UNIVERSITÉ PARIS DAUPHINE

Département MIDO(*)

MASTER MIDO
MENTION MMD(**)

SPÉCIALITÉ ACTUARIAT

Année Universitaire : 2007-2008

Mémoire d'Actuariat présenté en **Novembre 2008** devant l'Université Paris Dauphine et l'Institut des Actuaires

Par : ***Christelle ANDRIANTAVY*** Tuteur : ***Danielle BRIEGLEB***

Sujet : ***Tempêtes : prise en compte de la dépendance spatiale avec la théorie des copules***

Entreprise d'accueil : ***Groupama***

NON CONFIDENTIEL

JURY

Membres du Jury	**Fonctions / Entreprise**
Christian HESS	Président du jury, Professeur à l'Université Paris-Dauphine
Gérard CROSET	Membre du Jury de l'Institut des Actuaires

(*) MIDO : Mathématiques, Informatique, Décision, Organisation
(**) MMD : Mathématiques, Modélisation, Décision

Remerciements

Avant tout développement sur cette expérience professionnelle, je tiens à remercier sincèrement toutes les personnes qui ont contribué de près ou de loin à l'élaboration de ce mémoire.

Aussi, j'adresse mes remerciements à Christophe Graber et son équipe pour la confiance qu'ils m'ont accordée ainsi que pour leur sympathie, qui m'a aidé à m'intégrer rapidement dans l'équipe.

Je remercie plus particulièrement Danielle Briegleb pour sa gentillesse et pour toute l'attention qu'elle m'a apportée tout au long de ce stage.

J'exprime ma reconnaissance à Emmanuel Figuereau, Gilles André et Nikita Aksenov qui m'ont beaucoup aidé et avec qui j'ai pris plaisir à travailler pendant ces six mois.

Je remercie également toute l'équipe enseignante à l'université Paris-Dauphine pour la réussite de cette année universitaire, plus particulièrement Christian Hess pour sa disponibilité et pour les conseils qu'il m'a donnés pour l'élaboration de ce mémoire.

Je remercie enfin tous les proches et amis qui m'ont soutenue et encouragée au cours de la réalisation de ce mémoire.

Résumé

La réglementation française distingue, parmi les événements naturels, la tempête, le poids de la grêle et de la neige, obligatoirement inclus dans les contrats Dommages aux Biens. Les tempêtes Lothar et Martin, survenues en décembre 1999, ont constitué le plus gros sinistre de l'histoire de l'assurance française avec un coût global assuré de 6,8 milliards d'euros (source : FFSA).

Les programmes d'assurance et de réassurance ont été sérieusement atteints par ces sinistres, mettant en évidence la nécessité de modéliser la survenance et la gravité des tempêtes, afin de permettre aux assureurs et réassureurs de définir le prix de la couverture ainsi que son organisation.

Les mesures de risque sont au centre des problématiques de Solvabilité 2. La Value-at-Risk (VaR) et la Tail Value-at-Risk (TVaR) en sont deux mesures de risque acceptables. L'objectif essentiel de la solvabilité est de contrôler la probabilité de ruine : la VaR est un outil naturel pour le faire. La TVaR fournit une mesure de l'intensité de la ruine lorsque celle-ci survient.

Les simulations de Monte-Carlo sont une des méthodes qui permettent d'évaluer la VaR et la TVaR. Il s'agit de générer des coûts pour avoir une distribution empirique, puis d'en extraire un quantile à $\alpha\%$. Puisque la société est composée de plusieurs entités réparties dans toute la France, il faudrait générer des coûts pour chaque entité, tout en prenant compte d'une certaine dépendance entre les différentes entités. Les copules sont un des outils permettant de modéliser la dépendance entre les risques.

La première partie de ce mémoire montre comment est modélisé le risque tempête. Il introduit ensuite les notions de mesures de risque puis la théorie des copules. Des simulations de Monte-Carlo permettront enfin de mesurer l'impact de la prise en compte ou non de la dépendance sur les différentes mesures de risques calculées.

Abstract

In French rule, windstorms, weight of snow and hail are considered apart from natural catastrophes and must be included in property damage insurance policies. Windstorms Lothar and Martin, which occurred in December 1999, represent the largest insurance loss in France with a total of 6.8 billion insured losses.

Insurance and reinsurance programs have been seriously affected by this event. That is why it is substantial to model occurrence and severity of windstorms, so that insurers and reinsurers can price and organize the coverage.

Risk measures are major concern for the framework Solvency II. Value-at-Risk (VaR) and Tail Value-at-Risk (TVaR) are examples of measures that can be calculated. One essential target for solvency is to control the ruin probability: VaR can be used for it. TVaR tells us about how big will be the ruin if the VaR is exceeded.

Monte-Carlo simulations are a way to assess VaR and TVaR. One can obtain an empirical distribution by simulating costs, and then extract a quantile from it. Because the company is divided into several entities in the whole France, one must generate costs for each entity and take into account dependence between them. Copulas can be used to model dependence between risks.

Windstorm modelling will be explained in the first part of this thesis. Then, risk measures and copula theory will be introduced. Finally, Monte-Carlo simulations will allow us to measure the impact of dependence on the different risk measures.

Table des matières

Première partie : Modélisation du risque tempête

Les tempêtes qui s'abattent en Europe représentent un énorme potentiel de sinistres, comme l'ont démontré les tempêtes Lothar et Martin, survenues en décembre 1999. Parallèlement, les récentes conférences internationales sur le changement climatique expliquent le regain d'intérêt pour les tempêtes. Elles seraient plus intenses à cause du dérèglement climatique dû à l'effet de serre.

Pour que les assureurs et les réassureurs puissent définir le prix de la couverture et l'organisation de celle-ci, la survenance et la gravité des événements doivent être modélisées de manière correcte. Des outils de modélisation de risques naturels tels RMS (Risk Management Solutions) permettent de déterminer l'impact financier des dommages.

1. Principe de fonctionnement des modèles

Les modèles dits modèles CAT se composent de plusieurs éléments. En particulier, pour le risque tempête :

- le *module d'événements stochastiques* contient une base de tempêtes simulées. Chaque tempête est décrite par ses paramètres physiques et sa fréquence. Ceux-ci sont déterminés grâce à une librairie de tempêtes historiques qui ont impacté l'Europe depuis plusieurs dizaines d'années : à partir des événements antérieurs, on génère des milliers d'événements possibles en faisant varier différents paramètres physiques.
- le *module de hasard* détermine l'intensité et la direction du vent à chaque localisation pour chaque événement stochastique susceptible de causer des pertes à cette localisation. Les régions les plus exposées ou les plus susceptibles d'être touchées sont sujettes à des calculs plus complexes. De plus, ce module tient compte de la rugosité du sol et de la topographie.
- le *module de vulnérabilité* fournit des courbes de vulnérabilités permettant de déduire le taux d'endommagement moyen (mean damage ratio) ainsi qu'un coefficient de variation (coefficient of variation) associé, pour prendre en compte l'incertitude sur le montant des dommages à une vitesse de vent donnée. Ce module contient des informations relatives aux biens assurés afin de pouvoir entreprendre une modélisation des sinistres. Pour pouvoir évaluer l'impact financier du sinistre, il faut également connaître la valeur des biens assurés.
- le *module d'analyse financière* calcule les pertes selon différentes perspectives financières, en considérant les structures d'assurance et de réassurance. En effet, les conditions d'assurance comme les pleins de conservation ou les limites sont, pour l'assureur, des outils essentiels pour limiter ses engagements.

Une combinaison de ces quatre modules permettra de calculer le sinistre annuel attendu (annuel expected loss) ou une mesure des sinistres catastrophiques extrêmes. A la fin du processus de modélisation, on obtient une liste des événements catastrophiques susceptibles de se produire dans une année et les fréquences associées à ces événements : les Event Loss Table ou ELT.

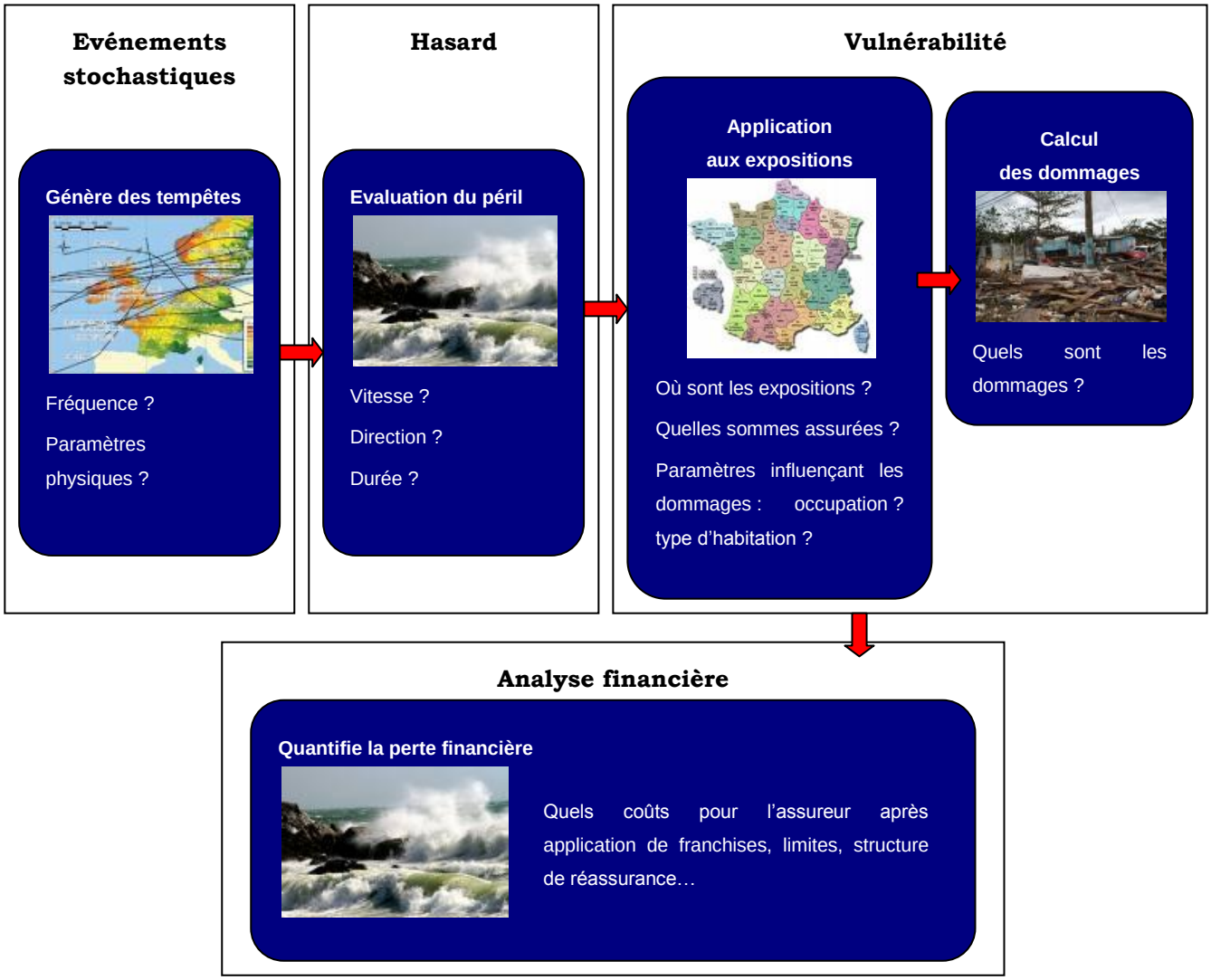

Figure 1: Modélisation du risque tempête

Le calcul des sommes assurées est fondamental pour estimer les expositions aux catastrophes naturelles.

La somme assurée est le montant pour lequel les biens sont couverts. En France, les contrats ne prévoient en général pas de calcul explicite de valeur assurée : les primes sont déterminées à partir du nombre de pièces ou de la surface du bien. Les indemnisations en cas de sinistre sont basées sur la valeur de remplacement ou de reconstruction vétusté déduite, ou encore sur une valeur à neuf du bien sinistré.

Les modèles CAT utilisent cette valeur comme base de calcul pour appliquer les taux d'endommagement aux portefeuilles. Même si on ne parle ici que de tempête, il va de soi que les sommes assurées sont aussi utiles pour divers risques comme la conflagration, les inondations ou la grêle. Elles peuvent également être un outil pour le risk management.

2. Les systèmes d'information géographique (SIG)

Les systèmes d'information géographique sont utiles au domaine de l'assurance : ils permettent, entre autres, aux assureurs de segmenter leur tarification et de spatialiser la vulnérabilité de leurs portefeuilles d'assurance. Les assureurs appuient leur gestion sur des cartes de risques, d'exposition, d'engagements… L'information géographique devient donc fondamentale pour les assureurs, notamment en ce qui concerne les catastrophes naturelles.

Les assureurs ont à leur disposition différentes données sur les biens assurés : localisation, mode d'occupation du bien assuré, type du bien, nombre de pièces… Ces informations seront utiles à la gestion assurantielle dans la mesure où elles permettent par exemple d'établir la carte des expositions.

La carte des expositions montre la répartition spatiale du portefeuille d'une compagnie d'assurance. Elle présente la répartition des sommes assurées, le plus souvent agrégées à l'échelle de la zone Cresta (Catastrophe Risk Evaluating and Standardizing Target Accumulations). Dans notre cas, les zones Cresta sont les différents départements. L'évaluation des sommes assurées est effectuée par les assureurs en fonction des données dont ils disposent. Parmi ces informations, la précision de la localisation est fondamentale, car elle permet une évaluation plus exacte des sommes assurées.

3. Présentation du contexte

Groupama étant composé de 11 compagnies d'assurance à part entière, les sommes assurées sont évaluées de différentes manières par chaque caisse. Un logiciel de SIG permet d'illustrer les disparités entre les différentes caisses régionales (CR) :

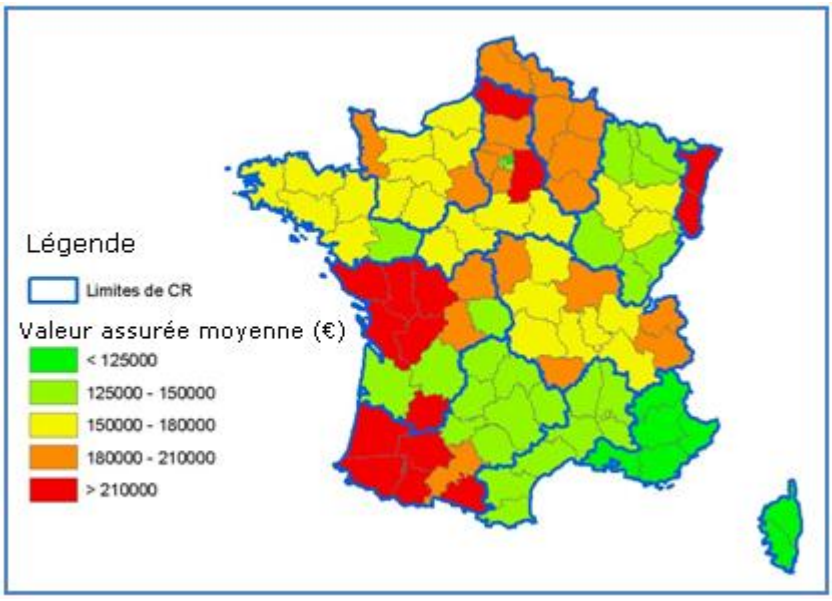

Figure 2: Valeur assurée moyenne d'un risque MRH par département

Cette carte représente les sommes assurées moyennes par département dans le portefeuille Groupama. Les lignes en bleu représentent les limites des 11 différentes caisses régionales. On constate une certaine hétérogénéité entre les caisses. A priori il

n'y a aucune raison que le coût de la reconstruction varie du simple au double d'un département à un autre. Puisque les sommes assurées sont essentielles à la modélisation du risque tempête, il est indispensable de les évaluer de manière correcte. Le paragraphe suivant propose une méthodologie de calcul.

4. Méthodologie de calcul des sommes assurées

a. Segmentation du portefeuille

La segmentation vise à répartir les risques dans des groupes homogènes qui seront traités de manière indépendante. C'est pourquoi le portefeuille est segmenté en risques de particuliers, risques agricoles, collectivités territoriales, artisans et commerçants et enfin risques industriels.

b. Les informations nécessaires à la valorisation

- La localisation du risque

La valeur du coût de la reconstruction moyen d'une habitation peut varier d'une commune à une autre, selon le prix de la main d'œuvre ou des matériaux de construction. Il faut également prendre en compte le fait que dans chaque commune, il peut y avoir des types de construction différents : si certaines ont plus de villas, d'autres pourraient avoir plus de logements sociaux comme les HLM.

- La qualité juridique de l'assuré

On entend ici par qualité juridique le statut d'occupation de l'assuré. Trois catégories sont ainsi considérées, à savoir locataire, propriétaire non occupant ou propriétaire occupant. Cette distinction sert à déterminer le type de couverture : si un locataire s'assure uniquement sur le contenu, un propriétaire non occupant n'assure que le bâtiment. Un propriétaire occupant assurera alors à la fois le bâtiment et le contenu.

- Le type d'habitation

La ventilation entre « maisons individuelles » et « appartements collectifs » permet d'affiner les vulnérabilités appliquées par les modèles au portefeuille : une maison isolée est plus exposée à la tempête qu'un immeuble dans un milieu urbain. De plus, une maison individuelle est en général plus spacieuse et contient plus de pièces qu'un simple appartement situé dans un immeuble. Cette information pourra être nécessaire lorsqu'il s'agira de rehausser certaines données.

- Le nombre de pièces et/ou la surface

Les sommes assurées sont clairement fonction du nombre de pièces ou de la surface que l'on assure. Une pièce ici est une pièce à usage d'habitation : sont donc exclues la salle de bain et la cuisine, sauf si cette dernière dépasse une certaine surface. Il s'agit d'un point délicat car chaque Caisse Régionale ne compte pas les pièces de la même manière : la surface apparaît donc comme la donnée la plus juste pour les calculs lorsqu'elle est renseignée.

c. Inventaire des données manquantes et des anomalies

Avant de commencer tout calcul, il est nécessaire de voir dans chaque portefeuille la proportion d'observations exploitables en mettant à part les mal ou non renseignées. Il s'agit par exemple de nombre de pièces aberrants, ou d'informations manquantes. Dans les portefeuilles étudiés, si pour le risque résidentiel, cette proportion de données exploitables est de l'ordre de 90%, elle est nettement inférieure pour le risque agricole.

Une étude sur la distribution des valeurs observées des différents paramètres conduit à définir la notion d'anomalie et à proposer une approche de correction permettant de rehausser des données.

d. Rehaussement et correction des données

Le rehaussement de données vise à compléter les informations non renseignées et corriger les anomalies.

Il est en général difficile et irréaliste de faire des hypothèses sur la localisation en absence de ces données. Heureusement, ce type d'anomalie est rare dans la base de données à disposition.

En ce qui concerne le reste, les données de l'INSEE et les données « correctes » de la base permettent de la corriger et de la compléter.

- Qualité juridique de l'assuré

Lorsque le statut d'occupation n'est pas renseigné, on peut se baser sur le type d'habitation pour remplir cette information. On dispose en effet d'un nombre considérable d'observations par département à partir desquelles on peut calculer la répartition des propriétaires et des locataires selon le type d'habitation. Puisqu'une part représentative du portefeuille est renseignée, on peut se baser sur les observations pour compléter les informations manquantes.

Dans les grandes agglomérations comme Paris, on a une plus grande proportion de propriétaires habitant dans des appartements. Dans les villes plus petites, les propriétaires vivent le plus souvent dans des maisons. Il est donc plus prudent de prendre en compte les spécificités de chaque département pour faire cette répartition.

En se basant toujours sur les données du portefeuille, on conclut que 3% des propriétaires sont non occupants.

