

Introduction aux copules

Régis LEBRUN

EADS IW
SE – MA

18 mars 2010

Plan de l'exposé

- 1 Cadre industriel
- 2 Incertitudes probabilistes
- 3 Copules
- 4 Autres concepts de dépendance
- 5 Conclusion

Objectifs

Que retenir de cette présentation ?

- La notion de dépendance stochastique est **exactement** couverte par le concept de copule ;
- Traiter cette notion à l'aide de **corrélations linéaires seules** est (en général) une **très mauvaise idée** ;
- **Tout** modèle probabiliste multivarié possède (au moins) une copule...
- ... mais ce n'est pas forcément l'objet le mieux adapté aux manipulations et aux modélisations probabilistes (quoique...) !

Contexte industriel

Passer du calcul «sûr» au calcul «incertain»

- Les chaînes de calcul industrielles sont paramétrées par un ensemble de données associées à des valeurs numériques **bien déterminées**.
 - Les valeurs effectives des données d'entrée sont souvent **mal connues**.
 - Pour prendre en compte cette incertitude, la démarche standard est l'approche **pire cas**, qui est souvent **très pessimiste**, et offre une **fausse certitude**.
- ⇒ Une autre modélisation des grandeurs incertaines est nécessaire.

Cadre probabilistes

Modélisation probabiliste des incertitudes

- Les données d'entrée de la chaîne de calcul sont regroupées dans un vecteur aléatoire \underline{X} .
- La chaîne de calcul est représentée par une fonction numérique vectorielle $f \in \mathcal{F}(\mathbb{R}^N, \mathbb{R}^P)$, où souvent $P = 1$.
- La variable de décision \underline{Y} s'écrit donc :

$$\underline{Y} = f(\underline{X}) \quad (1)$$

Variables aléatoires, vecteurs aléatoires

Rappels

Soit (Ω, \mathcal{F}) un espace mesurable et $(\mathbb{R}^N, \mathcal{B})$ l'espace mesurable numérique \mathbb{R}^N muni de la tribu des boréliens. On a :

- Un **vecteur aléatoire \underline{X} réel** de dimension N est une application de (Ω, \mathcal{F}) dans $(\mathbb{R}^N, \mathcal{B})$ telle que :

$$\forall B \in \mathcal{B}, \underline{X}^{-1}(B) \in \mathcal{F} \quad (2)$$

- Une **variable aléatoire X réelle** est un vecteur aléatoire de dimension $N = 1$.

Un vecteur aléatoire \underline{X} de dimension N **ne se résume pas** à la donnée des N variables aléatoires marginales X_k !

Modélisation probabiliste

Loi jointe et copule

- Modèle probabiliste = fonction de répartition $F_{\underline{X}}$ du vecteur \underline{X} .
- On sépare la modélisation des lois marginales (**fonctions de répartition 1D**) et de la structure de dépendance (**copule**)
- Le modèle probabiliste est alors une conséquence du théorème de Sklar : la loi de \underline{X} est obtenue par **assemblage** des lois marginales et de la copule.

Pourquoi vous avez pu vous en passer jusqu'à aujourd'hui...

A quoi bon les copules ?

Vous calculez des probabilités de dépassement de seuil depuis très longtemps sans parler de copule, car...

- Vos variables sont **TOUJOURS** indépendantes !
- Quand elles ne le sont pas, vous avez une **boîte noire magique** qui vous prend une poignée de corrélations linéaires pour décrire la dépendance, puis le calcul se fait... magiquement !
- Vous avez toujours lu qu'il était illusoire de vouloir identifier une loi multivariée...

