Cours 1* : *Copules et leurs propriétés

1. Introduction

La copule est un outil relativement innovant pour modéliser la structure de dépendance de plusieurs variables aléatoires. Le concept de copule a été introduit par Sklar (1959) pour résoudre un problème de probabilité énoncé par Maurice Fréchet.
Le terme copule (copula) vient du mot latin "*copũlae*", qui signifie au sens figuré, liaison, lien, alliance ou union. D'une façon explicite, les copules sont des fonctions de répartition particulières, qui lient les fonctions de répartition multivariées de lois de probabilité dans $\mathbb{R}^d$, pour $d \geq 2$, aux fonctions de répartition marginales de leurs coordonnées. La caractéristique des copules permet de séparer les distributions marginales de la structure de dépendance.

2. Les mesures de dépendance

Une des premières mesures de dépendance introduite dans l'étude des statistiques est probablement le coefficient de corrélation de Pearson, une mesure de dépendance linéaire entre deux vas quantitatives. Par contre, lorsque nous travaillons avec des variables aléatoires continues non normales, l'utilisation de notions basées sur la linéarité n'est pas toujours appropriée. De plus, contrairement au coefficient de corrélation de Pearson, certaines mesures sont invariantes sous transformations strictement croissantes et valent 1 pour la borne supèrieure de Fréchet et -1 pour la borne infèrieure de Fréchet, ce qui, en ce sens, les rendent plus intéressantes. C'est le cas entre autres du tau de Kendall τ et du rhô Spearman ρ, deux mesures d'association bivariées qui mesurent la concordance. Elles donnent une mesure de la corrélation entre les rangs des observations, à la différence du coefficient de corrélation linéaire qui mesure la corrélation entre les valeurs des observations. Elles offrent par ailleurs l'avantage de s'exprimer simplement en fonction de la copule associée au couple de variables aléatoires.

2.1 Le coefficient de Corrélation linéaire de Pearson:

Soient X et Y deux variables aléatoires. La covariance de X et Y est défini comme suit :

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$

Définition 1 *Soient X et Y deux v.a continues des variances finies $var(X)$ et $var(Y)$ respectivement. Le coefficient de Pearson entre X et Y est donné par :*

$$\rho(x,y) = \frac{cov(x,y)}{\sqrt{var(X)var(Y)}}, \rho \in [-1,1]$$

Remarque 1.

Il est important de rappeler que la dépendance et la corrélation sont des notions différentes. En effet, si X et Y sont des variables indépendantes elles sont non corrélées mais la réciproque est fausse .

$$X \text{ et } Y \text{ sont indépendantes} \Longrightarrow \rho = 0$$

Concepte de concordance

Définition 2 *Soient (X_1, Y_1), (X_2, Y_2) deux observations d'un couple de vas continus (X,Y) sont dites que concordantes ou discordantes si :*

- ***Concordantes :*** $(x_1 - x_2)(y_1 - y_2) > 0 \iff (x_1 < x_2 \text{ et } y_1 < y_2)$ *ou* $(x_1 > x_2 \text{ et } y_1 > y_2)$.

- ***Discordantes :*** $(x_1 - x_2)(y_1 - y_2) < 0 \iff (x_1 < x_2 \text{ et } y_1 > y_2)$ *ou* $(x_1 > x_2 \text{ et } y_1 < y_2)$.

Définition 3 *(Fonction de concordance)*

La fonction de concordance est la différence entre la probabilité de concordance et celle de discordance entre deux couples (X_1, Y_1) *et* (X_2, Y_2)*: Elle est donnée par :*

$$Q = P\left[(X_1 - X_2)(Y_1 - Y_2) > 0\right] - P\left[(X_1 - X_2)(Y_1 - Y_2) < 0\right]$$

2.2 Le tau de kendall

Le tau de Kendall est une mesure de la dépendance entre deux vas, mais contrairement au coefficient de corrélation linéaire, il ne dépend pas des lois des deux variables étudiés mais que de leur ordre. Il est d'une importance fondamentale dans l'étude des copules, car il existe une relation analytique entre le taux de Kendall et le copule.

