative entre les risques décrit un besoin en fonds propres moins élevé par rapport à une hypothèse d'indépendance entre les risques. Les écarts (allant de -4 à 3%) sont non-négligeables et reflètent un réel impact des dépendances sur le besoin en fonds propres.

L'étude de risques multiples du portefeuille, par la méthode de produits de copules bivariées (PBC), a permis d'analyser les dépendances des couples de risques dans un cadre multivarié. Il apparaît alors, du fait principalement de la présence de paires de marginales indépendantes, une dépendance plus faible au sein des différents couples de risques de la copule multivariée par rapport aux dépendances mesurées par les copules paramétriques bivariées. Néanmoins, la présence d'asymétrie dans les queues de distribution subsiste dans le cadre multirisques, ce qui doit retenir l'attention de l'assureur dans son choix final de modélisation des risques (une modélisation par copule multivariée gaussienne ou par copule indépendante supposerait l'absence de dépendances de queues). L'impact des dépendances sur besoin en fonds propres est moins significatif dans le cadre multirisques, on note cependant un léger besoin en capital supplémentaire dans le cas dépendant, dû en grande partie à la présence de dépendances de queue.

Il ressort de cette étude, en comparaison à une modélisation des dépendances entre risques de lois marginales suivant des lois normales, une estimation du besoin en fonds propres plus fine lorsque l'on modélise la dépendance entre les risques du portefeuille à l'aide de copules paramétriques. En général, l'assureur aura un coût en capital plus élevé lorsqu'il construira un modèle de dépendances à l'aide de copules. Le point intéressant traité dans ce mémoire est l'effet des dépendances négatives qui généreront un coût en capital moins élevé, compensant ainsi les coûts plus élevés liés aux dépendances positives. Il est intéressant également de remarquer un léger affaiblissement des dépendances dans un cadre multirisques, dues principalement à la présence de couples de risque indépendants. Le coût en capital sera de facto moins élevé mais pourra subsister si une dépendance des queues est observée.

A noter également que l'assureur peut avoir des réticences à la modélisation des dépendances par les copules paramétriques du fait de biais possibles dans la calibration du paramètre de copule (ajustement des lois marginales, biais dans les méthodes de maximum de vraisemblance). L'approche semi-paramétrique sur laquelle est basé ce mémoire à l'aide de la méthode du pseudo maximum de vraisemblance participe dans ce contexte à une réduction de l'erreur de calibration du modèle.

Le mémoire ne fait pas figure d'exhaustivité dans la solvabilité d'un portefeuille non-vie et présente des limites et des améliorations.

L'étude se base sur un modèle simplifié avec un besoin en fonds propres exprimé selon l'unique année 2014 et une hypothèse d'indépendance temporelle entre sinistres.

Enfin, le manque d'observations de sinistres sur l'ensemble des risques ne nous a malheureusement pas permis d'explorer la totalité des risques du portefeuille. En effet, cela permettrait d'établir des corrélations plus significatives et ainsi des modélisations de copules bivariées plus pertinentes. Surtout, cela permettrait d'établir un besoin en fonds propres fidèle au portefeuille donné et de mesurer de meilleure façon l'erreur commise par le calcul du besoin en fonds propres selon l'indépendance entre les risques.

Annexes

Annexe 1

Démonstration - Bornes multivariées de Fréchet-Hoeffding

Posons $U_i = F_i(X_i)$ pour tout $i = 1, \dots, d$, d'après la proposition 1.1 on a $\mathbb{P}(U_i \leq u_i) = u_i$ pour tout $i = 1, \dots, d$.

— La première inégalité provient du fait que

$$\begin{aligned}\mathbb{P}\left(\bigcap_{1 \leq i \leq d} (U_i \leq u_i)\right) &= 1 - \mathbb{P}\left(\bigcup_{1 \leq i \leq d} (U_i \geq u_i)\right) \\ &\geq 1 - \sum_{i=1}^d \mathbb{P}(U_i > u_i) \\ &\geq 1 - \sum_{i=1}^d (1 - u_i) \\ &\geq 1 - d + \sum_{i=1}^d u_i.\end{aligned}$$

Et comme $\mathbb{C}(\cdot)$ est définie sur $[0, 1]^d$, alors

$$\mathbb{C}(u) \geq \max(0, 1 - d + \sum_{i=1}^d u_i)$$

— La seconde inégalité provient du fait que

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

ainsi $\mathbb{C}(u) = \mathbb{P}\left(\bigcap_{1 \leq i \leq d} (U_i \leq u_i)\right) \leq \mathbb{P}(U_i \leq u_i) = u_i$ pour tout $i = 1, \dots, d$.