La boîte magique et les copules

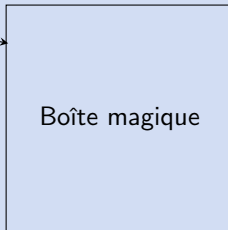
Copule inside

Boîte magique

La boîte magique et les copules

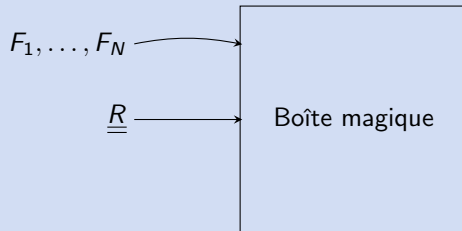
Copule inside

F_1, \dots, F_N



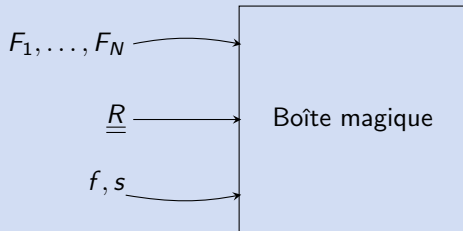
La boîte magique et les copules

Copule inside



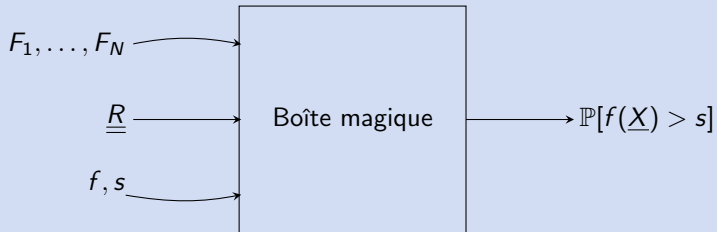
La boîte magique et les copules

Copule inside



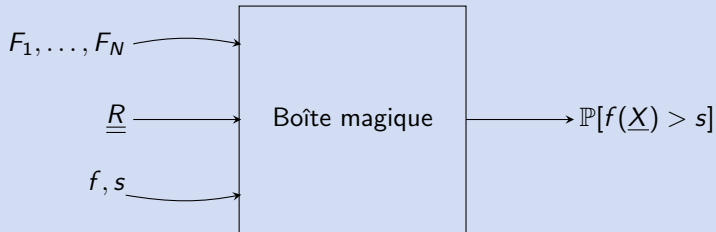
La boîte magique et les copules

Copule inside



La boîte magique et les copules

Copule inside

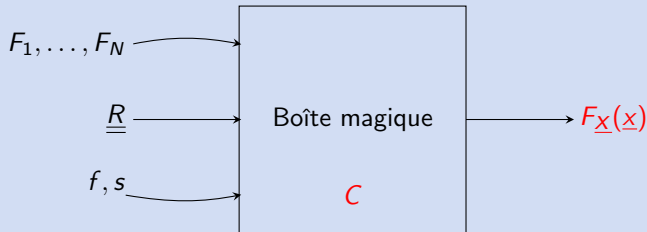


Où est la magie? Prendre $f = \prod_{k=1}^N \mathbf{1}_{\{X_k \leq x_k\}}$ et $s = 1/2 \dots$

La boîte magique calcule la fonction de répartition de \underline{X} !

La boîte magique et les copules

Copule inside



Où est la magie? Prendre $f = \prod_{k=1}^N \mathbf{1}_{\{X_k \leq x_k\}}$ et $s = 1/2 \dots$

La boîte magique calcule la fonction de répartition de \underline{X} !

La copule est cachée dans la boîte!

Copules : perspective historique

- 1940 Hoeffding : mesures de dépendance, corrélation linéaire, lois multivariées à marginales uniformes sur $[-1/2, 1/2]$.
 - 1951 Fréchet : lois multivariées à marginales fixées.
 - 1959 Sklar et Schweizer : espaces métriques probabilistes, première apparition du terme de copule.
 - 1979 Deheuvels : tests d'indépendance, estimation non paramétrique multivariée.
 - 1992 Darsow, Nguyen et Olsen : formalisation des processus de Markov en termes de copule.
 - 1999 Embrechts : modélisation du risque en finance et en assurance.
- Depuis, travaux en statistiques, simulation, modélisation etc.

Définition

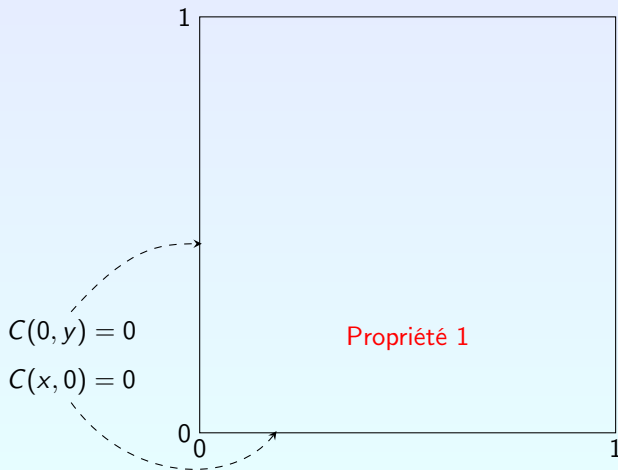
Qu'est-ce qu'une copule ?

Définition «pratique» : Une *copule* est une fonction de répartition définie sur $[0, 1]^N$ dont les lois marginales sont égales à la loi uniforme sur $[0, 1]$.

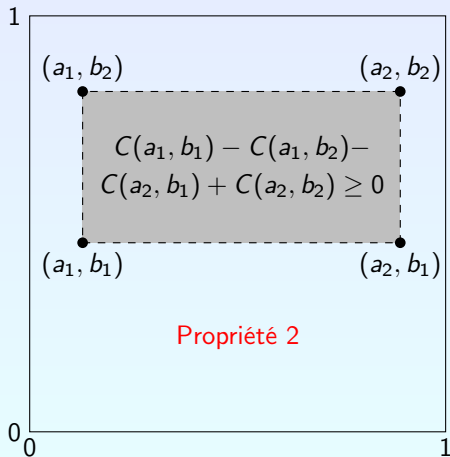
Définition «technique» : Une *copule* est une fonction C définie sur $[0, 1]^N$ vérifiant :

- 1 Pour tout \underline{u} ayant au moins une composante nulle, $C(\underline{u}) = 0$;
- 2 C est N -croissante :
$$\sum_{i_1=1}^2 \cdots \sum_{i_N=1}^2 (-1)^{i_1+\cdots+i_N} C(x_{1i_1}, \dots, x_{Ni_N}) \geq 0$$
 avec
 $x_{j1} = a_j$ et $x_{j2} = b_j$ pour tout $j \in \{1, \dots, N\}$ et $\underline{a}, \underline{b} \in [0, 1]^N$,
 $\underline{a} \leq \underline{b}$.
- 3 Pour tout \underline{u} ayant toutes ses composantes égales à 1 sauf éventuellement u_k , $C(\underline{u}) = u_k$.

