

Tempêtes : Etude des dépendances entre les branches Automobile et Incendie à l'aide de la théorie des copulas

Topic 1 Risk evaluation

Belguise Olivier
ACM
34 rue du Wacken
67000 Strasbourg France
tel : 33 (0)3 88 14 78 46
fax : 33 (0)3 88 14 96 20
belguiol@acm.fr

Charles Levi
Guy Carpenter
47/53 rue Raspail
92594 Levallois Perret Cedex France
tel : 33 (0)1 56 76 48 06
fax : 33 (0)1 41 27 93 03
charles.levi@guycarp.com

ABSTRACT

This paper presents a model of the dependence structure of storm impact on both “motor” and “fire” insurance using copulas.

The data are taken from a French Insurance Company.

The paper applies some new techniques, invented by Gary Venter, which allow to select the more appropriate copula among a set of copulas.

A few functions are introduced in order to highlight different properties of the various copulas. As these functions can also be approximated from the data, they are used to assess which copulas more closely capture features of the data.

Thanks to this, we are able to demonstrate that the dependence structure between “motor” and “fire” insurance can be described with the HRT copula (also called the Clayton “survival copula”). This copula has low correlation in the left tail but high correlation in the right, *ie* for large losses.

Firstly we estimate the copula parameter without any assumption on the parametric shape of the marginal distributions.

The same procedure as in the Klugman-Parsa paper “Fitting bivariate loss distributions with copulas” is also used .

Parameters of the bivariate distribution (two for each marginal and one for the copula) are also estimated with a MLE technique. Goodness of fit tests are performed.

The main difference with Klugman-Parsa paper is that our model is not a “ground-up model”. The marginals which are linked by the copula are only defined above a certain amount in each line of business. Thus we may achieve greater accuracy in the tail of the distribution.

Parametric estimation is used to derive some reinsurance premiums. Keeping the same HRT dependence structure, we change the dependence coefficient to evaluate its impact on the premium of different reinsurance contracts.

Keywords: copulas, windstorms, reinsurance

INTRODUCTION

La fin du 20^{ème} siècle a été marquée par de nombreux événements climatiques de toute première importance.

Les tempêtes de 1990, dont tout le monde s'accordait à l'époque à dire qu'il s'agissait des tempêtes du siècle, en sont un premier exemple. Que dire alors de «Lothar» et de «Martin» survenus fin 1999 et qui touchèrent la France de plein fouet. «Tempête du millénaire» ou «événement exceptionnel» furent les qualificatifs les plus souvent employés.

Toujours est-il que ces événements ont amené assureurs, et surtout réassureurs, à se pencher de manière plus approfondie sur ce type de phénomène.

Même s'il est vrai que lorsque surviennent des tempêtes, les bâtiments sont les principaux biens assurés endommagés, l'assurance automobile constitue également une part non négligeable du montant total imputable aux tempêtes.

Afin de coter certains contrats de réassurance prenant en compte l'impact des tempêtes sur les branches «Automobile» et «Incendie-Tempête-Grêle-Neige», une estimation de la distribution conjointe de ces deux types de dommages s'avère nécessaire voire impérative. Il est évident que nous ne pouvons supposer l'indépendance entre ces deux branches sous peine d'aboutir à des résultats erronés.

Jusqu'à présent la plupart des techniques utilisées présentaient certains inconvénients.

Nous pouvons, à titre d'exemple, citer l'utilisation de distributions bidimensionnelles comme la Pareto bivariée. Mais dans ce cas précis, les lois marginales seront nécessairement des lois de Pareto. L'idéal serait donc de pouvoir créer des distributions multivariées et a fortiori bivariées en choisissant nous-mêmes nos lois marginales.

La théorie des copulas va nous permettre de répondre à toutes ces attentes.

Le mot copula, dont la signification en latin est littéralement *lien*, a été employé pour la première fois en 1959 par Abe Sklar dont le théorème constitue la clef de voûte de toute la théorie.

De manière générale les copulas sont des fonctions qui vont constituer un véritable *lien* entre la distribution multivariée et les lois marginales.

La première partie de cet article présentera la théorie des copulas dans ses grandes lignes, notamment le théorème de Sklar, clarifiera la notion de dépendance. Pour des informations plus précises au sujet de ce concept, le lecteur pourra se reporter à l'article de Embrechts-McNeil-Strauman (1999).

Cette partie mettra également en avant les copulas utilisées au cours de notre étude, et donnera toutes les formules nécessaires à l'estimation de paramètres par la méthode du maximum de vraisemblance.

D'autre part, des techniques nous permettant de choisir la copula susceptible de correspondre au mieux à une certaine série de données y seront également exposées.

Dans la seconde partie, nous utiliserons les copulas dans le but d'évaluer une distribution bivariée des montants « automobile » et « incendie ».

Les données qui serviront de base à notre étude nous ont été fournies par une grande compagnie d'assurance française. La base de données utilisée comprend tous les sinistres dus aux tempêtes sur une période de 11 ans, à compter du 1er janvier 1990 jusqu'au 31 décembre 2000.

Nous déterminerons dans un premier temps la ou les copulas que nous allons juger susceptibles de convenir grâce aux techniques exposées dans la première partie. Nous déterminerons ensuite la ou les lois marginales permettant la meilleure adéquation possible dans les branches « Automobile » et « Incendie ».

Enfin, nous estimerons les paramètres de notre distribution bivariée avec les lois marginales et les copulas retenues. Ces estimations seront complétées par les tests usuels pour voir si notre adéquation est acceptable.

L'une des finalités de cette adéquation à une loi bivariée est l'évaluation de contrats de réassurance par la méthode de Monte-Carlo.

Les outils utilisés pour mener à bien cette étude ont été les logiciels SAS (modules Base, Stat) et Excel.

1 Théorie des copulas

L'objet de cette partie est de donner des bases théoriques sommaires mais néanmoins nécessaires à la bonne compréhension des copulas.

Nous présenterons tout d'abord le concept de copula et nous donnerons quelques exemples de celles-ci qui nous serviront par la suite. Nous aborderons enfin la notion de dépendance, et nous présenterons des mesures de dépendance qui peuvent être exprimées à partir des copulas. Nous terminerons en donnant quelques exemples de fonctions permettant d'affiner le choix d'une copula pour une certaine série de données.

1.1 Définition d'une copula et théorème de Sklar

1.1.1 Définition d'une copula

Une copula est définie comme étant une fonction de répartition multivariée ayant des lois marginales uniformes sur $[0,1]$.

Nous nous concentrerons dans cet article sur le cas bidimensionnel, même si ces définitions peuvent être étendues à un nombre de dimensions plus élevé.

$C(u,v)=P(U\leq u, V\leq v)$ est donc une fonction définie de $[0,1]^2$ vers $[0,1]$ qui vérifie les trois propriétés suivantes:

- 1 $C(u,v)$ est une fonction croissante pour chacune de ses composantes u et v
- 2 $C(u,1)=u$ et $C(1,v)=v$
- 3 Pour tout $a_1\leq a_2$ et $b_1\leq b_2$, nous avons
 $C(a_2,b_2)-C(a_1,b_2)-C(a_2,b_1)+C(a_1,b_1)\geq 0$

1.1.2 Théorème de Sklar

1.1.2.1 Soit F une fonction de distribution en 2 dimensions ayant des marginales F_X, F_Y

Alors il existe une copula C telle que pour tout $(x,y) \in [-\infty, \infty]^2$

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

De plus, si les lois marginales sont continues, alors cette copula est unique.

Les copulas permettent donc la création de distributions bivariées ayant des lois marginales définies.

Si C est une copula et F_X et F_Y des lois marginales alors $C(F_X(x), F_Y(y))$ est une distribution bivariée.

La copula pourrait donc se comprendre comme étant la fonction de répartition bivariée des variables aléatoires $F_X(X)$ et $F_Y(Y)$, qui sont, est il besoin de le rappeler, uniformément réparties.

1.1.2.2 Le corollaire de ce théorème s'énonce comme suit :

Définissons F, C, F_X, F_Y comme précédemment et notons

$F_X^{(-1)}, F_Y^{(-1)}$ les fonctions quantiles respectives de F_X et F_Y

Alors, pour tout vecteur $(u,v) \in [0,1]^2$

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

1.2 Notion de Survival Copula

La notation $S(x)$ est très souvent employée pour décrire la fonction de survie $P(X > x)$.

La fonction de survie conjointe $S(x,y) = P(X > x, Y > y)$ n'est pas $1 - F(x,y)$ comme certains auraient pu le penser (en fait il s'agit de la probabilité $P(X > x$ ou $Y > y$)) mais

$$S(x,y) = 1 - F_X(x) - F_Y(y) + F(x,y)$$

Par analogie, pour une copula nous savons que $C(u,v) = P(U < u, V < v)$, la fonction de survie est $C_s(u,v) = P(U > u, V > v) = 1 - u - v + C(u,v)$.

Comme $C(F_X(x), F_Y(y)) = F(x,y)$, nous obtenons $C_s(F_X(x), F_Y(y)) = S(x,y)$.

C_s n'est pas une copula dans la mesure où elle n'est pas nulle au point $(0,0)$, mais que en revanche elle est égale à zéro en $(1,1)$.

$$\begin{aligned} \text{Nous définissons alors } C_f(u,v) &= C_s(1-u, 1-v) = 1 - (1-u) - (1-v) + C(1-u, 1-v) \\ &= u + v - 1 + C(1-u, 1-v) \end{aligned}$$

Ainsi $C_f(S_x(x), S_y(y)) = C_s(F_X(x), F_Y(y)) = S(x,y)$.

La fonction C_f est, quant à elle, une Copula. Nous l'appellerons donc "flipped" copula (on peut également la retrouver sous le nom de "survival copula" dans la littérature), et, en appliquant celle-ci aux fonctions de survie marginales, nous obtenons alors la distribution de survie conjointe.

Néanmoins, les "flipped copula" peuvent être appliquées aux fonctions de distributions marginales, et avoir ainsi des propriétés inverses à la copula originale.

Un exemple sera présenté dans la suite de cet article.

1.3 Densité des copulas

Nous savons que si elle existe, la densité f d'une fonction de distribution F est définie comme suit:

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}$$

L'expression de la densité de la copula que nous noterons c ou C_{12} s'exprime donc comme suit

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

La densité f de F s'exprime alors

$$f(x,y) = c(F_x(x), F_y(y)) \times f_x(x) \times f_y(y)$$

Cette formule prendra toute son importance lorsque nous effectuerons des estimations paramétriques par la méthode du maximum de vraisemblance.

1.4 Distribution conditionnelle et Copulas

Les distributions conditionnelles peuvent également s'exprimer à l'aide des copulas.

