Gestion des risques multiples en assurance

introduction à la théorie des copulas

ARTHUR CHARPENTIER arthur.charpentier@ensae.fr

La fonction de répartition du couple (X,Y) est

$$F_{XY}(x,y) = P(X<x,Y<y)$$

avec
$$F_X(x) = P(X<x) = F_{XY}(x,+\infty)$$
$$F_Y(y) = P(Y<y) = F_{XY}(+\infty,y)$$

On notera $f_{XY}(x,y)$ la densité *(si elle existe)*, définie par

$$f_{XY}(x,y) = \partial^2 F_{XY}(x,y) \,/\, \partial x \partial y$$

Étant données deux fonctions de répartition F_X et F_Y on notera $\Im(F_X,F_Y)$ l'ensemble des fonctions de répartition de couples (X,Y) dont les fonctions de répartition marginales sont F_X et F_Y

$$\Im(F_X,F_Y)=\{\ F_{XY} \text{ fdr} \mid F_X(x)=F_{XY}(x,+\infty),\ F_Y(y)=F_{XY}(+\infty,y)\ \}$$

RQ F(x,y) est une fonction de répartition en dimension 2 ssi
- $F(x,t) \to 0$ et $F(t,y) \to 0$ quand $t \to -\infty$
- $F(x,y) \to 1$ quand $x \to +\infty$ et $y \to +\infty$
- pour tout $x_1<x_2$ et $y_1<y_2$,

$$F(x_1,y_1) + F(x_2,y_2) \geq F(x_1,y_2) + F(x_1,y_2)$$

Cette dernière propriété est analogue à $\partial^2 F(x,y)/\partial x\partial y \geq 0$

Def : X et Y sont indépendantes si et seulement si

$$F_{XY}(x,y) = F_X(x)\, F_Y(y)$$

Gestion des risques multiples en assurance

Le cas de variables dépendantes se définie comme contraposée de l'hypothèse d'indépendance,

$$F_{XY}(x,y) \neq F_X(x)\, F_Y(y)$$

- <u>**Les structures de dépendance**</u>

• Le couple (X,Y) est dit **PQD** (dépendance positive par quadrant) si et seulement si, pour tout x,y

$$P(X>x,Y>y) \geq P(X>x)P(Y>y)$$

ou de façon équivalente, si et seulement si l'une des relation suivantes est vérifie, pour tout x,y

(i) $P(X>x|Y>y) \geq P(X>x)$
(ii) $P(Y>y|X>x) \geq P(Y>y)$
(iii) $(f(X),g(Y))$ est PQD pour f et g croissantes
(iv) $\text{cov}(f(X),g(Y)) \geq 0$ pour f et g croissantes
(v) $P(X\leq x,Y\leq y) \geq P(X\leq x)P(Y\leq y)$

Rq si ϕ est une fonction supermodulaire, i.e.
$\phi(x_1,y_1) - \phi(x_1,y_2) - \phi(x_2,y_1) + \phi(x_2,y_2) \geq 0$ où $x_1\leq x_2$, $y_1\leq y_2$
[c'est à dire $\partial^2\phi\ \partial x\partial y \geq 0$ si la dérivée seconde existe]
alors, si (X,Y) est PQD,

$$E(\phi(X,Y)) \geq E(\phi(X^\perp,Y^\perp))$$

où $(X^\perp,Y^\perp)$ est une version indépendante du couple (X,Y)

En dimension d, on définir de façon analogue la dépendance par orthant supérieur (**PUOD**) par

$$P(X_1>x_1,\ldots,X_d>x_d) \geq P(X_1>x_1)\ldots P(X_d>x_d)$$

c'est à dire $P(X_1>x_1,\ldots, X_d>x_d) \geq P(X_1^\perp >x_1,\ldots, X_d^\perp >x_d)$
et la dépendance par orthant inférieur (**PLOD**) par

$$P(X_1\leq x_1,\ldots,X_d\leq x_d) \geq P(X_1^\perp\leq x_1,\ldots, X_d^\perp\leq x_d)$$

Gestion des risques multiples en assurance

Rq En dimension $d \geq 3$ ces deux notions ne sont pas équivalentes

• Le couple (X,Y) est dit **associé** si et seulement si, pour toutes fonctions croissantes ϕ et ψ

$$\text{cov}(\phi(X,Y),\psi(X,Y)) \geq 0$$

(dès lors que la covariance existe).

Cette relation se définie aisément en dimension d, et on a le résultat suivant

$$(X_1,\ldots,X_d) \text{ associé} \Rightarrow (X_1,\ldots,X_d) \text{ POD (PLOD et PUOD)}$$

• La variable Y est dite croissante conditionnellement à X (**CI**) si et seulement si

$$P(X>x|Y=y) \text{ est croissante en } y$$

Rq Y croissante conditionnellement à X $\Rightarrow$ (X,Y) associé

• Soit (X,Y) un vecteur aléatoire possédant une densité f_{XY}. Le vecteur (X,Y) est dit TP_2 si et et seulement si

$$f_{XY}(x_1,y_1)f_{XY}(x_2,y_2) \geq f_{XY}(x_1,y_2)\, f_{XY}(x_2,y_1)$$

où $x_1 \leq x_2$ et $y_1 \leq y_2$.

Rq (X,Y) $TP_2 \Rightarrow$ Y croissante conditionnellement à X

• Le couple (X,Y) est dit **dépendant par mélange** s'il existe une variable aléatoire Θ telle que, pour tout θ, les variables $(X|\Theta=\theta)$ et $(Y|\Theta=\theta)$ soient indépendantes

Rq si (X,Y) est dépendant par mélange et si $F_X(x|\Theta=\theta) \geq F_X(x|\Theta=\theta')$ et $F_Y(y|\Theta=\theta) \geq F_Y(y|\Theta=\theta')$ pour tout $\theta \leq \theta'$, alors le couple (X,Y) est associé

Arthur Charpentier *arthur.charpentier@ensae.fr*

Gestion des risques multiples en assurance

A partir de ces notions, il est possible de définir des relations d'ordre sur l'ensemble des vecteurs aléatoires

Rq En dimension 1, les relations d'ordre usuelles sont, pour des variables aléatoires X et Y positives :

• ***comparaison à l'ordre 1*** X est dominé par Y
$X \leq_1 Y$ ssi $E(u(X)) \leq E(u(Y))$ pour u croissante

$X \leq_1 Y$ ssi $F_X(t) \geq F_Y(t)$ pour tout $t \in R$

• ***comparaison à l'ordre 2*** X est préféré à Y à l'ordre 2,
$X \leq_2 Y$ ssi $E(u(X)) \leq E(u(Y))$ pour u croissante convexe

$X \leq_2 Y$ ssi $E((X-\omega)^+) \leq E((Y-\omega)^+)$ pour tout ω

Si $E(X) \leq E(Y)$ et s'il existe $c \geq 0$ telle que

$F_X(t) \leq F_Y(t)$ pour tout $0 \leq t < c$

$F_X(t) \geq F_Y(t)$ pour tout $c < t$

alors $X \leq_2 Y$

Il existe davantage de relations d'ordre en dimension d

• ***dominance par orthant inférieur***

$X \leq_{LO} Y$ ssi $F_X(t) \geq F_Y(t)$ pour tout $t \in R^d$

• ***dominance par orthant supérieur***

$X \leq_{UO} Y$ ssi $\overline{F}_X(t) \leq \overline{F}_Y(t)$ pour tout $t \in R^d$

• ***dominance 'à l'ordre 1'***

$X \leq_1 Y$ ssi $E(\phi(X)) \leq E(\phi(Y))$ pour ϕ croissante

• ***dominance 'à l'ordre 2'***

$X \leq_2 Y$ ssi $E(\phi(X)) \leq E(\phi(Y))$ pour ϕ croissante convexe

Arthur Charpentier *arthur.charpentier@ensae.fr*

• ***dominance par transformée de Laplace***

$X \leq_{LT} Y$ ssi $E(\exp[-u'X]) \geq E(\exp[-u'Y])$ pour tout $u \in R^d$

Rq Il est possible de trouver des relations analogue au cas unidimensionnel en introduisant les fonctions supermodulaires :

En dimension d, la fonction ϕ sera dite supermodulaire si, pour tout hypercube [x,y], $V_\phi([x,y]) \geq 0$ où

$$V_\phi([x_1,y_1]\times\ldots\times[x_d,y_d]) = \sum_{z=(z_1,\ldots,z_d)} \sigma(z)\phi(z_1,\ldots,z_d)$$

où la somme se fait sur l'ensemble des sommets de l'hypercube, et où la fonction $\sigma(z)$ vaut +1 si $z_i=x_i$ pour un nombre pair i, -1 sinon.

Ex En dimension 3, ϕ sera dite supermodulaire si, pour tout $x_1 \leq x_2$, $>y_1 \leq y_2$ et $z_1 \leq z_2$,

$$\phi(x2,y2,z2) - \phi(x1,y2,z2) - \phi(x2,y1,z2) - \phi(x2,y2,z1) + \phi(x2,y1,z1) + \phi(x1,y2,z1) + \phi(x1,y1,z2) - \phi(x1,y1,z1) \geq 0$$

Prop $X \leq_{UO} Y$ ssi $E(\phi(X)) \leq E(\phi(Y))$ pour ϕ supermodulaire

D'autres formes de dominance ont été introduites, prenant en compte des portefeuilles de risques :

• $X \leq_{LCX} Y$ ssi $a'X \leq_2 a'Y$ pour tout $a \in R^d$

• $X \leq_{PLCX} Y$ ssi $a'X \leq_2 a'Y$ pour tout $a \in R^{+d}$

si $X \leq_{LCX} Y$, alors $E(X)=E(Y)$ et $V(Y)-V(X)$ est semidéfinie et positive

si $L(X_i)= L(Y_i)$ et si $E(\phi(X)) \leq E(\phi(Y))$ pour toute fonction ϕ supermodulaire, alors $X \leq_{PLCX} Y$

Gestion des risques multiples en assurance

- <u>Les Bornes de Fréchet</u>

Pour tout couple (X,Y), la fonction de répartition jointe est borné, pour tout x,y, par

$$\max\{F_X(x)+F_Y(y)-1,0\} \leq F_{XY}(x,y) \leq \min\{F_X(x),F_Y(y)\}$$

• La **borne supérieure** est atteinte pour le couple $(F_X^{-1}(U),F_Y^{-1}(U))$ où $U\sim U([0,1])$

(X,Y) admet pour fonction de répartition $\min\{F_X(x),F_Y(y)\}$ si et seulement il existe des fonctions strictement croissantes f et g telles que

(X,Y) = (f(Z),g(Z)) pour une certaine variable aléatoire Z

• La **borne inférieure** est atteinte pour le couple $(F_X^{-1}(U),F_Y^{-1}(1-U))$ où $U\sim U([0,1])$

(X,Y) admet pour fonction de répartition $\min\{F_X(x),F_Y(y)\}$ si et seulement il existe des fonctions f strictement croissante et g strictement décroissante telles que

(X,Y) = (f(Z),g(Z)) pour une certaine variable aléatoire Z

- <u>Inégalité de Tchen</u>

Pour tout couple (X,Y) de variables à valeurs positives, et pour toute fonction ϕ continûment différentiable, telle que $\partial^2\phi\, \partial x \partial y \geq 0$,

$$E(\phi(F_X^{-1}(U),F_Y^{-1}(1-U)) \leq E(\phi(X,Y)) \leq E(\phi(F_X^{-1}(U),F_Y^{-1}(U))$$

[ce résultat restant vérifie si ϕ est supermodulaire]

Ex Soit X le prix du d'un actif sous-jacent à la date t, et considérons les options européennes de prix d'exercice d

Soit C le payoff du call associé, C=max(0,X-d) et P le payoff du put, P=max(0,d-X)

La distribution du couple (X,C) est donnée par la borne supérieure de Fréchet

Les distributions des couples (X,P) et (C,P) est donnée par la borne inférieure de Fréchet

Ex Considérons une couverture Excess of Loss, définie par

$$X_{a,a+h} = \begin{cases} 0 & \text{si} \quad 0 \le X \le a \\ X - a & \text{si} \quad a \le X \le a + h \\ h & \text{si} \quad a + h \le X \end{cases}$$

où a est la franchise, et h la limite. La fonction de survie associée à $X_{a,a+h}$ est donnée par

$$\overline{F}_{X_{a,a+h}}(t) = \begin{cases} \overline{F}_X(a+t) & \text{si} \quad t < h \\ 0 & \text{si} \quad h \le t \end{cases}$$

Les variables $X_{a,a+h}$ et $X_{b,b+h}$ sont comonotoniques

Rq Si des risques $X_1,\ldots,X_n$ sont indépendants, de même loi (marginale) et de variance finie σ^2, alors

$$\mathrm{Var}\left(\frac{1}{n}[X_1 + \ldots + X_n]\right) = \frac{\sigma^2}{n} \to 0$$

Si les risques $X_1,\ldots,X_n$ sont comonotoniques, alors

$$\mathrm{Var}\left(\frac{1}{n}[X_1 + \ldots + X_n]\right) = \sigma^2$$

Arthur Charpentier *arthur.charpentier@ensae.fr*

Rq La comonotonie correspond au cas de dépendance « parfaite », et donne des bornes sur la distribution du couple (X,Y) : $P(X \le x, Y \le y) = F_{XY}(x,y) \le \min\{F_X(x),F_Y(y)\}$

MAIS cela ne donne pas de résultats, par exemple, sur la somme X+Y : VaR(X+Y) n'est pas maximale quand X et Y sont comonotoniques.