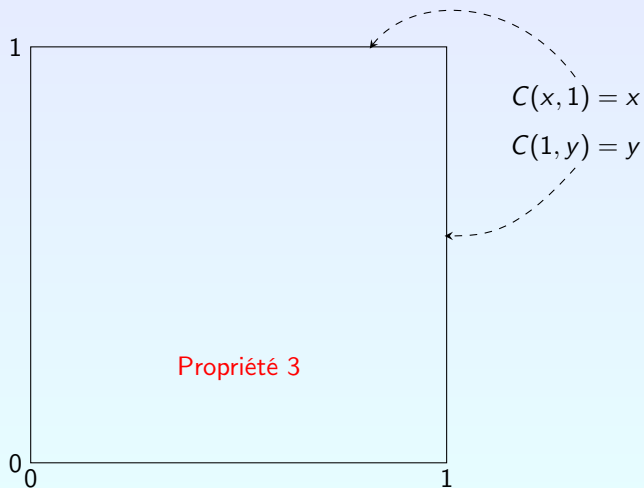
Illustration



Illustration



Illustration



Exemples de copules

Copule indépendante :

$$C(\underline{u}) = u_1 u_2$$

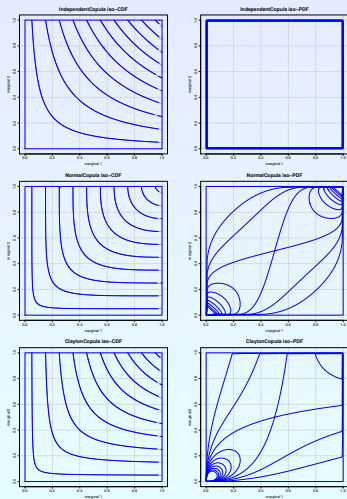
Copule normale :

$$C(\underline{u}) = \Phi_R^N(\Phi^{-1}(u_1), \Phi^{-1}(u_2))$$

Copule de Clayton :

$$C(\underline{u}) = \varphi^{-1}(\varphi(u_1) + \varphi(u_2))$$

avec $\varphi(t) = t^\theta - 1$



Les copules pour la modélisation

Théorème de Sklar

Théorème (Sklar, 1959) : Soit F une fonction de répartition N -dimensionnelle dont les lois marginales sont F_1, \dots, F_N . Il existe une copule C de dimension N telle que pour $\underline{x} \in \overline{\mathbb{R}}^N$ on ait :

$$F(x_1, \dots, x_N) = C(F_1(x_1), \dots, F_N(x_N)). \quad (3)$$

Si les **lois marginales** F_1, \dots, F_N sont **continues**, la **copule** C est **unique**, sinon elle est déterminée de manière unique sur $Im(F_1) \times \dots \times Im(F_N)$.

Dans le cas de lois marginales continues, pour tout $\underline{u} \in [0, 1]^N$ on a :

$$C(\underline{u}) = F(F_1^{-1}(u_1), \dots, F_N^{-1}(u_N)) \text{ et}$$

$$p(\underline{x}) = c(F_1(x_1), \dots, F_N(x_N)) \prod_{i=1}^N p_i(x_i), \text{ où } c \text{ est la densité associée à } C.$$

Lois composées : exemples

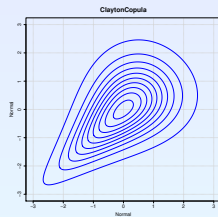
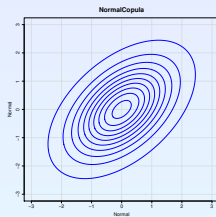
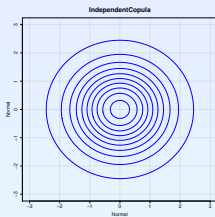
Copule

Indépendante

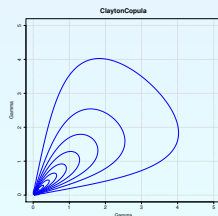
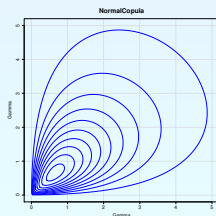
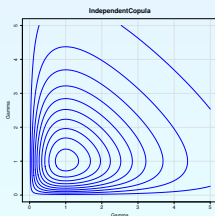
Normale