Notons $C_1(u,v)$ la dérivée de $C(u,v)$ par rapport à u .

$$C_1(u,v) = \frac{\partial C(u,v)}{\partial u}$$

Si la distribution jointe de X et Y est $F(x,y) = C(F_x(x), F_y(y))$ alors la distribution conditionnelle de $Y | X=x$ est donc $F_{Y|X}(y) = C_1(F_x(x), F_y(y))$.

Nous verrons ultérieurement que si C_1 est inversible, alors la simulation de probabilités jointes pourra être faite en utilisant cette propriété.

Cette formule revêt donc également une importance capitale.

1.5 Le coefficient de corrélation de Kendall et le concept de Tail dependence

Lorsque l'on évoque la dépendance, le premier mot qui vient généralement à l'esprit est celui de coefficient de corrélation linéaire.

Ce coefficient de corrélation linéaire, également connu sous le nom de coefficient de corrélation de Pearson, est en effet le plus couramment utilisé.

Ce coefficient est tout à fait approprié lorsque nous étudions des distributions normales ou de Student multivariées, mais celui-ci perd de son intérêt si le modèle est différent.

Malheureusement, les distributions dans le domaine de l'assurance suivent très rarement de telles lois.

Nous présenterons ici un autre coefficient de corrélation, à savoir le coefficient de Kendall qui présente l'avantage de demeurer inchangé sous l'hypothèse d'une transformation strictement croissante des variables aléatoires.

Une mesure de dépendance doit en fait nous permettre de nous faire une idée de la structure de dépendance entre deux variables aléatoires, et ceci exprimé à l'aide d'un seul nombre.

Nous examinerons également les notions de "tail dependence" ou de dépendance de queue, très importantes dans l'étude des dépendances entre les valeurs extrêmes et directement associés aux copulas.

1.5.1 Le coefficient de corrélation de Kendall

Soit (X,Y) un vecteur aléatoire et notons (X',Y') un vecteur en tout point identique.

Le τ de Kendall est tout simplement la probabilité de concordance moins celle de discordance, à savoir

$$\tau(X,Y) = P((X-X')(Y-Y') > 0) - P((X-X')(Y-Y') < 0)$$

Ce coefficient ne dépend donc que du rang de chaque observation, c'est pourquoi il est invariant par transformation croissante des variables aléatoires, que celle-ci soit linéaire ou pas.

Si (X, Y) a pour copula C , alors ce coefficient se détermine uniquement en fonction de la copula par :

$$\tau(X, Y) = 4 \int_0^1 \int_0^1 C(u, v) c(u, v) du dv - 1 = 4E[C(U, V)] - 1$$

1.5.2 Le concept de Tail dépendence.

Cette notion est très importante dans l'étude de la dépendance asymptotique entre deux variables aléatoires. Cela va nous permettre de voir le niveau de dépendance dans les valeurs extrêmes (upper tail dependence) et dans les valeurs petites (lower tail dependence). Nous n'aborderons ici que le premier cas de figure.

L'objet est donc l'étude de la dépendance dans la queue commune de la distribution bivariée. Prenons donc deux variables aléatoires continues X et Y ayant pour fonction de distribution respectives F_x et F_y .

Le coefficient d' "upper tail dependence" de X et Y est défini par

$$\lambda = \lim_{u \rightarrow 1^-} P\left(Y > F_Y^{-1}(u) \mid X > F_X^{-1}(u)\right) = \lim_{u \rightarrow 1^-} \frac{C(u, u) - 2u + 1}{1 - u}$$

si toutefois cette limite $\lambda \in [0, 1]$ existe.

La quantité λ est une fonction de la copula et est donc invariante par transformation croissante.

Si $\lambda \in] 0, 1]$, il existe alors une dépendance asymptotique.

Si $\lambda = 0$, on dit qu'il y a indépendance asymptotique.

A titre d'exemple, pour les copulas de Gumbel ou HRT que nous présenterons dans les pages qui suivent, le coefficient d'upper tail dependence sera respectivement de $2-2^{1/a}$ et $1/2^a$.

1.6 Exemples de copulas

Les principales propriétés des copulas ayant été définies, nous pouvons désormais en présenter quelques exemples.

L'échantillon proposé ne sera bien entendu pas exhaustif, mais nous tâcherons de présenter les Copulas les plus célèbres et celles qui présentent un intérêt dans l'étude des distributions de montant de sinistres.

Si l'on fait abstraction des copulas triviales comme la copula d'indépendance ($C(u, v) = uv$) la copula de dépendance totale positive ($C(u, v) = \min(u, v)$) ou négative ($C(u, v) = \max(u + v - 1, 0)$), nous pouvons citer :

1.6.1 La copula Normale

Cette copula fait partie de la famille des copulas elliptiques (de même que la Copula de Student que nous ne présenterons pas). La copula d'une distribution normale bivariée (avec ρ comme coefficient de corrélation) est définie comme suit:

$$C_\rho^{\text{Ga}}(u, v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{t_1^2 + t_2^2 - 2\rho t_1 t_2}{2(1-\rho^2)}\right\} dt_1 dt_2$$

avec, $-1 < \rho < 1$ et Φ la fonction de distribution de la loi normale.

La densité s'écrit alors

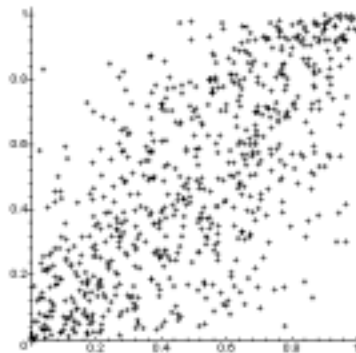
$$c_{\rho}^{Ga}(u,v) = \frac{1}{\sqrt{1-\rho^2}} \exp\left(\frac{1}{2} \left(\frac{\rho^2 \Phi^{-1}(u)^2 + \rho^2 \Phi^{-1}(v)^2 - 2\rho \Phi^{-1}(u)\Phi^{-1}(v)}{1-\rho^2} \right)\right)$$

Il est très facile de simuler une paire de variable aléatoire (U,V) ayant pour distribution jointe la Copula Normale.

La première étape consiste à simuler une paire de variable aléatoire (X,Y) ayant pour distribution la loi normale bvariée N(0,R). R étant la matrice de corrélation ayant ρ comme élément non diagonal, et 1 sur la diagonale.

En prenant ensuite $U=\Phi(X)$ et $V=\Phi(Y)$, alors (U,V) aura pour distribution la copula Normale. Le coefficient de corrélation de Kendall est

$$\tau=(2/\pi)\arcsin(\rho).$$



Simulation de 800 paires ayant pour distribution la copula $C_{0.6}^{Ga}$ et des lois marginales uniformes

1.6.2 Les copulas archimédiennes

Il s'agit d'une famille de copulas assez particulière, puisque les éléments de cette famille sont générés à partir d'une fonction φ que l'on appelle également générateur de la copula.

Cette fonction φ est continue, strictement décroissante de $[0,1]$ vers $[0,\infty]$, telle que $\varphi(1)=0$ et convexe.

Nous obtenons une copula en prenant

$$C_{\varphi}(u,v) = \begin{cases} \varphi^{-1}(\varphi(u)+\varphi(v)) & \text{si } \varphi(u)+\varphi(v) \leq \varphi(0) \\ 0 & \text{sinon} \end{cases}$$

En fait, ce n'est que si

$$\lim_{t \rightarrow 0^+} \varphi(t) = +\infty$$

que $C_{\varphi}(u,v) > 0$ (sauf lorsque u ou v est égal à 0).

Nous en présenterons ici trois d'entre elles, à savoir la copula de Frank, la copula de Gumbel et celle de Clayton.

1.6.2.1 La copula de Frank

Prenons

$$\varphi(t) = -\ln\left(\frac{e^{-at} - 1}{e^{-a} - 1}\right)$$

avec $a \neq 0$.

Nous obtenons ainsi la copula de Frank qui est donc définie comme suit:

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

Les cas limites sont donc les suivants:

$$\lim_{a \rightarrow -\infty} C_a = C^l, \lim_{a \rightarrow 0} C_a = C^{\text{ind}}, \lim_{a \rightarrow \infty} C_a = C^u$$

En définissant $g_z = e^{-az} - 1$ nous obtenons:

$$C_1(u, v) = \frac{\partial C(u, v)}{\partial u} = \frac{g_u g_v + g_v}{g_u g_v + g_1}$$

et finalement

$$c(u, v) = -a g_1 \left(\frac{1 + g_{u+v}}{(g_u g_v + g_1)^2} \right)$$

Nous rappelons à nouveau que le calcul de la densité est très important dans la mesure où celle-ci va être utilisée ultérieurement pour des estimations de paramètres par la méthode du maximum de vraisemblance.

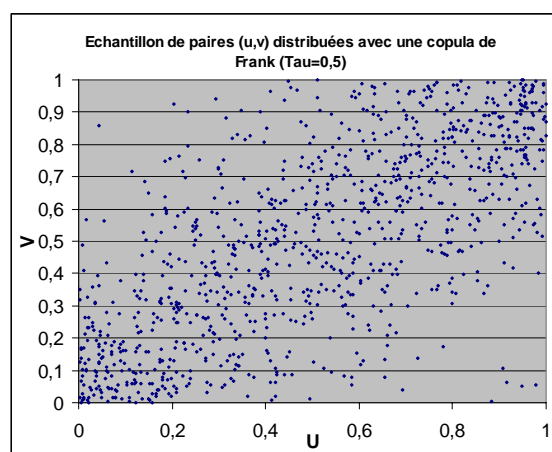
La copula de Frank étant la seule copula Archimédienne à respecter l'équation $C(u, v) = C_1(u, v)$ (cf 1.3), celle-ci risque de ne pas être toujours appropriée à des applications dans le domaine de l'assurance.

Pour simuler des paires (u, v) ayant comme distribution la Copula de Frank, nous pourrions utiliser la distribution conditionnelle.

La démarche à suivre consiste donc à simuler indépendamment u et p toutes deux issues de la loi uniforme sur $[0, 1]$. Mais ici p sera compris comme une réalisation de la distribution conditionnelle de $V|u$. Comme cette distribution n'est plus ni moins que C_1 , v sera déterminé en résolvant l'équation $v = C_1^{-1}(p|u)$.

C_1 étant inversible, v se détermine alors ainsi

$$v = -\frac{1}{a} \ln \left(1 + \frac{p g_1}{1 + g_u (1 - p)} \right)$$



Une fois u et v simulés, nous pouvons simuler des réalisations des variables X et Y en inversant tout simplement les distributions marginales, c'est à dire en prenant $x = F_X^{-1}(u)$ et $y = F_Y^{-1}(v)$.

Le coefficient de corrélation de Kendall que nous définirons dans la partie suivante, se définit, pour la copula de Frank de la manière suivante:

$$\tau(a) = 1 - \frac{4}{a} + \frac{4}{a^2} \int_0^a \frac{t}{e^t - 1} dt$$

La dernière partie de l'expression est également connue sous le nom de fonction de Debye.