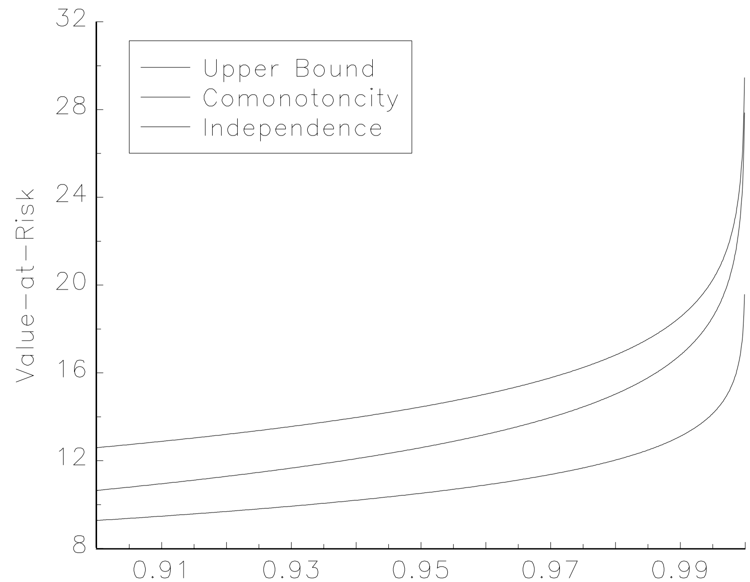
Ex Soient $X,Y \sim \mathrm{Gamma}(a,\theta)$, et G_a la cdf associée

(i) dans le cas comonotonique : $\mathrm{VaR}(X+Y,\alpha)=2G_a^{-1}(\alpha)$

(ii) dans le cas indépendant : $\mathrm{VaR}(X+Y,\alpha)=2G_{2a}^{-1}(\alpha)$

(iii) la borne supérieur de la $\mathrm{VaR}(X+Y,\alpha)$ est donnée par

$$\inf_{u+v-1=\alpha}\left\{G_a^{-1}(u) + G_a^{-1}(v)\right\} \approx 2G_a^{-1}\left(\frac{\alpha+1}{2}\right)$$



Arthur Charpentier *arthur.charpentier@ensae.fr*

La corrélation linéaire de Pearson

$$\mathrm{cov}(X,Y) = E[(X-m_X)(Y-m_Y)] = E(XY) - m_X.m_Y$$

$$r(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y}$$

RQ X et Y indépendantes $\Rightarrow$ X et Y non corrélées, mais la réciproque est fausse

r(.,.) n'intègre que la dépendance ***linéaire*** entre X et Y :

$$r = +1 \text{ ssi } Y = aX+b \text{ avec } a>0$$
$$r = -1 \text{ ssi } Y = aX+b \text{ avec } a<0.$$

Aussi, r est invariante pour des transformations linéaires

$$r(aX+b,cY+d) = r(X,Y) \text{ pour a et c de même signe}$$

mais n'est pas invariante par transformation croissante,
$r(T(X),T(Y)) \neq r(X,Y)$ où T est une fonction croissante.

En particulier, $r(\log(X),\log(Y)) \neq r(X,Y)$
$r(\exp(X),\exp(Y)) \neq r(X,Y)$
$r(F_X(X), F_Y(Y)) \neq r(X,Y)$

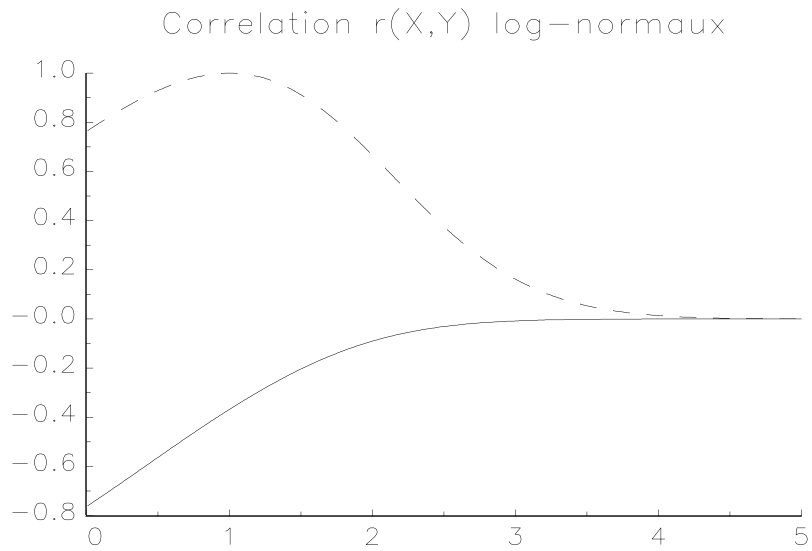
RQ La corrélation dépend des marginales, et toutes les valeurs comprises entre -1 et +1 ne sont généralement pas atteintes.

Etant données deux fonctions de répartition F_X et F_Y, il est parfois impossible de trouver une fonction de répartition F_{XY} telle que $r(X,Y) = 0.7$

D'après l'inégalité de Tchen, on a l'encadrement suivant

$$\frac{\mathrm{cov}\left(F_X^{-1}(U),F_Y^{-1}(1-U)\right)}{\sqrt{\mathrm{Var}(X_1)\mathrm{Var}(X_2)}} \leq r(X,Y) \leq \frac{\mathrm{cov}\left(F_X^{-1}(U),F_Y^{-1}(U)\right)}{\sqrt{\mathrm{Var}(X_1)\mathrm{Var}(X_2)}}$$

Ex Si $X \sim LN(0,1)$ et $Y \sim LN(0,s^2)$, alors r(X,Y) ne peut pas prendre toutes les valeurs comprises entre -1 et +1
si s=1 : $-0.37 \leq r \leq +1.00$
si s=2 : $-0.09 \leq r \leq +0.67$
De façon plus générale, les bornes des intervalles des corrélations atteignables, en fonction de s, ont l'allure suivante



Présentation probabiliste des copulas

Def En dimension 2, un copula est une application
$C : [0,1]\times[0,1] \to [0,1]$ telle que

- $C(u,0) = C(0,v) = 0$ pour tout $u,v \in [0,1]$
- $C(u,1) = u$ et $C(1,v) = v$ pour tout $u,v \in [0,1]$
- pour tout $u_1<u_2$ et $v_1<v_2$, dans $[0,1]$

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

On a alors les résultats suivants

- $C(u,v)$ est 1-lipschitzienne

$$|C(u_1,v_1) + C(u_2,v_2)| \leq | u_1-u_2 | + | v_1-v_2 |$$

donc C est absolument continue

- $C(u,v)$ est borné : pour tout u,v

$$C^-(u,v) \leq C(u,v) \leq C^+(u,v) \quad \textit{[bornes de Fréchet]}$$

où $C^-(u,v) = \max\{u+v-1,0\}$ et $C^+(u,v) = \min\{u,v\}$

Théorème de Sklar

si X et Y sont des variables aléatoires de fonctions de répartitions F_X et F_Y respectivement, avec pour loi jointe F_{XY}, alors il existe un copula C tel que

$$F_{XY}(x,y) = C(F_X(x),F_Y(y))$$

donné par $C(u,v) = F_{XY}(F_X^{-1}(u),F_Y^{-1}(v))$
où la notation $^{-1}$ désigne l'inverse généralisé, au sens où

$$\varphi^{-1}(t)=\inf\{s,\varphi(s)\leq t\},$$

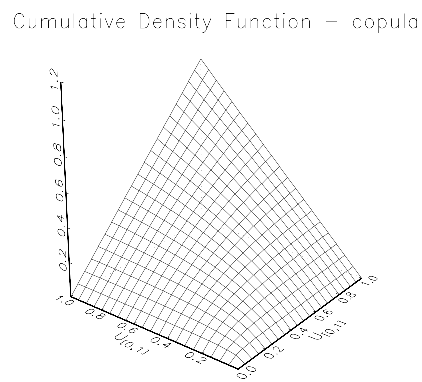
Arthur Charpentier *arthur.charpentier@ensae.fr*

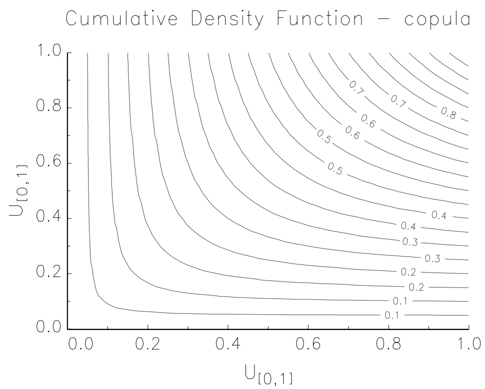
Gestion des risques multiples en assurance

- si X et Y sont indépendants, alors $C(u,v)=uv$,
- si $X=\varphi(Y)$ ou si $Y=\varphi(X)$ où φ est une fonction strictement croissante, alors $C(u,v)=\min(u,v)$ (*borne supérieure de Fréchet*) : cas de comonotonie,
- si $X=\psi(Y)$ ou si $Y=\psi(X)$ où ψ est une fonction strictement décroissante, alors $C(u,v)=\max(u+v-1,0)$ (*borne inférieure de Fréchet*) : cas d'anticomonotonie.
- Le copula du couple (X,Y) est la fonction de répartition du couple (U,V) où $U=F_X(X)$ et $V=F_Y(Y)$. Les variables U et V sont appelées transformations quantiles
- Le copula du couple (f(X),g(Y)) est le même que celui de (X,Y) si les transformation f et g sont croissantes
- L'espace des fonctions copulas est convexe

Propriétés de copulas

Les copulas C(u,v) sont des fonctions de répartition dont les lois marginales sont uniformes

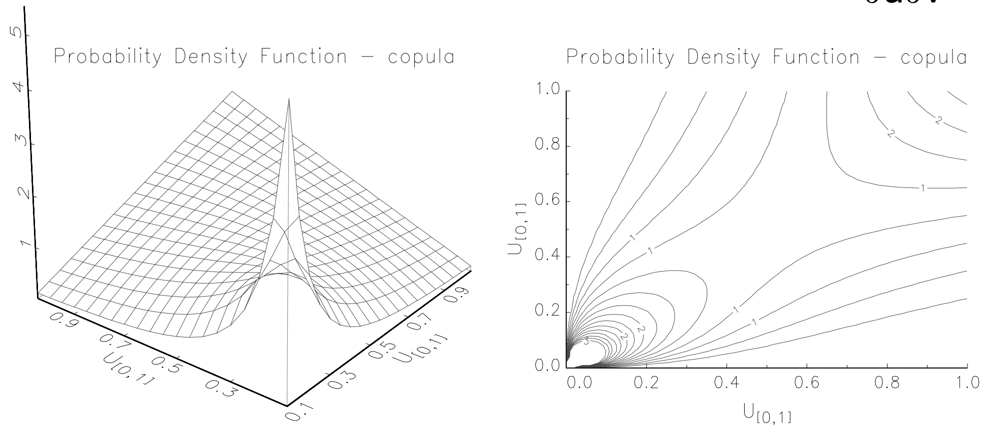




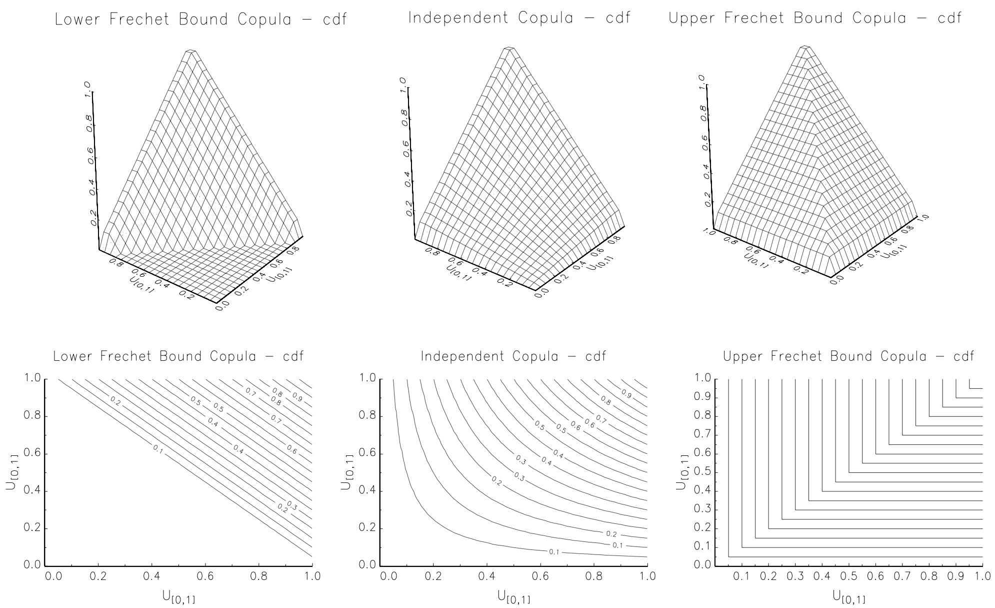
Arthur Charpentier *arthur.charpentier@ensae.fr*

Gestion des risques multiples en assurance

Densité du copula c(u,v) (*si elle existe*) $c(u,v)=\dfrac{\partial^2 C(u,v)}{\partial u \partial v}$



Copulas particuliers : $C^-(u,v)$, $C^\perp(u,v)$ et $C^+(u,v)$



Gestion des risques multiples en assurance

Rq En dimension d, la notion de '*croissance*' du copula C se fait de la façon suivante : soient [x,y] un hypercube de $[0,1]^d$, on appelle C-volume de l'hypercube [x,y]

$$V_C([x_1,y_1]\times\ldots\times[x_d,y_d])=\sum_{z=(z_1,\ldots,z_d)}\sigma(z)C(z_1,\ldots,z_d)$$

où la somme se fait sur l'ensemble des sommets de l'hypercube, et où la fonction $\sigma(z)$ vaut +1 si $z_i=x_i$ pour un nombre pair i, -1 sinon. C sera dite supermodulaire sur V_C est positive pour tout hypercube [x,y]

Ex En dimension 3, la condition de croissance s'écrit

$$C(x2,y2,z2) - C(x1,y2,z2) - C(x2,y1,z2) - C(x2,y2,z1)+ C(x2,y1,z1)+C(x1,y2,z1)+C(x1,y1,z2) - C(x1,y1,z1) \geq 0$$

Prop Le théorème de Sklar reste vrai en dimension d,

$F(x) = C(F_1(x_1),\ldots,F_d(y_d))$ et $C(u) = F(F_1^{-1}(u_1),\ldots,F_d^{-1}(u_d))$
dans le cas où les lois marginales sont continues.