Définition 4 *Soient* (X_1, Y_1) *et* (X_2, Y_2) *deux vecteurs aléatoires continues, (iid) de fonction de répartition conjointe* H *et de lois marginales* F *et* G. *(*F *pour* X_1, X_2 *et* G *pour* Y_1, Y_2 *).*
Le tau de Kendall du vecteur aléatoire (X, Y)*, noté* τ *est défini par :*

$$\tau = P\left[(X_1 - X_2)(Y_1 - Y_2) > 0\right] - P\left[(X_1 - X_2)(Y_1 - Y_2) < 0\right]$$

2.3 Rho de spearman

Le coefficient de corrélation rho de Spearman représente l'une des mesures les plus connues pour quantiffier le degré d'association entre deux variables aléatoires. La valeur de ce coefficient dénotée par ρ_s est équivalente au coefficient de corrélation de Pearson. Le rho de Spearman de deux variables aléatoires X et Y est égal au coefficient de corrélation entre les variables $F(X)$ et $G(Y)$ distribuées selon la loi uniforme sur l'intervalle $I = [0, 1]$ tel que :

$$\rho_s (X, Y) = \rho (F(X), G(Y))$$

Définition 5 *Soient $(X_1, Y_1), (X_2, Y_2)$ et (X_3, Y_3) des copies indépendantes de vecteur aléatoire (X, Y). Le rho de Spearman noté ρ_s, est définie par :*

$$\rho_s = 3 (P [(X_1 - X_2) (Y_1 - Y_2) > 0] - P [(X_1 - X_2) (Y_1 - Y_2) < 0])$$

3. Copules et leurs propriétés

Les copules constituent un outil statistique assez récent permettant de modéliser la dépendance entre des variables aléatoires. La fonction copule permet de relier la fonction de densité conjointe aux fonctions de densités marginales. La construction des copules repose sur les propriétés des fonctions de répartition. On rappelle ci-dessous quelques propriétés importantes des fds bivariées.

• Soit (X, Y) un couple de variables aléatoires (v.a), la loi du couple est caractérisée par la fonction de répartition bivariée H définie comme suit :

$$\forall (x, y) \in \mathbb{R}^2 : H (x, y) = P [X \leq x, Y \leq y]$$

• On appelle lois marginales, les lois de X et de Y prises séparément. On peut exprimer les fonctions de répartition de ces lois marginales en fonction de H. Par exemple, pour X et Y on obtient

$$F_X(x) = \lim_{y \longrightarrow \infty} H(x, y)$$
$$G_Y(y) = \lim_{x \longrightarrow \infty} H(x, y)$$

- On rappelle que les v.a X et Y sont indépendantes ssi :

$$\forall (x, y) \in \mathbb{R}^2 : H(x, y) = F(x)G(y)$$

où F et G sont les lois marginales de X et Y . On note par la suite :

$$F^{-1}(t) = \inf \{x \in \mathbb{R} : F(x) \geq t\}, \text{ pour } t \in [0, 1]$$

la fonction quantile qui est l'inverse de la fonction de répartition F. On rappelle le lemme d'inversion qui est couramment utilisé pour générer des échantillons d'une loi dont la fonction quantile est connue: si U suit une loi uniforme sur l'intervalle $I = [0, 1]$, alors $F^{-1}(U)$ suit la méme loi que X. Les fonctions de répartitions des lois marginales sont alors des fonctions continues et les v.a $F(x)$ et $G(y)$ suivent des lois uniformes sur l'intervalle I.

3.1 Définition d'une copule et le théorème d'existence:

Définition 6 *On appelle copule bivariée toute fonction C de $I^2 \longrightarrow I$ qui possède les propriétés suivantes :*

1. *La copule C est attachée (grounded), c-à-d :*

$$\forall (u, v) \in I^2, C(u, 0) = C(0, v) = 0 \tag{1}$$

2. *Les marges sont uniformes, c-à-d:*

$$\forall (u, v) \in I^2, C(u, 1) = u \text{ et } C(1, v) = v \tag{2}$$

3. *La copule C est 2-croissante c-à-d : $\forall (u_1, u_2), (v_1, v_2) \in I^2$ avec $u_1 \leq u_2$ et $v_1 \leq v_2$ on a:*

$$C(u_1, v_1) - C(u_1, v_2) - C(u_2, v_1) + C(u_2, v_2) \geq 0 \tag{3}$$

Cette définition signifie que C est une distribution des marginales uniformes.