1.6.2.2 La copula de Gumbel

Prenons comme générateur

$$\varphi(t) = (-\ln(t))^a$$

avec $a > 1$. Cette fonction satisfaisant à toutes les conditions du théorème sur les copulas archimédiennes, nous pouvons générer la copula de Gumbel en prenant:

$$C_a(u, v) = \exp\left(-\left[(-\ln(u))^a + (-\ln(v))^a\right]^{1/a}\right)$$

Les cas limites suivant la valeur du paramètre sont donc:

$$\lim_{a \rightarrow \infty} C_a = C^u, C_{a=1} = C^{\text{ind}}$$

La dérivée par rapport à la composante u s'écrit:

$$C_1(u, v) = C(u, v) [(-\ln u)^a + (-\ln v)^a]^{-1+1/a} (-\ln u)^{a-1}/u$$

Et la densité

$$c(u, v) = C(u, v) u^{-1} v^{-1} [(-\ln u)^a + (-\ln v)^a]^{-2+2/a} [(\ln u) (\ln v)]^{a-1} \{1 + (a-1)[(-\ln u)^a + (-\ln v)^a]^{-1/a}\}$$

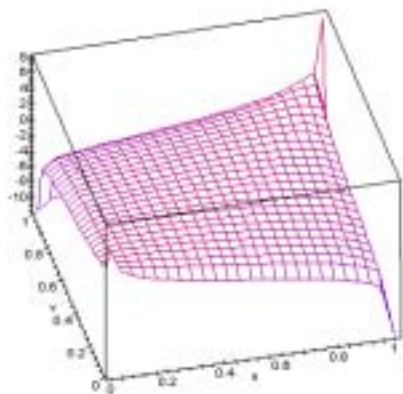
et $\tau(a) = 1 - 1/a$

En revanche, la distribution conditionnelle C_1 n'est pas inversible. Il existe néanmoins une technique de simulation propre aux copulas archimédiennes.

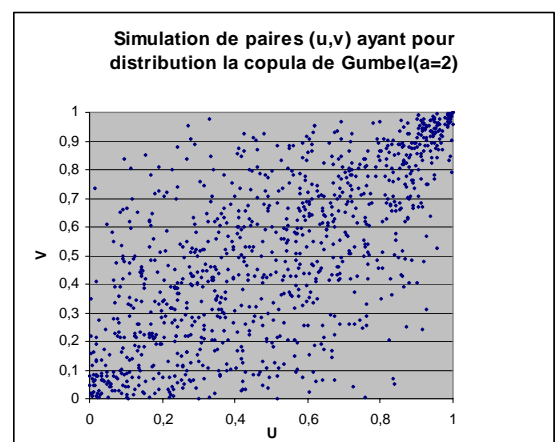
Définissons tout d'abord pour tout $t \in [0, 1]$ la fonction:

$$K(t) = t - \frac{\varphi(t)}{\varphi'(t^+)}$$

Supposons qu'un vecteur aléatoire (U, V) ait pour fonction de répartition une Copula archimédienne, alors cette fonction K sera tout simplement la fonction de répartition de la variable aléatoire $C(U, V)$.



La densité de la copula de Gumbel ($a=2$) sur une échelle logarithmique



Dans le cas de la Copula de Gumbel, cette fonction sera donc:

$$K(t) = t \left(1 - \frac{1}{a} \ln(t)\right)$$

La procédure à suivre pour simuler des paires (u,v) est donc

- Simuler indépendamment s et q suivant la loi uniforme U(0,1)
- Déterminer $t = K_c^{-1}(q)$
- Prendre $u = \varphi^{[-1]}(s \varphi(t))$ et $v = \varphi^{[-1]}((1-s) \varphi(t))$

La détermination de t ne peut se faire que par résolution numérique de l'équation, et donc par une méthode itérative.

1.6.2.3 La copula de Clayton

Prenons

$$\varphi(t) = a \left(t^{-1/a} - 1 \right)$$

avec $a > 0$.

Nous obtenons ainsi la copula de Clayton

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

Les cas limites suivant la valeur du paramètre sont donc:

$$\lim_{a \rightarrow \infty} = C^{\text{ind}}, \lim_{a \rightarrow 0} = C^u$$

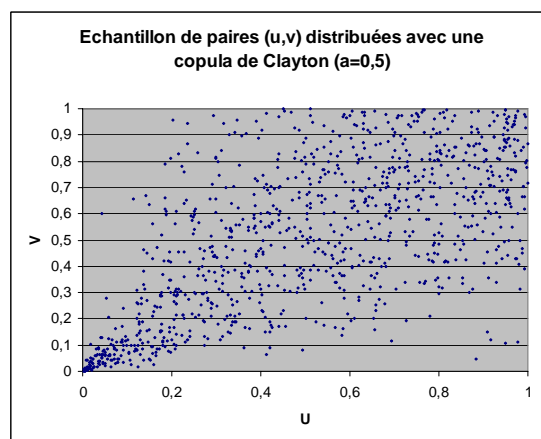
La dérivée première par rapport à u s'écrit:

$$C_1(u, v) = u^{-1-1/a} \left[u^{-1/a} + v^{-1/a} - 1 \right]^{a-1}$$

et ainsi la densité s'exprime:

$$c(u, v) = \left(1 + \frac{1}{a}\right) (uv)^{-1-1/a} \left(u^{-1/a} + v^{-1/a} - 1 \right)^{-a-2}$$

et $\tau(a) = 1/(2a+1)$.



La technique de simulation utilisée est la même que pour la copula de Frank, sauf que dans ce cas là $v = \left[(pu^{1+1/a})^{1/(-a-1)} - u^{-1/a} + 1 \right]^{-a}$.

Il est visible sur cet échantillon qu'il y a une importante concentration de points près de (0,0). Cette copula aurait donc tendance à "corrélér" entre eux les petits sinistres et pas du tout les gros. Cela va à l'encontre de ce que nous recherchons lorsque nous étudions des distributions de montants de sinistres.

L'idée serait donc de définir une copula présentant des propriétés opposées. Nous avons défini dans les pages précédentes la notion de "survival Copula". Nous allons donc appliquer cette formule pour obtenir la copula HRT "heavy right tail".

1.6.3 La copula HRT

Cette copula est donc définie comme étant la "survival copula" de la copula de Clayton. Elle s'écrit donc:

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

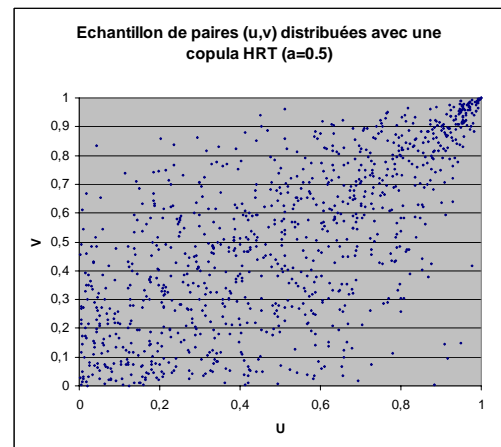
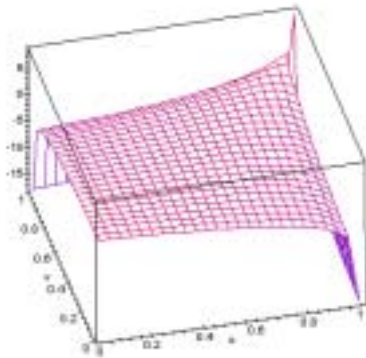
et par suite

$$C_1(u, v) = 1 - \left[(1-u)^{-1/a} + (1-v)^{-1/a} - 1 \right]^{-a-1} (1-u)^{-1-1/a}$$

la densité étant donc

$$c(u, v) = \left(1 + \frac{1}{a}\right) \left[(1-u)^{-1/a} + (1-v)^{-1/a} - 1 \right]^{-a-2} \left[(1-u)(1-v) \right]^{-1-1/a}$$

et $\tau(a) = 1/(2a+1)$, comme pour la copula de Clayton.



La densité de la copula HRT sur une échelle logarithmique (a=0.5)

Comme pour la plupart des autres copulas présentées jusqu'à présent, la distribution conditionnelle C_1 est inversible.

Dans ce cas précis, v est déterminé à partir de u et p en prenant :

$$v = 1 - \left[(1 - (1-u)^{-1/a} + [(1-p)(1-u)^{1+1/a}]^{-1/(1+a)})^{-a} \right]$$

La copula HRT ayant été conçue dans cette optique, la concentration de points près de (1,1) est la plus importante. Avec cette structure de dépendance, les gros sinistres auront tendance à survenir ensemble.

Bien qu'il s'agisse d'une Survival Copula, elle peut être utilisée comme une copula à part entière (Venter 2001).

Comme le montrent la simulation effectuée et le graphique représentant la densité de cette copula, celle-ci présente une densité de probabilité plus élevée aux abords des points (0,0) et surtout (1,1).

Cette copula pourrait donc éventuellement convenir dans le cadre d'études de distributions de montants de sinistres (avec dépendance pour les montants importants).

1.7 Le choix de la bonne copula

Au vu de toute la théorie exposée précédemment, nous sommes en droit de nous demander quelle copula pourrait le mieux correspondre à une certaine série de données.

Deux copulas ayant le même coefficient de corrélation de Kendall peuvent en effet avoir des comportements tout à fait différents. Il suffit pour s'en convaincre de jeter un bref coup d'œil aux graphiques des simulations présentés dans la partie précédente.

Nous allons donc dans cette partie donner quelques méthodes nous permettant de faire le tri parmi les copulas afin de ne choisir que celles qui pourraient présenter des caractéristiques semblables à celles de notre série de données.

En fait, nous allons exposer plusieurs fonctions, dont certaines ont été exposées par Gary Venter (2001) et qui auront des caractéristiques tout à fait différentes selon la copula choisie.

Ces fonctions pouvant également être établies de façon empirique, uniquement à partir du rang de chaque observation, une simple comparaison graphique nous permettra de ne retenir qu'une ou deux familles de copulas pour la poursuite de notre étude.

1.7.1 La fonction $K(z)$

Cette fonction, qui a déjà été évoquée dans notre présentation de la copula de Gumbel, n'est ni plus ni moins que la fonction de répartition de la variable aléatoire $C(U,V)$.

Il a été démontré que pour une copula de type archimédienne, cette fonction se définissait comme suit :

$$K(z) = z \frac{\varphi(z)}{\Phi(z)}$$

Dans le cadre des copulas archimédiennes présentées auparavant cette fonction $K(z)$ est donc la suivante

Gumbel

$$K_a(z) = z \left(1 - \frac{1}{a} \ln(z)\right)$$

Frank

$$K_a(z) = z + \frac{1}{a} \ln \left[\frac{1 - e^{-az}}{1 - e^{-a}} \right]$$

Clayton

$$K_a(z) = z + az \left(1 - z^{1/a}\right)$$

Supposons maintenant que nous disposions d'un échantillon d'observations $(x_1, y_1) \dots (x_n, y_n)$ issu d'un vecteur aléatoire (X, Y) .