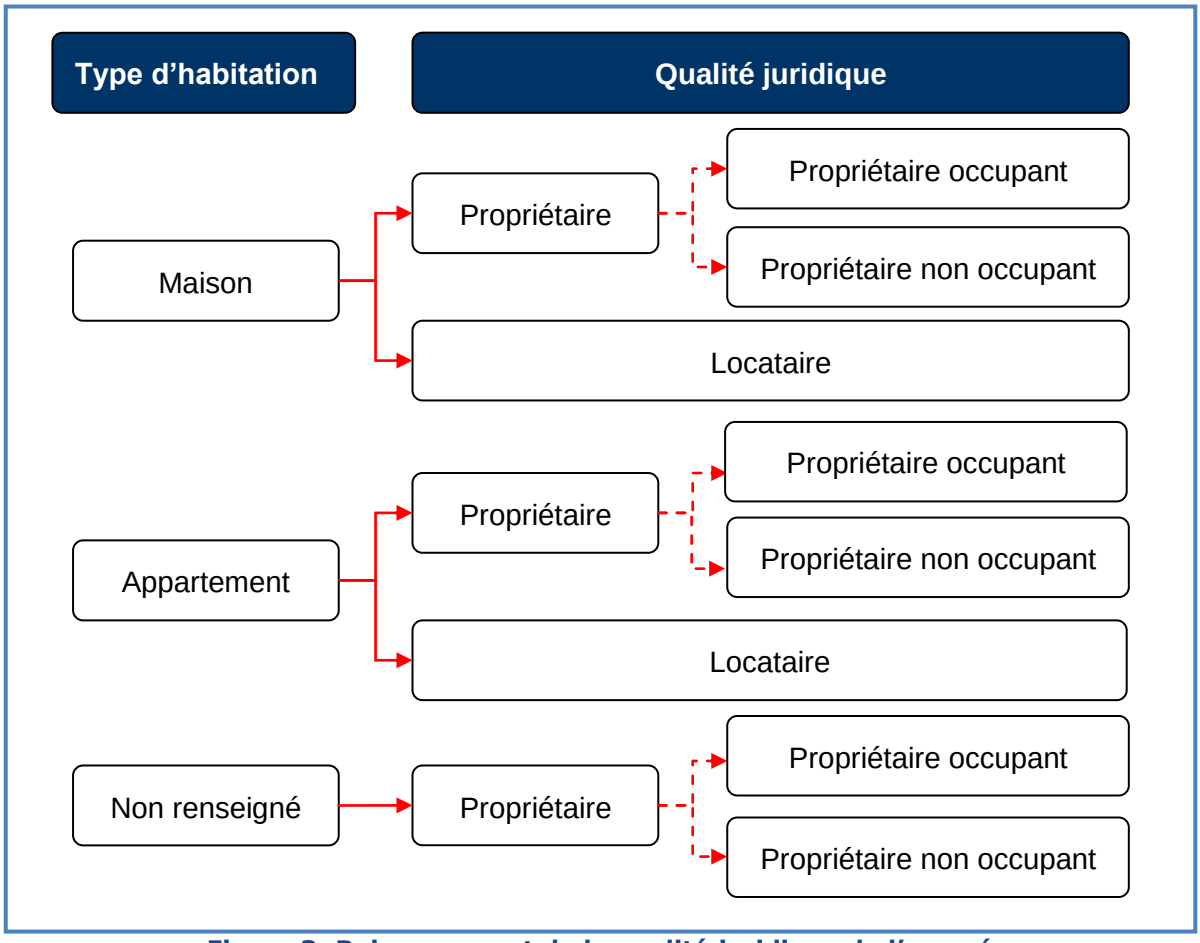

Figure 3: Rehaussement de la qualité juridique de l'assuré

Le portefeuille de Groupama se distingue par une proportion importante de propriétaires. C'est la raison pour laquelle on utilise comme qualité juridique par défaut le statut de propriétaire.

Procédure :

Soit $\mathrm{p_{prop}}$ la part de propriétaires pour un type d'habitation et un département donnés.

On tire une variable aléatoire U selon une loi uniforme $\mathcal{U}(0;1)$.

Si $0 \leq \mathrm{u} \leq 0.03 * \mathrm{p_{prop}}$ alors on considère que c'est un propriétaire non occupant.

Si $0.03 * \mathrm{p_{prop}} < u \leq \mathrm{p_{prop}}$ alors c'est un propriétaire occupant

Si $\mathrm{p_{prop}} < u \leq 1$ alors c'est un locataire.

- Le type d'habitation

De la même manière que précédemment, on peut se baser sur la qualité juridique pour remonter l'information sur le type d'habitation.

Dans les villes plus petites, les maisons sont souvent occupées par des propriétaires. En revanche, il y a plus de maisons louées dans les grandes villes.

Comme plus haut, on tient compte de la spécificité pour chaque département pour répartir les habitants dans les types d'habitation.

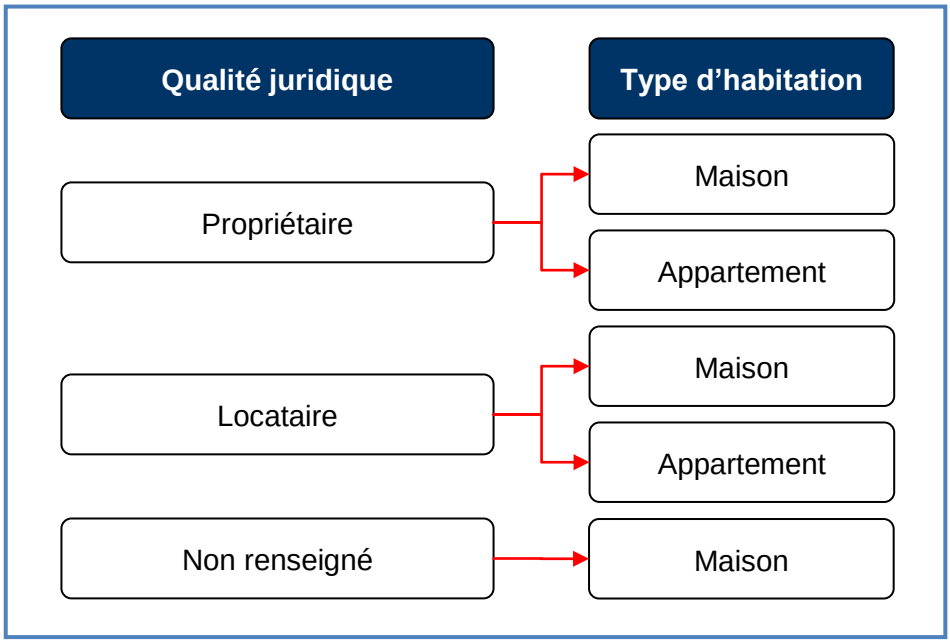

Figure 4:Rehaussement du type d'habitation

Puisque le portefeuille se distingue par une proportion importante de maisons individuelles, c'est cette dernière qu'on prend comme type d'habitation par défaut.

Procédure :

Soit $\mathrm{p_{mais}}$ la part de maisons pour une qualité juridique et un département donnés.

On tire une variable aléatoire U selon une loi uniforme $\mathcal{U}(0;1)$.

Si $0 \leq \mathrm{u} \leq \mathrm{p_{mais}}$ alors on considère que c'est une maison.

Si $\mathrm{p_{mais}} < u \leq 1$ alors c'est un appartement.

e. Calcul des sommes assurées

La cote annuelle des valeurs vénales immobilières et foncières donne les coûts de la construction au mètre carré par département et par type de standing, pour une maison ou pour un appartement.

Pour estimer un coût moyen de la reconstruction par commune, il faut considérer les différents types de standing et calculer la répartition de ces différents types de standing dans chaque commune. C'est ainsi qu'on obtient un coût moyen de la reconstruction par commune, aussi bien pour les maisons que pour les appartements.

Formellement, soient A, B et C les trois types de standing.

Pour une commune donnée, on a la répartition de ces 3 types de standing :

$$part_A + part_B + part_C = 1$$

On a également les coûts au mètre carré pour chaque type de standing, par exemple pour une maison $cout_A^{maison}, cout_B^{maison}$ et $cout_C^{maison}$.

Le coût moyen de la reconstruction d'une maison pour cette commune est alors donné par :

$$cout\ moyen = part_A . cout_A^{maison} + part_B . cout_B^{maison} + part_C . cout_C^{maison}$$

Il est parfois nécessaire de prendre les informations au département, notamment lorsque les communes sont trop petites, ce qui met en jeu la fiabilité des résultats.

Soient donc :

- n le nombre de pièces d'une résidence dans une commune donnée
- s la surface moyenne d'une pièce dans le département considéré
- c le coût moyen du m² (différent selon qu'il s'agisse d'une maison ou d'un appartement)

Alors, pour un risque donné, la somme assurée bâtiment se calcule comme suit :

$$BSI = \begin{cases} n * s * c \ \ pour\ un\ propriétaire\ non\ occupant\ ou\ occupant \\ 0 \ \ pour\ un\ locataire \end{cases}$$

Le contenu assuré *CSI* est défini de manière contractuelle : les informations fournies par les différentes caisses régionales sont donc exploitées pour obtenir ce contenu assuré. Il est à noter que ce terme est nul lorsqu'il s'agit d'un propriétaire non occupant.

La somme assurée totale pour un risque donné est donc :

$$TSI = BSI + CSI$$

Une fois les sommes assurées calculées, les franchises et limites fixées, des tempêtes peuvent être simulées sur le portefeuille pour estimer les engagements de l'assureur vis-à-vis des assurés.

5. En sortie du modèle : les Event Loss Table

Les Event Loss Table sont des tables qui résultent de la modélisation. Elles contiennent des informations sur les différents événements susceptibles d'impacter le portefeuille. Ainsi, chaque tempête se caractérise par un identifiant, une fréquence annuelle, le coût moyen qu'elle engendre, l'écart-type de ce coût ainsi que le coût maximal, c'est-à-dire l'exposition. Nous verrons dans le chapitre suivant que l'écart-type se compose en 2 parties.

Afin d'illustrer les calculs, considérons cet ELT dans une monnaie donnée:

Evénement	Fréquence	Coût	Ecart-type	Exposition
Tempête 1	0.01	1 500 000	800 000	5 500 000
Tempête 2	0.01	3 000 000	2 000 000	15 000 000
Tempête 3	0.02	6 500 000	5 000 000	50 000 000
Tempête 4	0.03	8 000 000	6 000 000	90 000 000
Tempête 5	0.03	10 000 000	7000 000	95 000 000

Tableau 1 : Echantillon de Event Loss Table

Chaque tempête a ses caractéristiques. Ainsi d'après cette table, une tempête telle que la Tempête 1 a une fréquence annuelle de 0.01, et apparaît donc en moyenne une fois tous les 100 ans. Cet événement est susceptible de détruire une exposition de

5 500 000 : sur les 5 500 000 exposés à cette tempête, on estime que le montant des sinistres s'élèvera à 1 500 000. Cependant, on doit aussi considérer le fait que le coût peut être inférieur ou supérieur aux 1 500 000 estimés car l'écart-type est supérieur à 0. C'est un point important quant à la génération des coûts.

6. Prise en compte de l'incertitude secondaire

Avant de parler d'incertitude secondaire, il faut d'abord définir ce qu'est l'incertitude primaire. L'incertitude primaire est l'incertitude sur le fait qu'un événement ait lieu ou non, et si un événement a lieu, quel événement ce sera. Par exemple, pour le risque tempête : il peut ne pas y avoir de tempête, comme il peut y en avoir plus d'une. Y aura-t-il encore des tempêtes telles que Lothar et Martin ? L'incertitude due au nombre ou au type d'événement qui peut survenir est appelée incertitude primaire.

Les pertes financières dues aux catastrophes naturelles peuvent varier à cause de différents facteurs comme la taille de l'événement, sa localisation et les types et valeurs des bâtiments affectés. S'il est évident que différents événements engendrent différents niveaux de perte, on considère rarement le fait qu'un événement spécifique peut produire différents niveaux de perte.

Les modèles sont plus fiables quand ils intègrent les effets de l'incertitude secondaire et ne se basent pas uniquement sur la perte moyenne.

- Qu'est-ce que l'incertitude secondaire ?

Si un événement se produit aujourd'hui, la perte engendrée est rarement égale à la perte moyenne donnée par le modèle. Pour chaque événement dans le modèle, le potentiel de pertes est représenté par une distribution, avec une perte moyenne et une dispersion par rapport à cette moyenne. La dispersion de la perte autour de la moyenne est représentée par un coefficient de variation, et reflète ce qu'on appelle l'incertitude secondaire sur le coût. Le coefficient de variation est le rapport entre l'écart-type et la moyenne.

L'incertitude secondaire est l'incertitude sur la taille de la perte, sachant qu'un événement a eu lieu, c'est-à-dire son coût. Les coûts sont variables, certaines valeurs pouvant être plus probables que d'autres. Deux événements qui sont identiques (dans un sens à définir) peuvent avoir des coûts différents.

Ainsi, l'incertitude primaire concerne la survenance ou non d'événements inconnus. L'incertitude secondaire, quant à elle, concerne la taille de la perte pour les événements connus.

- Les sources d'incertitude secondaire

  - Incertitude sur le hasard

En ce qui concerne la tempête, l'incertitude sur le hasard vient principalement de l'incertitude sur la vitesse et la direction du vent.

De plus, l'environnement joue un rôle dans la vitesse du vent pour un endroit donné. Les montagnes, les arbres et les grands bâtiments tendent à réduire la vitesse

du vent. Cependant, il y a une incertitude sur la manière dont est réduite la vitesse après que la tempête soit passée sur ces structures.

- Incertitude sur la vulnérabilité

Il s'agit de l'incertitude sur le montant des dommages : la résistance d'un bâtiment à une tempête dépend de la vitesse du vent. La qualité et le type de construction des bâtiments ont aussi un rôle à jouer, et déterminent comment peuvent résister les bâtiments à une intensité et une durée données. Ceci constitue également une source d'incertitude.

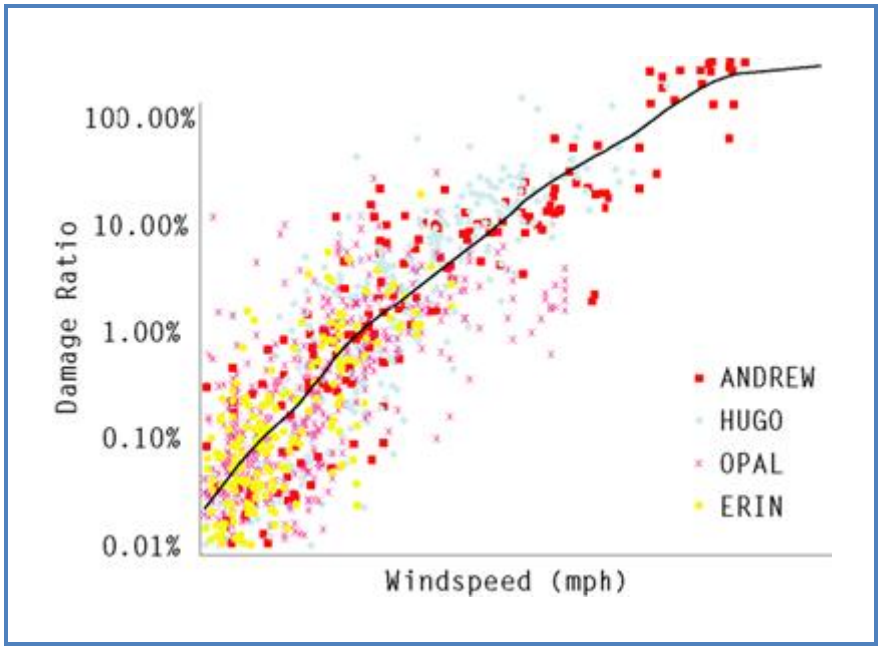

Figure 5 : Incertitude sur les dommages sur les structures en bois (ouragans)

Ce graphe est un exemple qui illustre l'incertitude sur les coûts générés par des ouragans survenus aux Etats-Unis. Il trace les taux d'endommagement des bâtiments par rapport à la vitesse des ouragans Andrew, Hugo, Opal et Erin. La courbe de tendance montre que le taux d'endommagement est fonction croissante de la vitesse. On voit que les variations sont importantes autour de cette courbe. Cela signifie qu'il y a une incertitude sur le coût à une certaine vitesse donnée, et sur la manière dont les dommages s'accroissent avec la vitesse.

- Incertitude sur les caractéristiques du modèle

L'incertitude sur les coûts est aussi impactée par les caractéristiques et la précision du modèle. Un modèle moins détaillé produit un coût qui a un niveau d'incertitude plus grand. On peut par exemple modéliser avec plus ou moins de détail la résolution géographique ou les différents types de construction.

- Incertitude sur les données du portefeuille

Il s'agit de la manière dont sont détaillées les données en entrée du modèle. Cette incertitude est similaire à la précédente mais cette dernière relève de l'incertitude sur les paramètres du modèle, et non des données fournies au modèle. Par exemple, si les données sont fournies à la commune et non au département, cela réduit l'incertitude secondaire.

- Prise en compte de la corrélation spatiale

Lorsque l'on analyse un portefeuille, on doit considérer la corrélation spatiale. Considérons un exemple simple de portefeuille où les risques sont répartis sur 2 localisations A et B. Quand un événement touche 2 endroits, le coût moyen de cet événement pour le portefeuille est calculé simplement en additionnant les coûts moyens pour ces 2 endroits. Par contre l'écart-type ne se calcule pas de la même manière, sauf dans un cas particulier : si les coûts aux 2 endroits sont complètement corrélés, alors on peut faire la somme des écart-types.

La corrélation et l'indépendance, dans le cas de l'agrégation de A et B en un seul portefeuille, reflètent comment sont reliés les dommages entre ces localisations. S'il y a corrélation complète entre A et B, les dommages causés à un bâtiment dans A sont proportionnels aux dommages causés à un bâtiment dans B. S'il y a indépendance complète entre A et B, il n'y a aucune relation entre les dommages causés à ces 2 bâtiments.

La corrélation ou l'indépendance entre les différentes localisations affectent directement l'incertitude secondaire de chaque événement pour un portefeuille. Plus la corrélation est grande, plus l'incertitude secondaire est grande. Il y a deux extrêmes :

- Considérer que tous les bâtiments sont complètement corrélés entre eux : dans ce cas, l'écart-type du portefeuille est égal à la somme des écart-types aux différents endroits

$$\begin{aligned} \mathbb{V}(X_A + X_B) &= \mathbb{V}(X_A) + \mathbb{V}(X_B) + 2\,cov(X_A, X_B) \\ &= \sigma_A^2 + \sigma_B^2 + 2\rho_{A,B}.\sigma_A.\sigma_B \quad \text{avec} \quad \rho_{A,B} = 1 \\ \sigma_{A+B}^2 &= (\sigma_A + \sigma_B)^2 \end{aligned}$$

D'où :

$$\sigma_{A+B} = \sigma_A + \sigma_B$$

- Considérer que tous les bâtiments sont complètement indépendants de tous les autres bâtiments : dans ce cas, l'écart-type du portefeuille est égal à la racine de la somme des variances.

$$\begin{aligned} \mathbb{V}(X_A + X_B) &= \mathbb{V}(X_A) + \mathbb{V}(X_B) + 2\,cov(X_A, X_B) \\ &= \sigma_A^2 + \sigma_B^2 + 2\rho_{A,B}.\sigma_A.\sigma_B \quad \text{avec} \quad \rho_{A,B} = 0 \\ \sigma_{A+B}^2 &= \sigma_A^2 + \sigma_B^2 \end{aligned}$$

D'où :

$$\sigma_{A+B} = \sqrt{\sigma_A^2 + \sigma_B^2}$$

Puisque les coûts à différents endroits sont corrélés partiellement, la réalité se trouve entre les 2 extrêmes. L'écart-type est alors calculé en donnant plus ou moins de poids à la partie corrélée, selon le type de catastrophe considéré. Ainsi, l'écart-type se divise en une composante indépendante et une composante corrélée. Les événements qui sont plus concentrés sont modélisés avec un poids de corrélation plus élevé que les

événements qui sont plus dispersés. Un tremblement de terre aura donc un poids de corrélation plus élevé qu'un ouragan.

En agrégeant pour le portefeuille, et en généralisant les 2 formules ci-dessous, les écart-types corrélé et décorrélé s'obtiennent donc comme suit :

Composante corrélée :

$$\sigma_{c,port} = w \sum_{loc} \sigma_{loc}$$

Composante indépendante :

$$\sigma_{i,port} = (1 - w) \sqrt{\sum_{loc} \sigma_{loc}^2}$$

où w est le poids de corrélation.