Examples

- **Copule produit:** La copule produit, noté Π est définie par

$$C(u,v) = \Pi(u,v) = uv \ , \ \forall u,v \in I$$

Comme on observe cette copule vérifie les propriétés (1) et (2), telle que:

$$\begin{aligned} C(u,0) &= \Pi(u,0) = u0 = 0 \text{ et } C(0,v) = \Pi(0,v) = 0v = 0 \ , \\ C(u,1) &= \Pi(u,1) = u1 = u \text{ et } C(1,v) = \Pi(1,v) = 1v = v \ . \end{aligned}$$

pour la propriété (3), on a :

$$\Pi(u_1,v_1) - \Pi(u_1,v_2) - \Pi(u_2,v_1) + \Pi(u_2,v_2) = u_1v_1 - u_1v_2 - u_2v_1 + u_2v_2 \geq 0$$

où

$$\text{si } u_2 \geq u_1 \Longrightarrow \begin{cases} u_2v_2 \geq u_1v_2 \\ u_2v_1 \geq u_1v_1 \end{cases}$$

alors

$$u_2v_2 - u_2v_1 \geq u_1v_2 - u_1v_1 \Longrightarrow u_2v_2 - u_2v_1 - u_1v_2 - u_1v_1 \geq 0$$

- **Copule min:** La copule min, noté M, est définie par:

$$M(u,v) = \min(u,v), \quad \forall \ u,v \in I.$$

- **Copule max:** La copule max, noté W, est définie par:

$$W(u,v) = \max(u+v-1,0), \quad \forall \ u,v \in I.$$

3.2 Théorème d'existence :théorème de Sklar

Ce théorème représente la base de la théorie de copules, car il valide l'existence de la copule C qui permet de relier les lois marginales de deux vas pour obtenir la distribution jointe. Ce théorème a été développé par Sklar (1959) [?] et porte son nom. La relation entre la distribution jointe, les fonctions de distribution marginales et la copule C est montré dans le théorème suivant.

Théorème 7 *Soient X, Y deux v.a de fonction de répartition jointe H et des marginales F et G. Alors il existe une copule C telle que*

$$\forall (x, y) \in \mathbb{R}^2 : H(x, y) = C(F(x), G(y)) \tag{4}$$

Preuve 1 *Soit C la fonction de répartition du vecteur (U, V) où $U = F(X)$ et $V = G(Y)$:*

$$\begin{aligned} H(x, y) &= P[X \leq x, Y \leq y] \\ &= P\left[F^{-1}(U) \leq x, G^{-1}(V) \leq y\right] \\ &= P[U \leq F(x), V \leq G(y)] \\ &= C(F(x), G(y)) \end{aligned}$$

Corollaire 8 *(Inverse de théorème de Sklar)*

Soit H une fonction de répartition 2-dimentionnelle dont les marginales sont F et G. Alors la copule C associée à H est donnée par

$$C(u, v) = H(F^{-1}(u), G^{-1}(v)), \text{ pour tout } (u, v) \in I^2$$

Preuve 2 *Voir TD*

Remarque 2.

Si F et G sont continues, alors C est unique.

Sinon, lorsque les marginales ne sont pas continues, il est toujours possible de définir une copule mais elle ci n'est plus unique.

Relation entre copule et mesures de dépendance

Proposition 9 *Soit (X,Y) un couple de vas continues de copule C. Si les couples sont identiquement distribués i.e $H_1 = H_2 = H$, l'expression du tau de Kendall en terme de copules est donnée par :*

$$\tau_C = Q\left(C,C\right) = 4\int_0^1\int_0^1 C(u,v)dC(u,v) - 1$$

Comme les couples sont uniformément distribuées sur I, alors :

$$\int_0^1\int_0^1 C(u,v)dC(u,v) = E[C(u,v)]$$

Ainsi, on peut écrire :

$$\tau_C = 4E[C(u,v)] - 1$$

Théorème 10 *Soient X , Y deux vas continues dont la copule est C. Le rho de Spearman de X et Y est défini par :*

$$\begin{aligned}
\rho_C &= 3Q\left(C,\Pi\right) = 3\left(4\int_0^1\int_0^1 C\left(u,v\right)d\Pi\left(u,v\right) - 1\right) \\
&= 12\int_0^1\int_0^1 C\left(u,v\right)d(uv) - 3 \\
&= 12\int_0^1\int_0^1 C\left(u,v\right)dudv - 3
\end{aligned}$$