Clayton

Marginale
 $\mathcal{N}(0, 1)$



Marginale
 $\Gamma(2, 1)$



Quelques propriétés

Bornes de Fréchet-Hoeffding

Bornes de Fréchet-Hoeffding : On définit les fonctions suivantes sur $[0, 1]^N$:

$$W^N(\underline{u}) = \max(u_1 + \dots + u_N - N + 1, 0) \text{ et } M^N(\underline{u}) = \min(u_1, \dots, u_N). \quad (4)$$

Pour toute copule C et tout $\underline{u} \in [0, 1]^N$, on a :

$$W^N(\underline{u}) \leq C(\underline{u}) \leq M^N(\underline{u}). \quad (5)$$

Remarque : W^N est une copule pour $N = 2$ mais pas pour $N > 2$, alors que M^N est une copule pour tout $N \geq 2$. M^N est la copule d'un vecteur dont les composantes X_i sont presque sûrement l'image par une application strictement croissante d'une même variable aléatoire V .

Quelques propriétés

Copules marginales et conditionnelles

$$C_k(u_1, \dots, u_k) = C(u_1, \dots, u_k, 1, \dots, 1)$$
$$C_k(u_k | u_1, \dots, u_{k-1}) = \frac{\partial^{k-1} C_k(u_1, \dots, u_k)}{\partial u_1 \dots u_{k-1}} / \frac{\partial^{k-1} C_{k-1}(u_1, \dots, u_{k-1})}{\partial u_1 \dots u_{k-1}} \quad (6)$$

Effet d'une transformation croissante

Soit \underline{X} un vecteur aléatoire continu de copule C et $\alpha_1, \dots, \alpha_N$ des fonctions strictement croissantes. Le vecteur $(\alpha_1(X_1), \dots, \alpha_N(X_N))$ a également C pour copule.

Simulation d'une loi assemblée

Démarche générale

La simulation d'un vecteur \underline{X} de lois marginales F_1, \dots, F_N et de copule C se fait en deux étapes :

- 1 On simule une réalisation \underline{u} de la copule C ;
- 2 On obtient une réalisation \underline{x} de \underline{X} grâce à la transformation :

$$\underline{x} = (F_1^{-1}(u_1), \dots, F_N^{-1}(u_N)) \quad (7)$$

La principale difficulté consiste à simuler C .

Simulation d'une copule

Algorithme à usage général

1. On simule une réalisation u_1 selon $\mathcal{U}(0, 1)$;
2. On simule une réalisation u_2 selon $C_{2|1}(\cdot|u_1)$;
- ⋮
- N . On simule une réalisation u_N selon $C_{N|1,\dots,N-1}(\cdot|u_1, \dots, u_{N-1})$.

Cet algorithme n'est pas forcément le plus efficace selon le type de copule considéré. Il est toutefois raisonnablement efficace quand les fonctions $C_{k|1,\dots,k-1}^{-1}(\cdot|u_1, \dots, u_k)$ ont une expression analytique.

Lois composées : simulation

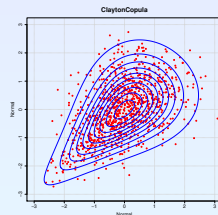
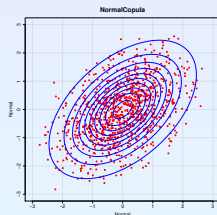
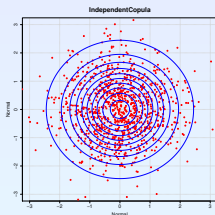
Copule

Indépendante

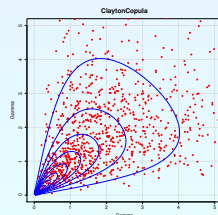
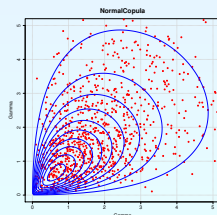
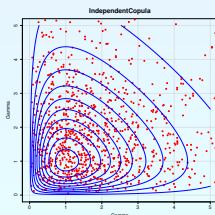
Normale

Clayton

Marginale
 $\mathcal{N}(0, 1)$



Marginale
 $\Gamma(2, 1)$



Composition de copules

Composition produit

Soient C_1, \dots, C_k k copules de dimensions respectives n_1, \dots, n_k . On pose $N = \sum_{i=1}^k n_i$. La fonction C définie sur $[0, 1]^N$ par :

$$C(u_1, \dots, u_N) = C_1(u_1, \dots, u_{n_1}) \times \dots \times C_k(u_{N-n_k+1}, \dots, u_N) \quad (8)$$

est une copule de dimension N .

On peut ainsi construire la structure de dépendance globale d'un vecteur aléatoire constitué de sous vecteurs indépendants.