Par exemple, pour la tempête en Europe, RMS utilise un poids de corrélation w=8%.

Ainsi l'écart-type s'obtient en additionnant les composantes indépendante et corrélée :

$$\sigma_{port} = w \sum_{loc} \sigma_{loc} + (1 - w) \sqrt{\sum_{loc} \sigma_{loc}^2}$$

Prenons un exemple simple d'un portefeuille où les risques sont localisés dans 2 endroits A et B. Nous avons donc 2 ELT par localisation :

Evénement	Fréquence	Coût	Ecart-type	Exposition
Tempête 1	0.04	8 000	4 000	100 000
Tempête 2	0.03	7 000	5 000	100 000
Tempête 3	0.02	9 000	2 700	100 000
Tempête 4	0.09	10 000	5 000	100 000
Tempête 5	0.01	10 000	2 000	100 000

Tableau 2 : Event Loss Table à la localisation A

Evénement	Fréquence	Coût	Ecart-type	Exposition
Tempête 1	0.04	6 000	4 000	120 000
Tempête 2	0.03	5 000	2 500	120 000
Tempête 3	0.02	10 000	3 700	120 000
Tempête 4	0.09	3 600	4 000	120 000
Tempête 5	0.01	6 000	1 200	120 000

Tableau 3 : Event Loss Table à la localisation B

Lorsqu'on considère le portefeuille, on doit prendre en compte les différentes localisations. Le coût moyen s'obtient en additionnant les 2 coûts moyens. Puisque les coûts aux deux endroits ne sont ni complètement corrélés ni complètement indépendants, l'écart-type se calcule de manière un peu plus compliquée.

Avec un poids de corrélation $w = 8\%$, la composante corrélée pour la tempête 1 est donnée par :

$$\sigma_c = 0{,}08 * (4000 + 4000) = 640$$

La composante indépendante, quant à elle, vaut :

$$\sigma_i = 0{,}92 * \sqrt{4000^2 + 4000^2} = 5204$$

L'écart-type s'obtient en additionnant ces deux composantes, soit 5844.

L'exposition se calcule simplement en additionnant les expositions de A et de B.

Evénement	Fréquence	Coût	Ecart-type (I)	Ecart-type (C)	Exposition
Tempête 1	0.04	14 000	5 204	640	220 000
Tempête 2	0.03	12 000	5 143	600	220 000
Tempête 3	0.02	19 000	4 214	512	220 000
Tempête 4	0.09	13 600	5891	720	220 000
Tempête 5	0.01	16 000	2 146	256	220 000

Tableau 4 : Event Loss Table avec incertitude secondaire

C'est donc de ce type d'Event Loss Table dont on dispose pour la suite de l'étude.

Plus précisément, on dispose de 11 ELT correspondant aux 11 caisses régionales. On fait donc passer 14900 tempêtes sur la France, une caisse pouvant être touchée ou non par une tempête donnée. De plus, une même tempête peut se retrouver dans plusieurs caisses de par son étendue.

7. Loi du nombre de sinistres

La distribution de la fréquence est la distribution du nombre de survenances dans une année. Pour la modéliser, nous utilisons une loi de Poisson de paramètre λ. Ce paramètre peut être calculé à partir des ELT comme étant la somme des fréquences de tous les événements.

Si nous reprenons le Tableau 4, nous obtenons :

$$\lambda = 0{,}04 + 0{,}03 + 0{,}02 + 0{,}09 + 0{,}01 = 0{,}19$$

Le paramètre λ peut être interprété comme la fréquence moyenne. Ainsi, si $\lambda = 0{,}19$, cela voudrait dire qu'il y a 19 événements en 100 ans.

On suppose qu'il y a indépendance entre la survenance des différents événements.

La probabilité d'avoir exactement n survenances dans une année est donnée par :

$$\mathbb{P}(N = n) = \frac{e^{-\lambda}\lambda^n}{n!}$$

Dans le cadre du stage, 14900 tempêtes ont été modélisées, ce qui donne une fréquence annuelle moyenne de 10,59.

8. Loi du coût des sinistres

Le taux d'endommagement est le rapport entre le coût moyen et l'exposition.
On suppose que ce taux d'endommagement suit une loi Beta de paramètres α et β.
La densité de probabilité de la loi Beta s'écrit :

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha;\beta)}$$

où

$$B(\alpha;\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$$

Prenons l'exemple de la Tempête 1 recensée dans le Tableau 4.
Les coûts générés par cette tempête sont, en moyenne, égaux à 14 000. Cependant, les coûts sont répartis autour de cette moyenne comme le montre la figure suivante :

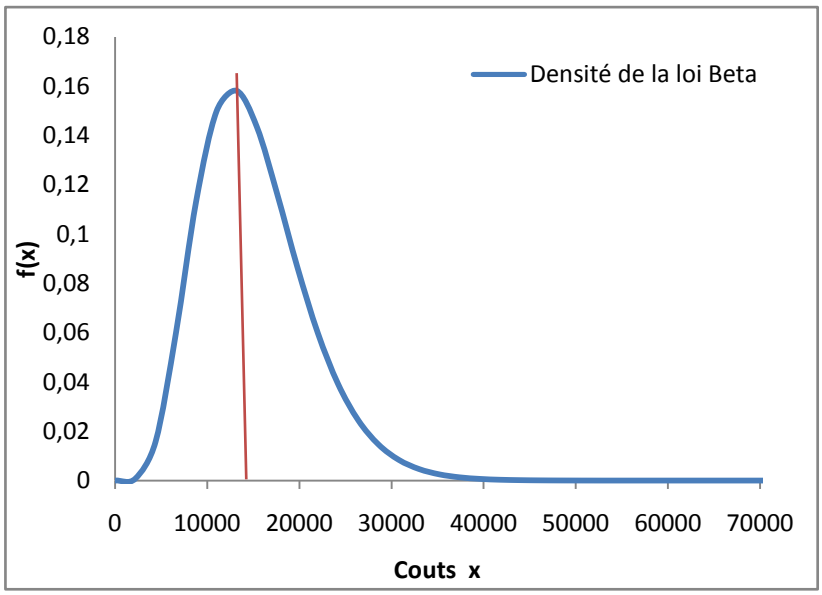

Figure 6 : Densité de la loi Beta

La droite verticale représente la valeur attendue des coûts pour l'événement. La courbe bleue trace la densité de la loi Beta, et représente l'écartement des coûts par rapport à la moyenne.

L'axe des abscisses correspond aux montants. Il est à noter toutefois que ce n'est pas le coût en lui même qui suit une Beta, mais le taux d'endommagement : à des fins illustratives le taux d'endommagement a été multiplié par l'exposition afin d'avoir les coûts sur l'axe des abscisses.

Pour une tempête donnée, la moyenne μ et l'écart-type σ de la loi Beta peuvent être calculées à partir des ELT :

$$\mu = \frac{coût\ moyen}{exposition} = \frac{\alpha}{\alpha+\beta}$$

$$\sigma = \frac{Ecart - type}{Exposition} = \sqrt{\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}}$$

Ainsi, en utilisant la méthode des moments, nous pouvons déduire les paramètres de la loi Beta :

$$\alpha = \frac{\mu^2(1-\mu)}{\sigma^2} - \mu$$

$$\beta = \frac{\alpha(1-\mu)}{\mu}$$

Dans les simulations, pour générer les coûts, il s'agira alors dans un premier temps de simuler un taux d'endommagement puis de le multiplier ensuite par l'exposition.

Evénement	Fréquence	Coût	Ecart-type	Exposition	μ	σ	α	β
Tempête 1	0.04	14 000	5 844	220 000	0,06364	0,02656	5,31015	78,13503
Tempête 2	0.03	12 000	5 743	220 000	0,05455	0,02610	4,07332	10,60421
Tempête 3	0.02	19 000	4 726	220 000	0,08636	0,02148	14,68067	155,30599
Tempête 4	0.09	13 600	6 611	220 000	0,06182	0,03005	3,90855	59,31796
Tempête 5	0.01	16 000	2 402	220 000	0,07273	0,01092	41,07079	523,65261

Tableau 5 : Caractéristiques des lois Beta

La distribution des coûts dépend donc de chaque tempête considérée.

Deux raisons motivent le choix de la loi Beta :

- Le taux d'endommagement varie de 0% à 100%. C'est également le domaine de définition de la loi Beta.
- Cette distribution peut décrire une grande variété de courbes comme le montre le graphe suivant :

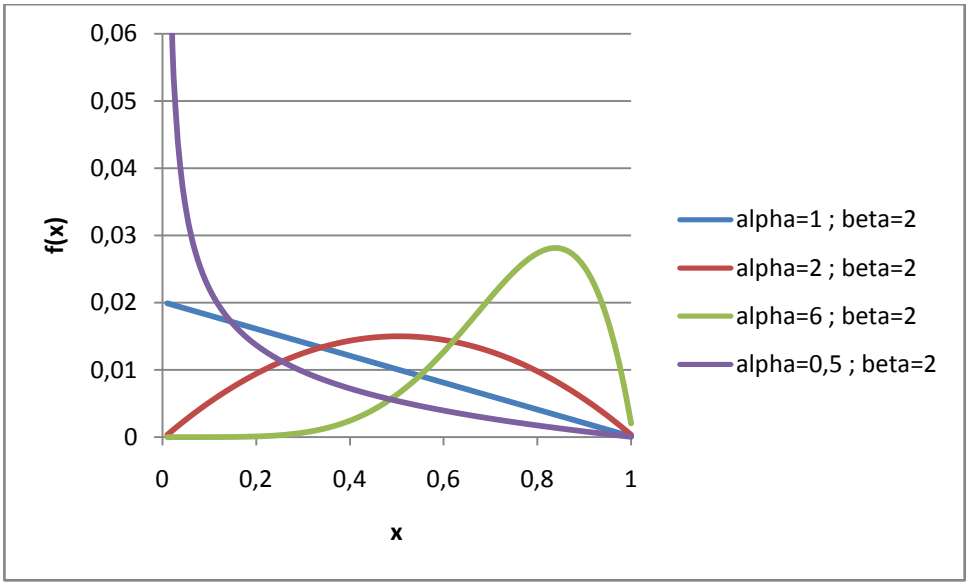

Figure 7 : Exemples de densités pour la loi Beta

Le but est donc de générer via des simulations de Monte Carlo des coûts pour les 11 caisses régionales, qu'on agrègera ensuite au niveau national pour avoir le coût total. Si 2 caisses sont proches, leurs coûts peuvent être corrélés. C'est de cette corrélation qu'on va tenir compte lors de la génération des coûts. On verra ensuite comment la prise en compte ou non de cette corrélation entre les différentes caisses influence différentes mesures de risque.

Deuxième partie : Mesures de risque

L'activité d'assurance repose sur le concept de transfert de risque : moyennant une prime, l'assuré se protège d'un risque financier aléatoire. Afin d'être solvable, un assureur doit disposer d'un certain niveau de fonds propres, et de même ajouter un chargement de sécurité à la prime pure. Pour ce faire, il a besoin de mesurer le risque assuré.

1. Définitions

Soit Ω l'ensemble fini des états de nature possibles. On appelle **variable aléatoire** réelle une fonction X qui à un état de nature ω associe le réel $X(\omega)$.

La variable aléatoire X peut être caractérisée par sa fonction de répartition F_X :

$$F_X(x) = P\{\omega / X(\omega) \leq x\} = P\{X \leq x\}$$

On appelle **mesure de risque** une fonction ρ associant à un risque X un réel positif $\rho(X)$.

D'après cette définition, nous pouvons dire que l'espérance, la variance ou l'écart-type, lorsqu'ils existent, sont des mesures de risque. Si beaucoup d'applications répondent à cette définition, pour être jugée « satisfaisante », une mesure de risque doit pouvoir vérifier un certain nombre de propriétés élémentaires.

Soit $\mathcal{U}$ un ensemble de variables aléatoires réelles définies sur l'espace de probabilité $(\Omega, \mathcal{A}, P)$ et $\rho : \mathcal{U} \to \mathbb{R}$.

ρ est **sous-additive** si :

$$X \in \mathcal{U}, \quad Y \in \mathcal{U}, \quad X + Y \in \mathcal{U} \Rightarrow \rho(X+Y) \leq \rho(X) + \rho(Y)$$

Cette propriété exprime le fait que la diversification tend à réduire le risque. Ainsi si l'on a deux caisses régionales, leur fusion ne crée pas de risque supplémentaire, mais peut même réduire le risque.

ρ est **positivement homogène** si :

$$\lambda > 0, \quad X \in \mathcal{U}, \quad \lambda X \in \mathcal{U} \Rightarrow \rho(\lambda X) = \lambda \rho(X)$$

Nous venons de voir qu'une fusion ne crée pas de risque supplémentaire. Cet axiome dit qu'une fusion sans diversification ne modifie pas le risque.

ρ est **monotone** si :

$$X \in \mathcal{U}, \quad Y \in \mathcal{U}, \quad X \leq Y \Rightarrow \rho(X) \geq \rho(Y)$$

Si les pertes encourues avec la caisse régionale X sont supérieures à celles encourues avec la caisse régionale Y, alors X est plus risquée que Y.

ρ est **invariante par translation** si :

$$a \in \mathbb{R}, \quad X \in \mathcal{U}, \quad X + a \in \mathcal{U} \Rightarrow \rho(X + a) = \rho(X) - a$$

Si on ajoute un montant certain au résultat d'une caisse, alors le risque décroitra de ce montant là.

Une mesure de risque est dite **cohérente** si elle est :

- invariante par translation
- sous-additive
- positivement homogène
- monotone.

Une autre propriété que doivent que doivent vérifier les mesures de risque est de contenir un chargement de sécurité.

ρ contient un **chargement de sécurité** si :

$$X \in \mathcal{U} \Longrightarrow \rho(X) \geq \mathbb{E}(X)$$

Le capital minimal doit excéder la perte attendue, sinon la probabilité de ruine deviendrait certaine, sous les conditions de la loi des grands nombres.

2. Mesures de risque usuelles

Dans ce paragraphe nous introduirons les mesures de risques les plus utilisées. La Value-at-Risk et la Tail Value-at-Risk sont vraisemblablement à la base de la détermination du niveau prudentiel des provisions techniques et du besoin en fonds propres préconisés dans le nouveau système Solvency II.

a. L'écart-type et la variance

Ce sont les premières mesures de risque à avoir été utilisées. Elles ont par exemple été utilisées dans la théorie de Markowitz (critère moyenne-variance), à la base des premières théories d'évaluation des actifs financiers.

Cependant, ce critère n'est pas bien adapté au domaine de l'assurance dans le sens où il est symétrique : on pénalise aussi bien les situations favorables que les situations défavorables. Pourtant, il serait plus correct de n'éliminer que les situations défavorables.

b. La Value-at-Risk

La notion de Value-at-Risk s'est à l'origine développée dans les milieux financiers, avant d'être reprise dans le domaine de l'assurance.

Soit X la perte entre deux dates t et $t+h$.

On a donc $X = X_t - X_{t+h}$

Notons q_α le α-quantile de la distribution de probabilité de X.

La **Value-at-Risk** de niveau α associée au risque X est donnée par :

$$VaR_\alpha(X) = \mathrm{q}_\alpha(\mathrm{X})$$

Autrement dit, la VaR est définie par :

$$\mathbb{P}(X_t - X_{t+h} \leq VaR) = \alpha$$

Pour un niveau de probabilité α donné, la VaR est la perte maximale que peut subir une société, sur une période de temps donnée.

Si on prend $\alpha = 0{,}95$, il y a une probabilité de $0{,}95$ pour que la perte soit inférieure à la VaR, ou une probabilité $0{,}05$ pour qu'elle soit supérieure à la VaR.

L'avantage de la VaR c'est qu'elle repose sur un concept simple et facile à calculer. Cependant, elle n'est pas cohérente dans la mesure où elle n'est pas sous-additive : elle ne favorise pas systématiquement la diversification. C'est d'ailleurs pourquoi elle fait l'objet de plusieurs critiques.

Si la VaR renseigne à propos du niveau de perte maximal (associé à une probabilité donnée) que peut subir une société, elle ne donne aucune information sur ce qui se passe lorsque la VaR est dépassée. Ce genre d'information est par contre donné par la TVaR qui est présentée dans le chapitre suivant.

c. La Tail Value-at-Risk

La **Tail Value-at-Risk** au niveau de probabilité α est définie par :

$$TVaR_{\alpha}(X) = \frac{1}{1-\alpha}\int_{\alpha}^{1} VaR_{\rho}(X)d\rho$$

La TVaR apparait donc comme la moyenne des VaR de niveau supérieur à α.

En plus d'être cohérente, la TVaR a l'avantage de contenir un chargement de sécurité, ce qui n'est pas le cas de la VaR.

3. Solvency II

Solvency II est une réforme réglementaire européenne du monde de l'assurance. Son objectif est de permettre aux assureurs et réassureurs de comprendre les risques inhérents à leur activité afin de pouvoir allouer suffisamment de capital pour les couvrir.

Ce système doit donc couvrir non seulement la mesure quantitative de la solvabilité, mais également des aspects qualitatifs influençant l'exposition au risque de la société.

Solvency II repose sur 3 piliers ayant chacun un objectif :

- Pilier 1 : Exigences en capital : exigence quantitative

Le premier pilier a pour objectif de définir des seuils quantitatifs aussi bien pour les provisions techniques que pour les fonds propres. Ces seuils deviendront des seuils réglementaires.

Deux niveaux de fonds propres seront définis: MCR (Minimum Capital Requirement ou Capital Minimum Requis) et SCR (Solvency Capital Requirement ou Capital Cible) :

- MCR représente le niveau minimum de fonds propres en-dessous duquel l'intervention de l'autorité de contrôle sera automatique.

- SCR représente le capital cible nécessaire pour absorber le choc provoqué par une sinistralité exceptionnelle.

- Pilier 2 : Cadre du contrôle : Exigence qualitative de la gestion des entreprises

Le deuxième pilier concerne les exigences qualitatives et les règles de contrôle applicables aux compagnies d'assurance et de réassurance. Ces dernières doivent être en mesure de mettre en place un système de gouvernance efficace afin de pouvoir apprécier par elles-mêmes la mesure et la gestion de leurs risques.

L'autorité de contrôle aura les pouvoirs de vérifier la qualité des données et des procédures d'estimation, des systèmes mis en place pour mesurer et maîtriser les risques probables.

L'identification des sociétés "les plus risquées" est un objectif et les autorités de contrôle auront en leur pouvoir la possibilité de réclamer à ces sociétés de détenir un capital plus élevé que le montant suggéré par le calcul du SCR et/ou de réduire leur exposition aux risques.

- Pilier 3 : Information et discipline de marché

Le troisième pilier a pour objectif de définir l'ensemble des informations détaillées que les autorités de contrôle jugeront nécessaires pour exercer leur pouvoir de surveillance. Il s'agit notamment d'un rapport sur la solvabilité, d'un rapport sur la situation financière et une description du système de gestion des risques.

Le capital de solvabilité (SCR) requis résulte de l'application d'une formule standard ou d'un modèle interne dans une hypothèse de continuité de l'activité. Ce capital de solvabilité requis correspond à la VaR des fonds propres de base de la société, avec un niveau de confiance de 99,5% sur un horizon un an.

On associe généralement au calcul de la VaR, la notion de période de retour d'un événement. Il s'agit de l'inverse de la probabilité de survenance du phénomène. Le montant donné par la VaR à 99,5% est communément appelé événement bi-centennal, c'est-à-dire un événement qui a une période de retour de 200 ans.

Troisième partie : Théorie des copules

Le concept des copules a été introduit en 1959 par Abe Sklar. Les copules sont devenues un élément essentiel dans la modélisation des distributions multivariées en finance et en assurance.

Si la distribution normale multivariée a souvent servi de modèle, c'est pour la simplicité relative des calculs et le fait que la dépendance est entièrement déterminée par le coefficient de corrélation. Malheureusement la loi normale s'ajuste mal à certains problèmes en économie.