Pour établir un estimateur non-paramétrique de la fonction K à partir de cet échantillon, la procédure à suivre consiste à :

1 Définir la pseudo-observation z_i pour chaque $i=1 \dots n$

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

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

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

Cette estimation non-paramétrique de K pourra ensuite être comparée graphiquement aux versions paramétriques de K pour les différentes copulas archimédiennes.

Le paramètre « a » de la copula pourra être établi, par exemple à partir du coefficient de Kendall empirique de l'échantillon.

En effet nous avons vu qu'il existe une *relation directe* entre le coefficient de Kendall et le paramètre de la copula. Il suffit donc simplement de résoudre une équation pour déterminer le paramètre de la copula, même si cette opération peut se révéler un peu plus délicate dans le cas de la copula de Frank par exemple.

Une autre méthode pour déterminer ce paramètre pourra être celle du maximum de vraisemblance que nous utiliserons plus loin.

1.7.2 La fonction J(z) ou de Tau cumulatif

Cette fonction est basée sur le coefficient du τ de Kendall dont nous rappelons ici la formule :

$$\tau = -1 + 4 \int_0^1 \int_0^1 C(u, v) c(u, v) du dv$$

L'idée a ensuite consisté à définir une fonction de τ cumulatif en prenant :

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

Il est clair que $J(1) = \tau$.

On constate des différences notables dans le comportement de la fonction $J(z)$ pour les différentes copulas, la fonction $J(z)$ peut démarrer avec une faible corrélation (HRT) ou une corrélation plus importante (Gumbel). Nous n'avons pas représenté la copula de Clayton dans la mesure où $J(z)$ est représentée très rapidement par une ligne horizontale dont la valeur est bien entendu égale à τ . Ceci s'explique aisément dans la mesure où la copula de Clayton possède un coefficient de « right tail dependence » et donc la « dépendance est surtout concentrée près du point (0,0) ».

L'intérêt de cette fonction réside là encore dans le fait qu'une version empirique de celle-ci puisse être établie.

Nous allons, comme pour la fonction $K(z)$, donner la procédure pour établir la fonction $J(z)$ de manière empirique. Considérons le même échantillon d'observations $(x_1, y_1) \dots (x_n, y_n)$. Notons rang x_i le rang de la i -ème observation dans l'ensemble des valeurs x_1, \dots, x_n . La procédure à suivre consiste à

1 Transformer les données (x_i, y_i) en (u_i, v_i) en prenant $(u_i, v_i) = (\{\text{rang } x_i\}/(n+1), \{\text{rang } y_i\}/(n+1))$

2 Définir $C(z, z)$ en prenant

$$C(z, z) = \frac{\text{nombre de paires } \{u_j, v_j\} \text{ telles que } u_j < z \text{ et } v_j < z}{n}$$

ce qui revient à prendre la fonction de répartition empirique des données transformées en (1).

3 Définir la pseudo-observation z_i pour chaque $i=1 \dots n$ comme précédemment

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

4 Définir ensuite la quantité

$$I(z) = \frac{1}{n} \sum_{i=1}^n (z_i \times 1\{u_i < z \text{ et } v_i < z\})$$

5 Nous définissons ensuite $J(z)$ par

$$J(z) = -1 + \frac{4I(z)}{C(z, z)^2}$$

Il ne reste plus qu'à comparer la version « empirique » ainsi obtenue à la version « paramétrique » de $J(z)$ pour différentes copulas.

1.7.3 La fonction $M(z)$

Posons

$$M(z) = E(V|U < z) = \frac{\int_{u=0}^z \int_{v=0}^1 vc(u, v) dudv}{z}$$

Comme $E[V]=1/2$, quelle que soit la copula choisie, la fonction M vérifiera $M(1)=1/2$.

La différence entre les copulas se fera donc au niveau des petites valeurs de z et à l'allure de la courbe au fur et à mesure que l'on approche du point $z=1$.

La procédure à suivre pour calculer cette fonction à partir des données $(x_1, y_1) \dots (x_n, y_n)$ consiste à :

1 Définir $D(z)$ en prenant

$$D(z) = \sum_{i=1}^n 1\{u_i < z\}$$

2 Définir $N(z)$ suivant la formule

$$N(z) = \sum_{i=1}^n 1\{u_i < z\} \times v_i$$

3 Soit $M(z) = N(z)/D(z)$.

Il est évident $M(1)=1/2$ puisque

$$N(1) = \frac{1}{n+1} \sum_{i=1}^n i = \frac{1}{n+1} \times \frac{n(n+1)}{2} = \frac{n}{2}$$

Et comme $D(1)=n$, on retrouve bien $M(1)=1/2$.

1.7.4 Les fonctions $L(z)$ et $R(z)$ ou “Tail concentration functions”

Nous nous sommes déjà penchés précédemment sur les concepts de “upper tail” et “lower tail dependence”. Les fonctions que nous allons présenter maintenant sont totalement basées sur ces concepts.

Pour tout z appartenant à $(0,1)$, nous définissons

$$L(z)=P(U<z,V<z)/z^2 \text{ et } R(z)=P(U>z,V>z)/(1-z)^2$$

Ces fonctions mettent donc en valeur la “concentration de probabilité” aux abords des points $(0,0)$ et $(1,1)$. Celles-ci peuvent donc se traduire très facilement en termes de copulas, puisque:

$$L(z)=C(z,z)/z^2 \text{ et } R(z)=(1-2z+C(z,z))/(1-z)^2.$$

Nous avons expliqué précédemment (lors de la présentation de la fonction de tau cumulatif) comment calculer $C(z,z)$ à partir des données. Etablir la version non-paramétrique de ces fonctions ne pose donc a priori aucun problème.

Les graphiques de ces fonctions sont présentés ci-dessous pour différentes valeurs du tau de Kendall.

Nous savons que $R(0)=L(1)=1$, donc ces fonctions pourront très facilement être différenciées l’une de l’autre.

Une constatation s’impose d’emblée au vu de ces graphiques: plus le τ de Kendall est important, plus les valeurs de ces fonctions sont élevées aux abords des points dits intéressants (à savoir 0 pour $L(z)$ et 1 pour $R(z)$)

Comme nous pouvions nous y attendre, les graphiques des copulas HRT et Clayton sont parfaitement symétriques.

La copula de Gumbel présente une forte concentration dans les deux queues, même si celle de droite est de loin la plus épaisse. Au fur et à mesure que la corrélation diminue, l’importance relative de la queue à gauche décroît par rapport à celle de droite.

La copula HRT présente une concentration à droite aussi importante que la copula de Gumbel, mais à gauche celle-ci est quasiment nulle.

La copula de Frank est, quant à elle, totalement symétrique.

Nous venons donc de passer en revue les différentes fonctions qui serviront de préliminaire à notre étude. Nous sommes maintenant en mesure d’étudier le phénomène des tempêtes en utilisant ces copulas.

2 Adéquation à une loi bivariée grâce aux copulas

2.1 Présentation des données utilisées

Nos données sont issues des fichiers « sinistres tempêtes individuels » d’une grande compagnie d’assurance française ventilées entre les branches « Incendie » et « Automobile ».

Notre historique se base sur la période du 1^{er} janvier 1990 au 31 décembre 2000, soit une période de 11 ans et comprend des sinistres majeurs comme la tempête du 3 février 1990 et les deux tempêtes de fin décembre 1999 (Lothar et Martin).

Nous avons ensuite procédé à un regroupement des sinistres individuels par période de trois jours conformément à la clause horaire en vigueur en matière de réassurance.

Nous avons enfin réalisé une actualisation de ces montants de manière à les homogénéiser. Nous avons pour ce faire tenu compte, entre autres, de l'inflation, de l'évolution du nombre de polices du portefeuille et des variations des montants des franchises.

Un indice d'inflation propre à chaque branche a été utilisé, à savoir l'indice FFB pour la branche « Incendie » et un indice du coût de réparation automobile pour la seconde branche.

Nous avons, sur la base de travaux internes de C.Levi reconstitué le montant des événements à niveau de franchise égal. En effet une diminution ou une augmentation de la franchise a un impact tant au niveau de la fréquence que du coût moyen des sinistres individuels. Il convenait donc d'effectuer un réajustement pour estimer le coût des tempêtes pour un niveau de franchise équivalent.

Nos données ne sont donc en aucun cas censurées, puisque nous disposons de l'ensemble des événements et du coût occasionné à la compagnie.

2.2 Modèle utilisé .

Le principal problème auquel nous nous sommes retrouvés confrontés a été le nombre impressionnant d'événements «tempête » ainsi reconstitué.

En fait, nous avons dénombré plus de 1100 tempêtes, soit à peine moins que le nombre de périodes de 3 jours compris dans une période de 11 ans.

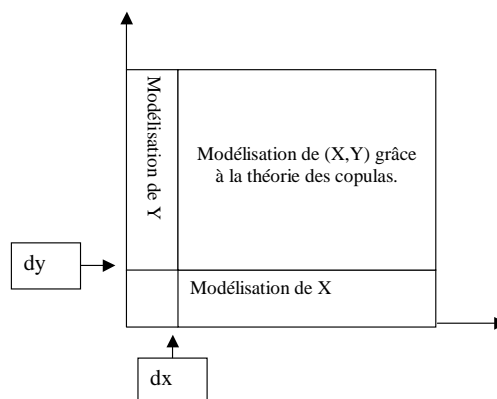
La garantie tempête est donc très souvent touchée.

Nous noterons (x_i, y_i) les montants occasionnés par ces tempêtes aux branches « Incendie » et « Automobile ». x_i correspondra donc au montant occasionné par la $i^{\text{ème}}$ tempête à la branche « Incendie » alors que y_i représentera les dommages subis par la partie « Automobile ».

Nous noterons par la suite

- X la branche incendie .
- Y la branche automobile.

Il n'était pas raisonnable de prendre en compte la totalité de ces événements, certains de ces événements ne résultant en fait que d'un simple coup de vent tout à fait localisé et n'ayant de fait pas droit à l'appellation de tempête au sens météorologique du terme. C'est pourquoi nous avons adopté le schéma de modélisation suivant



dx et dy représentent des niveaux de sélection.

Nous avons en fait choisi de retenir les deux niveaux de sélection suivants:

- Tempêtes dépassant le montant de 1 000 Francs en Auto **et** en incendie. (cas1)
- Tempêtes dépassant 17 500 Francs en Auto **et** 570 000 Francs en incendie. (cas2)

Nous présenterons dans cet article que les résultats obtenus pour le premier niveau de sélection, c'est à dire en conservant un nombre de données important. Mais la méthode à suivre demeure exactement la même quel que soit le nombre de données retenues.