On obtient ainsi une représentation creuse de la structure de dépendance

Copules et processus de Markov

Produits de copules, définitions

- ❶ Soient A et B deux copules bidimensionnelles. On définit le produit $C = A * B$ de ces deux copules par :

$$C(u_1, u_2) = \int_0^1 A_{1|2}(u_1, t) B_{2|1}(t, u_2) dt \quad (9)$$

- ❷ Soient A une copule de dimension m et B une copule de dimension n . On définit le produit $C = A \star B$ de ces deux copules par :

$$C(u_1, \dots, u_{m+n-1}) = \int_0^{u_m} A_{1, \dots, m-1|m}(u_1, \dots, u_{m-1}, t) \times \quad (10) \\ \times B_{2, \dots, m|1}(t, u_{m+1}, \dots, u_{m+n-1}) dt$$

On a la relation $A * B(u, v) = A \star B(u, 1, v)$

Copules et processus de Markov

Produits de copules, propriétés

On a les propriétés suivantes (avec $\bullet \in \{*, \star\}$) :

- C est une copule (de dimension 2 pour $*$, de dimension $m + n - 1$ pour \star) ;
- Ces produits sont continus par rapport à A et à B : si $(A_n)_{n \in \mathbb{N}} \rightarrow A$ et $(B_n)_{n \in \mathbb{N}} \rightarrow B$, $A_n \bullet B \rightarrow A \bullet B$ et $A \bullet B_n \rightarrow A \bullet B$;
- Ces produits sont associatifs : $(A \bullet B) \bullet C = A \bullet (B \bullet C)$;
- Ces produits sont distributifs à gauche et à droite pour les combinaisons convexes de copules.

Copules et processus de Markov

Copules et équations de Chapman-Kolmogorov (Darsow, Nguyen, Olsen)

Soit $X_t, t \in T$ un processus stochastique réel, et pour tout $s, t \in T$, soit C_{st} la copule du vecteur (X_s, X_t) . On a l'équivalence entre :

- 1 Les probabilités de transition $\mathbb{P}(s, x, t, A) = \mathbb{P}(X_t \in A | X_s = x)$ du processus vérifient les équations de Chapman-Kolmogorov :

$$\mathbb{P}(s, x, t, A) = \int_{-\infty}^{\infty} \mathbb{P}(u, \xi, t, A) \mathbb{P}(s, x, u, d\xi) \quad (11)$$

pour tout borélien A , tout $s < t$ dans T , tout $u \in]s, t[\cap T$ et presque tout $x \in \mathbb{R}$;

- 2 Pour tout $s, u, t \in T$ vérifiant $s < u < t$,

$$C_{st} = C_{su} * C_{ut} \quad (12)$$

Copules et processus de Markov

Copules et processus de Markov (Darsow, Nguyen, Olsen)

Un processus stochastique réel $X_t, t \in T$ est un processus de Markov si et seulement si pour tout entier $n > 0$ et pour tout $t_1, \dots, t_n \in T$ vérifiant $t_1 < \dots < t_n$ on a :

$$C_{t_1 \dots t_n} = C_{t_1 t_2} \star C_{t_2 t_3} \star \dots \star C_{t_{n-1} t_n} \quad (13)$$

où $C_{t_1 \dots t_n}$ est la copule de $(X_{t_1}, \dots, X_{t_n})$ et $C_{t_k t_{k+1}}$ celle de $(X_{t_k}, X_{t_{k+1}})$. Ce résultat peut être généralisé ([Ibragimov]) à un processus de Markov d'ordre k , i.e. tel que :

$$\mathbb{P}(X_t < x | X_{t_1}, \dots, X_{t_{n-k}}, X_{t_{n-k+1}}, \dots, X_{t_n}) = \mathbb{P}(X_t < x | X_{t_1}, \dots, X_{t_{n-k}}) \quad (14)$$

pour tout $t, t_i \in T$ tels que $t_1 < \dots < t_{n-k} < t_{n-k+1} < \dots < t_n < t$ et $x \in \mathbb{R}$.

Mesures d'association

Motivation

La donnée d'une copule comme modèle de dépendance d'un vecteur aléatoire est très riche.

La notion de **mesure d'association** sert à résumer cette structure de dépendance dans une collection de scalaires.

Mesures d'association

Définition

Une **mesure d'association** r entre deux variables aléatoires X_1 et X_2 est une fonction scalaire de X_1 et X_2 telle que :

- 1 r est définie pour tout couple (X_1, X_2) .
- 2 $r(X_1, X_2) \in [-1, 1]$, $r(X_1, X_1) = 1$, $r(X_1, -X_1) = -1$.
- 3 Si X_1 et X_2 sont indépendantes, $r(X_1, X_2) = 0$.
- 4 Si g et h sont deux fonctions strictement croissantes, $r(X_1, X_2) = r(g(X_1), h(X_2))$.