Dans la théorie des copules, si on considère un vecteur aléatoire $X = (X_1, \dots, X_d)$ on considère que toute l'information à propos de la structure de dépendance de X est contenue dans une fonction appelée la copule. Ce concept a l'avantage d'être un outil puissant et flexible dans la mesure où il permet de modéliser la dépendance sans tenir compte de l'effet du comportement des marges, c'est-à-dire des fonctions de répartition $F_1, \dots, F_d$ des variables aléatoires $X_1, \dots, X_d$ prises individuellement.

1. Rappels sur les distributions multivariées

Un bref rappel sur les distributions multivariées peut s'avérer nécessaire avant d'aborder les copules proprement dites.

a. Fonction de répartition conjointe

Soit un vecteur aléatoire $X = (X_1, \dots, X_d)$ de dimension d.

La **fonction de répartition conjointe** de X est la fonction de d variables définie par :

$$F(x_1, \dots, x_d) \mapsto \mathbb{P}(X_1 \leq x_1, \dots, X_d \leq x_d)$$

b. Lois marginales

Les **fonctions de répartition marginales** sont les fonctions de répartition $F_1, \dots, F_d$ des variables aléatoires $X_1, \dots, X_d$ et sont définies par :

$$F_i(x_i) = \mathbb{P}(X_i \leq x_i) = F(\infty, \dots, \infty, x_i, \infty, \dots, \infty)$$

Cette notation doit se comprendre comme la limite de F lorsque chacune des composantes x_j $(j \neq i)$ tend vers l'infini.

c. Notion d'indépendance

Les composantes du vecteur aléatoire X sont 2 à 2 indépendantes si pour tout $(x_1, \dots, x_d)$ dans $\mathbb{R}^d$ on a :

$$F(x_1, \dots, x_d) = \prod_{i=1}^{d} F_i(x_i)$$

ou encore si X a pour densité :

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

d. Les bornes de Frechet-Hoeffding

On définit la **classe de Fréchet** associée à $F_1, \dots, F_d$ par l'ensemble :

$$\mathcal{F}(F_1, \dots, F_d) = \{F: \mathbb{R}^d \to [0;1], fonction\ de\ répartition\ dont\ les\ marginales\ sont\ F_1, \dots, F_d\}$$

Deux éléments de cette classe sont :

- la borne supérieure de Fréchet définie par :

$$W(x_1, \dots, x_d) = \min(F_1(x_1), \dots, F_d(x_d))$$

- la borne inférieure de Fréchet définie par :

$$M(x_1, \dots, x_d) = \max\left(\sum_{i=1}^{d} F_i(x_i) + 1 - d\ ;\ 0\right)$$

Toute fonction de répartition multivariée est comprise entre ces 2 bornes :

$$M(x_1, \dots, x_d) \leq F(x_1, \dots, x_d) \leq W(x_1, \dots, x_d)$$

2. Définition

Une **copule C de dimension d** est une fonction de répartition multivariée définie sur $[0;1]^d$ et ayant des lois marginales uniformes.

Ainsi,

$$C: \quad [0;1]^d \to [0;1]$$

$$(u_1, \dots, u_d) \mapsto \mathbb{P}(U_1 \leq u_1, \dots, U_d \leq u_d)$$

Puisque C est une fonction de répartition, il en découle les propriétés suivantes :

- $C(u_1, \dots, u_d)$ est croissante en chaque composante u_i
- $C(1, \dots, 1, u_i, 1, \dots 1) = u_i, \quad i \in [\![1, d]\!], \quad u_i \in [0;1]$
- la probabilité $\mathbb{P}(U_1 \in [a_1, b_1], \dots, U_d \in [a_d, b_d])$ ne doit pas être négative pour tous $(a_1, \dots, a_d), (b_1, \dots, b_d)$ dans $[0;1]^d$ tels que $a_i \leq b_i$:

$$\sum_{i_1=1}^{2} \dots \sum_{i_d=1}^{2} (-1)^{i_1 + \dots + i_d} * C(u_{1,i_1}, \dots, u_{d,i_d}) \geq 0$$

où $u_{j,1} = a_j$ et $u_{j,2} = b_j$.

Réciproquement, une fonction vérifiant ces trois propriétés est une copule.

3. Théorème de Sklar

Soit F une fonction de répartition de dimension d, et dont les marginales $F_1, \dots, F_d$ sont continues. Alors il existe une copule unique telle que :

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

Remarque :

Lorsque les marginales ne sont pas continues, il est toujours possible de définir une copule, mais celle-ci n'est plus unique et perd beaucoup d'intérêt. On peut en effet toujours poser :

$$C(u_1, \dots, u_d) = F(F_1^-(u_1), \dots, F_d^-(u_d))$$

où F_i^- représente l'inverse généralisée :

$$F_i^-(u) = \inf\{t | F_i(t) \geq u\}$$

Ce résultat est important en pratique pour les applications à l'assurance et à la finance dans le sens où il indique que dans le cadre d'une analyse d'une problématique multivariée, celle-ci peut être faite en 2 étapes :

- l'identification des distributions marginales
- l'analyse de la structure de dépendance entre les composantes.

4. Densité d'une copule

Si elle existe, la densité d'une copule se définit par :

$$c(u_1, \dots, u_d) = \frac{\partial C(u_1, \dots, u_d)}{\partial u_1 \dots \partial u_d}$$

On en déduit la densité f de la variable aléatoire X :

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

5. Mesures de dépendance

L'objet de ce paragraphe est de définir ce qu'est une mesure de dépendance et de présenter les mesures les plus utilisées.

Une **mesure de dépendance** est une application qui associe à deux variables aléatoires un réel permettant de quantifier la force de la dépendance qui lie les deux variables aléatoires.

Pour être jugée satisfaisante, une mesure de dépendance doit satisfaire un nombre de propriétés, entre autres être une mesure de concordance.

Soient X et Y deux variables aléatoires de fonctions de répartition respectives F_X et F_Y.

Une mesure de dépendance δ est une **mesure de concordance** si elle possède les propriétés qui suivent :

- $\delta(X,Y) = \delta(Y,X)$
- $-1 \leq \delta(X,Y) \leq 1$
- $\delta(X,Y) = 1 \Leftrightarrow (X,Y) =_{loi} \left(F_X^{-1}(U), F_Y^{-1}(U)\right)$ où $U \sim \mathcal{U}([0;1])$
- $\delta(X,Y) = -1 \Leftrightarrow (X,Y) =_{loi} \left(F_X^{-1}(U), F_Y^{-1}(1-U)\right)$ où $U \sim \mathcal{U}([0;1])$
- Si f est strictement monotone,

$$\delta(f(X),Y) = \begin{cases} \delta(X,Y) \; si \; f \; est \; croissante \\ -\delta(X,Y) \; si \; f \; est \; décroissante \end{cases}$$

a. Le coefficient de corrélation de Pearson

Le coefficient de corrélation de Pearson, également appelé le coefficient de corrélation linéaire, est la première mesure de dépendance à avoir été utilisée.

La **covariance** est donnée par :

$$\mathbb{C}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

Elle mesure le surcroît de variabilité (ou la diminution) de la somme de deux variables aléatoires par rapport à la somme de leur variance. Elle nous renseigne donc par son signe sur le sens de la covariation entre X et Y.

Le coefficient de corrélation linéaire en est une version normée, et a ainsi l'avantage d'être sans dimension.

Le **coefficient de corrélation de Pearson** est défini par :

$$\rho(X,Y) = \frac{\mathbb{C}(X,Y)}{\sqrt{\mathbb{V}(X)\mathbb{V}(Y)}}$$

Plus ce coefficient est grand en valeur absolue, plus la dépendance entre les variables est forte. Par contre, ce n'est pas une mesure de concordance.

Le coefficient de corrélation de Pearson est tout à fait adapté lorsque l'on étudie les distributions normales ou de Student multivariées. Cependant, ces distributions ne sont pas toujours appropriées aux problèmes de l'assurance. Pour y remédier, on peut recourir aux coefficients de corrélation de rang, comme le tau de Kendall et le rho de Spearman. Il pourrait être nécessaire de définir les notions de concordance et de discordance.

b. Notion de concordance/discordance

Soient (x,y) et (x',y') deux réalisations d'un vecteur aléatoire continu (X,Y).

(x,y)et (x',y') sont **concordantes** si $(x-x')(y-y') > 0$, autrement dit si $\{x < x' \; et \; y < y'\}$ ou $\{x > x' \; et \; y > y'\}$.

(x,y) et (x,y) sont **discordantes** si $(x-x')(y-y') < 0$, autrement dit si $\{x < x' \; et \; y > y'\}$ ou $\{x > x' et \; y < y'\}$

c. Le coefficient de corrélation de Kendall

Soit (X,Y) un vecteur aléatoire et (X',Y') un vecteur en tout point identique au premier. Le **tau de Kendall** se définit comme la différence entre la probabilité de concordance et la probabilité de discordance :

$$\tau(X,Y) = \mathbb{P}\big((X-X')(Y-Y') > 0\big) - \mathbb{P}\big((X-X')(Y-Y') < 0\big)$$

Le tau de Kendall compare ainsi la probabilité que deux couples indépendants mais de même loi soient en concordance et la probabilité qu'ils soient en discordance. Aussi, un τ positif signifiera qu'en probabilité, les couples considérés sont plus souvent en concordance qu'en discordance.

Si (X,Y) a pour copule C, alors le tau de Kendall s'écrit :

$$\tau(X,Y) = 4\int_0^1\int_0^1 C(u,v).c(u,v)dudv = 4\mathbb{E}[C(U,V)] - 1 \text{ où } U,V\sim[0;1]$$

On peut construire un estimateur du tau de Kendall à partir d'un échantillon $\{(x_1,y_1),\dots,(x_n,y_n)\}$ de (X,Y) de la manière suivante :

$$\hat{\tau}(X,Y) = \frac{2}{n(n-1)}\sum_{i=2}^{n}\sum_{j=1}^{i-1} sign\big[(x_i-x_j)(y_i-y_j)\big]$$

où $sign(z) = \begin{cases} 1 & si\ z \geq 0 \\ -1 & si\ z < 0 \end{cases}$

Le tau de Kendall est une mesure de concordance.

d. Le coefficient de corrélation de Spearman

Pour calculer ce coefficient, on prend les deux paires (X,Y) et (X',Y') identiques définies dans le paragraphe précédent, et on y rajoute une troisième paire (X^*,Y^*) qui leur est identique.

Etant donnés ces couples, le **rho de Spearman** est défini par :

$$\rho_S(X,Y) = 3\big\{\mathbb{P}\big((X-X')(Y-Y^*) > 0\big) - \mathbb{P}\big((X-X')(Y-Y^*) < 0\big)\big\}$$

Si on a un échantillon $\{(x_1,y_1),\dots,(x_n,y_n)\}$ de réalisations de (X,Y), on peut construire un estimateur du rho de Spearman :

$$\hat{\rho}_S = \frac{\sum_{i=1}^{n}(R_i-\bar{R})(S_i-\bar{S})}{\sqrt{\sum_{i=1}^{n}(R_i-\bar{R})^2}\sqrt{\sum_{i=1}^{n}(S_i-\bar{S})^2}}$$

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

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

6. Exemples de copules

a. Copule d'indépendance

La **copule d'indépendance**, notée Π, est donnée par

$$\Pi(u_1, \dots, u_d) = \prod_{i=1}^{d} u_i$$

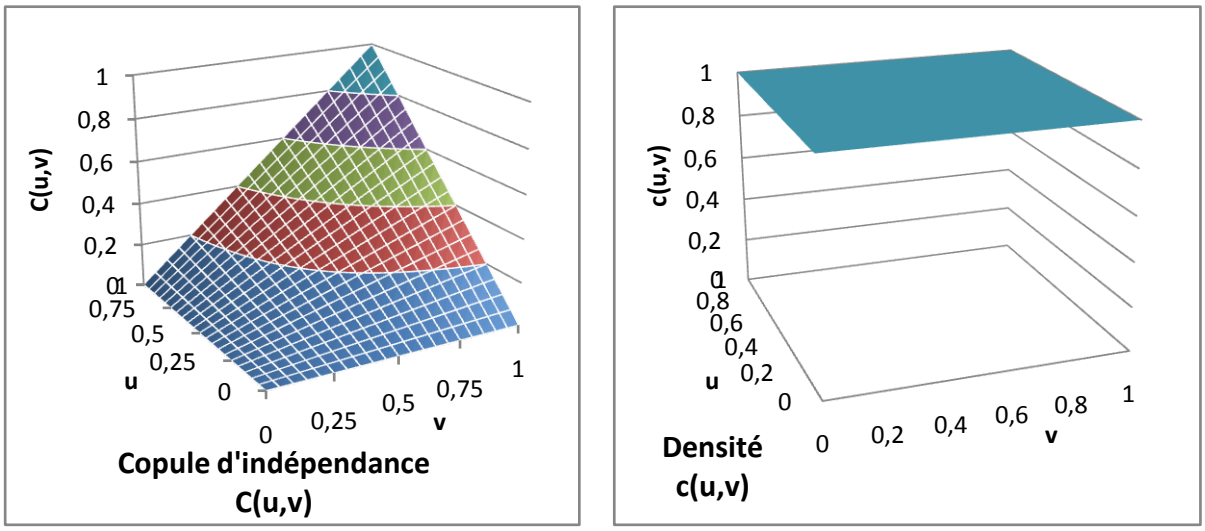

Figure 8 : Copule d'indépendance

La densité de cette copule est une constante.

b. Copule de la borne supérieure de Fréchet

La **copule comonotone**, notée W, est définie comme suit :

$$W(u_1, \dots, u_d) = \min(u_1, \dots, u_d)$$

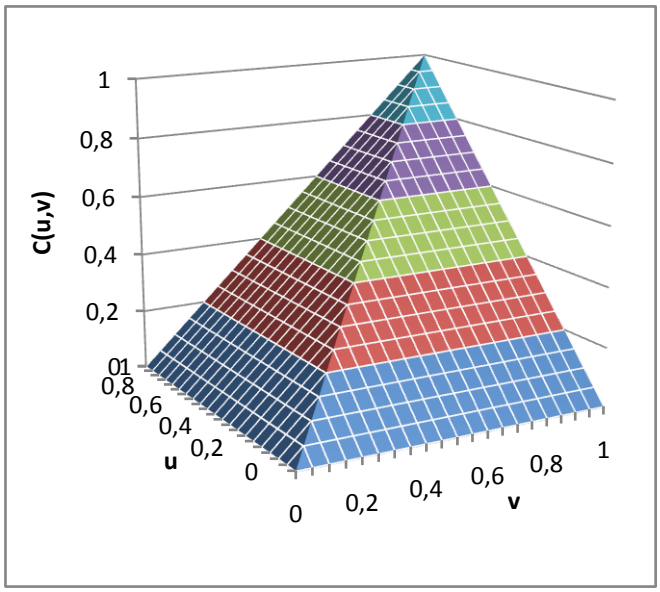

Figure 9 : Représentation de la copule comonotone

c. Copule de la borne inférieure de Fréchet

La **copule antimonotone,** notée M, se définit par :

$$M(u_1, u_2) = \max(u_1 + u_2 - 1; 0)$$

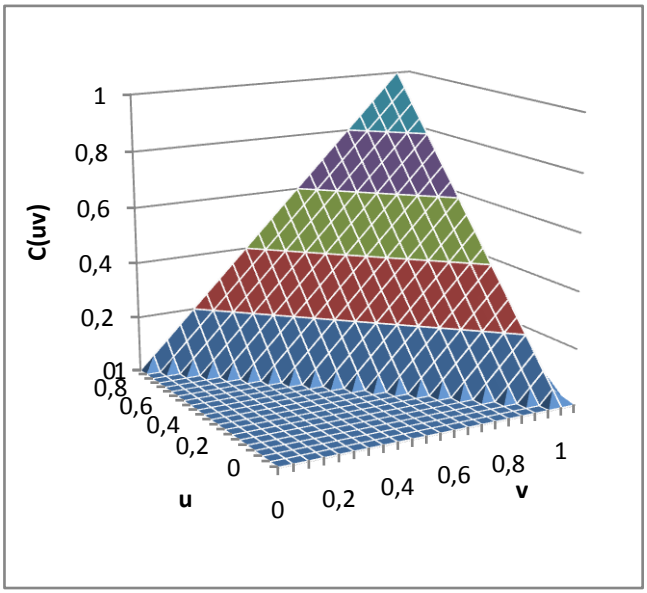

Figure 10 : Représentation de la copule antimonotone

Remarques :

- La borne inférieure de Fréchet n'est une copule que dans le cas où d=2.
- Les bornes de Fréchet n'admettent pas de densité.
- Toutes les copules sont comprises entre ces deux bornes.

d. Copule normale

La copule gaussienne fait partie de la famille des copules elliptiques (avec la copule de Student que nous ne présenterons pas ici). Seulement, les copules elliptiques sont moins bien adaptées en assurance dans le sens où elles sont symétriques. Si la copule normale est présentée ici, c'est parce qu'elle a l'avantage d'être facile à simuler.

La copule normale mutivariée de matrice de corrélation R s'écrit :

$$C_N(u_1, \dots, u_d) = \Phi_R\left(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d)\right)$$

où Φ_R est la fonction de répartition d-dimensionnelle de la loi normale, et Φ^{-1} l'inverse de la fonction de répartition de la loi normale standard.

$$\mathrm{C}_N(u_1, \dots, u_d) = \int_{-\infty}^{\Phi^{-1}(u_1)} \dots \int_{-\infty}^{\Phi^{-1}(u_d)} \frac{|R|^{-d/2}}{(2\pi)^{d/2}} * \exp\left(-\frac{1}{2} s^T R^{-1} s\right) ds$$

Remarque :

Pour la copule gaussienne, R doit nécessairement être la matrice de corrélation de Pearson. Cette dernière peut être calculée à partir des matrices de corrélation de Spearman ou de Kendall d'après les relations suivantes :

$$R_{ij} = 2 * sin\left(\frac{\pi}{6}\rho_{ij}^{spearman}\right)$$

$$R_{ij} = sin\left(\frac{\pi}{2}\tau_{ij}^{kendall}\right)$$

En dimension 2, la densité de la copule est définie par :

$$C_N(u,v) = \frac{1}{\sqrt{1-\rho^2}} exp\left\{\frac{-(\zeta_u^2 - 2\rho\zeta_u\zeta_v + \zeta_v^2)}{2(1-\rho^2)}\right\} exp\left\{\frac{\zeta_u^2 + \zeta_v^2}{2}\right\}$$

où :

$$\zeta_x = \Phi^{-1}(x)$$

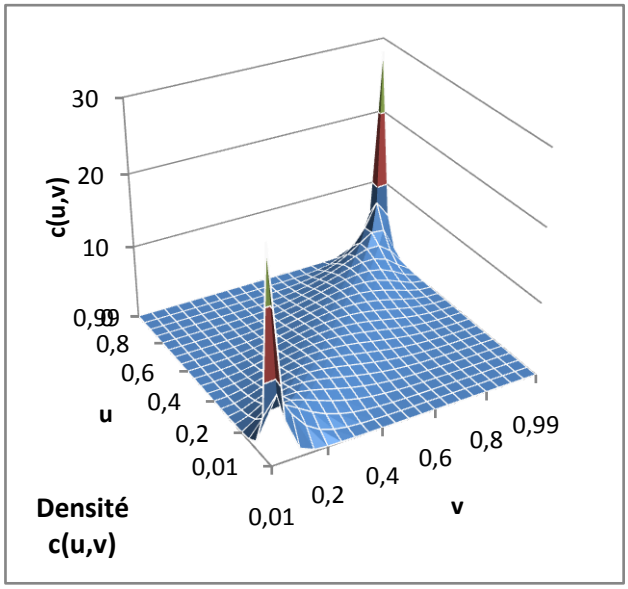

Figure 11 : Densité de la copule normale $(\tau = 0,7 \Leftrightarrow \rho = 0,89)$

Etant donné que c'est la plus facile à simuler, nous l'utiliserons dans la suite.

e. Les copules archimédiennes

Les copules archimédiennes sont construites à partir d'une fonction $\varphi: (0;1] \to [0;\infty]$, continue, décroissante et convexe telle que $\varphi(1) = 0$. Cette fonction est appelée **générateur**. Si de plus, on a $\varphi(0) = +\infty$, alors φ est appelée **générateur strict**.