La démarche que nous avons suivie pour modéliser (X,Y) a été la suivante :

- Identification de la ou des copulas à utiliser.
- Détermination de la ou des lois marginales correspondant au phénomène grâce à la méthode du maximum de vraisemblance.
- Vérification de la qualité de l'ajustement de ces lois grâce au test du Khi-Deux.
- Utilisation conjointe des copulas et lois marginales retenues pour estimer les paramètres de notre distribution bivariée par la méthode du maximum de vraisemblance.
- Vérification de la qualité de l'ajustement.

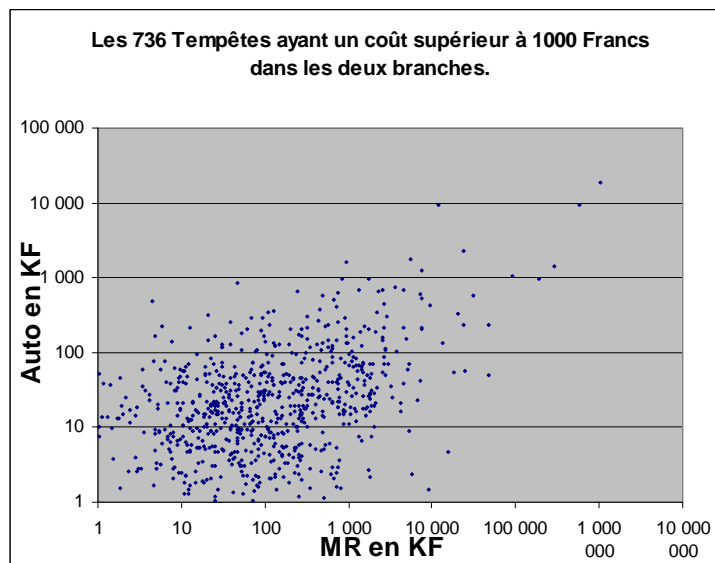
De plus nous avons également effectué une adéquation de la loi marginale de X lorsque Y est inférieur à dy, et vice-versa. Ces simples adéquations n'étant que d'un intérêt relatif, nous ne présenterons pas ces résultats.

2.3 Identification de la copula à utiliser

Nous ne retenons dans un premier temps, que les tempêtes dont le montant a dépassé 1 000 Francs à la fois dans la branche Auto et dans la branche Incendie.

Nous travaillerons cependant en milliers de francs (KF). Le nombre de tempêtes répondant à ces conditions est de 736.

A toutes fins utiles, le graphique représentant les montants dans les deux branches est le suivant:



Ce nuage de points montre de manière évidente qu'il existe bien une dépendance entre ces deux branches.

Le τ de **Kendall** empirique est de 0,243.

Le ρ de **Pearson** est de 0,877.

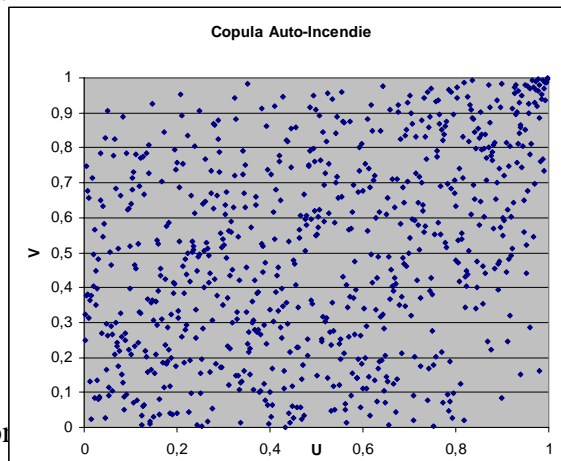
Ces deux coefficients sont significativement différents de 0.

Nous allons maintenant estimer le paramètre d'association pour différentes copulas sans faire d'hypothèses sur la forme paramétrique des lois marginales.

Notons donc n le nombre de nos observations (soit 736).

La méthode consiste donc à transformer notre série de données $(x_i, y_i) \{i=1 \dots n\}$ en (u_i, v_i) en utilisant pour cela la fonction de répartition empirique.

Il suffit, pour ce faire, de repérer le rang de chaque x_i et y_i dans leurs branches respectives et de le diviser par $(n+1)$.



Ensuite, nous pouvons utiliser la méthode du maximum de vraisemblance (*en fait CMLE : canonical maximum likelihood estimation*) pour estimer le paramètre de notre copula. Conformément aux notations adoptées dans notre seconde partie, nous noterons celui-ci « a ». Donc,

$$\hat{a} = \arg \max \sum_{i=1}^n \ln c(u_i, v_i; a)$$

où c représente la fonction de densité de la copula.

Nous allons donc estimer le paramètre pour les copulas de Frank, Gumbel, Clayton, HRT, et enfin la copula normale. Ces cinq copulas ont été retenues essentiellement pour des raisons de notoriété.

Les fonctions utilisées pour la densité sont celles qui ont été précédemment exposées.

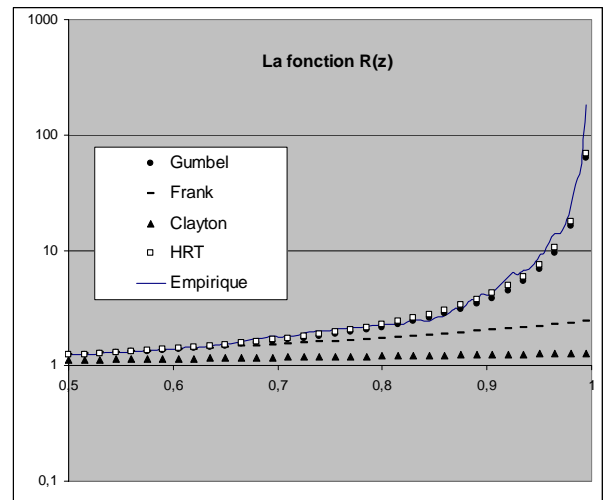
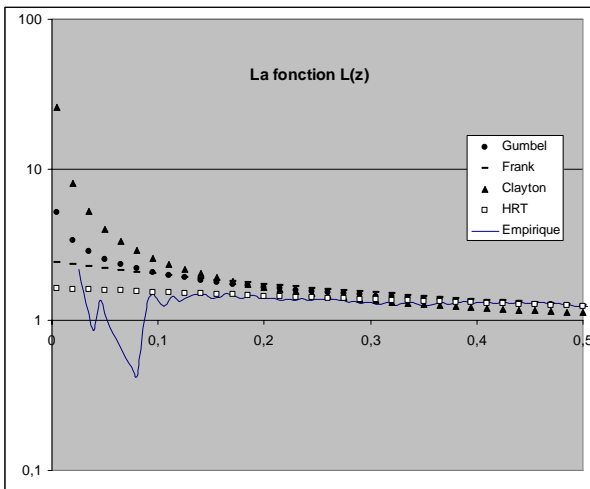
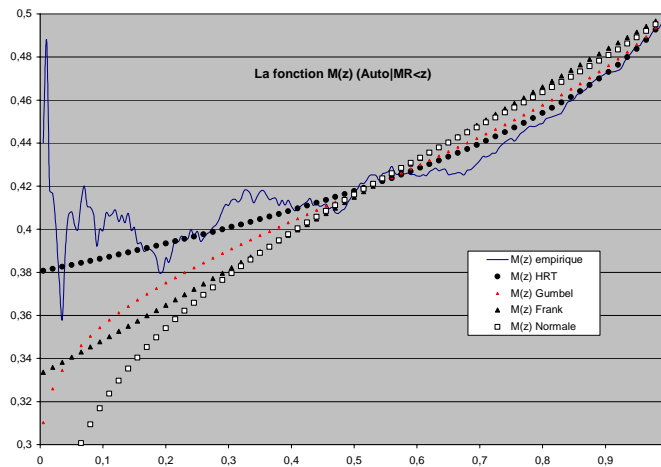
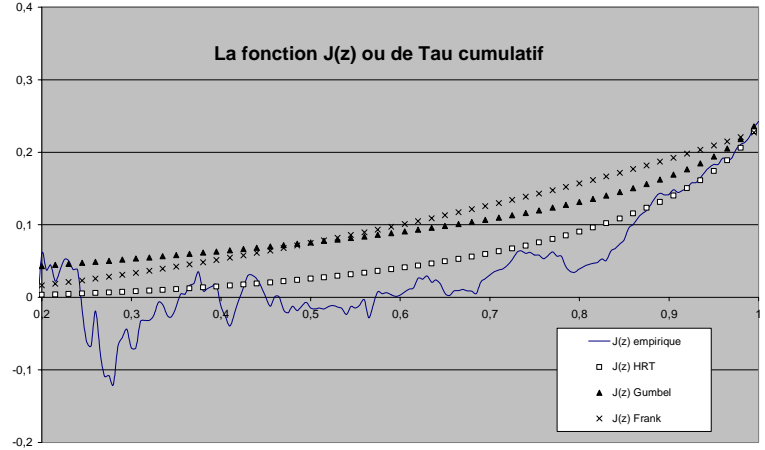
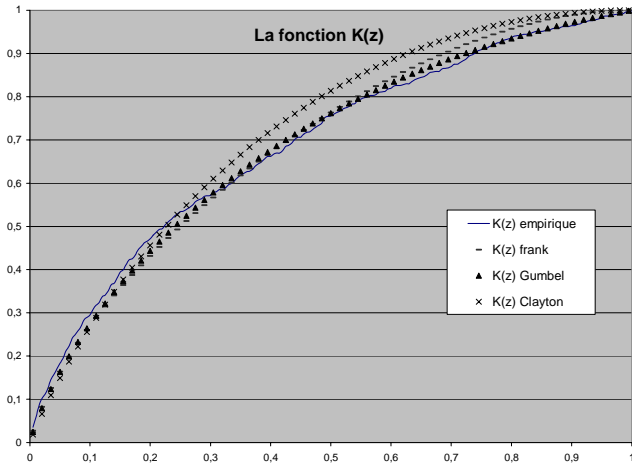
Les résultats obtenus par la méthode du maximum de vraisemblance sont donc les suivants :

	Gumbel	Normale	HRT	Frank	Clayton
Paramètre	1,323	0,378	1,445	2,318	3,378
Log Vraisemblance	77,223	55,428	84,070	50,330	16,447
τ de Kendall	0,244	0,247	0,257	0,245	0,129

Comme pour ces cinq copulas la relation entre le paramètre et le τ de Kendall est « bijective », nous avons également mentionné dans le tableau la valeur de τ résultant de cette maximisation.

Par ailleurs, nous avons établi les versions non-paramétriques des fonctions $J(z)$, $M(z)$, $K(z)$, $R(z)$ et $L(z)$ exposées précédemment.

Nous pouvons les comparer aux versions paramétriques obtenues en prenant les paramètres trouvés ci-dessus.



Au vu des graphes présentés et de la valeur de la fonction du logarithme de la vraisemblance à son maximum, il apparaît que deux copulas pourraient effectivement correspondre à nos données.