Prop Les bornes de Fréchet sont toujours les mêmes

$$C^-(u_1,\ldots,u_d) \leq C(u_1,\ldots,u_d) \leq C^+(u_1,\ldots,u_d)$$

où $C^-(u,v)=\max\{u_1+\ldots+u_d-d+1,0\}$
$C^+(u,v)=\min\{u_1,\ldots,u_d\}$

Rq C^- n'est pas un copula si d>2

Rq Il n'est pas équivalent de travailler sur le copula du vecteur $(X_1,\ldots,X_n)$ et sur l'ensemble des copulas des couples (X_i,X_j) pour $i\neq j$

Présentation statistique des copulas

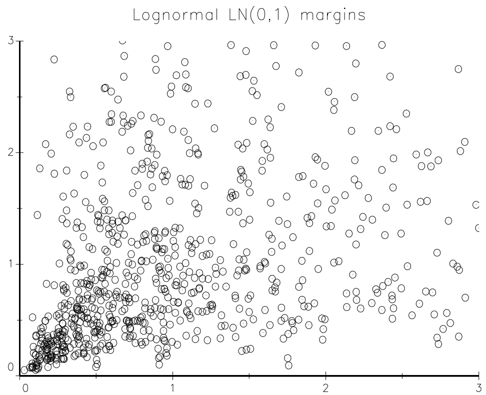
On ne travaille plus sur la variable X mais sur $U=F_X(X)$
On considère des couples d'observations

$$(X_1,Y_1), (X_2,Y_2), \ldots, (X_n,Y_n)$$

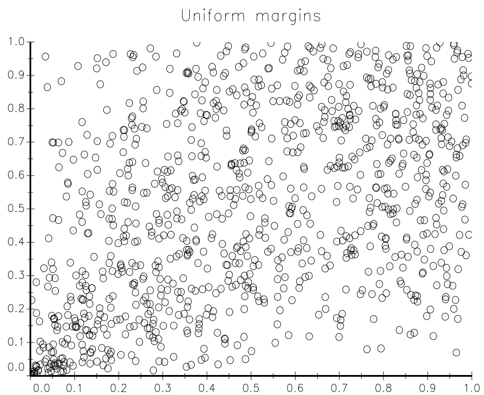
L'idée est de ne plus travailler en volume *(en montant ou en nombre)* sur les variables X_i ou Y_i mais en rang/centile *(en pourcentage)* R_i ou S_i

où R_i = rang de X_i au sein des $X_1,\ldots,X_n$
S_i = rang de Y_i au sein des $Y_1,\ldots,Y_n$

On travaille alors non plus sur les couples (X_i,Y_i) mais sur les couples (R_i,S_i), ou - à *une transformation linéaire près* – les couples (U_i,V_i), où $U_i=[R_i-1/2]/n$ et $V_i=[S_i-1/2]/n$



Echantillon (X_i,Y_i)



Echantillon (U_i,V_i)

Gestion des risques multiples en assurance

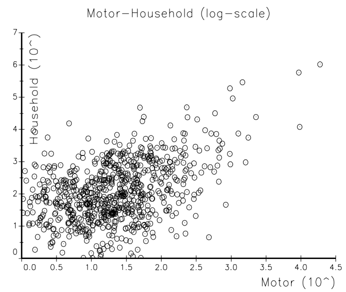
Rq Le copula du couple (X,Y) a été obtenu en rendant les lois marginales uniformes sur [0,1].

On pourrait, de façon analogue se ramener à d'autres types de lois (ex: N(0,1))

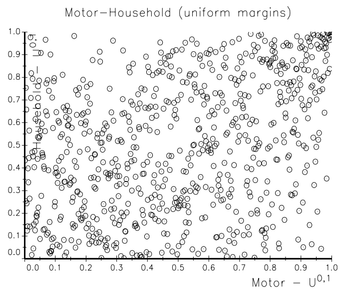
Ex : le couple (U,V) est défini par $U=F_X(X)$ et $V=F_Y(Y)$

on peut considérer le couple (X_N,Y_N) où $X_N=\Phi^{-1}(U)$ et $Y_N=\Phi^{-1}(V)$ où Φ est la fdr de la loi N(0,1)

Les (X_i,Y_i) correspondent à la charge totale journalière pour des contrats auto (X) et habitation (Y)

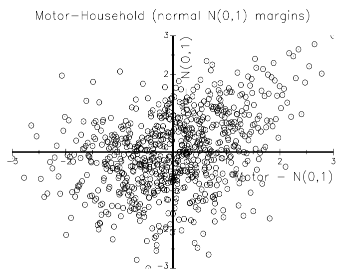


Les (U_i,V_i) sont les variables uniformes associées : $U_i=F_X(X_i)$ obtenues à l'aide des statistiques de rang



Les $(X_{N,i},Y_{N,i})$ sont les variables « normalisées » $X_{N,i}=\Phi^{-1}(U_i)$

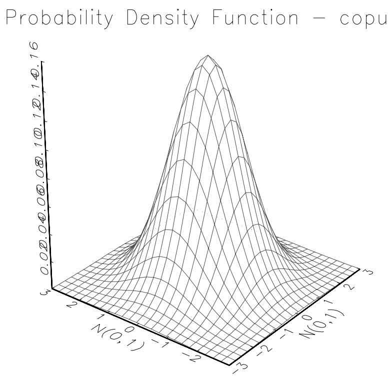
Cette transformation permet de comparer la *structure de dépendance* au cas d'un vecteur gaussien

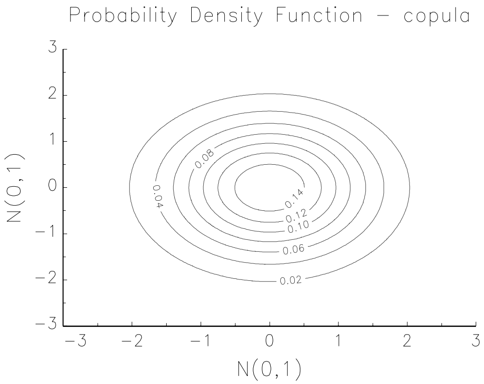


Arthur Charpentier *arthur.charpentier@ensae.fr*

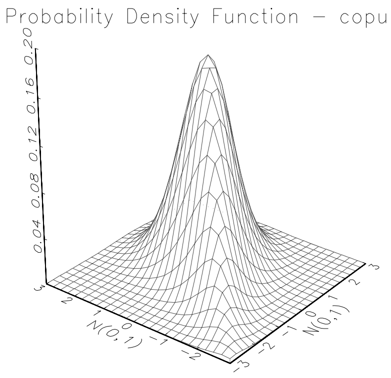
Gestion des risques multiples en assurance

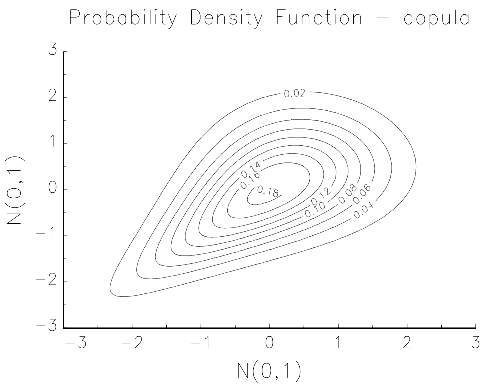
- cas **indépendant**





- cas **non-indépendant**





Cette transformation permet d'obtenir des représentations graphiques de la forme de la dépendance entre les variables X et Y plus pertinentes, à comparer avec les ellipses obtenues dans le cas gaussien.

le copula C est la fonction de répartition de $(F_X(X),F_Y(Y))$

Le « *copula normalisé* » est la fonction de répartition du couple $(\Phi^{-1}(F_X(X)), \Phi^{-1}(F_Y(Y)))$

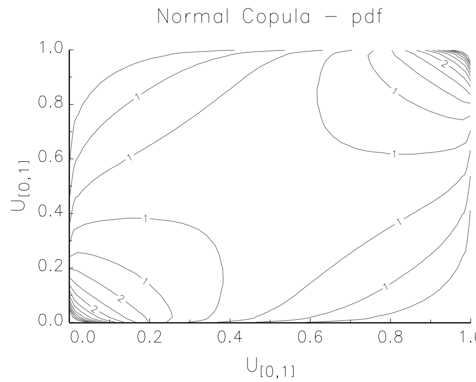
Arthur Charpentier *arthur.charpentier@ensae.fr*

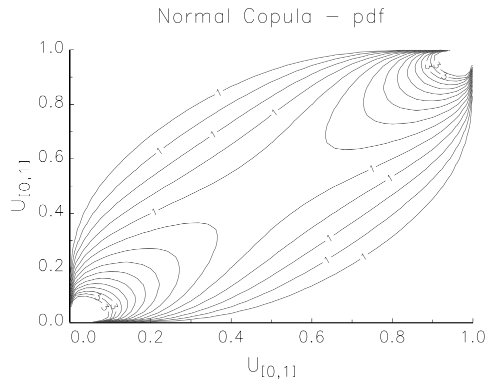
Gestion des risques multiples en assurance

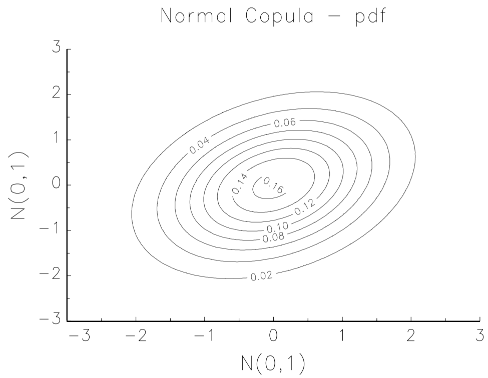
- copula **gaussien** ou normal

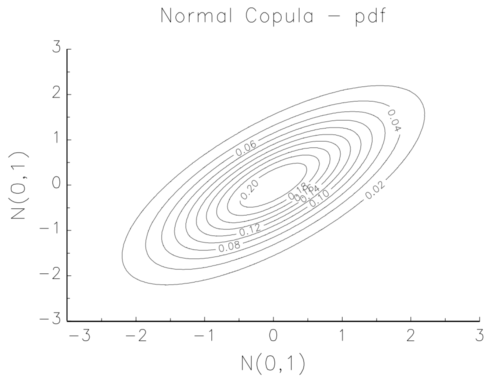
$$C(u,v)=\Phi_r(\Phi^{-1}(u), \Phi^{-1}(v)) \text{ pour } -1 \leq r \leq +1$$

où $\Phi_r(x,y)$ est la fonction de répartition de la loi normale centrée réduite bivariée, de coefficient de corrélation r, et $\Phi(x)$ est la fonction de répartition de la loi normale N(0,1)

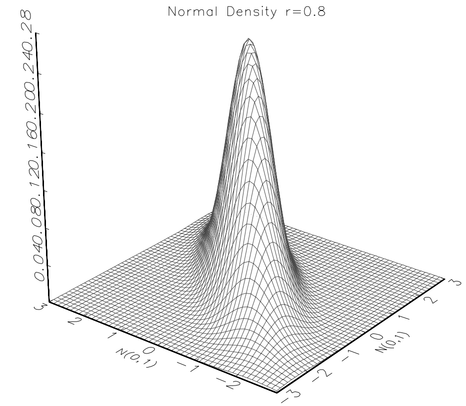


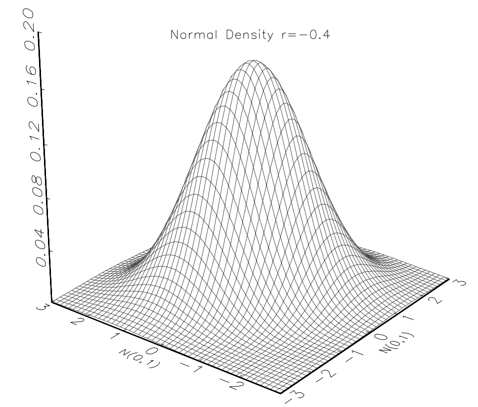


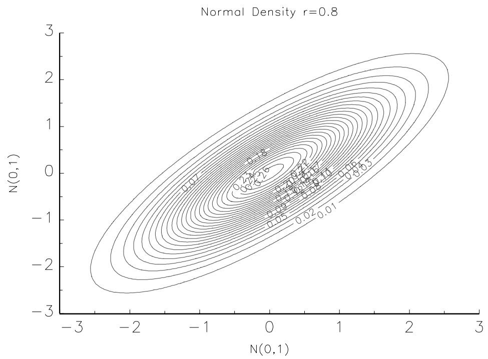


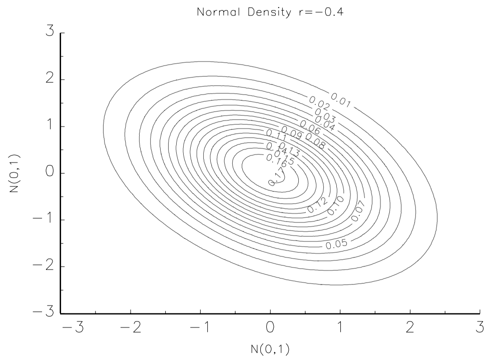


Arthur Charpentier *arthur.charpentier@ensae.fr*









- copula **gaussien** ou normal

Rq Le copula Gaussien a été utilisé dabs CreditMetrics *(JPMorgan)* pour simuler des probabilités de changement de rating d'émetteur

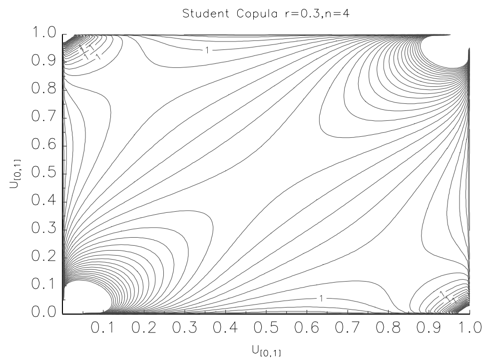
La loi normale multivariée est LA loi continue multivariée la plus connu, elle est simulable facilement, et il existe des méthodes d'inférence statistique pour estimer tous les paramètres.

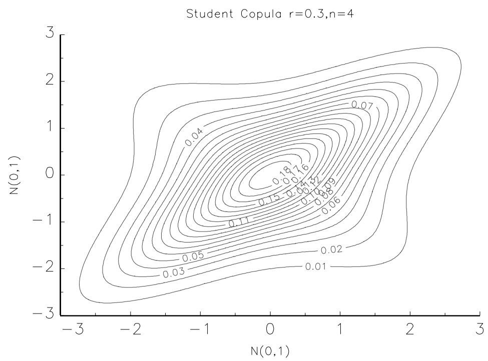
On connaît $X \sim N(0,I)$, on pose $Y = \mu + \Sigma^{-1/2} X \sim N(\mu,\Sigma)$, où Σ est une matrice de covariance.