Soit φ un générateur strict tel que φ^{-1} soit monotone sur [0,∞] et que pour tout k dans $[\![1;d]\!]$ on ait :

$$(-1)^k \frac{\partial^k(\varphi^{-1})}{\partial s^k}(s) > 0.$$

Alors la **copule archimédienne** C de dimension d est la fonction définie par :

$$C(u_1, \dots, u_d) = \varphi_\theta{}^{-1}(\varphi_\theta(u_1), \dots, \varphi_\theta(u_d))$$

Les générateurs permettent également de calculer le tau de Kendall. Pour cela on définit d'abord la fonction :

$$K(\theta,t) = t + \sum_{i=1}^{d-1} \frac{(-1)^i}{i!} \{\varphi_\theta(t)\}^i f_i(\theta,t)$$

où

$$f_i(\theta,t) = \frac{d^i}{dx^i} \varphi_\theta^{-1}(x)\Big|_{x=\varphi_\theta(t)}$$

sous la condition que $\{\varphi_\theta(t)\}^i f_i(\theta,t)$ tende vers 0 quand t tend vers 0 pour tout i.

On note au passage que

$$f_{i+1}(\theta,t) = f_1(\theta,t)\frac{\partial}{\partial t} f_i(\theta,t) \quad i \in [\![1,d]\!]$$

Pour d=2, la relation se simplifie et on a :

$$K(\theta,t) = t - \frac{\varphi_\theta(t)}{\varphi'_\theta(t)}$$

Le tau de Kendall se calcule à partir la relation suivante :

$$\tau(\theta) = \left(\frac{2^d-1}{2^{d-1}-1}\right) - \left(\frac{2^d}{2^{d-1}-1}\right)\int_0^1 K(\theta,t)dt$$

ou encore :

$$\tau(\theta) = 1 - \left(\frac{2^d}{2^{d-1}-1}\right)\sum_{i=1}^{d-1} \frac{(-1)^i}{i!}\int_0^1 \{\varphi_\theta(t)\}^i f_i(\theta,t)dt$$

En dimension 2, l'expression se réduit à :

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

Cette relation nous permettra d'exprimer le paramètre θ de la copule en fonction de τ. Ce paramètre mesure le degré de dépendance entre les risques. Une valeur élevée de θ traduit une dépendance forte entre les risques. De plus, une valeur positive indique une dépendance positive.

Nous ne présenterons ici que les copules archimédiennes les plus utilisées, entre autres la copule de Gumbel, la copule de Clayton et la copule de Frank.

Dans la suite, l'expression de la copule est explicitée en dimension d, mais par souci de simplification, le reste sera fait en dimension 2.

- La copule de Gumbel

Le générateur de la copule de Gumbel est donné par :

$$\varphi(t) = (-\ln(t))^{\theta}, \qquad \theta > 1$$

On en déduit la copule de Gumbel :

$$C_G(u_1, \dots, u_d) = exp\left\{-\left(\sum_{i=1}^{d}(-\ln(u_i))^{\theta}\right)^{1/\theta}\right\}$$

Dans le cas d'une copule bivariée, on a donc :

$C_G(u,v)$	$exp\left\{-\left[(-\ln(u))^{\theta} + (-\ln(v))^{\theta}\right]^{1/\theta}\right\}$
$c_G(u,v)$	$C_G(u,v) * [(-\ln(u))(-\ln(v))]^{\theta-1} * \frac{1}{uv} * \left[(-\ln(u))^{\theta} + (-\ln(v))^{\theta}\right]^{\frac{1}{\theta}-2} * \left\{\left[(-\ln(u))^{\theta} + (-\ln(v))^{\theta}\right]^{\frac{1}{\theta}} - 1 + \theta\right\}$
$K(\theta,t)$	$t\left(1 - \frac{1}{\theta}\ln(t)\right)$
$\tau(\theta)$	$1 - \frac{1}{\theta}$

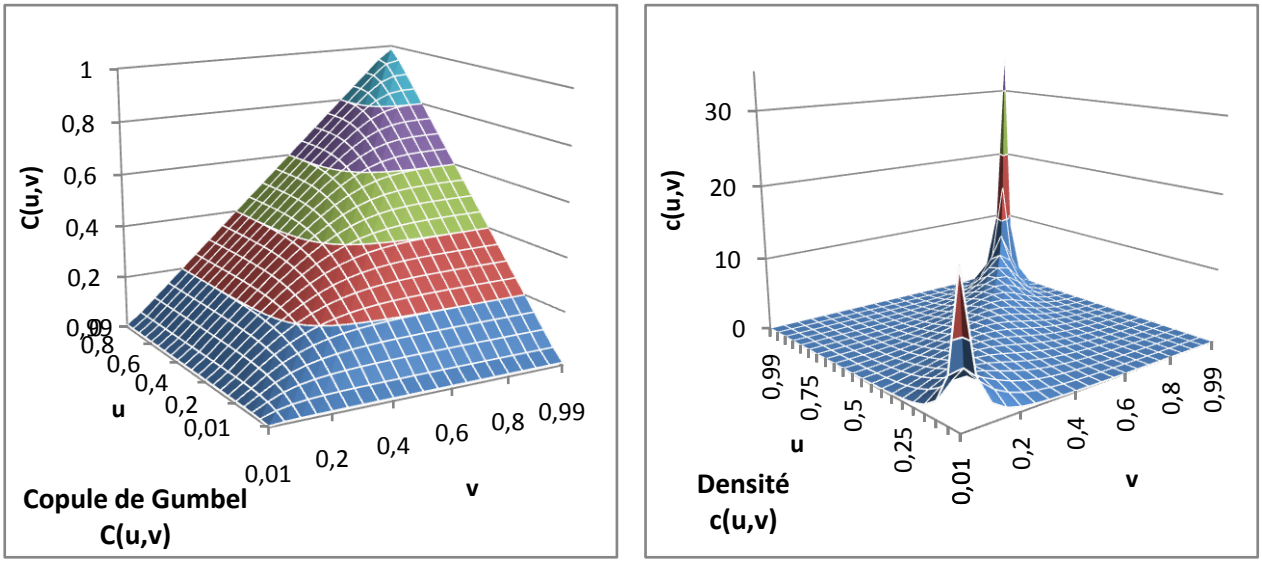

Figure 12 : Représentation de la Copule de Gumbel $(\tau = 0,7 \Leftrightarrow \theta = 3,33)$

Le paramètre θ de la copule de Gumbel étant nécessairement supérieur à 1, celle-ci n'appréhende que les dépendances positives. Elle représente bien les risques ayant une structure de dépendance plus accentuée que la queue. C'est pourquoi elle est particulièrement adaptée en assurance et en finance, notamment lorsque l'on étudie la dépendance entre les événements de forte intensité.

➢ La copule de Clayton

Le générateur de la copule de Clayton se définit comme suit :

$$\varphi_\theta(t) = \frac{(t^{-\theta} - 1)}{\theta} \qquad \theta > 0$$

La copule de Clayton est donc :

$$C_C(u_1, \dots, u_d) = \left(\sum_{i=1}^{d} u_i^{-\theta} - d + 1 \right)^{-\frac{1}{\theta}}$$

En dimension 2, on a les résultats suivants :

$C_C(u,v)$	$\left(u^{-\theta} + v^{-\theta} - 1\right)^{-\frac{1}{\theta}}$
$c_C(u,v)$	$\theta\left(\frac{1}{\theta} + 1\right)\left(u^{-\theta} + v^{-\theta} - 1\right)^{-\frac{1}{\theta}-2}(uv)^{-\theta-1}$
$K(\theta,t)$	$\frac{t^{\theta+1} - t}{\theta}$
$\tau(\theta)$	$\frac{\theta}{\theta+2}$

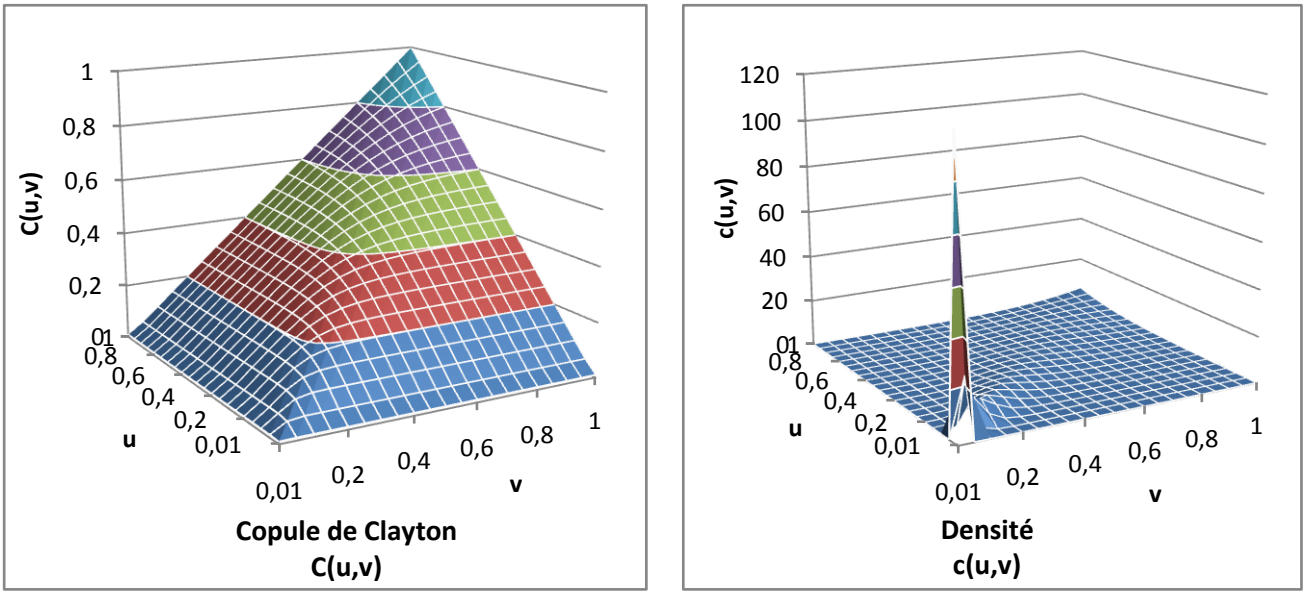

Figure 13 : Représentation de la copule de Clayton $(\tau = 0,7 \Leftrightarrow \theta = 4,67)$

Comme la copule de Gumbel, la copule de Clayton ne permet également de ne modéliser les dépendances positives. En revanche, elle permet d'étudier la dépendance entre les événements de faible intensité.

- La copule de Frank

La copule de Frank a pour générateur la fonction :

$$\varphi_\theta(t) = -ln\left(\frac{e^{-\theta t}-1}{e^{-\theta}-1}\right)$$

On en déduit l'expression de la copule :

$$C_F(u_1,\dots,u_d) = -\frac{1}{\theta}ln\left(1-\frac{\prod_{i=1}^{d}(1-e^{-\theta u_i})}{(1-e^{-\theta})^{d-1}}\right) \qquad \theta \neq 0$$

La copule de Frank bivariée a les caractéristiques suivantes :

$C_F(u,v)$	$-\frac{1}{\theta}ln\left(1-\frac{(1-e^{-\theta u})(1-e^{-\theta v})}{(1-e^{-\theta})}\right)$
$c_F(u,v)$	$\theta\frac{e^{-\theta(u+v)}(1-e^{-\theta})}{(e^{-\theta(u+v)}-e^{-\theta u}-e^{-\theta v})^2}$
$K(\theta,t)$	$t-\frac{ln\left(\frac{e^{-\theta u}-1}{e^{-\theta}-1}\right)}{\theta}*\left(e^{\theta u}-1\right)$
$\tau(\theta)$	$1-\frac{4}{\theta}+\frac{4}{\theta^2}\int_0^\theta \frac{t}{e^t-1}dt$

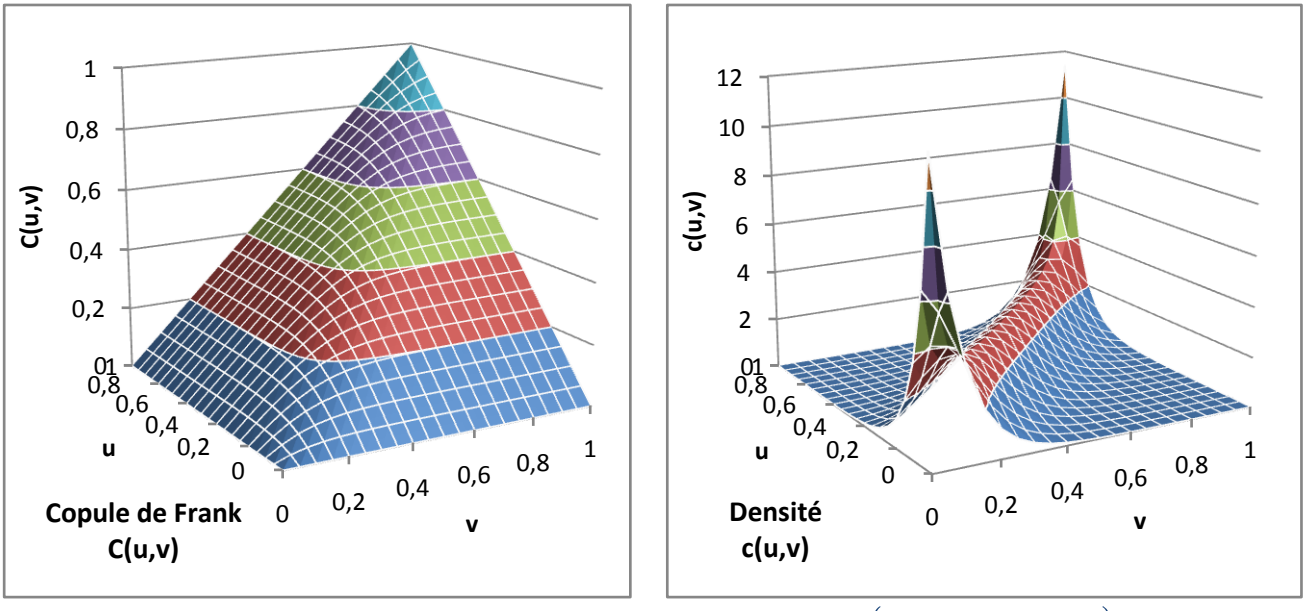

Figure 14 : Représentation de la Copule de Frank $\left(\tau = 0,7 \Leftrightarrow \theta = 11,4\right)$

La copule de Frank est symétrique dans la queue inférieure et supérieure, et a donc tendance à corréler les petits sinistres comme les gros sinistres.

Le domaine de définition de θ lui permet de représenter aussi bien les dépendances négatives que les dépendances positives.

f. Survival Copula

La **fonction de survie** pour la loi d'un vecteur $U = (U_1, \dots, U_d)$ est définie par :

$$S(t_1, \dots, t_d) = \mathbb{P}(U_1 > t_1, \dots, U_d > t_d)$$

Dans le cas univarié, si on note F la fonction de répartition de la variable aléatoire U, alors on a :

$$S(t) = 1 - F(t)$$

La **copule de survie** d'une loi multivariée de marginales ayant les fonctions de survie $(S_i)_{i=1,\dots,d}$ est la fonction qui vérifie :

$$S(t_1, \dots, t_d) = \hat{C}(S_1(t_1), \dots, S_d(t_d))$$

L'expression suivante donne la relation entre la copule de survie et la copule :

$$\hat{C}(u_1, \dots, u_d) = C(1 - u_1, \dots, 1 - u_d) + \sum_{i=1}^{d} u_i - 1$$

- Exemple : La copule HRT

Nous avons vu que la copule de Clayton a tendance à corréler entre eux les petits sinistres. Pourtant nous cherchons le plus souvent à étudier les gros sinistres ainsi que la manière dont ils sont corrélés.

L'idée est donc de définir une copule qui présente des propriétés opposées.

La copule HRT est définie comme la copule de survie de la copule de Clayton.

On a donc les résultats suivants en dimension 2 :

$C_{HRT}(u, v)$	$u + v - 1 + \left[(1-u)^{-\theta} + (1-v)^{-\theta} - 1\right]^{-\frac{1}{\theta}}$
$c_{HRT}(u, v)$	$(1+\theta)\left[(1-u)^{-\theta} + (1-v)^{-\theta} - 1\right]^{-\frac{1}{\theta}-2}\left[(1-u)(1-v)\right]^{-1-\theta}$
$\tau(\theta)$	$\dfrac{\theta}{\theta + 2}$

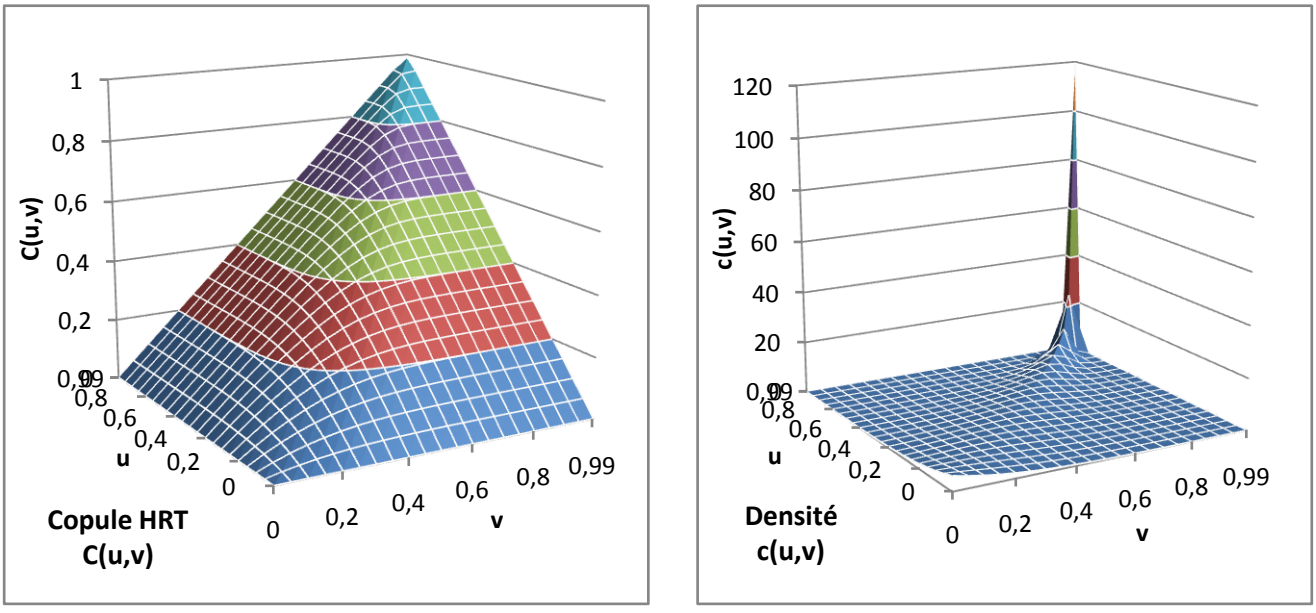

Figure 15 : Représentation de la copule HRT $(\tau = 0,7 \Leftrightarrow \theta = 4,67)$

Malheureusement cette copule n'est pas facile à simuler pour les grandes dimensions d'autant plus que ce n'est pas une copule archimédienne. Cependant, elle est présentée ici pour l'intérêt qu'elle porte sur l'étude des gros sinistres.

7. Dépendance de queue

La dépendance de queue reflète la structure de dépendance entre des événements extrêmes. C'est une notion intéressante dans la mesure où elle permet d'étudier la survenance simultanée de valeurs extrêmes. L'estimation de la VaR que nous calculerons dans le cadre de Solvency II est en effet directement liée à l'estimation des queues de distribution, événements rares et valeurs extrêmes.

Soient donc deux variables aléatoires continues X et Y, ayant pour fonctions de répartition respectives F_X et F_Y, et de copule C.

Le **coefficient de dépendance de queue inférieure** ou lower tail dependence est défini par :

$$\lambda_L = \lim_{u \to 0^+} \mathbb{P}\left[X \leq F_X^{-1}(\alpha) \middle| Y \leq F_Y^{-1}(\alpha)\right] = \lim_{u \to 0^+} \frac{C(u,u)}{u}$$

si cette limite existe.