On montre que r est une fonction de la copule de (X_1, X_2) seule

Corrélation linéaire

Définition

La **corrélation linéaire** ρ entre deux variables aléatoires X_1 et X_2 telles que $\text{Var}(X_1) = \sigma_1^2 < \infty$ et $\text{Var}(X_2) = \sigma_2^2 < \infty$ est définie par :

$$\begin{aligned} \rho(X_1, X_2) &= \frac{\text{Cov}(X_1, X_2)}{\sqrt{\text{Var}(X_1)\text{Var}(X_2)}} \\ &= \frac{1}{\sigma_1\sigma_2} \iint_{\mathbb{R}^2} F_{12}(x_1, x_2) - F_1(x_1)F_2(x_2) dx_1 dx_2 \end{aligned} \quad (15)$$

Propriétés

- $\rho(X_1, X_2) \in [-1, 1]$ avec $|\rho(X_1, X_2)| = 1 \iff \exists a, b \in \mathbb{R}, a \neq 0, X_2 = aX_1 + b$
- X_1, X_2 indépendantes implique $\rho(X_1, X_2) = 0$;
- $\rho(aX_1 + b, \alpha X_2 + \beta) = \text{sign}(a\alpha)\rho(X_1, X_2)$

Ce n'est pas une mesure d'association ! Elle n'est pas définie pour toutes les variables aléatoires, n'est pas invariante par transformation croissante et n'est pas une fonction de la copule seule.

Corrélation linéaire

Théorème de Fréchet-Hoeffding

Soit (X_1, X_2) un vecteur aléatoire de lois marginales F_1, F_2 données. Les valeurs possibles de $\rho(X_1, X_2)$ forment un intervalle inclu dans $[-1, 1]$, l'inclusion étant stricte en générale.

Conséquence : il est impossible de spécifier $\rho(X_1, X_2)$ indépendamment de F_1 et F_2 .

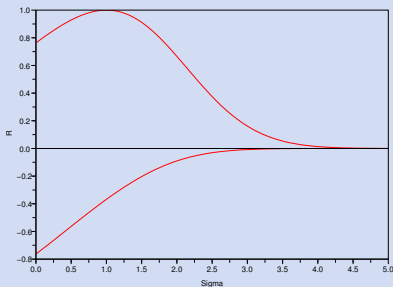
Corrélation linéaire

Illustration

Si $X_1 \hookrightarrow \mathcal{LN}(0, 1)$ et $X_2 \hookrightarrow \mathcal{LN}(0, \sigma^2)$, alors $\rho(X_1, X_2) \in [\rho_{min}, \rho_{max}] \subsetneq [-1, 1]$, avec

$$\rho_{min} = \frac{e^{-\sigma} - 1}{\sqrt{e^{-1} - 1} \sqrt{e^{\sigma^2} - 1}} \text{ et } \rho_{max} = \frac{e^{\sigma} - 1}{\sqrt{e^{-1} - 1} \sqrt{e^{\sigma^2} - 1}}.$$

Plage de corrélation pour un couple de lois LogNormal



On note que ρ_{min} et ρ_{max} tendent vers 0 quand σ tend vers $+\infty$. **Pour**
 $\sigma = 5, \rho \in [-3 \cdot 10^{-6}, 4 \cdot 10^{-4}]!$

Rho de Spearman

Définition

Le ρ_S de Spearman entre deux variables aléatoires X_1 et X_2 est défini par :

$$\rho_S(X_1, X_2) = \rho(F_1(X_1), F_2(X_2)) = 12 \iint_{[0,1]^2} C(u, v) du dv - 3 \quad (16)$$

où C est la copule de la loi jointe de (X_1, X_2) .

Propriétés

- $\rho_S(X_1, X_2) \in [-1, 1]$ avec
 $|\rho_S(X_1, X_2)| = 1 \iff \exists \varphi$ monotone telle que $X_2 = \varphi(X_1)$
- X_1, X_2 indépendantes implique $\rho_S(X_1, X_2) = 0$;
- $\rho_S(\varphi(X_1), \psi(X_2)) = \rho_S(X_1, X_2)$ pour toutes fonctions monotones φ et ψ de même monotonie.

Il s'agit bien d'une mesure d'association.

Tau de Kendall

Définition

Le τ de Kendall entre deux variables aléatoires X_1 et X_2 est défini par :

$$\begin{aligned}\tau(X_1, X_2) &= \mathbb{P}[(\hat{X}_1 - \tilde{X}_1)(\hat{X}_2 - \tilde{X}_2) > 0] - \mathbb{P}[(\hat{X}_1 - \tilde{X}_1)(\hat{X}_2 - \tilde{X}_2) < 0] \\ &= 4 \iint_{[0,1]^2} C(u, v) dC(u, v) - 1\end{aligned}$$

où (\hat{X}_1, \hat{X}_2) et $(\tilde{X}_1, \tilde{X}_2)$ ont la même loi que (X_1, X_2) , de copule C .

Propriétés

- $\tau(X_1, X_2) \in [-1, 1]$ avec $|\tau(X_1, X_2)| = 1 \iff \exists \varphi$ monotone telle que $X_2 = \varphi(X_1)$
- X_1, X_2 indépendantes implique $\tau(X_1, X_2) = 0$;
- $\tau(\varphi(X_1), \psi(X_2)) = \tau(X_1, X_2)$ pour toutes fonctions monotones φ et ψ de même monotonie.