On peut construire d'autres lois multivariées à partir de Y

- ***Loi de Student Multivariée***

Permet de définir le **copula de Student**, qui dépend d'une matrice de corrélation (R), et d'un nombre de degrés de liberté (n)

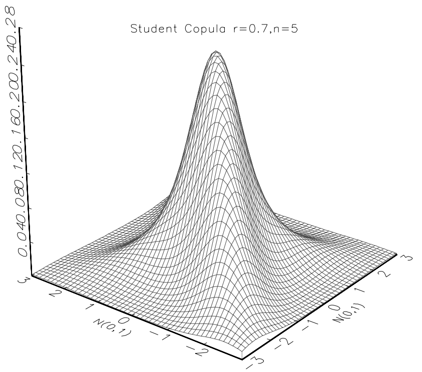


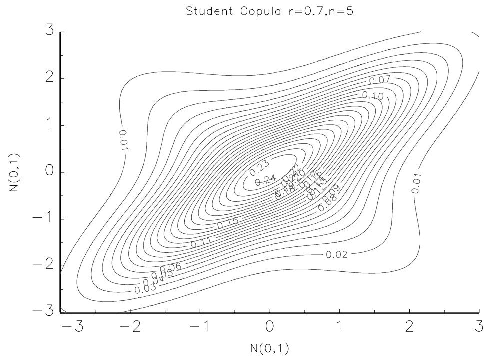


Gestion des risques multiples en assurance

- copula **gaussien** ou normal

Le **copula de Student**, associée à des lois marginales normales est représenté ci-dessous,

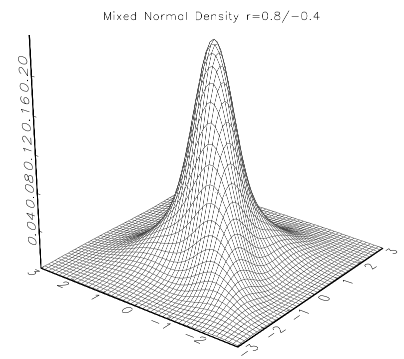


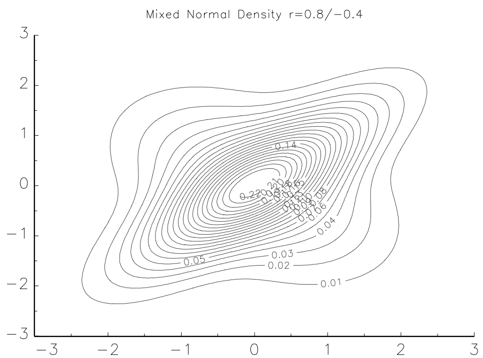


- ***Combinaisons linéaires de lois normales***

Une autre possibilité peut être de générer des lois multivariées en considérant des combinaisons linéaires de lois normales bivariées

Ex Z=α.X+(1- α).Y où X et Y sont de vecteurs normaux de matrices de covariances différentes

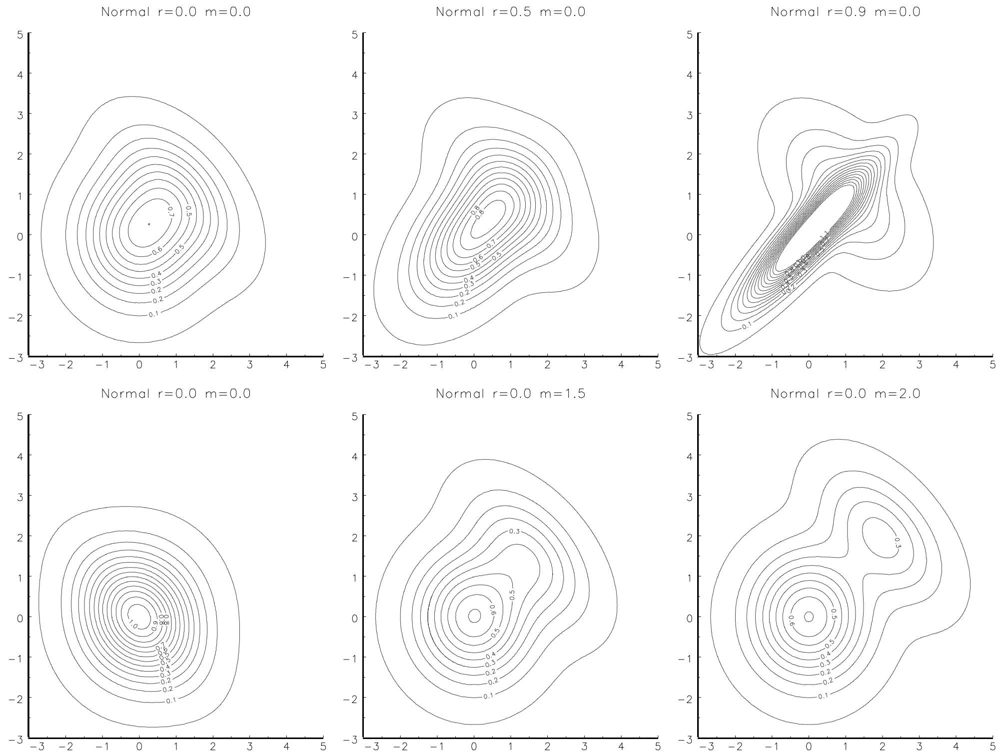


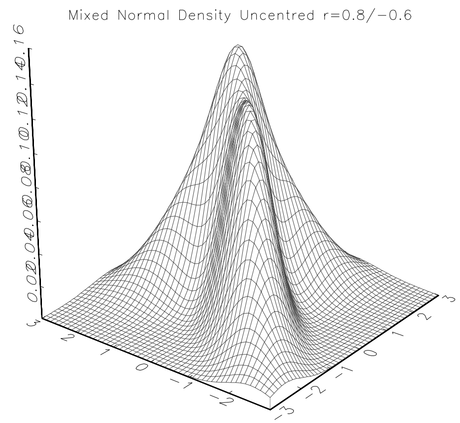


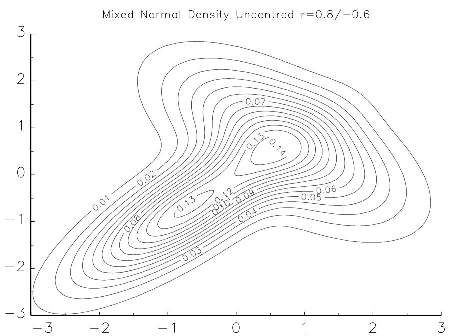
Arthur Charpentier *arthur.charpentier@ensae.fr*

Gestion des risques multiples en assurance

Ex Z=α.X+(1- α).Y où X et Y sont de vecteurs normaux de moyennes différentes permet d'engendrer toutes formes de distributions bivariées



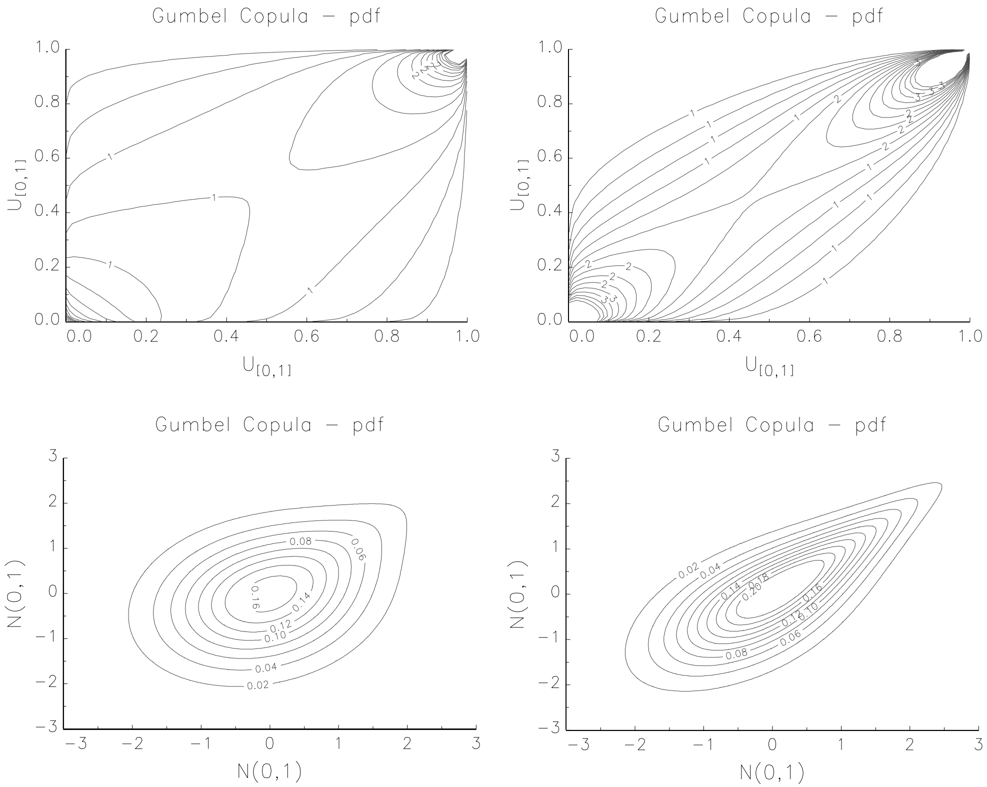




Arthur Charpentier *arthur.charpentier@ensae.fr*

• copula de **Gumbel**

$C(u,v,\theta)=\exp(-[[-\log(u)]^{\theta}+[-\log(v)]^{\theta}]^{1/\theta}$ pour $\theta \geq 1$

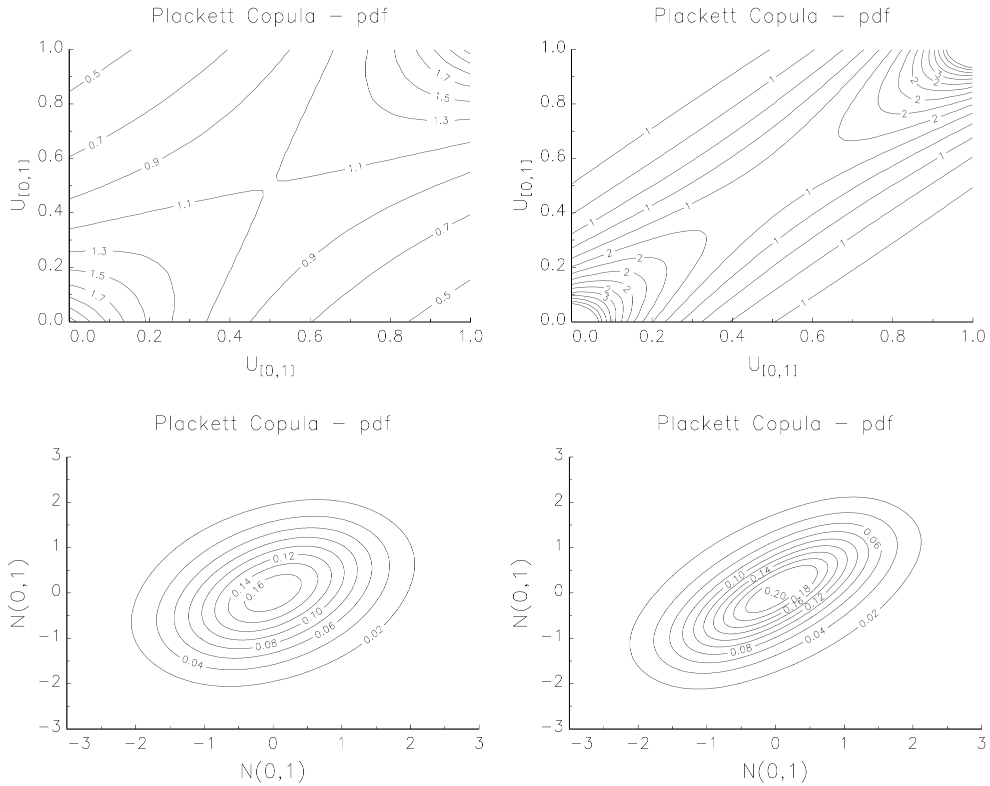


Rq ce copula est un copula extrême

Gestion des risques multiples en assurance

• copula de **Plackett**

$$C(u,v)=\frac{1}{2}\frac{1}{\theta-1}\left[1+(\theta-1)(u+v)-\sqrt{(1+[\theta-1](u+v))^2-4\theta(\theta-1)uv}\right]$$



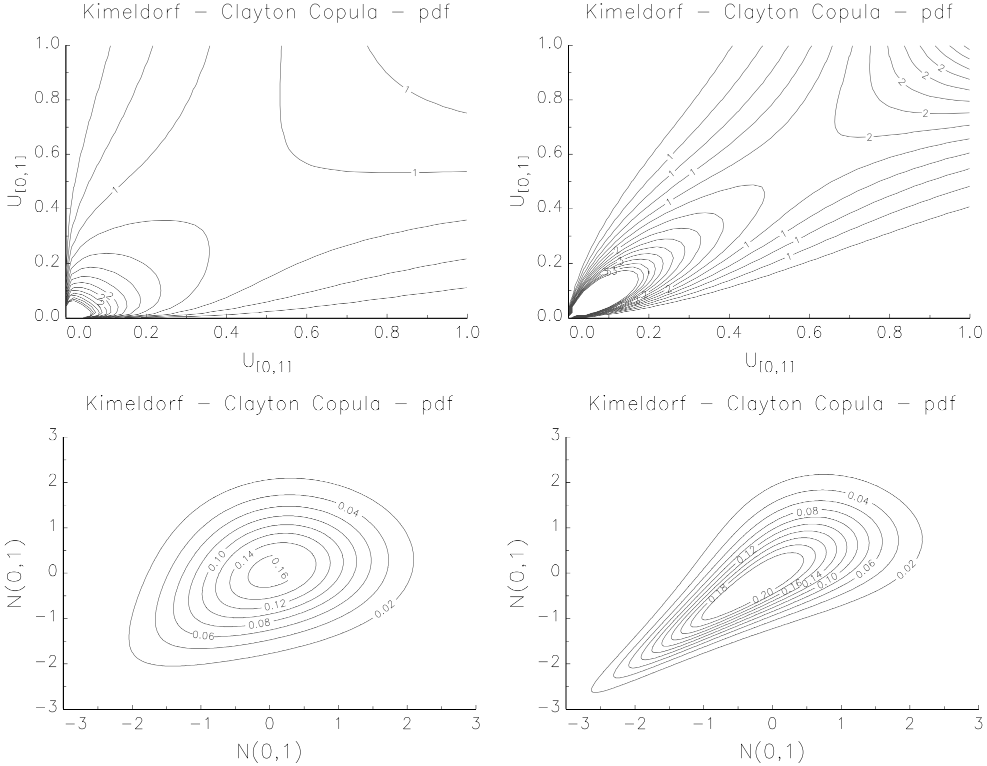
Ce copula vérifie la propriété suivante

$$\frac{P(X\leq x, Y\leq y).P(X>x, Y>y)}{P(X>x, Y\leq y).P(X\leq x, Y>y)}=\theta$$

Gestion des risques multiples en assurance

• copula de **Clayton** ou de Kimeldorf

$$C(u,v,\theta)=(u^{-\theta}+v^{-\theta}-1)^{-1/\theta} \text{ pour } \theta \geq 1$$