3.3 Propriétés des copules

1. **la continuité**

   Théorème 11 *Soit C une copule bivariée, alors $\forall u_1, u_2, v_1, v_2 \in I$ avec $u_1 \leq u_2$ et $v_1 \leq v_2$, on a :*

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

   les copules sont des fonctions continues. Plus précisément, elles vérifient une condition de Lipschitz.

2. **la différentiabilité**

Théorème 12 *Soit C une copule bivariée, $\forall u_1, u_2 \in I$*

1 les dérivées partielles $\frac{\partial C(u_1,u_2)}{\partial u_j}$ existe p.s et $0 \leq \frac{\partial C(u_1,u_2)}{\partial u_j} \leq 1, \forall$ $j = 1, 2$.

2 les fonctions $u \longrightarrow \frac{\partial C(u,v)}{\partial v}$ et $v \longrightarrow \frac{\partial C(u,v)}{\partial u}$ sont bien définies et décroissantes sur I p.s.

3. **la convexité et concavité:**

Soit (a, b), $(c, d) \in I^2$ et $\lambda \in I$.

• Une copule C est dite **concave** si :

$$C(\lambda a + (1 - \lambda)c, \lambda b + (1 - \lambda)d) \geq \lambda C(a, b) + (1 - \lambda)C(c, d)$$

• Une copule C est dite **convexe** si :

$$C(\lambda a + (1 - \lambda)c, \lambda b + (1 - \lambda)d) \leq \lambda C(a, b) + (1 - \lambda)C(c, d)$$

4. **L'ordre:**

Soit C_1,C_2 deux copules. On dit que C_1 est plus petite que C_2 ou C_2 est plus grande que C_1 et on note $C_1 < C_2$ si

$$\forall (u, v) \in I^2, C_1(u, v) < C_2(u, v)$$

comme exemple on a La copule $W = \max(u+v-1, 0)$ est la plus petite copule et $M = \min(u, v)$ est la plus grande copule.

3.4 Les bornes de Fréchet-Hoeffding:

Toute copule bivariée C est bornée par deux copules. Ces bornes sont déterminées dans le théorème suivant

Théorème 13 *Soit C une copule alors :*

$$\forall (u,v) \in I^2 : W(u,v) = \max(u+v-1,0) \leq C(u,v) \leq \min(u,v) = M(u,v)$$

Les copules W et M sont appelées borne infèrieure (respectivement borne supèrieure) de Fréchet-Hoeffding.

Preuve 3 *Soit $(u,v) \in I^2$ on a :*

$$\begin{cases} C(u,v) \leq C(u,1) = u \\ C(u,v) \leq C(1,v) = v \end{cases} \Rightarrow C(u,v) \leq \min(u,v) = M(u,v)$$

d'où la deuxième égalitée.

Pour montrer la première égalitée, nous utilisons le fait que $C(u,v) \geq 0$ et l'inégalité (3) :

$$\begin{aligned} C(1,1) - C(1,v) - C(u,1) + C(u,v) &\geq 0 \\ &\Longrightarrow 1 - v - u + C(u,v) \geq 0 \\ &\Longrightarrow C(u,v) \geq u + v - 1 \\ &\Longrightarrow C(u,v) \geq \max(u+v-1,0) \\ &\Longrightarrow C(u,v) \geq W(u,v) \end{aligned}$$

Donc: $W(u,v) \leq C(u,v) \leq M(u,v)$

Comme une conséquence de théorème de Sklar, si X, Y sont des variables aléatoires de fonction de répartition jointe H et des marginales F et G, respectivement, alors pour tout $x, y \in \mathbb{R}$

$$max(F(x) + G(y) - 1, 0) \leq H(x,y) \leq min(F(x), G(y)).$$

Car M et W sont des copules, les bornes ci-dessus sont des fonctions de répartition jointes et sont appelés les bornes de Fréchet-Hoefding pour une fonction de répartition jointe H et des marginales F et G.