Il s'agit bien d'une mesure d'association.

Utilisation des mesures d'association

Estimation paramétrique de copule

Les mesures d'association ρ_S et τ étant des fonctions de la copule seule, il est possible de les relier aux paramètres $\underline{\theta}$ de la copule $C_{\underline{\theta}}$. On construit ainsi un estimateur $\hat{\underline{\theta}}_n$ de $\underline{\theta}$ qui est robuste à l'effet des lois marginales sur l'échantillon multivarié de taille n .

Exemples :

- Copule normale $C_{\underline{R}}$: $R_{ij} = 2 \sin\left(\frac{\pi}{6}\rho_{S_{ij}}\right) = \sin\left(\frac{\pi}{2}\tau_{ij}\right)$
- Copule de Clayton C_{θ} : $\theta = \frac{2\tau}{1-\tau}$

Estimation statistique

Méthodologie générale

Soit $(\underline{X}_k)_{k \in \{1, \dots, N\}}$ un échantillon de taille N d'une loi multivariée. La démarche d'estimation de la copule sous-jacente à \underline{X} est :

- Identifier les fonctions de répartition marginales (estimation 1D) ;
- Transformer l'échantillon $(\underline{X}_k)_{k \in \{1, \dots, N\}}$ en un échantillon $(\underline{U}_k)_{k \in \{1, \dots, N\}}$ avec $U_k^j = F_j(X_k^j)$.
- Estimer la copule sur la base de $(\underline{U}_k)_{k \in \{1, \dots, N\}}$

Estimation statistique

Estimation des lois marginales

Toutes les techniques classiques sont possibles :

- Estimation paramétrique : $F_j^\theta \in \mathcal{L}(\theta)$, on estime θ par $\hat{\theta}_N(X_1^j, \dots, X_n^j)$ et on prend $\hat{F}_j = F_j^{\hat{\theta}_N(X_1^j, \dots, X_n^j)}$ comme modèle marginal.
- Estimation non-paramétrique : fonction de répartition marginale empirique, reconstruction à noyaux, histogramme etc.

Estimation statistique

Estimation de la copule

On distingue là encore plusieurs méthodes :

- Estimation paramétrique : $C^\theta \in \mathcal{C}(\theta)$, on estime θ par $\hat{\theta}_N(U_1^j, \dots, U_n^j)$ et on prend $\hat{C} = C^{\hat{\theta}_N}(X_1^j, \dots, X_n^j)$ comme estimation de la copule.
- Estimation non-paramétrique multivariée plus réciproque du théorème de Sklar : reconstruction à noyaux, histogramme etc. suivi du filtrage des marginales (légèrement) non uniformes obtenues.
- Estimation semi paramétrique dans une classe de copules à espace de paramétrage infini : copule archimédienne, copule elliptique, voir les travaux de P. Lambert, K. Kostadinov, A. Charpentier, J-D. Fermanian.

Test d'adéquation

test d'adéquation de copules

Ce domaine est encore en plein essor, depuis le travail pionnier de J-D. Fermanian basé sur une comparaison du modèle proposé avec une reconstruction à noyaux multivariée. Voir les travaux suivants :

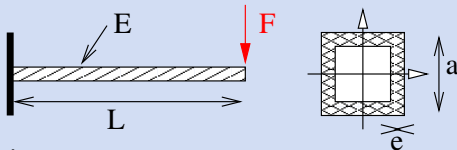
- C. Genest, B. Rémillard, D. Beaudoin, *Goodness-of-fit tests for copulas : A review and a power study*, Insurance Mathematics & Economics, in press.
- D. Fermanian, *Goodness-of-fit tests for copulas*, Journal of Multivariate Analysis 95 (2005) 119-152.

Bonne nouvelle : en 2D, les tests semblent performants dès $N = 150$ observations

Est-ce que tout ça en vaut la peine ?

Un petit exemple

On considère une poutre console élastique homogène isotrope dans l'hypothèse des petites perturbations :



Les données incertaines sont :

- E , F , L et I , l'inertie de flexion de la poutre.
- E et F sont indépendantes des autres variables.
- une étude statistique donne $\tau(L, I) = 0.5$.
- $E \sim \mathcal{TN}(\mu_E, \sigma_E, a_E, b_E)$, $F \sim \mathcal{U}(a_F, b_F)$, $L \sim \mathcal{U}(a_L, b_L)$, $I \sim \mathcal{B}(s_I, t_I, a_I, b_I)$

Objectif : calculer la probabilité pour que la déflexion d soit supérieure à s

Est-ce que tout ça en vaut la peine ?