Ce copula peut se généraliser en dimension supérieure à 2, sous la forme

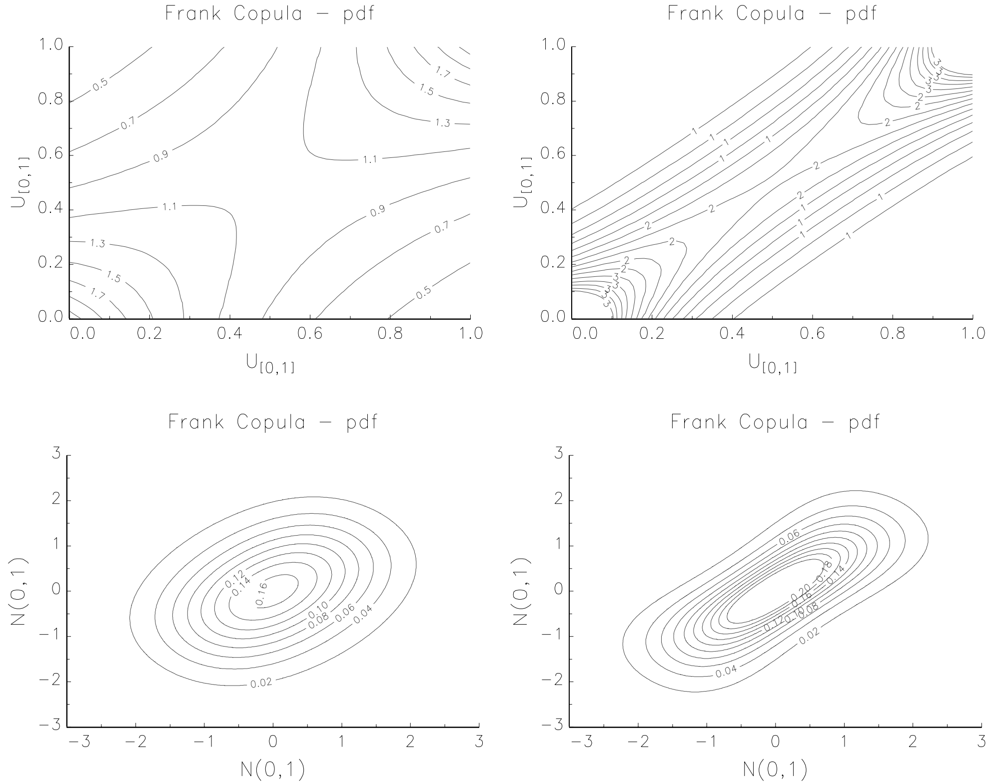
$$C(u_1,\ldots, u_p)=[u_1^{-\theta}+\ldots+u_p^{-\theta}-(n-1)]^{-1/\theta} \text{ pour } 0 \leq \theta$$

Sous cette forme, la corrélation est toujours positive. La corrélation négative est obtenue en effectuant la transformation U en 1-U

Gestion des risques multiples en assurance

• copula de **Frank**

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

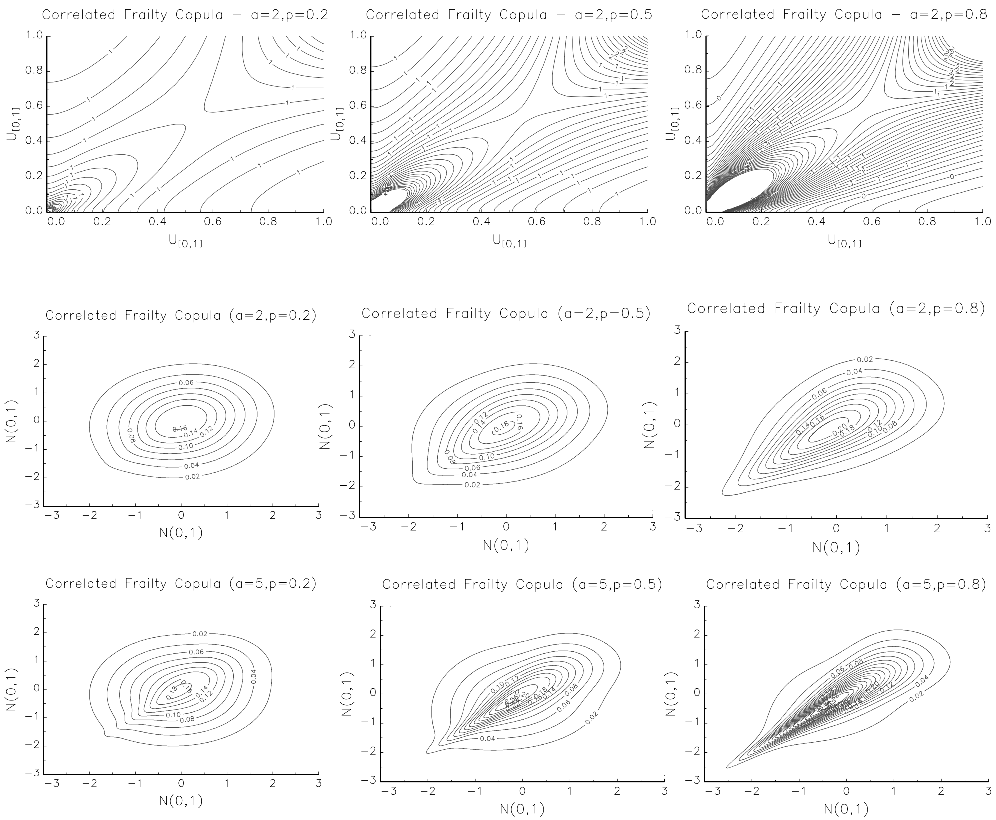


Gestion des risques multiples en assurance

- copula **correlated frailty**

$$C(u,v)=\frac{(uv)^{1-p}}{\left(u^{-a}+v^{-a}-1\right)^{p/a}}=C^{\perp}(u,v)^{1-p}C_a(u,v)^p$$

où $0<p<1$ et $a>0$, et où C_a est le copula de Clayton



Rq Les moyennes géométriques pondérées ne sont pas toujours des copulas

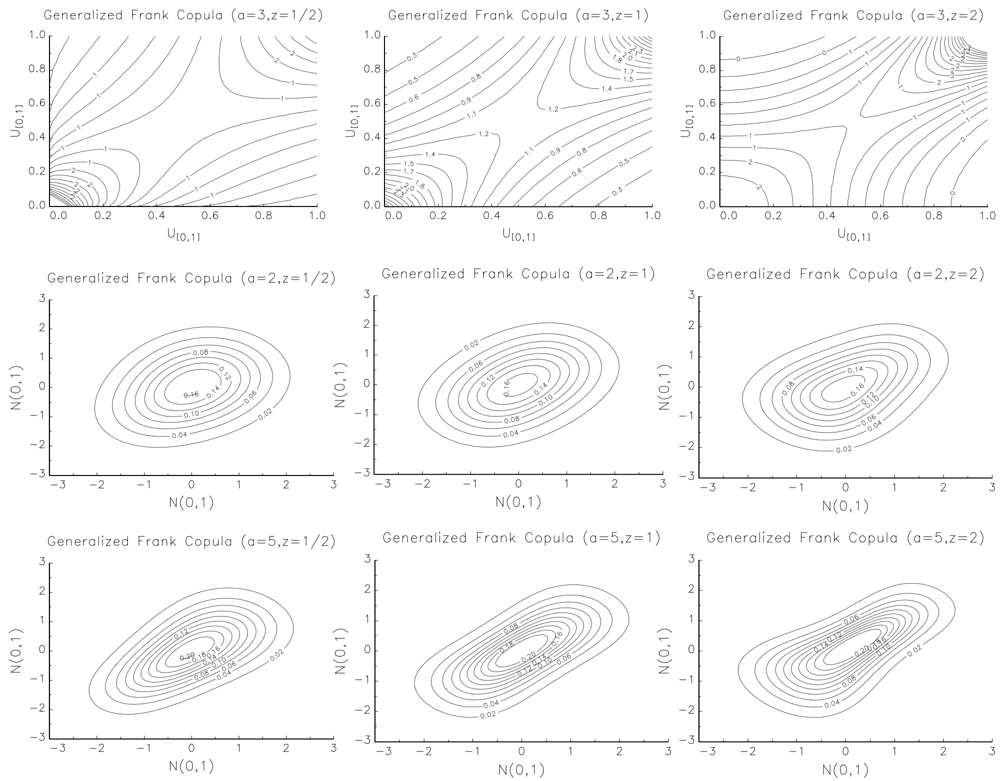
Arthur Charpentier *arthur.charpentier@ensae.fr*

Gestion des risques multiples en assurance

- copula de **Frank généralisé**

$$C(u,v)=\left[-\frac{1}{\theta}\log\left(1+\frac{(\exp(-\theta u^{\zeta})-1)(\exp(-\theta v^{\zeta})-1)}{\exp(-\theta)-1}\right)\right]^{1/\zeta}=C_\theta(u^{\zeta},v^{\zeta})^{1/\zeta}$$

où $\zeta\in[0,1]$, $\theta>0$, et où C_θ est le copula de Frank

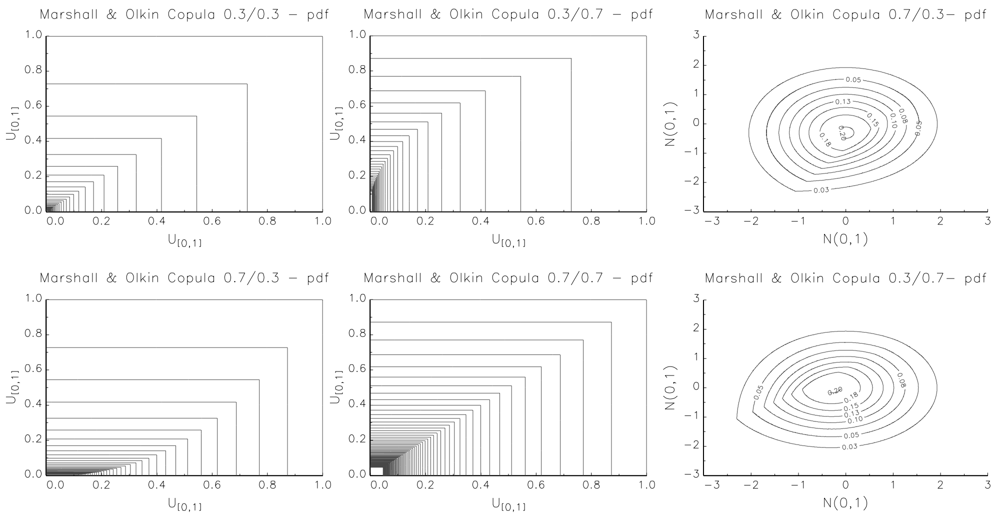


Rq Il est possible de construire de nouvelles familles de copulas en considérant des transformations de la forme $(u,v)\to C(u^{\zeta},v^{\zeta})^{1/\zeta}$. De façon plus générale, pour toute application a croissante bijective et dérivable, $(u,v)\to a^{-1}(C(a(u),a(v))$ défini un copula.

Arthur Charpentier *arthur.charpentier@ensae.fr*

• copula de **Marshall-Olkin**

$$C_{\alpha,\beta}(u,v)=\min\{vu^{1-\alpha}, uv^{1-\beta}\} \text{ où } \alpha,\beta\in[0,1].$$



Rq Ce copula est obtenu en considérant le copula de survie d'un vecteur (X,Y) où, étant donnés trois variables aléatoires indépendantes Z_1, Z_2, et Z_{12} distribuées suivant des lois exponentielles, on pose $X=\min\{Z_1,Z_{12}\}$, $Y=\min\{Z_2,Z_{12}\}$, puis $\alpha=\lambda_{12}[\lambda_1+\lambda_{12}]$ et $\beta=\lambda_{12}[\lambda_2+\lambda_{12}]$

Rq Il est possible de considérer des généralisations de la forme $C_{\phi,\psi}(u,v)=\min\{v\phi(u), u\psi(v)\}$ où ϕ et ψ sont des fonctions croissantes telles que $\phi(0)=\psi(0)=0$, $\phi(1)=\psi(1)=1$, et telles que $t\rightarrow\phi(t)/t$ et $t\rightarrow\psi(t)/t$ soient décroissantes sur]0,1].

• copula **Archimédien**

Cette famille de copulas est à paramètre fonctionnel : Soit $\varphi:[0,1]\rightarrow[0,+\infty]$ une fonction convexe et strictement décroissante, telle que $\varphi(1)=0$. La fonction

$$C(u,v)=\varphi^{-1}(\varphi(u)+\varphi(v)) \text{ pour tout } 0\leq u,v\leq 1$$

est un copula, appelé copula Archimédien. La fonction φ est appelée générateur du copula.

Ex $\varphi(t)=t^{-\theta}-1$: copula de Clayton
$\varphi(t)=\exp(-t^{1/\theta})$: copula de Gumbel
$\varphi\ \phi(t)=-\log[(e^{-\theta t}-1)/(e^{-\theta}-1)]$:copula de Frank

Rq La classe des copulas Archimédiens peut être définie en utilisant une représentation factorielle : soit Z une variable positive telle que X et Y soient indépendantes conditionnellement à Z, et telle que

$$P(X\leq x|Z)=G_X(x)^Z \text{ et } P(Y\leq y|Z)=G_Y(y)^Z$$

où G_X et G_Y sont des fonctions de répartition. En notant ψ la transformée de Laplace de Z, on peut alors montrer que le copula du couple (X,Y) peut s'écrire

$$C(u,v)=\psi(\psi^{-1}(u)+\psi^{-1}(v))$$

Compléments sur les structures de dépendance

• (X,Y) est PQD si et seulement si $C^{\perp}(u,v) \leq C(u,v)$ pour tout u,v dans [0,1]

• (X,Y) est CI, si X est conditionnellement croissante en Y et réciproquement (P(X>x|Y=y) est croissante en y ,et P(Y>y|X=x) est croissante en x) si et seulement si

$t \to C(t,v)$ et $t \to C(u,t)$ sont concaves pour tout u et v

Ex Un copula Archimédien de générateur ϕ deux fois différentiable tel que $\phi(0)=+\infty$ est CI si et seulement si

$$-\frac{\partial}{\partial u}\log\phi'(u) \geq \frac{\partial}{\partial u}\log\frac{\partial C(u,v)}{\partial v} \quad \text{et} \quad -\frac{\partial}{\partial v}\log\phi'(v) \geq \frac{\partial}{\partial v}\log\frac{\partial C(u,v)}{\partial u}$$

Rq une autre structure de dépendance entre X et Y peut être définie si C est TP2 (*totally positive of order 2*), c'est à dire si

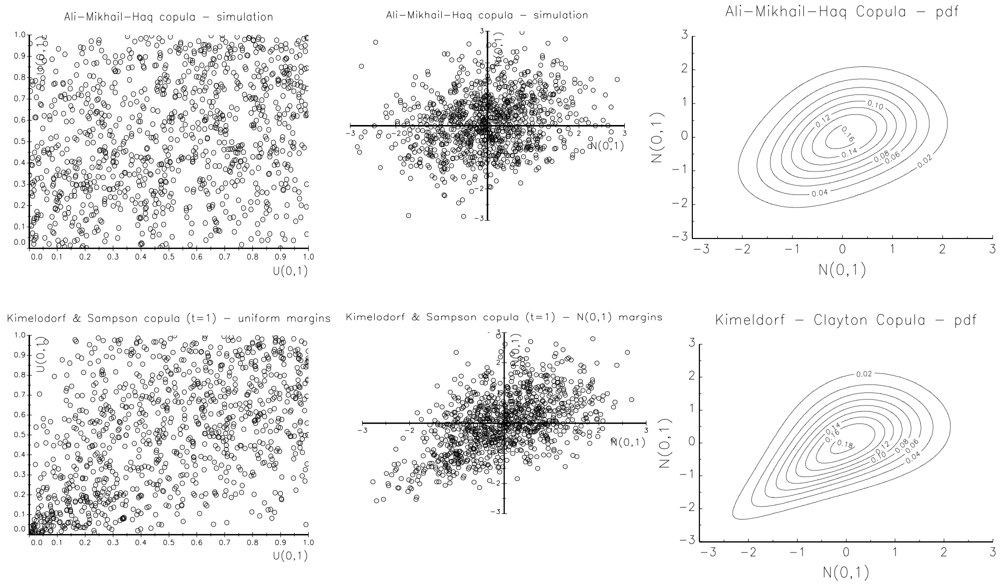
$$\det\begin{pmatrix} C(u_1,v_1) & C(u_1,v_2) \\ C(u_2,v_1) & C(u_2,v_2) \end{pmatrix} \geq 0 \text{ pour } u_1 \leq u_2,\ v_1 \leq v_2$$