La meilleure candidate est sans conteste la copula HRT puisque c'est elle qui maximise « le mieux » la fonction de vraisemblance, et se rapproche au plus près des différentes fonctions indicatrices calculées. Il semble d'ailleurs, au vu du graphique nommé « Copula Auto-Incendie » que la seule concentration visible de points se trouve aux abords de (1,1), ce qui corrobore le fait que cette copula soit la plus appropriée.

Nous allons néanmoins garder la copula de Gumbel, car au vu du graphe de $K(z)$ il s'agit de la meilleure copula Archimédienne, et en terme de vraisemblance elle se place en seconde position. La fonction $R(z)$ nous montre bien que son comportement « à droite » est similaire à celui de la copula HRT, même si « à gauche » elle ne retranscrit pas tout à fait le comportement exact du phénomène observé.

Ceci est dû au fait que la copula de Gumbel induit également une dépendance entre les petites valeurs de u et de v , et explique sûrement pourquoi la courbe de $J(z)$ passe légèrement au dessus de la courbe non-paramétrique.

2.4 Détermination des lois marginales

Comme nous n'avons pris en compte que les données à partir d'un certain seuil, il est nécessaire d'utiliser des lois tronquées.

Une loi tronquée à gauche au seuil dx est définie de la manière suivante :

Supposons que X soit une variable aléatoire continue ayant F_x comme fonction de répartition. Nous définissons la variable aléatoire tronquée T de fonction de répartition F_{T_x} en prenant $T_x = X$ si $X > dx$, et T_x est non définie dans le cas contraire.

La fonction de répartition de T est donc la suivante :

$$F_{T_x}(x) = P(X \leq x | X > dx) = \frac{F_x(x) - F_x(dx)}{1 - F_x(dx)} \quad \text{si } x > dx$$

et 0 sinon. Pour une loi donnée, et en notant $\theta = (\theta_1, \theta_2, \dots, \theta_n)$ le vecteur des paramètres de cette loi, la fonction de log-vraisemblance à maximiser est la suivante :

$$l(\theta) = \sum_{i=1}^n \ln \left(\frac{f_x(x_i; \theta)}{1 - F_x(dx; \theta)} \right)$$

Nous avons grâce à un logiciel développé chez Guy Carpenter effectué cette opération pour quatorze lois différentes.

Les lois marginales testées furent les suivantes :

- Lois à un paramètre : Exponentielle, Pareto simple, Generalized Extreme value limit.
- Lois à deux paramètres : Gamma, Weibull, Gamma Inverse, Log Gamma, Inverse Weibull, Loglogistique, Paralogistique, Inverse Paralogistique, Log-Normale.
- Lois à trois paramètres : Burr, Inverse Burr.

Le critère de sélection du logiciel est appelé HQ, il est égal à la valeur absolue de la fonction de log-vraisemblance à son maximum à laquelle on rajoute le produit entre le nombre de paramètres et le logarithme du nombre de données divisé par 2π . Ce critère, tout comme le critère SBIC $((-2 * \log\text{-vraisemblance}) + \text{nombre paramètres} * \ln(\text{nombre données}))$, pallie la faiblesse du critère AIC $((-2 * \log\text{-vraisemblance}) + 2 * (\text{nombre paramètres}))$, qui tend à accepter un modèle avec un grand nombre de paramètres si l'échantillon est large.

Les deux lois que nous avons retenues aussi bien pour la branche Automobile que pour la branche Incendie sont les lois Loglogistique et Paralogistique.
Les fonctions de répartition et les densités de ces lois sont les suivantes :

La loi Loglogistique

$$F_X(x) = 1 - \frac{1}{1 + \left(\frac{x}{\theta}\right)^\alpha} \quad f_X(x) = \frac{\alpha}{\theta} \frac{\left(\frac{x}{\theta}\right)^{\alpha-1}}{\left(1 + \left(\frac{x}{\theta}\right)^\alpha\right)^2}$$

La loi Paralogistique

$$F_X(x) = 1 - \frac{1}{\left(1 + \left(\frac{x}{\theta}\right)^{\sqrt{\alpha}}\right)^{\sqrt{\alpha}}} \quad f_X(x) = \frac{\alpha}{\theta} \frac{\left(\frac{x}{\theta}\right)^{\sqrt{\alpha}-1}}{\left(1 + \left(\frac{x}{\theta}\right)^{\sqrt{\alpha}}\right)^{\sqrt{\alpha}+1}}$$

En fait ces deux lois, ne sont que des cas spéciaux de la fonction de distribution Burr qui est elle plus connue et dépend de 3 paramètres.

La loi Burr

$$F_X(x) = 1 - \frac{1}{\left(1 + \left(\frac{x}{\theta}\right)^\beta\right)^{\frac{\alpha}{\beta}}} \quad f_X(x) = \frac{\alpha}{\theta} \frac{\left(\frac{x}{\theta}\right)^{\beta-1}}{\left(1 + \left(\frac{x}{\theta}\right)^\beta\right)^{\frac{\alpha}{\beta}+1}}$$

Cette loi est de type « heavy tailed distribution » ou loi à queue épaisse. La loi Paralogistique est donc une loi Burr avec la contrainte $\beta = \alpha^{1/2}$, et la loi Loglogistique satisfait la contrainte $\beta = \alpha$. Néanmoins, nous n'avons pas retenu la loi Burr, car la maximisation de la log-vraisemblance par cette loi nécessitait une amélioration d'au moins 4,76 ($= \ln(736/2\pi)$), ce qui n'a pas été le cas.

L'estimation des paramètres par la méthode du maximum de vraisemblance a donc conduit aux résultats suivants :

	Incendie		Automobile	
	Loglogistique	Paralogistique	Loglogistique	Paralogistique
estimateur de θ	97,690	71,574	18,245	19,245
estimateur de α	0,771	0,703	1,063	1,087
Max Log vraisemblance	-5 088,752	-5 089,080	-3 587,450	-3 587,442
Critère HQ	5 098,279	5 098,606	3 596,976	3 596,969

2.5 Vérification de la qualité des ajustements

Ce test est le test d'adéquation le plus communément employé.

Nous utiliserons les notations usuelles.

$$H_0 \quad F_X(x) = F(x; \theta)$$

$$H_1 \quad F_X(x) \neq F(x; \theta)$$

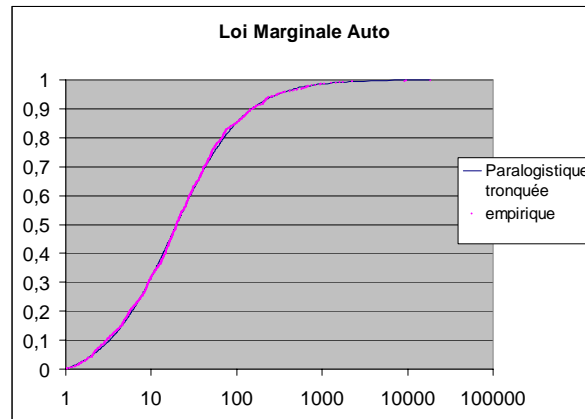
Nous avons décidé de faire une partition de notre échantillon en 16 intervalles aussi bien dans la branche Incendie que pour la branche Automobile.

Le nombre de degrés de liberté de notre loi du khi-deux est par conséquent de 13 puisque deux paramètres ont été estimés pour chaque loi et pour chaque branche.

La valeur critique au seuil 0,05 est donc de 22,36 dans tous les cas.

	Automobile		Incendie	
	Paralogistique	Loglogistique	Paralogistique	Loglogistique
Stat du Khi-deux	16,484	16,639	17,750	17,272
valeur critique	22,362	22,362	22,362	22,362

Au vu des résultats, l'hypothèse H_0 n'est pas rejetée quelque soit la branche. Le graphique ci dessous présente l'ajustement des données Automobile sur une loi paralogistique tronquée.



2.6 Détermination des paramètres de la distribution bivariée.

Nous allons donc conserver pour estimer au mieux les paramètres de notre distribution bivariée la copula HRT ainsi que la copula de Gumbel, et aussi les lois marginales paralogistique et loglogistique.

Nous avons vu, que, dans un cadre général, la densité de (X, Y) se présentait sous la forme:

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

où c représente la densité de la copula, F_X et F_Y les distributions marginales de X et Y respectivement, et f_X, f_Y leurs densités marginales.

Nous notons toujours θ le vecteur des paramètres. Dans le cas présent la dimension de ce vecteur est de 5×1 puisque nous avons en fait deux paramètres à estimer pour chaque loi marginale plus le paramètre de dépendance.

En notant θ_X et θ_Y les composantes du vecteur θ relatives aux lois marginales de X et Y , la fonction de log-vraisemblance devient par suite:

$$l(\theta) = \sum_{i=1}^n \ln(c(F_X(x_i; \theta_X), F_Y(y_i; \theta_Y); a)) + \sum_{i=1}^n \ln f_X(x_i; \theta_X) + \sum_{i=1}^n \ln f_Y(y_i; \theta_Y)$$

Comme dans notre cas nous avons utilisé des lois marginales tronquées, notre fonction de log-vraisemblance à maximiser prendra la forme :

$$l(\theta) = \sum_{i=1}^n \ln \left(c \left(\frac{F_X(x_i) - F_X(dx)}{1 - F_X(dx)}, \frac{F_Y(y_i) - F_Y(dy)}{1 - F_Y(dy)} \right) \right) + \sum_{i=1}^n \ln \frac{f_X(x_i)}{1 - F_X(dx)} + \sum_{i=1}^n \ln \frac{f_Y(y_i)}{1 - F_Y(dy)}$$

$$l(\theta) = \sum_{i=1}^n \ln(c(F_{T_X}(x_i), F_{T_Y}(y_i))) + \sum_{i=1}^n \ln f_{T_X}(x_i) + \sum_{i=1}^n \ln f_{T_Y}(y_i)$$

Par souci de clarté, nous n'avons pas fait apparaître dans la formule de log-vraisemblance les composantes du vecteur θ .

Il paraît peu probable que l'on puisse trouver une solution analytique à ce problème de maximisation. Nous ne pouvons utiliser que des techniques numériques itératives.

Nous avons opté pour le solveur d'Excel qui utilise la méthode de Newton-Raphson.

Comme le solveur procède par itérations, nous avons dans un premier temps utilisé comme valeurs de départ celles obtenues lors des maximisations précédemment effectuées, tant pour le paramètre de la copula que ceux des lois marginales. Puis, nous avons changé de valeurs de départ (au moins une dizaine de répétitions) pour nous assurer que nous retrouvions le même résultat. Ceci fut systématiquement le cas.

Les calculs ont néanmoins été effectués avec un ordinateur possédant 1 Go de Ram et deux processeur Pentium 3 agissant de concert.