Le **coefficient de dépendance de queue supérieure** ou upper tail dependence est défini par :

$$\lambda_U = \lim_{u \to 1^-} \mathbb{P}\left[X > F_X^{-1}(\alpha) \middle| Y > F_Y^{-1}(\alpha)\right] = \lim_{u \to 1^-} \frac{1 - 2u + C(u,u)}{1-u}$$

si cette limite existe.

La démonstration de ces formules est donnée en annexe.

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

Ainsi, si $\lambda_U > 0$, il y a une probabilité non nulle que deux événements de forte intensité surviennent simultanément.

Dans le cas unidimensionnel, la caractérisation des extrêmes dépend de la queue de distribution. Dans le cas multidimensionnel, elle dépend de la dépendance des queues de distribution.

Un exemple souvent donné pour illustrer cette notion est la copule de Gumbel, pour laquelle le coefficient de dépendance de queue supérieure vaut :

$$\lambda_U = 2 - 2^{\frac{1}{\theta}}$$

Donc la dépendance tend vers 1 quand θ tend vers l'infini.

La copule Normale et la copule de Frank n'ont pas de dépendance de queue.

8. Estimation des paramètres des copules

Ce paragraphe présente les principales méthodes permettant d'estimer le paramètre de la copule.

a. La méthode des moments

Cette méthode consiste à utiliser les mesures de dépendance.

Le paramètre θ est estimé par :

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

où $\tau(\theta)$ est le tau de Kendall multivarié défini dans le paragraphe **6.**e. et $\hat{\tau}$ sa version empirique.

Si les calculs se font aisément dans le cas d'une copule bivariée, ils sont en revanche plus complexes quand on passe aux dimensions supérieures.

A titre d'exemple, pour la copule de Clayton, on a les résultats suivants :

d	$\tau(\theta)$	$\tau^{-1}(x)$
2,3	$\frac{\theta}{\theta+2}$	$\frac{2x}{1-x}$
4	$\frac{3\theta(7\theta+4)}{7(\theta+2)(3\theta+2)}$	$\frac{28x-6+2\sqrt{49x^2+63x+9}}{21(1-x)}$
5	$\frac{\theta(9\theta+4)}{3(\theta+2)(3\theta+2)}$	$\frac{12x-2+2\sqrt{9x^2+15x+1}}{9(1-x)}$

L'expression de τ pour cette copule se généralise en dimension d de la manière suivante (Quessy 2005)

$$\tau(\theta) = \left(\frac{2^d}{2^{d-1}-1}\right)\frac{q(\theta,d,1)}{q(\theta,d,2)} - \frac{1}{2^{d-1}-1}$$

avec :

$$q(\theta,d,m) = \prod_{i=0}^{d-1}(m+i\theta)$$

La démonstration de cette formule se trouve en annexe.

Ainsi, après avoir trouvé l'estimateur du tau de Kendall, on peut trouver l'estimation du paramètre de la copule de Clayton en résolvant un polynôme en θ puis en choisissant la solution qui est strictement positive.

L'expression est relativement simple pour la copule de Clayton mais ce n'est pas le cas pour les autres copules.

b. La méthode du maximum de vraisemblance

Cette méthode consiste à choisir des lois marginales et une copule paramétrique, et à maximiser la vraisemblance des observations sur l'ensemble des paramètres (ceux des lois marginales et celui de la copule).

Nous avons vu au paragraphe **4.** l'expression de la densité de la variable aléatoire $X = (X_1, \dots, X_d)$:

$$f(x_1, \dots, x_d) = c(F_1(X_1), \dots, F_d(X_d)) \prod_{i=1}^{d} f_i(x_i)$$

Soit $\{(x_1^j, \dots, x_d^j)\}_{j=1}^{n}$ l'échantillon d'observations.

Soit $\theta = (\theta_1, \dots, \theta_d; \theta_C)$ le vecteur des paramètres à estimer.

La log-vraisemblance de l'échantillon considéré s'écrit alors :

$$l(\theta) = \sum_{j=1}^{n} ln\left(c\left(F_1\left(x_1^j; \theta_1\right), \dots, F_d\left(x_d^j; \theta_d\right); \theta_C\right)\right) + \sum_{j=1}^{n}\sum_{i=1}^{d} ln\left(f_i\left(x_i^j; \theta_i\right)\right)$$

On en déduit l'estimateur de θ, noté $\hat{\theta}_{ML}$:

$$\hat{\theta}_{ML} = \underset{\theta}{\operatorname{argmax}}\, l(\theta)$$

$\hat{\theta}_{ML}$ est sans biais, convergent et asymptotiquement normal sous certaines hypothèses que nous ne détaillerons pas ici (Durrleman, et al. 2000)

$$\sqrt{n}\left(\hat{\theta}_{ML} - \theta_0\right) \xrightarrow{loi} \mathcal{N}\left(0; I^{-1}(\theta_0)\right)$$

où $I^{-1}(\theta_0)$ représente la matrice d'information de Fisher.

Mais la maximisation de la vraisemblance est difficile lorsque le nombre de paramètres est très élevé, et peut alors engendrer des calculs très longs.

c. La méthode IFM (Inference Functions for Margins)

Cette méthode a été proposée par Joe et Xu (Joe et Xu 1996). Il s'agit de séparer l'estimation des paramètres des lois et ceux de la copule. La maximisation se fait alors en 2 étapes.

Dans un premier temps, on estime les paramètres de chaque marginale :

$$\hat{\theta}_i = \underset{\theta_i}{\operatorname{argmax}}\, l_i(\theta_i) = \underset{\theta_i}{\operatorname{argmax}} \sum_{j=1}^{n} ln\left(f_i\left(x_i^j; \theta_i\right)\right)$$

On utilise ensuite ces estimateurs pour estimer le paramètre de la copule :

$$\hat{\theta}_C = \underset{\theta_C}{\operatorname{argmax}}\, l_C(\theta_C) = \underset{\theta_C}{\operatorname{argmax}} \sum_{j=1}^{n} \ln\left(c\left(F_1\left(x_1^j; \hat{\theta}_1\right), \dots, F_d\left(x_d^j; \hat{\theta}_d\right); \theta_C\right)\right)$$

L'estimateur $\theta_{IFM} = \left(\hat{\theta}_1, \dots, \hat{\theta}_d, \hat{\theta}_C\right)$ est également convergent, asymptotiquement normal et sans biais. On a (Joe et Xu, 1996):

$$\sqrt{n}\left(\hat{\theta}_{IFM} - \theta_0\right) \xrightarrow{loi} \mathcal{N}\left(0; V^{-1}(\theta_0)\right)$$

où $V(\theta_0)$ représente la matrice d'information de Godambe.

Si on définit $g(\theta) = \left(\frac{\partial}{\partial\theta_1} l_1, \ldots, \frac{\partial}{\partial\theta_d} l_d\right)$, la matrice d'information de Godambe s'écrit (Joe 1997) :

$$V(\theta_0) = D^{-1} M (D^{-1})^T$$

avec :

$$D = \mathbb{E}\left[\frac{\partial}{\partial\theta} g(\theta)^T\right]$$
$$M = \mathbb{E}[g(\theta)^T g(\theta)]$$

La méthode IFM repose sur des calculs plus légers que ceux de la méthode du maximum de vraisemblance. Cependant, la détermination de la matrice de Godambe peut s'avérer très complexe à cause des multiples dérivées.

d. La méthode CML (Canonical Maximum Likelihood)

Cette méthode, proposée par Bouye, et al. (2000) ne nécessite aucune estimation sur les marginales. Il s'agit dans un premier temps de transformer les observations $\{(x_1^j, \ldots, x_d^j)\}_{j=1}^n$ en variables uniformes $\{(\hat{u}_1^j, \ldots, \hat{u}_d^j)\}_{j=1}^n$ en utilisant les fonctions de répartition empiriques $\hat{F}_i(.)$ définies comme suit :

$$\hat{F}_i(\blacksquare) = \frac{1}{n}\sum_{j=1}^{n} \mathbb{I}_{\{X_i^j \leq \blacksquare\}}$$

L'estimation du paramètre de la copule se calcule ensuite de la manière suivante :

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}} \sum_{j=1}^{n} \ln\left(c(\hat{u}_1^j, \ldots, \hat{u}_d^j; \theta)\right)$$

L'avantage de cette méthode est de pouvoir procéder à l'estimation paramétrique de la copule indépendamment des marginales. De plus, elle génère des temps de calcul limités.

9. Simulations

Ce paragraphe décrit des méthodes qui peuvent être utilisées pour générer des observations ayant une copule donnée.

Afin de pouvoir faire les comparaisons entre les différentes copules, nous nous placerons dans le cas bivarié et tracerons des graphes obtenus à partir des différentes copules, les lois marginales étant uniformes.

Pour commencer, le graphe ci-dessous montre le résultat de simulations de couples (u, v), dans le cas où u et v sont indépendants :

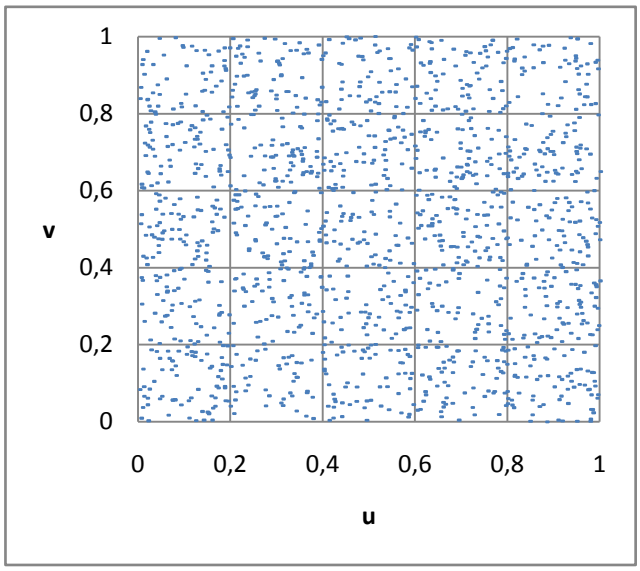

Figure 16 : Simulation de couples (u,v) en cas d'indépendance

Les points sont répartis uniformément dans le carré. Aucune structure de dépendance ne se dégage ici.

a. La méthode des distributions

Cette méthode est intéressante lorsque la distribution générée par la copule est plus facile que la copule elle-même. C'est par exemple le cas de la copule gaussienne : un vecteur gaussien de dimension d est facile à simuler alors que la copule gaussienne n'est pas simple à simuler directement.

La méthode consiste à simuler des réalisations de $X = (X_1, \dots, X_d)$ de loi F et à appliquer la transformation $U = (F_1(X_1), \dots, F_d(X_d))$.

- Copule normale

  - Trouver la décomposition de Cholesky L de la matrice de corrélation R.
    R étant une matrice définie positive, il s'agit de trouver une matrice triangulaire inférieure L telle que $R = LL^T$.
  - Simuler un vecteur $z = (z_1, \dots, z_d)^T$ de variables aléatoires indépendantes suivant une loi normale centrée réduite.
  - Poser $x = Lz$
  - Finalement, $(u_1, \dots, u_d) = (\phi(x_1), \dots, \phi(x_d))$, ϕ étant la fonction de répartition de la loi normale centrée réduite.

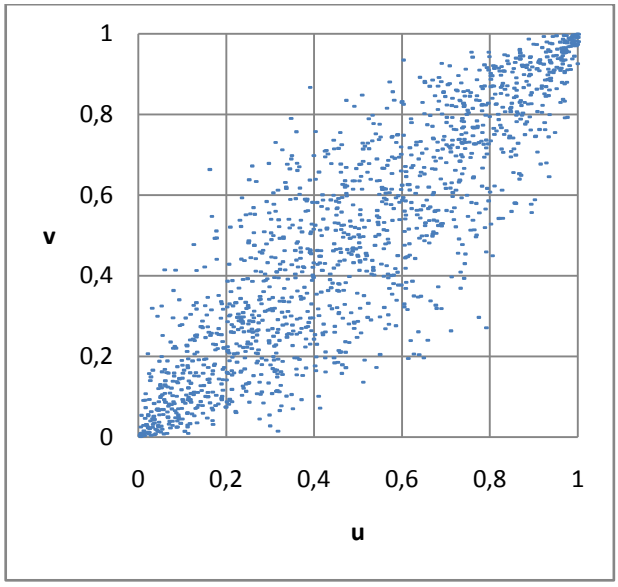

Figure 17 : Simulation de couples (u,v) avec la copule gaussienne $(\tau = 0,7 \Leftrightarrow \rho = 0,89)$

La loi multivariée simulée avec une copule normale est symétrique : la dépendance aux extrêmes est la même.

b. La méthode des distributions conditionnelles

Considérons, pour i appartenant à $\{1, \dots, d-1\}$ la fonction C_i définie par :

$$C_i(u_1, \dots, u_i) = C(u_1, \dots, u_i, 1, \dots, 1)$$

avec :

$$C_1(u_1) = u_1$$
$$C_d(u_1, \dots, u_d) = C(u_1, \dots, u_d)$$

La distribution conditionnelle de U_i sachant $U_1, \dots, U_{i-1}$ est donnée par :

$$C_i(u_i|u_1, \dots, u_{i-1}) = \mathbb{P}(U_i \leq u_i | U_1 = u_1, \dots, U_{i-1} = u_{i-1})$$
$$= \frac{\frac{\partial^{i-1}}{\partial u_1 \dots \partial u_{i-1}} C_i(u_1, \dots, u_i)}{\frac{\partial^{i-1}}{\partial u_1 \dots \partial u_{i-1}} C_{i-1}(u_1, \dots, u_i)}$$

sous les conditions que le numérateur et le dénominateur existent, et que le dénominateur soit non nul.

L'algorithme est le suivant :

- Simuler une variable aléatoire u_1 suivant une loi $\mathcal{U}(0;1)$
- Simuler une variable aléatoire u_2 suivant $C_2(.|u_1)$
- …
- Simuler une variable aléatoire u_d suivant $C_d(.|u_1, \dots, u_{d-1})$

Cette méthode n'est pas la plus efficace dans la mesure où quand on dépasse la dimension 2, les calculs deviennent plus complexes.

Considérons pour illustrer cette méthode les copules de Clayton, de Frank et HRT en dimension 2.

- Copule de Clayton

Rappelons l'expression de la copule de Clayton:

$$C_C(u,v) = \left(u^{-\theta} + v^{-\theta} - 1\right)^{-\frac{1}{\theta}}$$

On a :

$$C_2(v|u) = \frac{\partial C}{\partial u}(u,v)$$

On trouve alors :

$$C_2(v|u) = u^{-\theta-1}\left(u^{-\theta} + v^{-\theta} - 1\right)^{-\frac{1}{\theta}-2}$$

Soit l'équation $p = C_2(v|u)$

Il s'agit de déterminer l'expression de v en fonction de u et p.

On trouve alors:

$$\boldsymbol{v = \left[u^{-\theta}\left(p^{\frac{-\theta}{1+\theta}} - 1\right) + 1\right]^{-\frac{1}{\theta}}} \qquad \textbf{(9.1)}$$

L'algorithme pour obtenir le couple (u,v) est donc le suivant :

- Simuler 2 variables aléatoires indépendantes u et p suivant une loi $\mathcal{U}(0;1)$.
- Poser v tel qu'il est défini dans l'équation (9.1)

U et V ont pour distribution jointe la copule de Clayton.

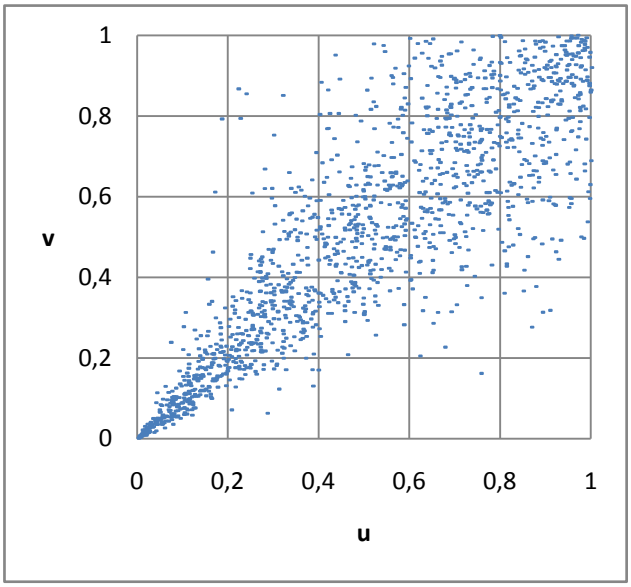

Figure 18 : Simulation de couples (u,v) avec la copule de Clayton $\left(\tau = 0,7 \Leftrightarrow \theta = 4,67\right)$

On voit d'après cette représentation graphique qu'il y a une importante concentration de points près de l'origine. La copule de Clayton aurait donc tendance à corréler entre eux les petits sinistres.

- Copule de Frank

Elle a pour expression :

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

En dérivant par rapport à u on obtient :

$$C_2(v|u) = \frac{e^{-\theta u}\left(e^{-\theta v} - 1\right)}{e^{-\theta} - 1 + (e^{-\theta u} - 1)(e^{-\theta v} - 1)}$$

En résolvant $p = C_2(v|u)$ on obtient :

$$\boldsymbol{v = -\frac{1}{\theta} ln\left\{1 + \frac{p(1-e^{-\theta})}{p(e^{-\theta u}-1)-e^{-\theta u}}\right\}} \qquad \textbf{(9.2)}$$

On en déduit l'algorithme :

- Simuler 2 variables aléatoires indépendantes u et p suivant une loi $\mathcal{U}(0;1)$.
- Poser v tel qu'il est défini dans l'équation (9.2)

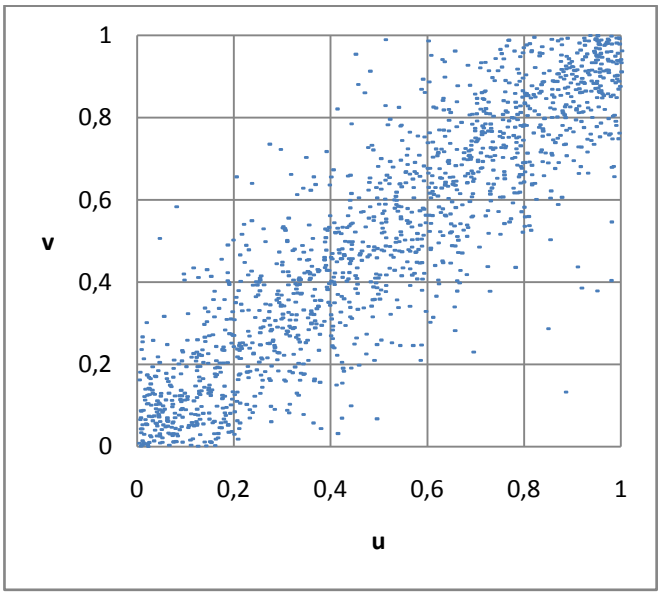

Figure 19 : Simulation de couples (u,v) avec la copule de Frank $(\tau = 0,7 \Leftrightarrow \theta = 11,4)$

On remarque dans cette représentation graphique une forte densité de points à l'origine comme au point $(1,1)$, ce qui illustre bien le fait que cette copule corrèle les petits sinistres comme elle corrèle les gros sinistres. La symétrie de cette copule est également remarquable.