(on dira alors que (X,Y) est LCSD – left corner set deacreasing)

Arthur Charpentier *arthur.charpentier@ensae.fr*

Gestion des risques multiples en assurance

Simulation de vecteurs aléatoires



Il existe des algorithmes pour simuler un certain nombre de copulas paramétriques

La méthode générale *(présentée ici en dimension2)* est

1) Simuler N couples de variables uniformes indépendantes $(u_n^{\perp}, v_n^{\perp})$

2) Poser $\begin{cases} u_n = u_n^{\perp} \\ v_n \text{ solution de } \partial_1 C(u_n,v_n)=v_n^{\perp} \end{cases}$

où $\partial_1 C(u,v)= \partial_1 C(u,v)= \partial C(u,v)/\partial u$

3) Poser $x_n = F_X^{-1}(u_n)$ et $y_n = F_Y^{-1}(v_n)$

Arthur Charpentier *arthur.charpentier@ensae.fr*

Copulas et théorie asymptotique

Lien avec la théorie des valeurs extrêmes

Considérons ici le couple $(X_{(n)},Y_{(n)})$ où $X_{(n)}=\sup(X_1,...,X_n)$ et $Y_{(n)}=\sup(Y_1,...,Y_n)$

Soit H_{XY} une fonction de répartition limite, au sens où il existe des suites a_n, b_n, α_n et β_n telles que

$$([X_{(n)}-a_n]/b_n \,,\, [Y_{(n)}-\alpha_n]/\beta_n)$$

converge en loi vers une variable de fonction de répartition H_{XY}. Soit C le copula associé à H_{XY}, alors C vérifie $C(u^\lambda,v^\lambda)=C(u,v)^\lambda$ pour tout u,v où λ est un réel positif donné.

Prop La copule de la loi limite s'écrit

$$C(u,v)=\exp\left(\log(uv).A\left(\frac{\log(u)}{\log(uv)}\right)\right)$$

où A est une fonction convexe, telle que $A(0)=A(1)=1$ et qui vérifie $\max\{w,1-w\}\leq A(w)\leq 1$.

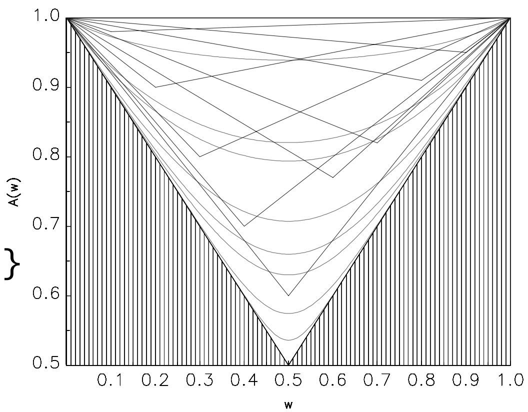
Ex Les copulas de Gumbel ont pour générateur

$$A(w)=(w^\alpha+(1-w)^\alpha)^{1/\alpha}$$

Les fonctions

$$A(w)=\max\{1-\alpha w,1-\beta(1-w)\}$$

génèrent les copulas de Marshall-Olkin



Arthur Charpentier *arthur.charpentier@ensae.fr*

Gestion des risques multiples en assurance

Estimation Statistique de copulas

Modélisation paramétrique

Utilisation de *mesures de dépendance*
On souhaite que la mesure a vérifie les propriétés suivantes
(i) symétrie $\alpha(X,Y) = \alpha(Y,X)$
(ii) normalisation $-1 \leq \alpha(X,Y) \leq +1$
(iii-1) indépendance $\alpha(X,Y)=0$ ***ssi*** X et Y indépendants
(iv) comonotonie $\alpha(X,Y)=+1$ ***ssi*** il existe ϕ croissante telle que $X=\phi(Y)$ ou telle que $Y=\phi(X)$
(v) anticomonotonie $\alpha(X,Y)=-1$ ***ssi*** il existe ϕ décroissante telle que $X=\phi(Y)$ ou telle que $Y=\phi(X)$
(vi) invariance pour f et g strictement monotones, $\alpha(f(X),g(Y)) = \alpha(X,Y)$

Toutefois, il n 'existe aucune mesure vérifiant la propriété d 'invariance et celle d 'indépendance. On se contente donc généralement de

- **(iii-2) indépendance** $\alpha(X,Y)=0$ ***si*** X et Y indépendants

Arthur Charpentier *arthur.charpentier@ensae.fr*

Gestion des risques multiples en assurance

Estimation Statistique de copulas

• La **corrélation linéaire** (Pearson) n'est pas utilisée parce qu'elle n'est pas invariante par transformation croissante des marges :

$$r(f(X),g(Y)) \neq r(X,Y)$$

où f et g sont des fonctions croissantes

On utilise généralement le rho de Spearman et de tau de Kendall qui peuvent s'exprimer en fonction du copula, et donc, sont invariants par transformation croissante des marges

• Le **tau de Kendall** est défini par

$$\tau(X,Y) = E(UV) \text{ où } U=F_X(X) \text{ et } V=F_Y(Y)$$

$$\tau(X,Y) = 4\int_{[0,1]\times[0,1]} C(u,v)dC(u,v)$$

• Le **rho de Spearman** par

$$\rho(X,Y) = 12 \int_{[0,1]\times[0,1]} uvdC(u,v) - 3$$

$$\rho(X,Y) = 12 \int_{[0,1]\times[0,1]} C(u,v)dudv - 3$$

Ces deux mesures ne dépendent pas des comportements des lois marginales, mais uniquement de la structure de dépendance C : pour toutes fonctions croissantes

$$\tau(f(X),g(Y))= \tau(X,Y) \text{ et } \rho(f(X),g(Y))=\rho(X,Y)$$

Arthur Charpentier *arthur.charpentier@ensae.fr*

Gestion des risques multiples en assurance

Estimation Statistique de copulas

Il est possible de montrer que

$$\rho(X,Y) = 12 \int_{[0,1]\times[0,1]} [C(u,v)-uv]dudv$$

$$\rho(X,Y) = 12 \int_{[0,1]\times[0,1]} [C(u,v)- C^{\perp}(u,v)]dudv$$

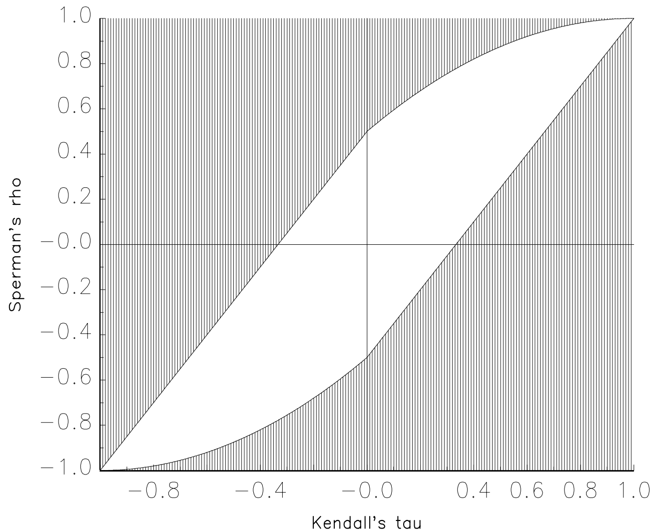
De façon générale, il est possible de définir des distances dans L^p, par

$$d_p(X,Y) = [K_p\int \int_{[0,1]\times[0,1]} |C(u,v)-C^{\perp}(u,v)|^p dudv]^{1/p}$$

Où la constante K_p est telle que $d_p(X,Y)=+1$ dans le cas comonotonique, c'est à dire $C = C^+$

Rq Ces deux mesures de dépendances (rho et tau) sont liées : il est possible d'avoir $\rho(X,Y)=0$ et $\tau(X,Y)=0.5$

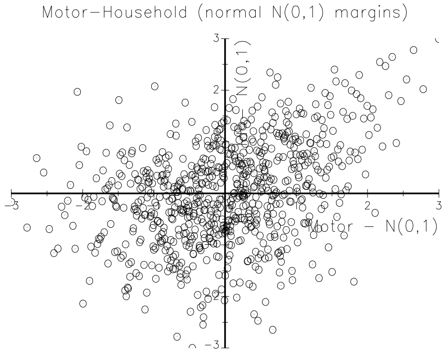
En fait, ρ et τ ont liés par une relation simple

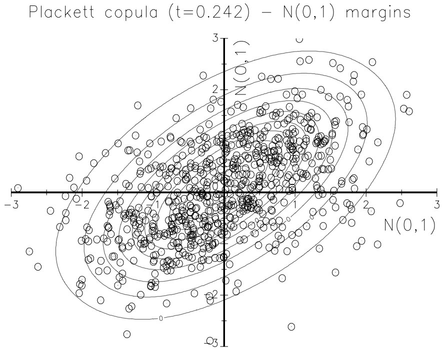


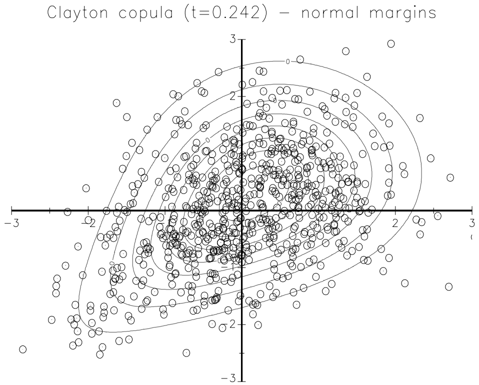
Arthur Charpentier *arthur.charpentier@ensae.fr*

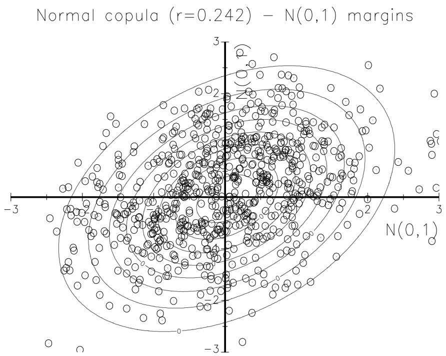
Estimation Statistique de copulas

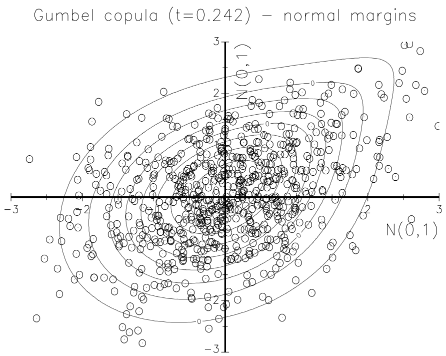
Ex Copula du couple charge totale journalière pour des contrats auto (X) et habitation (Y) normalisés







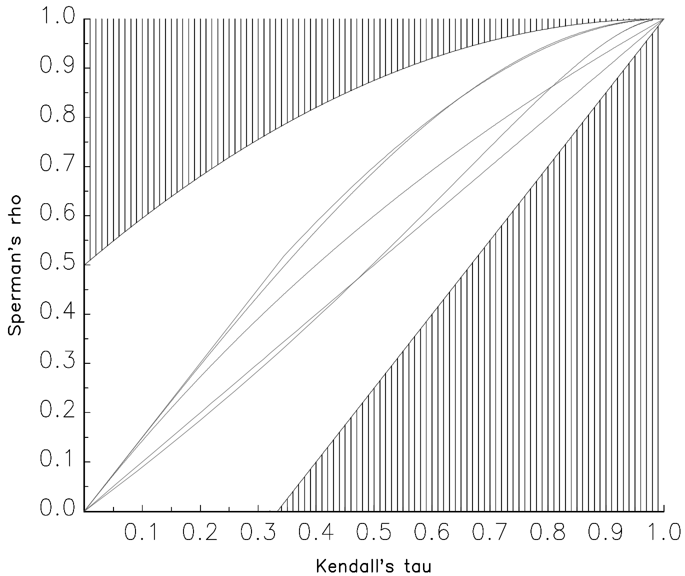




Gestion des risques multiples en assurance

Estimation Statistique de copulas

Pb Pour les copulas *usuels* à un paramètre, seuls quelques cas de dépendance peuvent être atteint



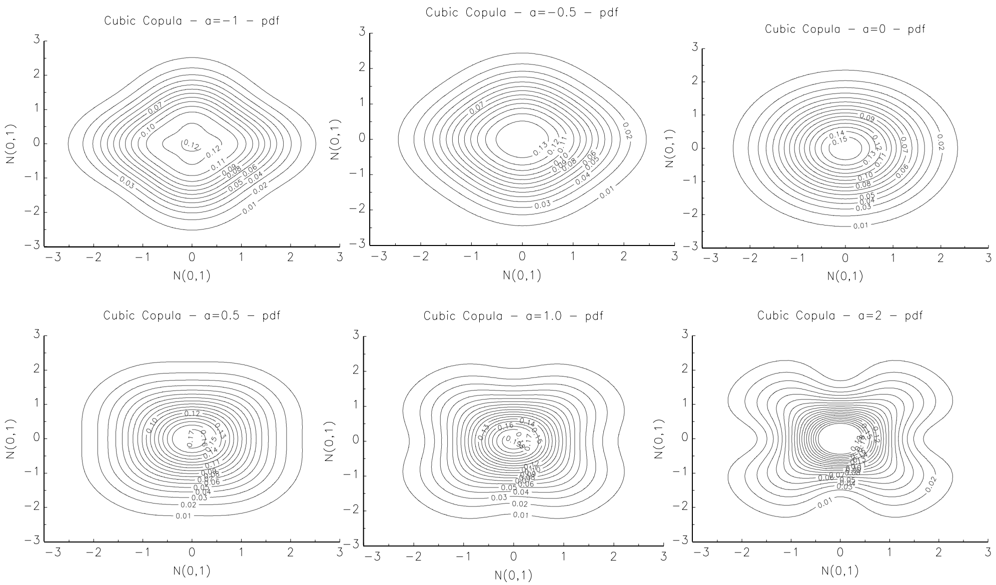
Quelles sont les régions atteignables pour des familles particulières de lois, par exemple, les copulas Archimédiens ?