Les résultats obtenus pour différentes combinaisons sont les suivants:

	copula HRT		copula HRT	
	Auto paralogistique	MR paralogistique	Auto paralogistique	MR loglogistique
estimateur de theta	19,127	73,532	19,277	98,916
estimateur de alpha	1,075	0,711	1,085	0,774
estimateur de a	1,448		1,472	
Max Log-vraisemblance	-8592		-8591	

	copula Gumbel		copula Gumbel	
	Auto paralogistique	MR paralogistique	Auto paralogistique	MR loglogistique
estimateur de theta	19,861	85,221	20,053	108,62
estimateur de alpha	1,115	0,766	1,129	0,833
estimateur de a	1,323		1,317	
Max Log-vraisemblance	-8613		-8615	

Les meilleurs résultats en termes de maximisation de la vraisemblance sont obtenus avec la **combinaison copula HRT, loi paralogistique pour l'automobile et loi loglogistique pour l'incendie**. La maximisation obtenue en supposant l'indépendance entre les deux lois marginales est tout simplement la somme des deux résultats obtenus pour les lois marginales soit à peu près -8676. L'utilisation des copulas pour modéliser le phénomène constitue donc une amélioration certaine.

2.7 Tests d'adéquation d'une distribution bivariée.

Nous allons exposer ici les différents tests d'adéquation possibles pour une distribution bivariée. Nous concentrerons nos efforts sur la présentation des résultats pour **la meilleure combinaison** possible au regard de la maximisation de la fonction de log-vraisemblance.

Avant d'effectuer nos tests d'adéquation nous allons tracer la ligne de régression médiane

2.7.1 La ligne de regression médiane

Pour une valeur donnée d'une tempête en incendie x , la valeur y sur cette ligne est la solution de l'équation $0,5 = F_{T_Y|T_X}(y|x)$, ce qui avec la copula HRT peut se réécrire :

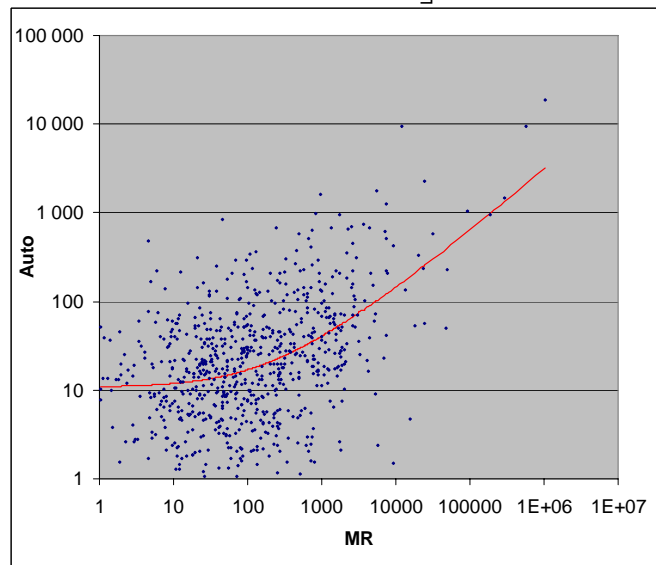
$$0,5 = C_1(F_{T_X}(x), F_{T_Y}(y)) = 1 - \left[(1 - F_{T_X}(x))^{-1/a} + (1 - F_{T_Y}(y))^{-1/a} - 1 \right]^{a-1} (1 - F_{T_X}(x))^{-1/a}$$

En résolvant par rapport à F_Y , nous obtenons

$$v = F_{T_Y}(y) = 1 - \left[1 - (1 - F_{T_X}(x))^{-1/a} + \left(\frac{1}{2} (1 - F_{T_X}(x))^{1+1/a} \right)^{-1/a} \right]^{-a}$$

Puis nous obtenons y en inversant v

$$y = F_{T_Y}^{-1}(v) = \theta \left[1 - \left(v(1 - F_Y(d_y)) + F_Y(d_y) \right)^{-1/\sqrt{\alpha}} - 1 \right]^{\frac{1}{\sqrt{\alpha}}}$$



Le résultat au premier coup d'œil semble plutôt bon, dans la mesure où la ligne obtenue semble bien passer au milieu de notre nuage de points.

2.7.2 Test du Khi-deux

Nous allons donc vérifier la qualité de notre distribution bivariée par l'intermédiaire du test du khi-deux dont le principe a été précédemment exposé.

Nous exposerons deux techniques différentes pour vérifier la justesse de l'adéquation

- la première consiste à étendre le test à deux dimensions
- la seconde permet de se ramener à deux tests à une dimension.

2.7.2.1 Extension du test à deux dimensions.

Nous avons en fait découpé l'espace en un certain nombre de pavés, dont certains sont de taille infinie. La taille des pavés diminue au fur et à mesure que l'on s'approche du mode.

Nous avons ensuite déterminé la probabilité théorique qu'une observation se retrouve dans ce pavé.

Nous savons que $P(x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2) = F(x_2, y_2) - F(x_2, y_1) - F(y_2, x_1) + F(x_1, y_1)$.

Ce qui se réécrit donc,

$$C(F_X(x_2), F_Y(y_2)) - C(F_X(x_2), F_Y(y_1)) - C(F_X(x_1), F_Y(y_2)) + C(F_X(x_1), F_Y(y_1))$$

Là encore cette formule est écrite dans le cas général. Nous avons à nouveau pris en compte le caractère tronqué de nos lois dans nos calculs.

$$C(F_{T_x}(x_2), F_{T_y}(y_2)) - C(F_{T_x}(x_2), F_{T_y}(y_1)) - C(F_{T_x}(x_1), F_{T_y}(y_2)) + C(F_{T_x}(x_1), F_{T_y}(y_1))$$

Nous avons donc découpé notre espace en 22 rectangles plus ou moins réguliers.

Le nombre de degrés de liberté de notre loi du khi-deux est de 22-5-1, soit 16.

Notre statistique du khi-deux a pour valeur 23,86 alors que notre valeur critique se situe à 26,3. Le test se révèle donc concluant.

A titre indicatif, cette même statistique a également été calculée avec la copula de Gumbel et les deux lois paralogistiques et a été évaluée à 26,3. Elle constitue donc un ajustement moins bon.

Nous avons également mené ce test en supposant l'indépendance entre les deux branches: la statistique ainsi déterminée a pris pour valeur 186.

2.7.2.2 Utilisation de deux tests uni-dimensionnels

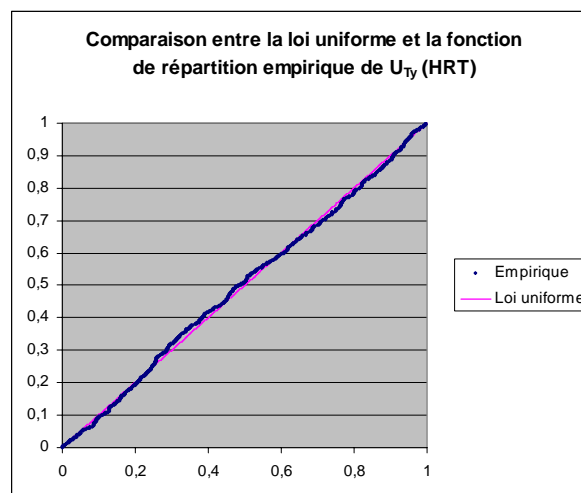
Cette méthode utilise le fait que les variables aléatoires $U_{T_x} = F_{T_x}(T_x)$ et $U_{T_y} = F_{T_y|T_x}(T_y|T_x)$ sont indépendantes et ont une distribution uniforme sur $[0,1]$.

Nous allons donc tester dans un premier temps le fait que U_{T_x} suive une loi uniforme, puis si cela s'avère être le cas, nous ferons de même pour U_{T_y} . Nous avons là encore besoin de la dérivée partielle de notre copula par rapport à u . (ie: $C_1(u,v)$)

Nous avons, dans un premier temps, divisé l'intervalle $[0,1]$ en 20 segments d'une longueur de 0,05 chacun, et utilisé le test du khi-deux avec un nombre de degrés de liberté de 14 (=20-5-1). Ceci revient à faire un test du Khi-deux sur une loi marginale, mais avec un nombre théorique égal dans chaque tranche.

Notre statistique est de 14,65 alors que la valeur critique est de 23,68. La première condition étant remplie, nous avons ensuite effectué le second test avec toujours 20 intervalles.

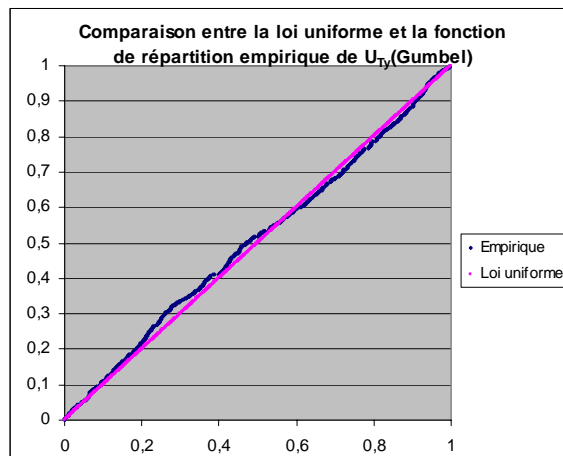
La statistique du Khi-deux pour ce second test est de 21.86, soit toujours en deçà de notre valeur critique.



Copula HRT-fonction de répartition empirique de U_{T_y}

Ce graphique permet de mieux se rendre compte de la qualité de l'adéquation, puisque la loi empirique est très semblable à la loi uniforme.

A titre de comparaison, ce même graphique, établi à partir de l'adéquation à la copula de Gumbel, et deux lois paralogistiques pour les lois marginales a l'allure suivante:



Copula de Gumbel-fonction de répartition empirique de U_{T_y}

L'adéquation est dans ce cas nettement moins bonne. Le premier test s'est révélé concluant, mais ceci ne fut pas le cas pour le second.

3 Exemple d'application : Etude d'un excédent de sinistre automobile dont le paiement est conditionné par la branche incendie.

Bien que nous ayons trouvé dans notre précédente partie une adéquation satisfaisant à tous les tests (HRT+para+log), celle-ci ne peut pas être considérée comme excellente aux niveaux qui intéressent habituellement les réassureurs. Nous avons donc réitéré notre opération avec un nombre de données plus restreint (*cf 2.2 cas 2*). Nous avons abouti à une adéquation sur une Pareto pour la branche Incendie et une Log-Normale Tronquée pour la branche automobile, et une structure de dépendance de type « Copula HRT » avec un paramètre de dépendance évalué à 1,365. Nous allons donc nous servir de cette dernière adéquation qui est de meilleure qualité au niveau des seuils auxquels travaillent les réassureurs. En ce qui concerne la modélisation de la fréquence annuelle, nous avons opté pour une loi de Poisson.