3.5 Théorème d'invariance:

L'un des théorèmes essentiels dans la théorie des copules est celui de l'invariance par transformations strictement croissantes.

Soient deux vas continues X et Y de marges F et G et de copule C_{XY} . Si α et β sont deux fonctions strictement croissantes, alors:

$$C_{\alpha(X),\beta(Y)} = C_{XY}$$

Preuve 4 *On peut démontrer ce théorème facilement à l'aide de lois de probabilité, comme suit :*

Soient F_1, G_1, F_2, G_2 les fds de X, Y, $\alpha(X)$ et $\beta(Y)$, respectivement où α et β sont strictement croissantes, alors:

$$F_2 = P\left[\alpha\left(X\right) \leq x\right] = P\left[X \leq \alpha^{-1}(x)\right] = F_1\left(\alpha^{-1}(x)\right)$$

aussi :

$$G_2 = P\left[\beta\left(Y\right) \leq y\right] = P\left[Y \leq \beta^{-1}(y)\right] = G_1\left(\beta^{-1}(y)\right)$$

on a donc :

$$\begin{aligned}
C_{\alpha(X),\beta(Y)}\left(F_2\left(x\right), G_2\left(y\right)\right) &= P\left[\alpha\left(X\right) \leq x, \beta\left(Y\right) \leq y\right] \\
&= P\left[X \leq \alpha^{-1}(x), Y \leq \beta^{-1}(y)\right] \\
&= C_{XY}\left(F_1\left(\alpha^{-1}(x)\right), G_1\left(\beta^{-1}(y)\right)\right) \\
&= C_{XY}\left(F_2\left(x\right), G_2\left(y\right)\right)
\end{aligned}$$

alors toute copule vérifie cette propriétée.

Exemple 14 *nous avons :*

$$\begin{aligned} C_{XY} &= C_{\ln XY} \\ &= C_{\ln X \ln Y} \\ &= C_{X \exp(Y)} \\ &= C_{\sqrt{X} \exp(Y)} \\ & \quad . \\ & \quad . \\ & \quad . \end{aligned}$$

Une transformation croissante ne modifie donc pas la copule, mais seulement les loi marginale.

3.6 Théorème d'indépendance

Soient X et Y deux vas continues et C_{XY} la copule associée.

Alors X et Y sont indépendantes ssi $C_{XY}(u, v) = uv = \Pi(u, v)$ pour tout $u, v \in I$.

Preuve 5

• *On suppose que* X *et* Y *sont indépendantes alors:* $\forall x, y \in \mathbb{R}, H(x, y) = F(x)G(y)$, *et on montre que* $C_{XY} = \Pi$

On a :

$$\begin{aligned} \text{Or } \forall x, y \in I, C_{XY}(u, v) &= H(F^{-1}(u), G^{-1}(v)) \\ &= F(F^{-1}(u))G(G^{-1}(v)) \\ &= uv = \Pi(u, v) \end{aligned}$$

- *Supposons maitenant que* $C_{XY} = \Pi$ *alors :*

  $\forall x, y \in I, C_{XY}(u, v) = \Pi(u, v) = uv$

On a :

$$\begin{aligned} \forall x, y \in \mathbb{R}, H(x, y) = C_{XY}(F(x), G(y)) \\ = \Pi(F(x), G(y)) = F(x)G(y) \end{aligned}$$

$\Longrightarrow$*Les vas* X *et* Y *sont donc indépendantes.*

3.7 Densité de la copule

Les copules admettent des densités de probabilités. Si la densité c associée à la copule C existe,alors elle est définie par :

$$c(u, v) = \frac{\partial^2 C(u, v)}{\partial u \partial v}$$

où $c : I^2 \to \mathbb{R}^+$Si la fonction de distribution jointe H est absolument continue, en utilisant le théorème de Sklar, on peut exprimer la densité d'un couple aléatoire (X, Y) en fonction de la densité de sa copule et de ses marginales f et g par :

$$h(x, y) = c(F(x), G(y))\ f(x)g(y).$$