Arthur Charpentier *arthur.charpentier@ensae.fr*

Estimation Statistique de copulas

Pb le fait d'avoir ρ=0 et τ=0 n'implique pas forcément que X et Y sont indépendante

Ex copula cubique



$$C(u,v,\theta)=uv+ \theta[u(u-1)(2u-1)]\,[v(v-1)(2v-1)]$$

où $\theta\in[-1,+2]$, alors, pour tout θ, ρ=0 et τ=0

• Le **lower tail dependence coefficient** est défini par

$$\lambda_L = \lim_{\alpha\to 0} P\left(X \le F_X^{-1}(\alpha) \middle| Y \le F_Y^{-1}(\alpha)\right)$$

(si la limite existe). Il peut se réécrire en utilisant le copula du couple (X,Y)

$$\lambda_L = \lim_{u\to 0} P\left(U \le u \middle| V \le u\right) = \lim_{u\to 0}\frac{C(u,u,)}{u}$$

Ou, de façon analogue, le **upper tail dependence coefficient**, défini par

$$\lambda_U = \lim_{\alpha\to 1} P\left(X > F_X^{-1}(\alpha) \middle| Y > F_Y^{-1}(\alpha)\right)$$

(si la limite existe), qui peut se réécrire

$$\lambda_U = \lim_{u\to 1} P\left(U > u \middle| V > u\right) = \lim_{u\to 1}\frac{\overline{C}(u,u,)}{1-u}$$

Rq Etant donnée un couple (X,Y), on pose $U=F_X(X)$ et $V=F_Y(Y)$, et le copula C du couple (X,Y) apparaît comme la fonction de répartition du couple (U,V) :

$$C(u,v) = P(U\le u,V\le v) = P(F_X(X)\le u,F_Y(Y)\le v)$$

de telle sorte que

$$P(X \le x, Y \le y) = F_{XY}(x,y) = C(F_X(x),F_Y(y))$$

Alors $P(X > x, Y > y) = \overline{F}_{XY}(x,y) = C^*(\overline{F}_X(x),\overline{F}_Y(y))$

où $C^*(u,v)=u+v-1+C(1-u,1-v)$ est appelé *survival copula* et C^* est un copula.

En revanche, la fonction de survie associée au couple (U,V) n'est pas un copula

$$\overline{C}(1-u,1-v) = P(U > u, V > v) = 1-u-v+C(u,v)$$

Estimation Statistique de copulas

Approche paramétrique

Dans un cadre paramétrique, on peut supposer qu'il existe deux types de paramètres :

- ceux associés aux lois marginales (α,β)

- ceux associés à la structure de dépendance (θ)

Estimation par maximum de vraisemblance

$$\left(\hat\alpha,\hat\beta,\hat\theta\right)=\operatorname{argmax}\left\{\sum_{i=1}^{n}\log f_X(x_i,\alpha)+\log f_Y(y_i,\beta)+\log c\left(F_X(x_i,\alpha),F_Y(y_i,\beta),\theta\right)\right\}$$

Méthode en deux étapes :

(i) Maximum de vraisemblance marginal

$$\hat\alpha=\operatorname{argmax}\left\{\sum_{i=1}^{n}\log f_X(x_i,\alpha)\right\}\text{ et }\hat\beta=\operatorname{argmax}\left\{\sum_{i=1}^{n}\log f_Y(y_i,\beta)\right\}$$

On pose alors $u_i=F_X\left(x_i,\hat\alpha\right)$ et $v_i=F_X\left(y_i,\hat\beta\right)$

(ii) Maximum de vraisemblance sur la loi de (U,V)

$$\hat\theta=\operatorname{argmax}\left\{\sum_{i=1}^{n}\log c\left(u_i,v_i,\theta\right)\right\}$$

Rq Cet estimateur *(IFM Inference Functions for Margins)* est convergent et asymptotiquement normal

Approche non-paramétrique

Dans un cadre non-paramétrique, on estime les fonctions de répartition marginales et la fonction de répartition empirique, puis on en déduit le copula à l'aide de la relation

$$\hat C(u,v)=\hat F_{XY}\left(\hat F_X^{-1}(u),\hat F_Y^{-1}(v)\right)$$

Gestion des risques multiples en assurance

Estimation Statistique de copulas

Approches non-paramétriques

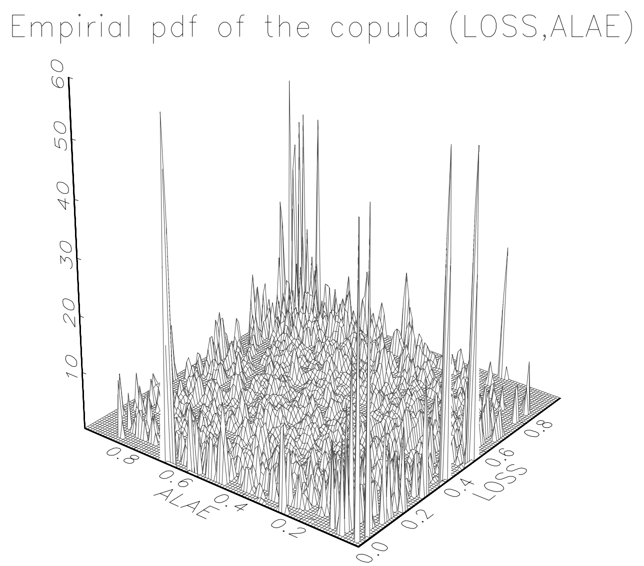
Soit (X_1,Y_1), (X_2,Y_2), …, (X_n,Y_n) un échantillon i.i.d. de fonction de répartition $F_{XY}(x,y)$ jointe, et de fonctions de répartition empiriques $F_X(x)$ et F_Y.

L'estimateur empirique *naturel* du copula est

$$\hat C(u,v)=\hat F_{XY}\left(\hat F_X^{-1}(u),\hat F_Y^{-1}(v)\right)$$

où les $\hat F$ sont les fonctions de répartition empiriques associées à l'échantillon. Si $\partial F_{XY}(x,y)/\partial x$ et $\partial F_{XY}(x,y)/\partial y$ existent et sont continues, on a des résultats asymptotiques sur la convergence de $\hat C\to C$

Ex le graphique ci-dessous correspond à la densité empirique du copula *(c'est à dire l'histogramme)*



Estimation Statistique de copulas

Il est possible d'utiliser un estimateur par noyau : soit K un noyau, en dimension d, supposé être un produit de noyaux de dimension 1

$$K(x=(x_1,\ldots,x_d),h=(h_1,\ldots,h_d))=K_1(x_1,h_1)\ldots K_d(x_d,h_d)$$

En dimension 2, la densité de X et Y sont estimées par

$$\hat{f}_X(x)=\frac{1}{nh_X}\sum_{k=1}^{n}K_X(x-X_k) \text{ et } \hat{f}_Y(y)=\frac{1}{nh_Y}\sum_{k=1}^{n}K_Y(y-Y_k)$$

et les fonctions de répartition empiriques sont alors

$$\hat{F}_X(x)=\int_{-\infty}^{x}\hat{f}_X(t)dt \text{ et } \hat{F}_Y(y)=\int_{-\infty}^{y}\hat{f}_Y(t)dt$$

La densité jointe est alors estimée par

$$\hat{f}_{XY}(x,y)=\frac{1}{n\|h\|}\sum_{k=1}^{n}K_X(x-X_k)K_Y(y-Y_k)$$

et la fonction de répartition est obtenue en intégrant

$$\hat{F}_{XY}(x,y)=\int_{-\infty}^{y}\int_{-\infty}^{x}\hat{f}_{XY}(s,t)dsdt$$

Le copula empirique lissé à l'aide des noyaux K_i est alors

$$\hat{C}(u,v)=\hat{F}_{XY}\left(\hat{F}_X^{-1}(u),\hat{F}_Y^{-1}(v)\right)$$

où les inverses vérifient

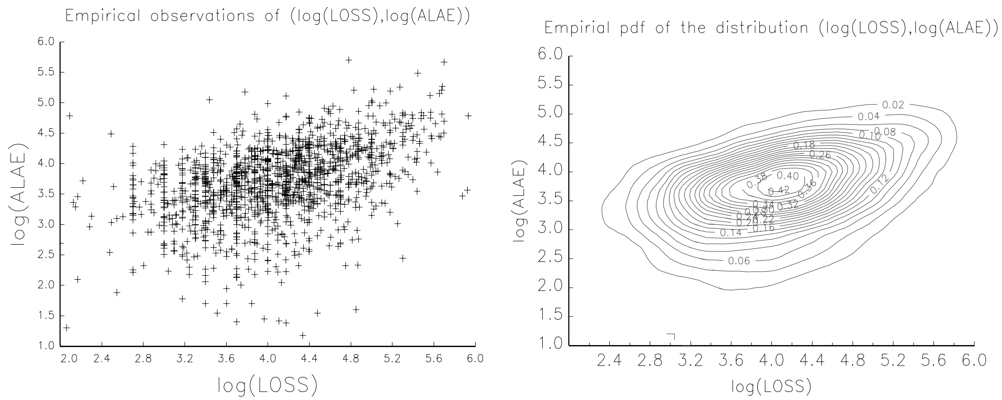
$$\hat{F}_X^{-1}(u)=\inf\{x,\hat{F}_X(x)\geq u\} \text{ et } \hat{F}_Y^{-1}(v)=\inf\{y,\hat{F}_Y(y)\geq v\}$$

(estimateurs par noyau des quantiles de niveau u et v associés aux X_i et aux Y_i)

Estimation Statistique de copulas

Rq On ne peut pas estimer directement la copule pour des problèmes de bord.

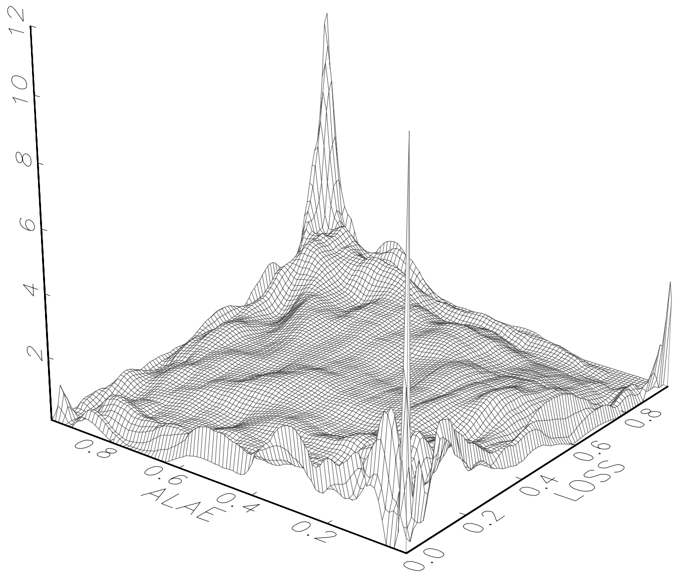
Ex L'exemple ci-dessous modélise les variables coût *(loss)* et frais associés *(ALAE-loss expenses)* pour des sinistre. La densité jointe est alors estimée par



En inversant les fonctions de répartitions marginales empiriques *(en estimant les quantiles empiriques)* il est alors possible de représenter la fonction de répartition empirique du copula.

Estimation Statistique de copulas

Empirial pdf of the copula (LOSS,ALAE)



Rq Il est possible d'approcher le copula empirique en utilisant les polynômes de Bernstein

$$B_n(C)(u,v)=\sum_{i=1}^{n}\sum_{j=1}^{n}B_{i,n}(u)B_{j,n}(v)C\left(\frac{i}{n},\frac{j}{n}\right)$$

où $B_{i,n}(x)$ correspond au polynôme de Berstein

$$B_{i,n}(x)=\binom{n}{i}x^i(1-x)^{n-i}$$

Arthur Charpentier *arthur.charpentier@ensae.fr*

Estimation Statistique de copulas

Prop pour tout copula C, $B_n(C)$ est un copula

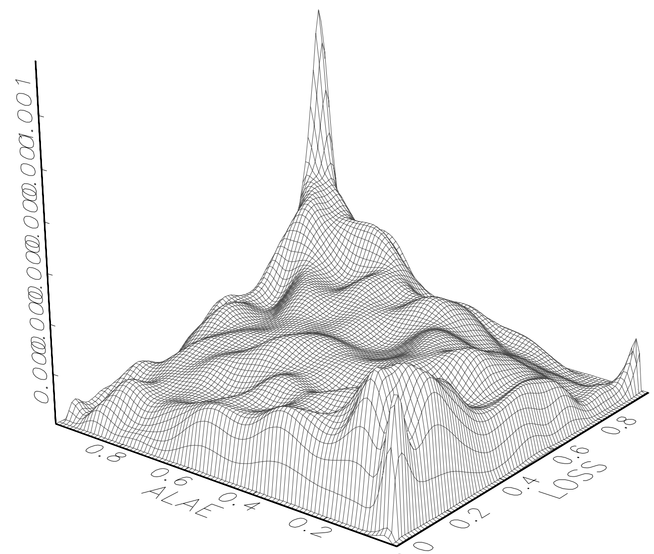
De plus, on a convergence uniforme sur [0,1]x[0,1],

$$\lim_{n\to\infty}\sup\{|B_n(C)(u,v)-C(u,v)|,(u,v)\in[0,1]^2\}$$

Étant donné un copula empirique $\hat{C}$, on peut alors considérer le copula de Berstein associé $B_n(\hat{C})$

On peut alors représenter la densité du copula de Bernstein, qui peut être vue comme une approximation de la fonction de répartition de la copule empirique

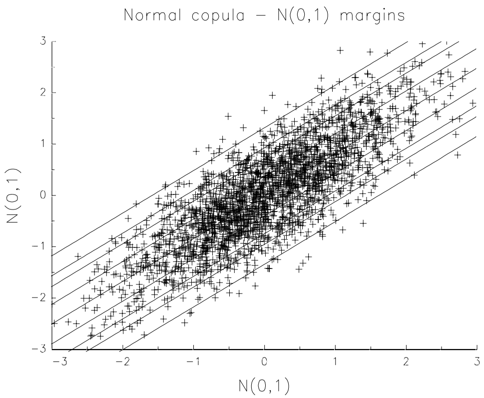
Pdf (Berstein) of the copula (LOSS,ALAE)

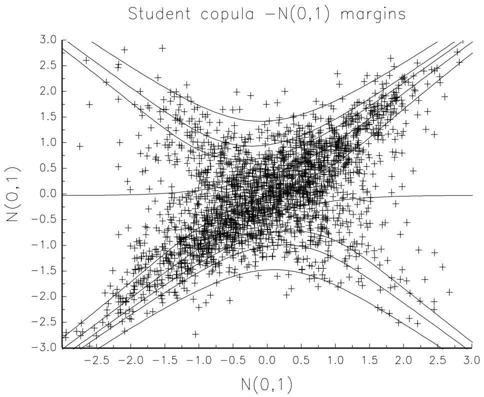


Arthur Charpentier *arthur.charpentier@ensae.fr*

Estimation Statistique de copulas

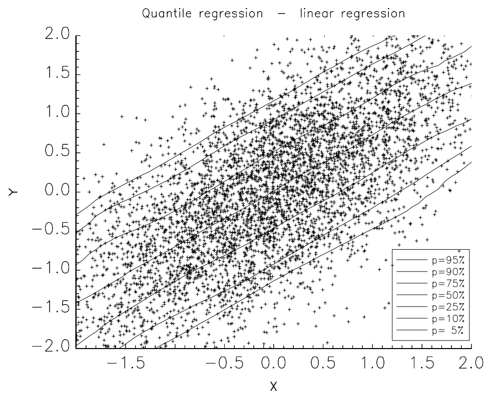
Utilisation de la fonction de régression quantile