Le principe du contrat étudié est simple. Le Réassureur de l'excédent de sinistres automobiles règle un sinistre automobile uniquement si, parallèlement, l'événement tempête a généré un montant supérieur à la priorité dans la branche incendie.

Nous allons donc fixer un plafond et une priorité dans chaque branche et évaluer la prime pour ce type de contrat. A titre indicatif, et pour mieux mettre en valeur la nécessité d'utiliser une distribution bivariee, nous fournirons également les résultats obtenus en supposant l'indépendance entre les deux branches.

Copula HRT $a=1,365$				
	Priorité Auto	Plafond Auto	Prime annuelle	Estimation de la fréquence de dépassement de la priorité.
Automobile	3 220 KF	32 000 KF	1 010 KF	0,11
Incendie	111 000 KF			0,2502

Par le vocable “estimation de la fréquence de dépassement de la priorité” nous entendons le nombre moyen d’années au cours de laquelle le réassureur va devoir effectuer un paiement sur ce type de contrat. Sur 25% des années simulées, il y a eu un événement dépassant la priorité de 111 000 KF en incendie, et seulement 11% des années ont vu un paiement effectué par le réassureur pour le contrat automobile. La valeur théorique du τ de Kendall entre les deux branches est de 0,268. Nous noterons cet exemple : cas1.

Les résultats obtenus en supposant l’indépendance entre les deux branches (cas 2):

Indépendance				
	Priorité Auto	Plafond Auto	Prime annuelle	Estimation de la fréquence de dépassement de la priorité.
Automobile	3 220 KF	32 000 KF	33 KF	0,0064
Incendie	111 000 KF			0,2502

Ceci a conduit, comme cela était prévisible, à une réduction considérable et erronée de la prime, et pourtant les lois marginales simulées sont strictement les mêmes.

Nous avons, par curiosité, calculé cette même prime pour des lois marginales reliées par la copula comotone ($\tau=1$, situation de parfaite dépendance, cas 3) : l’application numérique a conduit au résultat de 1621KF.

La différence absolue de prime est donc plus élevée entre la situation de totale indépendance (cas2) et le cas1 d’une part, qu’entre le cas1 et le cas3 d’autre part.

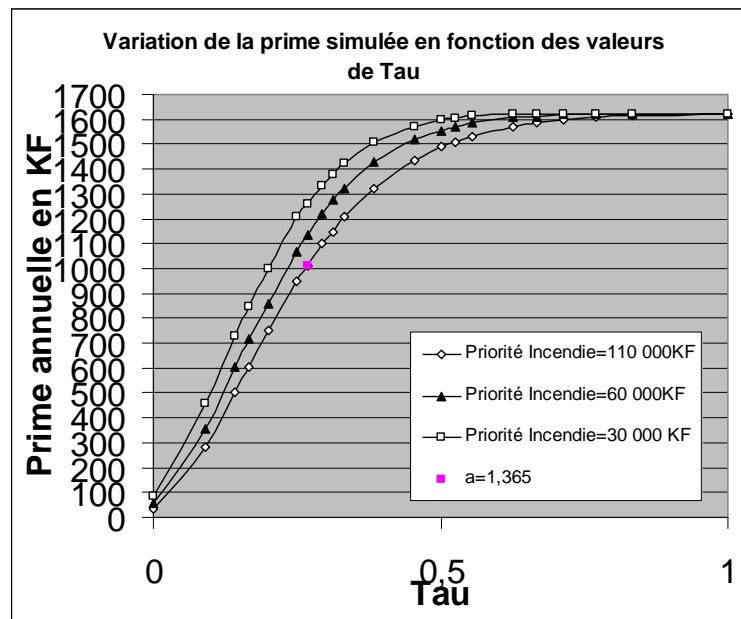
Cependant, ceci n’est pas conforme à ce que nous aurions pu escompter en nous intéressant à la différence en termes de τ de Kendall. Passer de 0 à 0,26 conduit à une augmentation de la prime équivalente à 1000KF, alors que passer de 0,26 à 1 fait seulement augmenter la prime de 600KF.

Ceci est du à deux facteurs agissant conjointement:

- Le premier est la structure de dépendance utilisée (copula HRT)
- Le second concerne les priorités choisies qui sont relativement élevées.

Or nous avons vu que la copula HRT possède un coefficient de “Right Tail Dependence” qui vaut $1/2^a$. Ainsi, à partir d’une certaine valeur de a et vu la structure de la copula HRT, tous les événements qui excèdent la priorité en incendie vont également conduire à un

dépassement de la priorité en automobile. En outre, nous avons choisi nos montants de priorité afin de les rendre homogènes dans les deux branches.(i.e une probabilité équivalente à 25% qu'au moins un événement auto ou incendie atteigne la priorité chaque année).



Le graphique ci-dessus a été établi en gardant les mêmes paramètres de lois marginales, la même priorité sur la branche automobile (soit 3220KF), la même structure de dépendance, mais nous avons fait simplement varier le paramètre de dépendance de la copula HRT et le montant de la priorité en incendie qui conditionne le paiement de la branche automobile.

Plus nous faisons tendre la priorité en Incendie vers 0, moins la convexité de la courbe sera sensible. Si nous fixons la priorité en Incendie à 570 KF, il n'y aura plus de conditionnement et la prime en Automobile sera de 1621 KF quelle que soit la valeur de τ .

Conclusion

Les copulas permettent donc l'adéquation à des fonctions de répartition en plusieurs dimensions. Les travaux présentés ci-dessus, reprenant en partie les techniques employées par Klugman-Parsa (1998) en sont donc un exemple concret et aisément reproductible. Toutefois, nos estimations paramétriques ne peuvent pas être considérées comme significatives au niveau qu'aurait retenu une cédante pour se couvrir par un traité en excédent de sinistre. C'est pourquoi, nous avons réitéré l'opération présentée ci-dessus mais en élevant nos seuils de sélection (dx et dy) et donc en restreignant le nombre de données prises en compte. Cependant, même si les lois marginales retenues ne furent pas les mêmes, la structure de dépendance entre l'Auto et l'Incendie demeurerait la copula HRT. Ceci s'est d'ailleurs vérifié après l'examen du portefeuille d'un autre Assureur, d'où il est également ressorti que la copula HRT était la meilleure des cinq, avec une structure de dépendance encore plus marquée.

A partir des estimations paramétriques ainsi obtenues, il est dès lors possible de coter toutes sortes de contrats de réassurance portant sur les branches Auto et Incendie en utilisant simplement la méthode de Monte-Carlo.

Cela nous a par exemple permis de déterminer la prime pure de certains contrats exotiques, dans lesquels le paiement de sinistres au titre d'un excédent de sinistre sur une branche est conditionné par la valeur prise des sinistres dans une autre branche. Nous avons également fait varier le coefficient de dépendance pour voir quel était l'effet de celle-ci sur la prime, en gardant bien entendu nos lois marginales inchangées. Les résultats obtenus allaient souvent à l'encontre de ce que notre intuition portait à croire.

Néanmoins, les copulas, même si elles demeurent un outil très puissant n'ont pour l'instant pas souvent été utilisées. Les exemples exposés dans cet article traitent de problèmes bivariés. En théorie, la méthode doit pouvoir être étendue à plus de deux dimensions. Mais, la plupart des articles parus à ce jour se contentent d'utiliser les copulas pour des travaux en deux dimensions. Lorsque le nombre de dimensions est supérieur à deux, les auteurs se ramènent toujours à une copula elliptique (Normale ou de Student) pour la simple et bonne raison que les densités de ces copulas sont aisées à calculer et que ces copulas sont simulables de manière simple, même lorsque le nombre de dimensions est très élevé. Or, ce type de copulas n'est pas forcément le plus adapté à l'étude de la sinistralité en assurance. Même si il existe des extensions, et encore sous certaines conditions, des copulas Archimédiennes à plus de deux dimensions, établir les dérivées nécessiterait des calculs tout à fait fastidieux.

En outre, l'utilisation des copulas deviendrait quasi-impossible en se restreignant à l'utilisation de logiciels de bureautique standard comme Excel. Ceci constitue donc pour l'instant un petit frein à leur utilisation, mais gageons que celui-ci sera certainement surmonté et que les prochains travaux sur les copulas seront multidimensionnels et sauront mettre en évidence leur apport dans de nombreux domaines.

Les applications peuvent être envisagées tant en Vie qu'en non Vie. Ainsi pour la branche Vie on peut étudier en utilisant les copulas les annuités viagères sur deux têtes. Un autre exemple pourrait être l'assurance décès avec un paiement à chaque décès, le paiement étant plus élevé au premier décès qu'au second. En non vie on peut citer des contrats d'assurance comportant plusieurs garanties (incendie et pertes d'exploitation) ou impliquant le paiement de plusieurs types de sinistres (en santé les remboursements pour différents actes comme les consultations et visites ainsi que la pharmacie ...).

L'expérience sinistres d'un assuré disposant de contrats dans différentes branches (Incendie, Automobile ...) pourrait faire également l'objet d'une modélisation par les copulas.

Bibliographie

1. **Belguise O ., (2001)**, *tempêtes :étude des dépendances entre les branches Auto et Incendie avec la théorie des copulas*, Mémoire magistère d'actuariat Université Louis Pasteur Strasbourg
2. **Embrechts P., McNeil A., Straumann D., (1999)**, *Correlation and dependence in risk management: properties and pitfalls*, document de travail ETHZ
3. **Foata D., Fuchs A., (1996)**, *Calcul des probabilités*, Masson
4. **Frees E., Valdez E.A , (1999)**, *Understanding relationship using copulas*, North American Actuarial Journal, vol 2(1), p1-23
5. **Groupe de recherche opérationnelle Crédit Lyonnais, (2000)** *Copulas for finance- A reading guide and some applications*
6. **Groupe de recherche opérationnelle Crédit Lyonnais, (2000)**, *Which copula is the right one?*
7. **Klugman S.A, Panjer H.H., Willmot G.E., (1998)** *Loss models*, Wiley-Interscience
8. **Klugman S.A, Parsa R., (1998)**, *Fitting bivariate loss distributions with copulas*, Insurance: Mathematics and economics, 24, p139-148
9. **Lindskog F., (2001)**, *Modelling dependence with copulas and applications to risk management*, document de travail ETHZ
10. **McNeil A., (2001)**, *Multivariate models:theory*, Cours magistère d'actuariat Université Louis Pasteur Strasbourg
11. **Nelsen R.B.(1999)**, *An introduction to copulas*, Lecture notes in statistics Springer
12. **Robert C., (2000)**, *Gestion de risques multiples*, Présentation lors de la journée « Statistique et Assurance Dommages » Société Française de Statistique
13. **Venter, G., (2001)**, *Tails of copulas*, 23ème ASTIN Colloquium, Washington D.C
14. **Wang S.,(1998)**, *Aggregation of correlated risk portfolios: models and algorithms*, Proceedings of the Casualty Actuarial Society vol LXXXV p848-938.