Modélisation de la dépendance et résultat

- On modélise la dépendance entre L et I en utilisant soit une copule normale, soit une copule de Clayton.
- Le paramétrage de ces copules est tel que $\tau(L, I) = 0.5$.
- On calcule la probabilité de l'événement $\{d > s\}$ par la méthode de Monté Carlo.
- On obtient :
 - $\mathbb{P}(d > s) = 32.10^{-7} \pm 4.10^{-7}$ avec une confiance de 95% pour la copule de Clayton
 - $\mathbb{P}(d > s) = 104.10^{-7} \pm 4.10^{-7}$ avec une confiance de 95% pour la copule normale.

Soit un facteur supérieur à 3 entre les deux calculs ! Il aurait sans doute été plus performant d'utiliser les données ayant conduit à $\tau(L, I) = 0.5$ à identifier directement la copule.

Utiliser les copules en pratique

OpenTURNS : une plateforme logicielle Open Source pour le traitement des incertitudes

- Le modèle de données de la plateforme **OpenTURNS** est axé sur la notion de fonction de répartition multivariée, ces lois étant principalement construites par **assemblage de lois marginales et de copules**.
- De nombreux exemples de mise en œuvre des copules dans le contexte de propagation probabiliste d'incertitudes sont fournis avec la plateforme.
- Une documentation scientifique et utilisateur permet de débiter très facilement dans l'usage effectif des copules pour la modélisation probabiliste des incertitudes.
- Rendez-vous sur **www.openturns.org**; -) !

Challenges scientifiques

Estimation non paramétrique et test d'adéquation

Les problèmes d'estimation non paramétrique et de test d'adéquation de copules restent des problèmes difficiles :

- Les tests sont d'autant moins puissants qu'on est en grande dimension $d > 4$.
- L'estimation non paramétrique est sensible à la manière d'estimer les marginales.

Ces problématiques sont au cœur des travaux sur l'estimation de risque en finance, et les progrès sont rapides. **L'enjeu est de rendre plus robuste la sélection d'une copule.**

Challenges scientifiques

Lien avec les processus stochastiques

- La notion de copule est très performante dans la modélisation de la loi d'un vecteur aléatoire.
- Un premier lien a été fait entre la théorie des copules et celle des processus de Markov vectoriels.
- Par contre, le lien entre la copule de la loi marginale d'un processus vectoriel quelconque et la loi d'un échantillon temporel de ce processus fait encore n'a pas encore été étudié en détail.

L'enjeu est de fournir de nouveaux outils théoriques et pratiques pour paramétrer la structure de dépendance de processus, notamment pour construire des processus non gaussiens vectoriels à structure de dépendance marginale donnée.

Conclusion

Une technologie efficace

- Toutes les étapes depuis l'estimation statistique à partir de données multivariées jusqu'à la simulation de Monté Carlo peuvent être réalisées via une modélisation à base de copules.
- A partir d'un ensemble de copules, il est possible d'en créer de nouvelles par assemblage de manière efficace.

Des challenges scientifiques

- Fléau de la dimension pour les aspects statistiques.
- Lien avec les processus.

Conclusion

Que retenir de cette présentation ?

- La notion de dépendance stochastique est **exactement** couverte par le concept de copule ;
- Traiter cette notion **uniquement** à l'aide de **corrélations linéaires** est (en général) une **très mauvaise idée** ;
- **Tout** modèle probabiliste multivarié possède (au moins) une copule...
- ... cependant, d'autres descriptions de la dépendance peuvent être plus adaptés au calcul ou à l'estimation statistique (processus, modèles bayésiens).

Pour en savoir plus...

Bases

- J.F. Delmas, *fonction de répartition et copules*, cours de la chaire Risques financiers (X, ENPC, SG), 2008.
- R. B. Nelsen, *An Introduction to Copula 2nd Ed.*, Springer, 2006.
- K.-T. Fang, S. Kotz, K. W. NG, *Symmetric Multivariate and Related Distributions*, Chapman and Hall, 1990.
- H. Joe, *Multivariate Models and Dependence Concepts*, Chapman & Hall, 1997.

Pour en savoir plus...

Estimation et tests

- P. Lambert, *Archimedean Copula Estimation using Bayesian Splines Smoothing Techniques*, IAP Statistics Network, Technical report 0550, 2005.
- K. Kostadinov, *Non-parametric Estimation of Elliptical Copulae With Application to Credit risk*, working paper.
- D. Berg, H. Bakken, *Copula Goodness-of-fit Tests : A Comparative Study*, working paper.
- C. Genest, B. Rémillard, D. Beaudoin, *Goodness-of-fit tests for copulas : A review and a power study*, Insurance Mathematics & Economics, in press.
- J.D. Fermanian, *Goodness-of-fit tests for copulas*, J. Multivariate Anal. 95 (2005), pp. 119–152.

Pour en savoir plus...

Processus

- W. F. Darsow, B. Nguyen, E. T. Olsen, *Copulas and Markov processes*, Illinois Journal of Mathematics, Vol. 36, N. 4, Winter 1992.
- R. Ibragimov, *Copula-based characterizations for higher-order Markov processes*, Econometric Theory, Cambridge University Press, 2009.

Merci pour votre attention !

Vos questions, remarques et
commentaires sont les bienvenus.