- La copule HRT

La copule HRT est définie par :

$$C_{HRT}(u,v) = u + v - 1 + \left((1-u)^{-\theta} + (1-v)^{-\theta} - 1\right)^{-\frac{1}{\theta}}$$

On trouve ensuite :

$$p = C_2(v|u) = 1 - \left[(1-u)^{-\theta} + (1-v)^{-\theta} - 1\right]^{-\frac{1}{\theta}-1}(1-u)^{-\theta-1}$$

On en déduit :

$$v = 1 - \left\{ \left[(1-p)(1-u)^{\theta+1} \right]^{-\frac{\theta}{1+\theta}} + 1 - (1-u)^{-\theta} \right\}^{-\frac{1}{\theta}} \qquad (9.3)$$

L'algorithme consiste donc à :

- Simuler 2 variables aléatoires indépendantes u et p suivant une loi $\mathcal{U}(0;1)$.
- Poser v tel qu'il est défini dans l'équation (9.3)

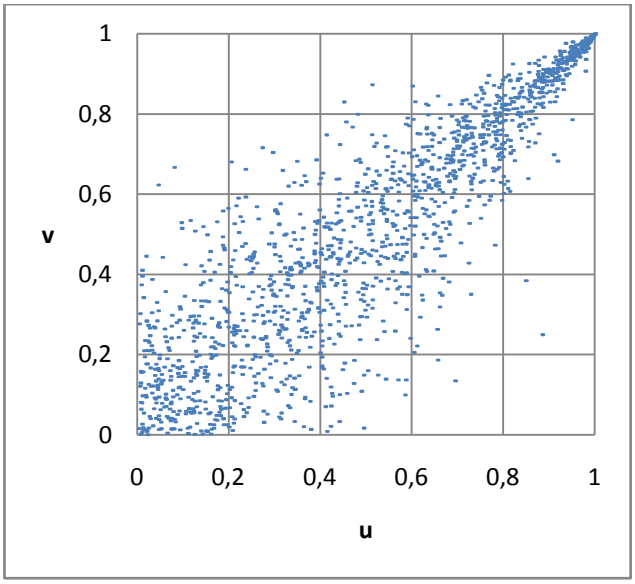

Figure 20 : Simulation de couples (u,v) avec la copule HRT $\left(\tau = 0,7 \Leftrightarrow \theta = 4,67\right)$

Comme étant la copule de survie de la copule de Clayton, la copule HRT trace une concentration importante de points près de (1,1). Les gros sinistres auraient donc tendance à survenir ensemble. Cette copule pourrait donc être intéressante quant à l'étude des événements de forte intensité.

c. La méthode de Marshall et Olkin

Contrairement aux 3 copules précédentes, la simulation de la copule de Gumbel par la méthode des distributions devient très compliquée.

Marshall et Olkin (1988) ont proposé une méthode propre aux copules archimédiennes. Cette dernière fait appel à la transformée de Laplace et son inverse.

La transformée de Laplace d'une variable positive Y est définie par :

$$\Psi(s) = \mathbb{E}[e^{-sY}] = \int_0^\infty e^{-st} dF_Y(t)$$

où F_Y représente la fonction de répartition de Y.

Marshall et Olkin (1988) montrent que la fonction de répartition jointe F peut s'écrire de la façon suivante :

$$F(x_1, \dots, x_d) = \Psi\{\Psi^{-1}(F_1(x_1)) + \cdots + \Psi^{-1}(F_d(x_d))\}$$

où F_i est la fonction de répartition marginale de la fonction de répartition jointe F, pour tout $i\epsilon\{1,\dots,d\}$.

Or, les générateurs des copules de Clayton, de Frank et de Gumbel sont les transformées de Laplace de variables aléatoires positives. Frees et Valdez (1998) proposent alors des algorithmes de simulation permettant de générer un vecteur de variables aléatoires $(u_1,\dots,u_d)$ suivant la copule de Clayton, de Frank ou de Gumbel. C'est donc cette méthode qui est plus appropriée lorsque l'on travaille avec de grandes dimensions.

- Copule de Clayton

Pour la copule de Clayton, Y suit une loi Gamma $\Gamma\left(\frac{1}{\theta},1\right)$.

L'algorithme est donc le suivant :

- Générer une variable aléatoire Y suivant une loi Gamma $\Gamma\left(\frac{1}{\theta},1\right)$.
- Simuler un vecteur de variables aléatoires indépendantes $(v_1,\dots,v_d)$ suivant une loi $\mathcal{U}(0,1)$, et indépendant de Y.
- Calculer $u_i = F_i^{-1}(v_i^*)$ pour tout i dans $\{1,\dots,d\}$, où :

$$v_i^* = \Psi\left(\frac{-\ln(v_i)}{Y}\right)$$

et $\Psi(s) = (1+s)^{-1/\theta}$

- Copule de Frank

Pour la copule de Frank, Y a une distribution de série logarithmique sur les nombres naturels.

Poser $\alpha = 1 - e^{-\theta}$

La distribution de Y est donnée par :

$$\mathbb{P}(Y=k) = -\frac{\alpha^k}{k.\ln(1-\alpha)}$$

L'algorithme est le suivant :

- Générer une variable aléatoire Y suivant une distribution de série logarithmique sur les nombres naturels.
- Simuler un vecteur de variables aléatoires indépendantes $(v_1,\dots,v_d)$ suivant une loi $\mathcal{U}(0,1)$, et indépendant de Y.
- Calculer $u_i = F_i^{-1}(v_i^*)$ pour tout i dans $\{1,\dots,d\}$, où :

$$v_i^* = \Psi\left(\frac{-\ln(v_i)}{Y}\right)$$

et $\Psi(s) = \theta^{-1} ln\left[1 + e^s(e^\theta - 1)\right]$

- Copule de Gumbel

Pour la copule de Gumbel, Y suit une loi stable positive de paramètre $\frac{1}{\theta}$:

- Simuler Y suivant la loi stable positive de paramètre $\frac{1}{\theta}$.

$$Y = \frac{\sin\left(\frac{s}{\theta}\right)}{\cos\,(s)^{\theta}}\left[cos\left(\frac{\left(1-\frac{1}{\theta}\right)s}{\xi}\right)\right]^{\theta-1}$$

où ξ suit une loi exponentielle de paramètre 1, s suit une loi uniforme $\mathcal{U}\left(-\frac{\pi}{2},\frac{\pi}{2}\right)$ et est indépendante de ξ.

- Simuler un vecteur de variables aléatoires indépendantes $(v_1,\dots,v_d)$ suivant une loi $\mathcal{U}(0,1)$, et indépendant de Y.
- Calculer $u_i = F_i^{-1}(v_i^*)$ pour tout i dans $\{1,\dots,d\}$, où :

$$v_i^* = \Psi\left(\frac{-\ln(v_i)}{Y}\right)$$

et $\Psi(s) = \exp\,(-s^{\frac{1}{\alpha}})$

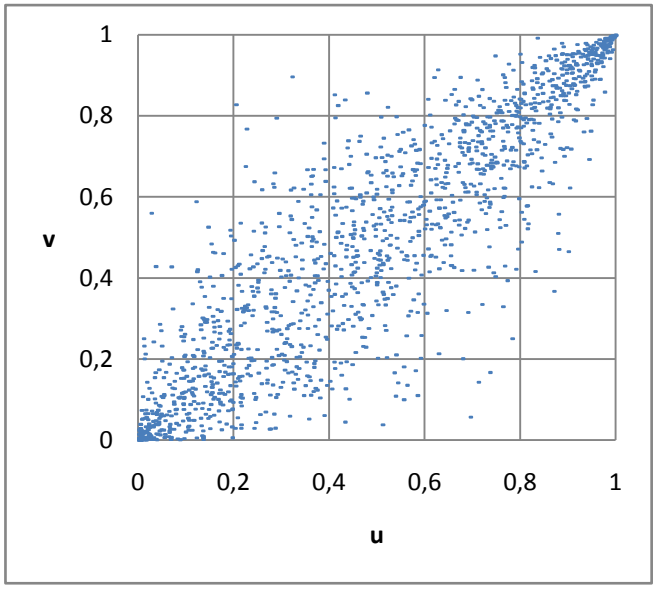

Figure 21 : Simulation de couples avec la copule de Gumbel $(\tau = 0,7 \Leftrightarrow \rho = 3,33)$

Contrairement à la copule de Clayton, cette copule a tendance à corréler entre eux les gros sinistres. On constate en effet une forte concentration des points près de (1,1).

Quatrième partie : Mise en œuvre

Cette partie aborde l'aspect pratique du travail qui a été fait. Le cas étudié concerne le portefeuille de risques de particuliers. Le but est de montrer comment la prise en compte de la dépendance spatiale influence la détermination du capital en risque de la société.

1. Présentation du contexte

Comme cité plus haut, l'étude porte sur le risque résidentiel. Le capital en risque évalué dans ce travail ne concerne donc que ce segment. La détermination du capital en risque de la société prise dans sa globalité devrait inclure les autres risques (agricoles et professionnels) ainsi que les portefeuilles des différentes filiales.

Pour chaque caisse régionale numérotée de 1 à 11, le risque tempête a été modélisé en simulant l'impact de 14900 scénarios possibles de tempêtes sur la France.

L'exploitation des 11 « Event Loss Table » (voir Première Partie, Chapitre 5) générés par le modèle RMS permettent de simuler les coûts moyens associés à chacun de ces événements.

On suppose que la loi du taux d'endommagement est une Beta, et que la fréquence des sinistres suit une loi de Poisson.

Toutes les simulations ont été implémentées sous SAS.

2. Nombre de sinistres

Rappelons les différents champs renseignés dans les tables, pour une tempête k :

- La fréquence annuelle f_k
- Le coût moyen c_k
- L'écart-type corrélé $\sigma_{c,k}$
- L'écart-type décorrélé $\sigma_{i,k}$
- L'exposition c_max_k

Puisqu'une tempête donnée n'affecte pas de la même manière chaque caisse régionale, une même tempête peut avoir un taux d'endommagement différent selon chaque caisse considérée. Ainsi, les paramètres α et β de la loi Beta du taux d'endommagement doivent être calculés dans chaque caisse.

La manière de les obtenir est explicitée dans la partie 1 au paragraphe **8**.

Afin de générer le nombre d'événements, il faut également faire l'inventaire des 14900 différentes tempêtes qui peuvent affecter le portefeuille et retenir leur fréquence annuelle. On commence alors par déterminer le nombre moyen d'événements dans une année en sommant les fréquences annuelles, tel qu'il a été expliqué dans la première partie, chapitre **7**. Il résulte des calculs une moyenne de 10,59 tempêtes par an.

Il s'agit ensuite de tirer aléatoirement un nombre suivant la loi de Poisson $\mathcal{P}(10{,}59)$.

Une fois qu'on aura tiré le nombre de sinistres n, il faut tirer ces n sinistres. Puisque l'on a les fréquences, on passe d'une approche déterministe à une approche probabiliste de la manière suivante :

$$p_k = \frac{f_k}{\sum_i f_i}$$

où p_k est la probabilité que la tempête k survienne.

La probabilité cumulée est donnée par :

$$\hat{p}_k = \sum_{i \leq k} p_i$$

$\hat{p}_k$ représente donc la probabilité que l'identifiant du sinistre tiré soit inférieur à k.

Pour tirer un sinistre au hasard, on commence par tirer une variable aléatoire $U \sim \mathcal{U}(0;1)$. On cherche ensuite le nombre k tel que $\hat{p}_{k-1} < u \leq \hat{p}_k$.
Alors la tempête tirée est la tempête n° k.

L'algorithme pour une itération est donc le suivant:

a. Tirer le nombre de sinistres n suivant une loi de Poisson $\mathcal{P}(10{,}59)$.
b. Tirer n variables aléatoires u_i suivant une loi uniforme $\mathcal{U}(0;1)$.
c. Pour i allant de 1 à n :
   - Trouver les nombres k_i tels que $\hat{p}_{k_i-1} < u \leq \hat{p}_{k_i}$

3. Coût des sinistres

Une fois le nombre de tempêtes tiré, la suite consiste à générer les coûts. Cependant, il faut tenir compte d'une certaine dépendance entre les caisses régionales. Cette dépendance est décrite dans les différentes copules.

➢ Cas d'indépendance

Pour pouvoir comparer les résultats, il semble nécessaire de commencer par simuler les coûts dans le cas où les 11 caisses régionales sont indépendantes.
Il suffit alors de tirer 11 variables aléatoires indépendantes suivant une loi uniforme $\mathcal{U}(0;1)$. La suite de l'algorithme est donc le suivant :

d. Pour i allant de 1 à n
   d.1. Tirer $u_i^1, \dots, u_i^{11}$ selon une loi $\mathcal{U}(0;1)$.
   d.2. Pour j allant de 1 à 11
      - Récupérer les paramètres α_i^j et β_i^j de la tempête n° k_i pour la caisse j
      - Récupérer l'exposition $c\_max_{k_i}^j$ de la tempête n° k_i pour la caisse j
      - Calculer le quantile $F^{-1}(u_i^j)$ de loi Beta $B(\alpha_i^j; \beta_i^j)$
      - Calculer les coûts corrélés :

$$C_i^j = \begin{cases} F^{-1}(u_i^j) * c\_max_{k_i}^j & si\ la\ caisse\ n°j\ est\ touchée\ par\ la\ tempête\ n°\ k_i \\ 0 & sinon \end{cases}$$

e. Pour j allant de 1 à 11, calculer le coût pour une année pour chaque caisse :

- $C^j = \sum_i C_i^j$

f. Calculer le coût global pour une année :

- $C = \sum_j C^j$

➢ Copule gaussienne

La copule gaussienne est plus facile pour une première approche. L'algorithme consiste donc à :

d. Calculer la matrice de corrélation R
e. Trouver L, la décomposition de Cholesky de R
f. Pour i allant de 1 à n :
   - Tirer un vecteur aléatoire $Z = (z_i^1, \dots, z_i^{11})$ où $z_i^j \sim \mathcal{N}(0;1)$
   - Calculer $X = LZ^T$
   - $(u_i^1, \dots, u_i^{11}) = (\Phi^{-1}(x_i^1), \dots, \Phi^{-1}(x_i^{11}))$
   - Reprendre l'étape d.2 du cas indépendant

L'algorithme est semblable pour les autres copules. La seule différence réside dans la manière de tirer le vecteur $(u_i^1, \dots, u_i^{11})$: une fois qu'il est tiré on reprend l'étape d.2 décrite dans le cas indépendant. Nous ne reprendrons donc pas les algorithmes qui ont déjà été détaillés dans le paragraphe **9** de la partie 3. Comme étant la plus facile à implémenter, c'est la méthode de Marshall & Olkin qui a été utilisée. Seule la détermination des paramètres des copules sera décrite dans la suite.

➢ Copule de Clayton

Pour déterminer le paramètre de la copule de Clayton, c'est la méthode des moments qui a été utilisée. Elle consiste dans un premier temps à calculer le tau de Kendall empirique.

En dimension d, le tau de Kendall empirique est défini par :

$$\hat{\tau}(\theta) = \left(\frac{2^d - 1}{2^{d-1} - 1}\right) - \left(\frac{2^d}{2^{d-1} - 1}\right) \int_0^1 K_n(\theta, t) dt$$

où $K_n(\theta, t)$ est la version empirique de $K(\theta, t)$.

Le calcul de cette version empirique de $K(\theta, t)$ est expliqué en annexe.

L'intégrale est ensuite approchée par la méthode de Newton-Cotes, également expliquée en annexe.

Dans notre cas d=11. Il en résulte que :

$$\hat{\tau}(\theta) = 0{,}41611$$

Rappelons l'expression de $\tau(\theta)$ citée dans le paragraphe **8**.a. de la partie 3 pour la copule de Clayton:

$$\tau(\theta) = \left(\frac{2^d}{2^{d-1} - 1}\right) \frac{q(\theta, d, 1)}{q(\theta, d, 2)} - \frac{1}{2^{d-1} - 1}$$

avec :

$$q(\theta, d, m) = \prod_{i=0}^{d-1}(m + i\theta)$$

On a :

$$\frac{q(\theta,11,1)}{q(\theta,11,2)} = \frac{1(1+\theta)(1+2\theta)(1+3\theta)\dots(1+9\theta)(1+10\theta)}{2(2+\theta)(2+2\theta)(2+3\theta)\dots(2+9\theta)(2+10\theta)}$$
$$= \frac{(1+6\theta)(1+7\theta)(1+8\theta)(1+9\theta)(1+10\theta)}{2^6(2+\theta)(2+3\theta)(2+5\theta)(2+7\theta)(2+9\theta)}$$

Il s'agit donc de résoudre :

$$\frac{\hat{\tau}(\theta) + \frac{1}{2^{d-1}-1}}{\frac{2^d}{2^{d-1}-1}} - \frac{(1+6\theta)(1+7\theta)(1+8\theta)(1+9\theta)(1+10\theta)}{2^6(2+\theta)(2+3\theta)(2+5\theta)(2+7\theta)(2+9\theta)} = 0$$

On trouve :

$$\theta_1 = 2{,}63903$$
$$\theta_2 = -0{,}22404$$
$$\theta_3 = -0{,}26204$$

La copule de Clayton n'étant définie que pour $\theta > 0$, on choisit la solution $\theta = 2{,}63903$.

- Copule de Frank

C'est la méthode CML qui est utilisée ici pour estimer le paramètre θ.
Savu et Trede (2004) ont proposé une méthode de calcul de la densité des copules archimédiennes :

$$c(u_1, \dots, u_d) = \varphi_\theta^{-1(d)}\left(\sum_{i=1}^{d}\varphi_\theta(u_i)\right) \times \prod_{i=1}^{d}\varphi'_\theta(u_i)$$

Ils montrent que pour la copule de Frank en dimension d, on a :

$$\varphi_\theta(t) = -ln\left(\frac{e^{-\theta t}-1}{e^{-\theta}-1}\right)$$

$$\varphi'_\theta(t) = \frac{\theta e^{-\theta t}}{e^{-\theta}-1}$$

$$\varphi_\theta^{-1(n)}(x) = \frac{(e^\theta - 1)e^{\theta+x}}{(e^\theta - e^{\theta+x} - 1)^n \theta} \times \sum_{k=1}^{n-1} B_{n-1,k} e^{(n-1-k)(x+\theta)}(e^\theta - 1)^{k-1}$$

où $B_{i,j}$ est défini par :

$$\begin{cases} B_{i,1} = B_{i,i} = 1 \\ B_{i,j} = (i-j-1)B_{i-1,j-1} + jB_{i-1,j} \end{cases}$$

Le solveur d'Excel nous permet d'obtenir la valeur de θ qui maximise la vraisemblance. On obtient $\theta = 2{,}41072$

- Copule de Gumbel

La formule de la densité proposée par Savu et Trede (2004) reste valable pour la copule de Gumbel. Malheureusement, l'expression de la dérivée n-ième de l'inverse du générateur n'a pas d'expression explicite : le calcul s'avère très fastidieux en dimension 11. Une valeur θ égale à 2,5 a été choisie de façon arbitraire, juste dans le but de comparer les résultats.

4. Résultats

Nous avons vu dans le paragraphe précédent comment simuler les engagements de l'assureur en une année. Il suffit de réitérer ce procédé N fois afin d'obtenir une distribution empirique des coûts. De là, on pourra déduire nos mesures de risque.

Les VaR et TVaR à 99,5% ainsi obtenus sont les suivantes :

	Indépendant	Normale	Clayton	Frank	Gumbel
VaR	711 092 561,83	792 263 511,55	763 028 918,98	769 948 086,66	837 420 210,73
Rapport	1,00	1,11	1,07	1,08	1,18
TVaR	1 103 929 272,65	1 166 041 102,20	1 194 572 248,04	1 151 651 392,85	1 274 864 737,14
Rapport	1,00	1,06	1,08	1,04	1,15

Tableau 6 : Résultats des simulations

La ligne « Rapport » indiquée dans le tableau désigne l'écart par rapport au cas indépendant. Ainsi, avec la copule Normale, la VaR augmente de 11% par rapport à la VaR où on n'a tenu compte d'aucune dépendance. Nous avons vu cependant que cette copule n'était pas toujours adaptée au domaine de l'assurance. En effet, la copule Normale ne présente pas de dépendance de queue et ne permet donc pas de caractériser une dépendance des valeurs extrêmes.

La copule de Clayton présente un coefficient de dépendance de queue inférieure, et a donc tendance à ne corréler que les valeurs minimales. En effet, on a pour cette copule :

$$\lambda_L = \frac{1}{2^{\frac{1}{\theta}}} = 0{,}7690$$

Puisqu'elle ne corrèle entre eux que les petits sinistres, cette copule est celle qui fait le moins augmenter la VaR. La copule de Clayton ne paraît donc pas adaptée à notre étude dans la mesure où par souci de prudence et afin de déterminer l'organisation de la couverture en réassurance, nous cherchons à appréhender les événements catastrophiques.

La copule de Frank, comme la copule normale, n'admet pas de dépendance de queue. Les extrêmes sont donc indépendants avec cette copule. Elle a donc tendance à sous-estimer les événements catastrophiques.