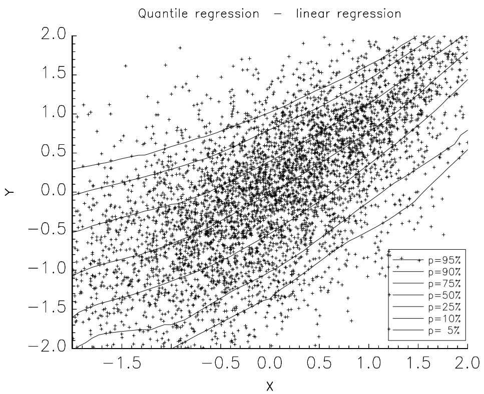




Étant donné un couple (X,Y), la fonction de régression au niveau $p \in]0,1[$ est la fonction $x \to g(x,p)$ où $g(x,p)$ est la solution y de l'équation $F_{Y|X}(y|X=x)=p$, c'est à dire

$$((\partial C)/(\partial u))(F_X(x),F_Y(g(x,p)))=p$$





Sommes *(ou fonctions)* de variables aléatoires

De façon générale, on cherche à expliciter le comportement de $\psi(X,Y)$ où ψ est une application (*par exemple* $\psi(X,Y)=X\pm Y$)

La convolution $F*G$, de où F et G sont des fonctions de répartition de variables positives est la fonction définie par

$$(F*G)(z)=\int_{[0,z]}F(z-x)dG(x) \text{ pour tout } x>0$$

qui peut se réécrire

$$(F*G)(z)=\int_{S(z)}dC^{\perp}(F(x),G(y)) \text{ où } S(x)=\{(x,y),x+y<z\}$$

Rq dans le cas où X et Y sont indépendantes (de copula $C^{\perp}$) alors F_X*F_Y est la fonction de répartition de X+Y

La fonction de répartition de $Z=\psi(X,Y)$ est donnée par

$$H(z)=\int_{V(z)}dC(F_X(x),F_Y(y)) \text{ où } V(z)=\{(x,y),\psi(x,y)<z\}$$

où C est le copula associé au couple (X,Y). Cette fonction sera également appelée σ-convolution pour l'opérateur ψ

Rq si $\psi(X,Y)=\max\{x,y\}$, alors $H(z)=C(F_X(z),F_Y(z))$

Sommes *(ou fonctions)* de variables aléatoires

On peut alors définir les deux fonctions suivantes

$$\tau_{C,\psi}(F_X,F_Y)(z)=\sup\{C(F_X(x),F_Y(y)),\psi(x,y)=z\}$$

$$\rho_{C,\psi}(F_X,F_Y)(z)=\inf\{C(F(x),G(y)),\psi(x,y)=z\}$$

Et, dans le cas où ψ correspond à la somme,

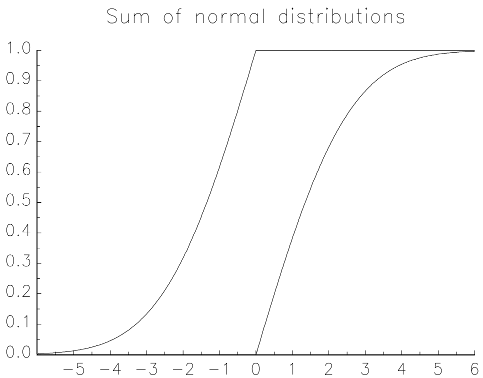
$$\tau_{C,+}(F_X,F_Y)(z)=\sup\{C(F_X(x),F_Y(y)),x+y=z\}$$

$$\rho_{C,+}(F_X,F_Y)(z)=\inf\{C(F(x),G(y)), x+y =z\}$$

Soient X et Y deux variables aléatoires, de fonction de répartition respectives F_X et F_Y. Alors $P(X+Y\leq z)$ est bornée par

$$\tau_{C,+}(F_X,F_Y)(z) \leq P(X+Y\leq z) \leq \rho_{C,+}(F_X,F_Y)(z)$$

Ex La fonction de répartition de la somme de deux lois N(0,1) est bornée par les fonctions représentées ci-contre

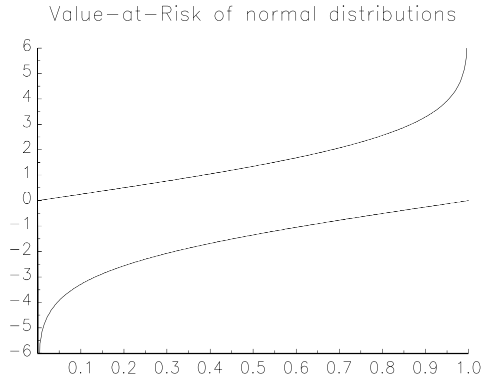


Gestion des risques multiples en assurance

Value-at-Risk de sommes de variables aléatoires

En notant que la Value-at-Risk correspond à l'inverse de la fonction de répartition, il est possible d'obtenir des bornes pour la VaR de X+Y

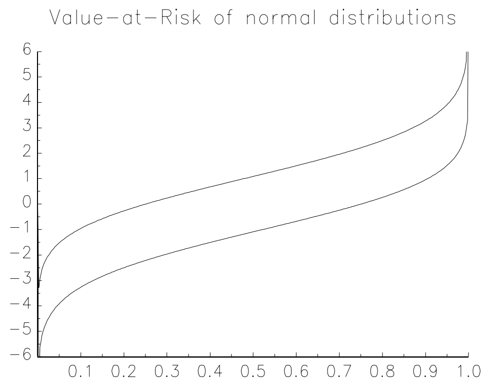
Ex La VaR de la somme de deux lois N(0,1) est bornée par les fonctions représentées ci-contre



Supposons que l'on connaisse deux bornes C_1 et C_2 telles que $C_1(u,v)\leq C(u,v)$ et $C^*(u,v)\leq C_2^*(u,v)$ pour tout $0\leq u,v\leq 1$, alors

$$\tau_{C1,+}(F_X,F_Y)(z) \leq P(X+Y\leq z) \leq \rho_{C2,+}(F_X,F_Y)(z)$$

Ex Supposons pour tout u,v, $C(u,v) \geq C^{\perp}(u,v)$ *(hypothèse PQD)* alors la VaR de la somme de deux lois N(0,1) est bornée par les fonctions représentées ci-contre



Ref : Williamson (1989) 'Probabilistic Arithmetic' (PhD)

Mesures de risques et risques multiples

Def Étant donné un cône convexe de variables aléatoires H, alors la fonction $\rho:H \to R$ est une **mesure de risque cohérente** si elle vérifie les 4 propriétés suivante

1) *monotonie* : si $X \leq Y$ p.s. alors $\rho(X) \leq \rho(Y)$

2) *sous-additivité* : $\rho(X+Y) \leq \rho(X)+\rho(Y)$

3) *homogénéité* : pour tout $\lambda \geq 0$, $\rho(\lambda X) = \lambda\rho(X)$

4) *invariance par translation* : $\rho(X+a) = a+\rho(X)$, où $a \in R$

En fait (Follmer et Scheid (2002)) ont proposé de relacher les hypothèses 2 et 3 pour l'hypothèse plus faible suivante :

5) *convexité* : $\rho(\lambda X+(1-\lambda)Y) \leq \lambda\rho(X)+(1-\lambda)\rho(Y)$, $\lambda \in [0,1]$

Les mesures vérifiant les axiomes 1, 4 et 5 sont appelées **mesures de risques convexes**

Ex Soit $u:R \to R$, alors $\rho(X) = E(u(X))$ est une mesure de risque convexe si et seulement si u est convexe

Ex La variance, l'écart-type, la Tail-VaR sont des mesures de risques convexes, mais pas la VaR

Considérons des vecteurs de risques X et Y de même copula : sous quelles conditions sur la mesure de risque ρ a-t-on le résultat suivant,

$$\text{si } \rho(X_i) \leq \rho(Y_i) \text{ pour tout i, alors } \rho(\sum\lambda_i X_i) \leq \rho(\sum\lambda_i Y_i)$$

Arthur Charpentier *arthur.charpentier@ensae.fr*

Gestion des risques multiples en assurance

Mesures de risques et risques multiples

Rq Soient $X=(X_1,X_2)'$ et $Y=(Y_1,Y_2)'$ tels que

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & -2 \\ -2 & 4 \end{pmatrix}\right) \text{ et } \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 4 & -4 \\ -4 & 4 \end{pmatrix}\right)$$

alors $C_X=C_Y=C^-$ (cas anti-comonotonique), avec $X_1 \sim N(0,1)$ et $X_2,Y_1,Y_2 \sim N(0,2^2)$. Il est aisé de voir que

$$X_1+X_2 \sim N(0,1) \text{ et } X_1+X_2=0 \sim N(0,0)$$

On a alors que $V(X_i) \leq V(Y_i)$ et $V(X_1+X_2) \geq V(Y_1+Y_2)$, et de même, $VaR(X_i) \leq VaR(Y_i)$ et $VaR(X_1+X_2) \geq VaR(Y_1+Y_2)$.

• Dans le cas où X et Y sont des vecteurs indépendants (i.e. $C_X=C_Y=C^{\perp}$), si $\rho(X)$ peut s'écrire $\rho(X) = g(E(u(X)))$ où u est une fonction convexe et g une fonction strictement croissante, alors la dominance est conservée pour les portefeuilles.

Rq Même dans le cas indépendant, même si $VaR(X_i) \leq VaR(Y_i)$ on peut avoir $VaR(X_1+X_2) \geq VaR(Y_1+Y_2)$.

• Dans le cas où X et Y sont des vecteurs ayant le même copula (i.e. $C_X=C_Y=C$), tel que C soit CI, alors la dominance est conservée pour les portefeuilles.

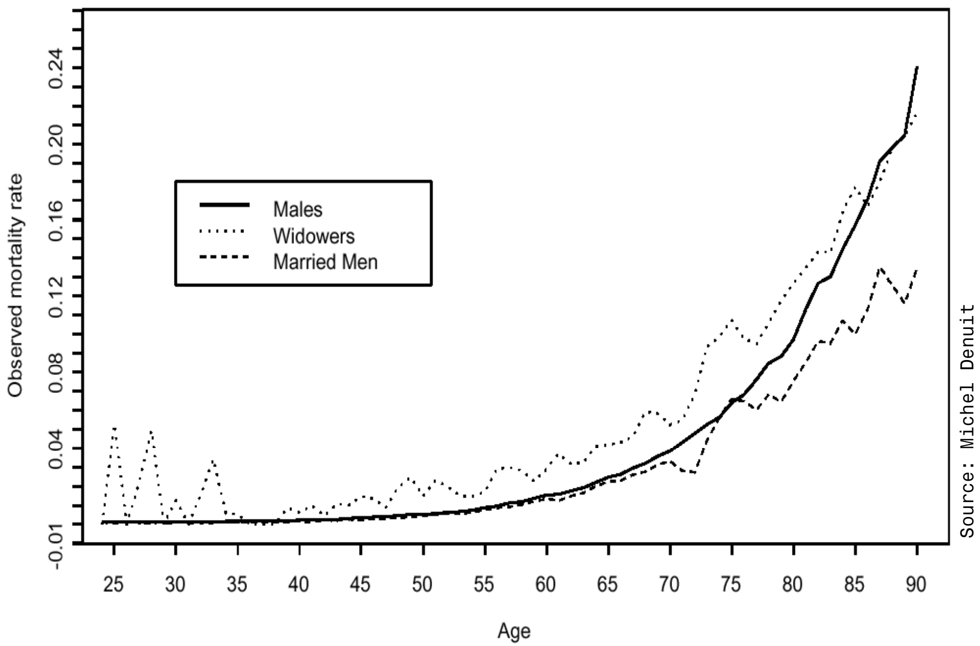
(en dimension 2, cette condition signifie que C est concave en chacune des composantes)

Ref : Scarsini (1998) '*Multivariate Convex Orderings, dependence, and Stochastic Equality*' (JAP)

Arthur Charpentier *arthur.charpentier@ensae.fr*

Application : assurance décès sur 2 têtes

La fonction de survie au sein d'un couple de personnes mariées est affectée par le décès d'un des conjoints *(syndrome du cœur brisé)*



Ex La fonction de survie *(le graphique ci-dessus représente la fonction de hazard)* est d'espérance moins inportante pour les veuf que pour les hommes mariés : le décès de l'épouse impacte la fonction de survie du veuf.

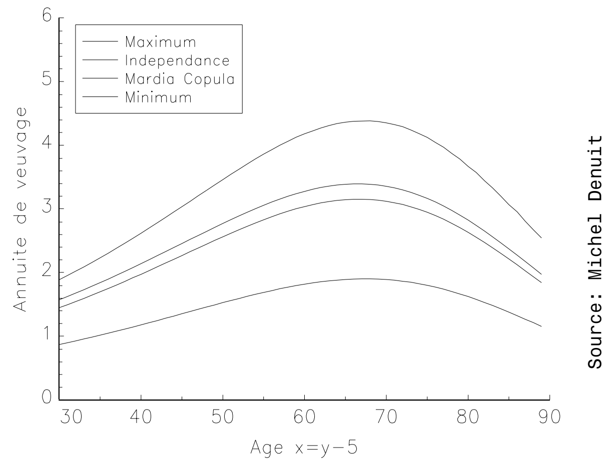
Application : assurance décès sur 2 têtes

Soient x et y les âges respectivement du mari et de son épouse, et notons T_x et T_y leurs fonctions de survies
On notera également ${}_tp_x$ la probabilité $P(T_x>t)$, et ${}_tp_y$ la probabilité $P(T_y>t)$. Enfin, on notera

$${}_tp_{xy}=P(\min\{T_x,T_y\}>t) = {}_tp_x + {}_tp_y - 1 + C(1-{}_tp_x, 1-{}_tp_y)$$

où C est le copula du couple (T_x,T_y).

Supposons que C soit un copula paramétrique.
L'estimation du copula permet de comparer, par exemple, la prime de veuvage, et la comparer au prix obtenu dans le cas où l'on fait l'hypothèse d'indépendance entre les deux fonctions de survie



On peut noter, sur cet exemple, que le calcul sous hypothèse d'indépendance surestime la prime pure de l'ordre de 10%.