Par conséquent

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

Annexe 2

Démonstration - théorème 1 sur la mesure de concordance

Puisque toutes les variables aléatoires sont continues, alors

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

ainsi $\mathcal{Q} = 2\mathbb{P}((X_1 - X_2)(Y_1 - Y_2) \geq 0) - 1$. Etudions le premier terme

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

ces quantités peuvent être calculées par intégration

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

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

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

D'une façon similaire

$$\begin{aligned} \mathbb{P}(X_1 < X_2, Y_1 < Y_2) &= \int_{\mathbf{R}^2} \mathbb{P}(X_2 > x, Y_2 > y) d\mathbf{C}_1(F(x), G(y)) \\ &= \int \int_{\mathbf{R}^2} (1 - F(x) - G(y) + \mathbb{C}_2(F(x), G(y))) d\mathbf{C}_1(F(x), G(y)) \\ &= \int_{[0,1]^2} (1 - u - v + \mathbb{C}_2(u, v)) d\mathbf{C}_1(u, v) \end{aligned}$$

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

$$\begin{aligned} \mathbb{P}(X_1 < X_2, Y_1 < Y_2) &= 1 - \frac{1}{2} - \frac{1}{2} + \int \int_{[0,1]^2} \mathbb{C}_2(u, v) d\mathbf{C}_1(u, v) \\ &= \int_{[0,1]^2} \mathbb{C}_2(u, v) d\mathbf{C}_1(u, v) \end{aligned}$$

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

En regroupant ces résultats on en déduit :

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

Bibliographie

- ALEIYOUKA, M. (2018), ‘Sur la dépendance des queues de distributions, physique mathématique normandie université.’.
- BELGUISE, O. (2001), ‘Tempêtes : Etude des dépendances entre les branches auto et incendie avec la théorie des copulas’.
- BRUNO, L. (2003), ‘Modélisation de la provision des sinistres automobile - apport des copules’.
- CADOUX, D. & LOIZEAU, J.-M. (2004), ‘Copules et dépendances : application pratique à la détermination du besoin en fonds propres d’un assureur non-vie’, *Bulletin Francais d’Actuariat* **(6)12**, 173–208.
- CHARPENTIER, A. (Novembre 2010), ‘Journées d’études statistique, mesures de risque’.
- CRENIN, F. & JEBBARI, K. (2012), ‘Modélisation des dépendances entre risques opérationnels à l’aide des copules’.
- DEHEUVELS, P. (1991), ‘On the limiting behavior of the pickands estimator for bivariate extreme-value distributions’, *Statistics and Probability Letters* **12**, 429–439.
- GENEST, C. & MACKAY, R. J. (1904), ‘The proof and measurement of association between two things’, *American Journal of Psychology* **15**, 72–101.
- GENEST, C. & MACKAY, R. J. (1963), ‘The advanced theory of statistics’, *Hodder Education* **1**.
- GENEST, C. & MACKAY, R. J. (1986), ‘The joy of copulas : Bivariate distributions with uniform marginals’, *The American Statistician* **40**, 280–283.
- HOEFFDING, W. (1948), ‘A class of statistics with asymptotically normal distribution’, *The Annals of Mathematical Statistics* **19**, 293–325.
- HOFERT, M. (2008), ‘Sampling archimedean copulas’, *Computational Statistics Data Analysis* **52**(12), 5163–5174.
- HU, S. & HAGAN, A. (2021), ‘Copula averaging for tail dependence in insurance claims data’.
- HURLIMAN, W. (2001), ‘"fitting bivariate cumulative returns with copulas", working paper, winterthur life and pensions.’.
- JOE, H. (2001), ‘Multivariate models and dependence concepts. chapman hall/crc’.
- JOE, H. & XU, J. (1996), ‘"the estimation method of inference functions for margins for multivariate models", department of statistics, university of british columbia, technical report.’.
- KOJADINOVIC, I. (2011), ‘Large-sample tests of extreme-value dependence for multivariate copulas’.

- KUMAR, P. (2010), ‘Probability distributions and estimation of ali-mikhail-haq copula’, *Applied Mathematical Sciences* **(4)14**, 657 – 666.
- LASZKIEWICZAND, M. (2021), ‘Copula-based normalizing flows’.
- LAURIA, D. (2021), ‘Global and tail dependence : A differential geometry approach’.
- LIEBSCHER, E. (2008), ‘Construction of asymmetric multivariate copulas. journal of multivariate analysis’.
- MAZO, G. (2014), ‘Construction et estimation de copules en grande dimension’.
- MISERAY, A. (2015), ‘Analyse du coût des sinistres dans le cadre de la modélisation ligne à ligne d’un portefeuille d’assurance automobile’.
- NELSEN, R. B. (1999), ‘An introduction to copulas’, *Lecture Notes in Statistic* **139**, 280–283.
- NELSEN, R. B. (2006), ‘An introduction to copulas. springer verlag, new york.’.
- RANNOU, S. (2012), ‘Modélisation du risque de défaut sur un portefeuille d’obligations d’entreprises en intégrant les approches par copules et valeurs extrêmes’.
- SCHMIDT, R. & STADMÜLLER, U. (2006), ‘Non parametric estimation of tail dependence’, *The Scandinavian Journal of Statistics* **33**, 307–335.
- SCHNEIDER, W. (2021), ‘On the distribution of the sum of dependent standard normally distributed random variables using copulas’.
- SHAO, J. (1999), ‘Mathematical statistics. springer’.
- VENTER, O. (2000), ‘Tails of copulas’.
- WIDDER, D. V. (1941), ‘The laplace transform. princeton mathematical series, v. 6. princeton university press, princeton, n. j.’.
- ZARI, T. (n.d.), ‘Contribution à l’étude du processus empirique de copule. mathématiques [math]. université pierre et marie curie - paris vi’.