La copule de Gumbel, contrairement à la copule de Clayton, est celle avec laquelle l'augmentation des mesures de risque est maximale. Elle a en effet une dépendance de queue supérieure donnée par :

$$\lambda_L = 2 - 2^{1/\theta} = 0{,}6805$$

Cette copule paraît donc plus adaptée à notre cas, dans la mesure où elle a la caractéristique de pouvoir représenter des risques dont la structure de dépendance est plus accentuée sur la queue supérieure.
On rappelle qu'en raison de la complexité des calculs, le paramètre de la copule a été choisi arbitrairement. Une prochaine étape serait donc d'approfondir ce point pour déterminer une valeur plus juste de θ.

Rappelons enfin l'expression de la copule HRT, fonction de survie de la copule de Clayton :

$$\begin{aligned} C_{HRT}(u,v) &= u + v - 1 + \left[(1-u)^{-\theta} + (1-v)^{-\theta} - 1\right]^{-\frac{1}{\theta}} \\ &= u + v - 1 + C_C(1-u, 1-v) \end{aligned}$$

Ainsi, si on pose $z = 1 - u$, on obtient :

$$\lim_{u \to 1^-} \frac{1 - 2u + C_{HRT}(u,u)}{1-u} = \lim_{u \to 1^-} \frac{C_C(1-u, 1-u)}{1-u} = \lim_{z \to 0^+} \frac{C_C(z,z)}{z}$$

Le coefficient de dépendance de queue supérieure de la copule HRT correspond donc au coefficient de dépendance de queue inférieure de la copule de Clayton. Cette assertion est d'ailleurs valable pour toutes les copules avec leur copule de survie. Etant donné qu'elle vise à rendre compte d'une dépendance sur les événements de forte intensité, la copule HRT pourrait convenir à notre étude. Cependant, aucune méthode de simulation n'a été présentée pour les dimensions d>2. Ceci constituerait également un point d'investigation.

De façon générale, les résultats nous montrent que la modélisation des dépendances avec les copules fait augmenter la Value-at-Risk et la Tail Value-at-Risk. La suite consisterait donc à sélectionner la copule qui correspondrait le plus à nos données. Le test d'adéquation du khi-deux est un outil qui permet de choisir la bonne copule. Cependant, ce test ne convient en général qu'aux distributions bivariées. Malheureusement, les articles parus à ce jour ne proposent pas de méthode de sélection de copule pour les dimensions supérieures. Même si certaines copules semblent convenir plus que d'autres, il faudrait approfondir ce problème d'adéquation.

Conclusion

Les tempêtes qui se sont abattues en Europe en 1999 ont démontré comment elles étaient un risque majeur pour les assureurs non-vie. Elles ont mis en évidence la nécessité pour les assureurs et réassureurs d'évaluer correctement le potentiel de sinistre correspondant à des catastrophes qui semblent être de plus en plus fréquentes.

La modélisation permet de fournir des points de repère pour déterminer l'étendue de la couverture de réassurance. Une étape indispensable à cette modélisation est l'évaluation des sommes assurées, qui représentent les engagements de l'assureur vis-à-vis des assurés. La qualité et la précision des données jouent un rôle fondamental dans cette évaluation car elles sont indispensables à une estimation correcte des sommes assurées. Un processus de rehaussement de données incomplètes est réalisé à l'aide des données correctes existant dans la base ainsi que des données nationales.

Les modèles disposent d'une librairie de tempêtes historiques qui ont impacté la France. Celles-ci sont ensuite déclinées suivant des variantes, où sont modifiées les trajectoires et les intensités des tempêtes historiques. A chacune de ses tempêtes est associée une probabilité annuelle de survenance. Des fonctions d'endommagement permettent ensuite de déterminer les dommages assurés à partir des valeurs assurées calculées.

Par ailleurs, le projet de règlementation Solvency II impose aux assureurs des niveaux réglementaires de capital permettant de couvrir les risques inhérents à leur activité. Parmi ces seuils, le SCR (Solvency Capital Requirement) ou capital cible correspond à la VaR à 99,5% sur un an, soit un événement ayant une période de retour de 200 ans.

Les simulations de Monte-Carlo sont utilisées pour évaluer de telles mesures de risque. On suppose pour cela que la loi du taux d'endommagement est une Beta, et que la fréquence des sinistres suit une loi de Poisson. Dans l'état actuel de leur développement, les modèles ne tiennent pas encore compte des dépendances possibles entre la survenance des événements, ce qui peut conduire à une sous-estimation de la charge annuelle.

Puisque une tempête a une grande étendue, les coûts subis par une caisse régionale donnée sont corrélés aux coûts subis par une caisse voisine. Il s'avère donc nécessaire de modéliser la dépendance entre les caisses régionales : les copules sont un outil flexible permettant de le faire.

Ce sont les copules Normale, de Clayton, de Frank et de Gumbel qui ont été utilisées dans le cadre de cette étude. Les résultats des simulations montrent que la modélisation de la dépendance avec les copules augmente les mesures de risques telles que la VaR et la TVaR. D'après leurs propriétés qui sont de bien caractériser une dépendance entre les événements catastrophiques, les copules de Gumbel et HRT sembleraient être les plus adaptées à l'étude. Cependant, la détermination du paramètre de la copule de Gumbel et la simulation de la copule HRT pour les grandes dimensions restent des axes d'investigation supplémentaires. Le test du Khi-Deux n'étant adapté qu'au cas bivarié, la suite consisterait à trouver une méthode d'adéquation des copules pour les grandes dimensions, afin de voir quelle copule correspondrait le mieux à nos données.

Annexes

A. Expression des coefficients de dépendance de queue

Soient X et Y deux variables aléatoires de fonctions de répartition respectives F_X et F_Y, et de fonction de répartition jointe F.

Le coefficient de dépendance de queue inférieure, par définition, s'écrit :

$$\begin{aligned}
\lambda_L &= \lim_{\alpha \to 1^-} \mathbb{P}(X \leq F_X^{-1}(\alpha) | Y \leq F_Y^{-1}(\alpha)) \\
&= \lim_{\alpha \to 1^-} \frac{\mathbb{P}(X \leq F_X^{-1}(\alpha), Y \leq F_Y^{-1}(\alpha))}{\mathbb{P}(Y \leq F_Y^{-1}(\alpha))} \\
&= \lim_{\alpha \to 1^-} \frac{F(F_X^{-1}(\alpha), F_Y^{-1}(\alpha))}{\alpha} \\
&= \lim_{\alpha \to 1^-} \frac{C\{F_X(F_X^{-1}(\alpha)), F_Y(F_Y^{-1}(\alpha))\}}{\alpha} \\
&= \lim_{\alpha \to 1^-} \frac{C(\alpha, \alpha)}{\alpha}
\end{aligned}$$

Le coefficient de dépendance de queue supérieure, par définition, s'écrit :

$$\begin{aligned}
\lambda_U &= \lim_{\alpha \to 0^+} \mathbb{P}(X > F_X^{-1}(\alpha) | Y > F_Y^{-1}(\alpha)) \\
&= \lim_{\alpha \to 0^+} \frac{\mathbb{P}(X > F_X^{-1}(\alpha), Y > F_Y^{-1}(\alpha))}{\mathbb{P}(Y > F_Y^{-1}(\alpha))} \\
&= \lim_{\alpha \to 0^+} \frac{1 - \mathbb{P}(X \leq F_X^{-1}(\alpha) \cup Y \leq F_Y^{-1}(\alpha))}{1 - \mathbb{P}(Y \leq F_Y^{-1}(\alpha))} \\
&= \lim_{\alpha \to 0^+} \frac{1 - \{\mathbb{P}(X \leq F_X^{-1}(\alpha)) + \mathbb{P}(Y \leq F_Y^{-1}(\alpha)) - \mathbb{P}(X \leq F_X^{-1}(\alpha), Y \leq F_Y^{-1}(\alpha))\}}{1 - \alpha} \\
&= \lim_{\alpha \to 0^+} \frac{1 - \{\alpha + \alpha - C(\alpha, \alpha)\}}{\alpha} \\
&= \lim_{\alpha \to 0^+} \frac{1 - 2\alpha + C(\alpha, \alpha)}{\alpha}
\end{aligned}$$

B. Expression du tau de Kendall pour la copule de Clayton

On part de la relation suivante :

$$K(\theta,t) = t + \sum_{i=1}^{d-1} \frac{(-1)^i}{i!}\{\varphi_\theta(t)\}^i f_i(\theta,t)$$

où

$$f_i(\theta,t) = \frac{d^i}{dx^i}\varphi_\theta^{-1}(x)\Big|_{x=\varphi_\theta(t)}$$

Pour la copule de Clayton, on voit facilement que :

$$f_i(\theta,t) = \frac{d^i}{dx^i}\varphi_\theta^{-1}(x)\Big|_{x=\varphi_\theta(t)} = (-1)^i q(\theta,i,1)t^{1+i\theta}$$

où :

$$q(\theta,d,m) = \prod_{i=0}^{d-1}(m+i\theta)$$

En effet, le générateur est donné par :

$$\varphi_\theta(t) = \frac{1}{\theta}\left(t^{-\theta}-1\right)$$

Son inverse est donc :

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

La dérivée première de cet inverse se calcule facilement :

$$\begin{aligned}[\varphi_\theta^{-1}(t)]' &= \left(-\frac{1}{\theta}\right)\theta(\theta t+1)^{-\frac{1}{\theta}-1}\\ &= -(\theta t+1)^{-\frac{1}{\theta}-1}\end{aligned}$$

Par récurrence, on montre que la dérivée n-ième s'écrit :

$$\varphi_\theta^{-1(n)}(t) = (-1)^n q(\theta,n,1)(\theta t+1)^{-\frac{1}{\theta}-n} \qquad (P_n)$$

Pour n=1, on a bien

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

Supposons alors (P_i) vrai pour tout $i \le n$. Montrons que (P_i) est vrai pour $i = n+1$

$$\begin{aligned}\varphi_\theta^{-1(n)}(t) &= (-1)^n q(\theta,n,1)\left(-\frac{1}{\theta}-n\right)\theta(\theta t+1)^{-\frac{1}{\theta}-n-1}\\ &= (-1)^{n+1} q(\theta,n,1)\left(\frac{1+\theta n}{\theta}\right)\theta(\theta t+1)^{-\frac{1}{\theta}-n-1}\\ &= (-1)^{n+1} q(\theta,n+1,1)(\theta t+1)^{-\frac{1}{\theta}-(n+1)}\end{aligned}$$

Donc (P_n) est vraie pour tout n.

Ainsi, on en déduit l'expression de $f_i(\theta,t)$:

$$\begin{aligned}f_i(\theta,t) &= (-1)^i q(\theta,i,1)\left[\theta\left(\frac{1}{\theta}(t^{-\theta}-1)\right)+1\right]^{-\frac{1}{\theta}-i}\\ &= (-1)^i q(\theta,i,1)\left[(t^{-\theta}-1)+1\right]^{-\frac{1}{\theta}-i}\end{aligned}$$

$$= (-1)^i q(\theta, i, 1) t^{1+\theta i}$$

On en déduit donc l'expression de $K(\theta, t)$:

$$K(\theta, t) = t + t \sum_{i=1}^{d-1} \left(\frac{1 - t^\theta}{\theta} \right)^i \frac{q(\theta, i, 1)}{i!}$$

Il faut l'intégrer pour obtenir le tau de Kendall. On part de la relation suivante :

$$\int_0^1 u^{x-1}(1-u)^{y-1} du = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$$

Soit le changement de variable : $u = t^\theta$

Donc $du = \theta t^{\theta-1} dt$

Alors l'intégrale s'écrit :

$$\int_0^1 u^{x-1}(1-u)^{y-1} du = \theta \int_0^1 t^{\theta(x-1)}(1-t^\theta)^{y-1} t^{\theta-1} dt$$

$$= \theta \int_0^1 t^{\theta x - 1}(1-t^\theta)^{y-1} dt$$

En prenant $x = \frac{2}{\theta}$ et $y = i + 1$, on trouve :

$$\theta \int_0^1 t\,(1-t^\theta)^i dt = \frac{\Gamma\left(\frac{2}{\theta}\right)\Gamma(i+1)}{\Gamma\left(\frac{2}{\theta} + i + 1\right)}$$

Soit :

$$\int_0^1 t\,(1-t^\theta)^i dt = \frac{\Gamma\left(\frac{2}{\theta}\right) \mathrm{i}!}{\theta\Gamma\left(\frac{2}{\theta} + i + 1\right)}$$

$$= \frac{\Gamma\left(\frac{2}{\theta}\right) i!}{\theta\left(\frac{2}{\theta} + i\right)\left(\frac{2}{\theta} + i - 1\right) \ldots \left(\frac{2}{\theta} + 1\right)\left(\frac{2}{\theta}\right)\Gamma\left(\frac{2}{\theta}\right)}$$

$$= \frac{i!}{\theta\left(\frac{2+\theta i}{\theta}\right)\left(\frac{2+\theta(i-1)}{\theta}\right) \ldots \left(\frac{2+\theta}{\theta}\right)\left(\frac{2}{\theta}\right)}$$

$$= \frac{i!}{\theta \frac{1}{\theta^{i+1}}(2+\theta i)\big(2+\theta(i-1)\big) \ldots (2+\theta)2}$$

$$= \frac{\theta^i i!}{q(\theta, i+1, 2)}$$

Donc

$$\int_0^1 K(\theta,t)dt = \frac{1}{2} + \sum_{i=1}^{d-1} \frac{q(\theta,i,1)}{q(\theta,i+1,2)}$$

Rappelons la version multivariée du tau de Kendall :

$$\begin{aligned} \tau(\theta) &= \left(\frac{2^d-1}{2^{d-1}-1}\right) - \left(\frac{2^d}{2^{d-1}-1}\right)\int_0^1 K(\theta,t)dt \\ &= \frac{2^d}{2^{d-1}-1}\left(1 - \int_0^1 K(\theta,t)dt\right) - \frac{1}{2^{d-1}-1} \end{aligned}$$

Des calculs que nous ne démontrerons pas ici montrent que :

$$1 - \int_0^1 K(\theta,t)dt = 1 - \left(\frac{1}{2} + \sum_{i=1}^{d-1} \frac{q(\theta,i,1)}{q(\theta,i+1,2)}\right) = \frac{q(\theta,d,1)}{q(\theta,d,2)}$$

On en déduit donc que :

$$\tau(\theta) = \frac{2^d}{2^{d-1}-1}\frac{q(\theta,d,1)}{q(\theta,d,2)} - \frac{1}{2^{d-1}-1}$$

C. Tau de Kendall empirique en dimension d

Le tau de Kendall multivarié s'écrit :

$$\tau(\theta) = \left(\frac{2^d - 1}{2^{d-1} - 1}\right) - \left(\frac{2^d}{2^{d-1} - 1}\right) \int_0^1 K(\theta, t) dt$$

Sa version empirique s'écrit :

$$\hat{\tau}(\theta) = \left(\frac{2^d - 1}{2^{d-1} - 1}\right) - \left(\frac{2^d}{2^{d-1} - 1}\right) \int_0^1 K_n(\theta, t) dt$$

avec :

$$K_n(t) = \frac{1}{n} \mathbb{I}_{\{V_{jn} \leq t\}}, \quad t \in [0,1]$$

où V_{jn} sont des pseudo-observations définies par :

$$V_{jn} = \frac{1}{n} \sum_{k=1}^{n} \mathbb{I}_{\{X_{1k} \leq X_{1j}, \ldots, X_{dk} \leq X_{dj}\}}$$

$$= \frac{1}{n} \sum_{k=1}^{n} \mathbb{I}_{\{R_{1k} \leq R_{1j}, \ldots, R_{dk} \leq R_{dj}\}},$$

R_{ij} étant le rang de X_{ij} parmi $X_{i1}, \ldots, X_{in}$

D. Méthode de Newton-Cotes

En analyse numérique, les méthodes de Newton-Cotes sont un ensemble de formules pour le calcul numérique d'une intégrale.
Soit $f:[a,b]\to\mathbb{R}$ une fonction continue. L'objectif est de calculer l'intégrale :

$$I(f)=\int_a^b f(x)dx$$

On partitionne l'intervalle $[a,b]$ en n intervalles $[x_i,x_{i+1}],\ i=0,\dots,n-1$ avec :

$$a=x_0<x_1<\cdots<x_n=b$$

On a alors :

$$I(f)=\int_a^b f(x)dx=\sum_{i=0}^{n-1}\int_{x_i}^{x_{i+1}} f(x)dx$$

On choisit les points $(x_i)_{i=0,\dots,n}$ équidistants :

$$x_i=x_0+hi$$

avec $h=\frac{b-a}{n}$.
Les méthodes consistent à remplacer un arc de courbe représentatif de f par une interpolation polynomiale de degré p au moyen des polynômes d'interpolation de Lagrange. Le cas $p=1$ est appelé méthode des trapèzes, et le cas $p=2$ représente la méthode de Simpson.

- Méthode des trapèzes

La méthode des trapèzes consiste à approximer la fonction entre deux abscisses successives par une droite.
$I(f)$ est donc approchée par :

$$I_h(f)=h\sum_{i=1}^{n-1}\frac{1}{2}(f(x_i)+f(x_{i+1}))$$

On peut majorer l'erreur d'approximation e_T à condition que f soit 2 fois dérivable. On pose :

$$M_T=\sup_{[a,b]}\left|f^{(2)}(x)\right|$$

Pour la méthode des trapèzes on a :

$$e_T\le\frac{(b-a)^3}{12n^2}M_T$$

- Méthode de Simpson

La méthode de Simpson consiste à remplacer la fonction par un polynôme de degré 2.
On approche $I(f)$ par :

$$I_h(f)=\frac{h}{6}\sum_{i=1}^{n}f(x_{i-1})+\frac{f(x_{i-1}+x_i)}{2}+f(x_i)$$

En supposant que f est 4 fois dérivable, si on pose :

$$M_S = \sup_{[a,b]} \left| f^{(4)}(x) \right|$$

alors l'erreur d'approximation e_S est majorée comme suit :

$$e_S = \frac{(b-a)^5}{2880n^4} M_S$$

Bibliographie

Abakarim, B. «Evaluation d'options sur plusieurs sous-jacents.» Mémoire, 2005.

Belguise, O. «Tempêtes : Etude des dépendances entre les branches Auto et Incendie avec la théorie des copulas.» Mémoire, 2001.

Bouye, E., V. Durrleman, A. Nikeghbali, G. Riboulet, et T. Roncalli. «Copulas for finance : a reading guide and some applications.» GRO, Crédit Lyonnais, Paris, Article, 2000.

Cadoux, D., et J.M. Loizeau. «Copules et dépendances : application pratique à la détermination du besoin en fonds propre d'un assureur non vie.», Article

Denuit, M., et A. Charpentier. «Mathématiques de l'assurance non-vie.» Vol. II : Principes fondamentaux de théorie du risque. Economica, 2004.

Durrleman, V., A. Nikeghbali, G. Riboulet, et T. Roncalli. «Copulas for finance: a reading guide and some applications.» Paris: GRO, Crédit Lyonnais, Article, 2000.

Frees, E.W., et E.A. Valdez. «Understanding relationship using copulas.» N. Amer. Actuarial, Article, 1998.

Joe, H. «Multivariate models and dependence concepts.» Chapman Hall, London, 1997.

Joe, H., et J.J. Xu. «The estimation method of inference functions for margins sor multivariate models.» Technical Report,Department of Statistics, University of British Columbia, 1996.

Marshall, A.W., et I. Olkin. «Families of multivaraite distributions.» J. Amer. Statist. Assoc., Article, 1988.

Planchet, F., P. Thérond, et J. Jacquemin. «Modèles financiers en assurance - Analyses de risques dynamiques.» Economica, 2005.

Quessy, J.F. «Méthodologie et application des copules: tests d'adéquation, tests d'indépendance, et bornes pour la valeur-à-risque.» , Thèse, 2005.

Savu, C., et M. Trede. «Goodness-of-fit tests for parametric families of Archimedian copulas.» University of Münster, 2004.

Schmidt, T. «Coping with Copulas.» Department of Mathematics, University of Leipzig, 2006.

Nelsen, R. B. « An introduction to copulas.» Springer, 2006