ERM
M2A, année 2022 - 2023

TD 2 : COPULES

Exercice 1
Soit (U, V) un vecteur Gaussien centré, dont les variances sont $Var(U) = a^2$ et $Var(V) = b^2$, et de corrélation $r \in [-1, 1]$. On pose

$$X = \frac{U}{a} + \frac{V}{b} \text{ et } Y = \frac{U}{a} - \frac{V}{b}.$$

Donner la copule du couple (X, Y).

Exercice 2
Déterminer les distributions marginales $F(x)$ et $G(x)$ associées à la distribution $H(x, y)$ puis construire la copule $C(u, v)$ associée à H par le théorème de Sklar.

$$H(x, y) = \frac{1}{1 + e^{-x} + e^{-y}} \text{ ; } pour\ x, y \in \mathbb{R}$$

Exercice 3
On considère deux variables alatoires indpendantes X_1 et X_3 telles que $X_1 \sim Exp(1)$ et $X_3 \sim Exp(2)$. On pose alors $X_2 = X_1 + X_3$ et $X = (X_1, X_2)$.

(1) **Copules**
   (a) Déterminer la fonction de répartition de X_2.
   (b) Déterminer la fonction de répartition jointe de X. Vérifier alors le résultat de (a).
   (c) On note désormais F_1 et F_2 les fonctions de répartition de X_1 et X_2. Déterminer la fonction copule associée à F_X, ie. la fonction de C de $[0, 1]^2$ dans $[0, 1]$ telle que :

$$F_X(X_1, X_2) = C(F_1(X_1), F_2(X_2)).$$

(2) **Corrélations**
   (a) Calculer le ρ de Spearmann de X, $\rho(X, Y) = 12 \iint_{[0,1]^2} C(u, v) dudv - 3$
   (b) Calculer le τ de Kendall de X de trois manières différentes :
      (i) A l'aide de la définition $\tau = \mathbb{P}[(X_1 - X_1')(X_2 - X_2') > 0] - \mathbb{P}[(X_1 - X_1')(X_2 - X_2') < 0]$
      (ii) A l'aide de la formule $\tau = 4\ \mathbb{E}[F_X(X_1, X_2)] - 1$
      (iii) A l'aide des copules $\tau(X, Y) = 4 \iint_{[0,1]^2} C(u, v) dC(u, v) - 1$

c) Déterminons la fonction copule associée à F_X.

$X = (X_1, X_2) \quad X_2 = X_1 + X_3$

On a $F_2(x) = (1 - e^{-x})^2$

$$F(x_1, x_2) = (1 + e^{-2x_2}) - (e^{-\min(x_1, x_2)} + e^{\min(x_1, x_2) - 2x_2})$$

$$F_1(x) = (1 - e^{-x})$$

- Inverse de $F_1(x)$.

$F_1(x) = 1 - e^{-x}$

$(1 - e^{-x}) = y$

$\Rightarrow x = -\ln(1-y)$

$\Rightarrow F_1^{-1}(u) = -\ln(1-u)$

$F_2(x) = (1 - e^{-x})^2$

$(1 - e^{-x})^2 = y$

$\Rightarrow x = -\ln(1 - \sqrt{y})$

$\Rightarrow F_2^{-1}(v) = -\ln(1 - \sqrt{v})$

on a : $C(u, v) = C(F_1(F_1^{-1}(u)), F_2(F_2^{-1}(v)))$.

$\Rightarrow$ Calculons $\min[-\ln(1-u), -\ln(1-\sqrt{v})]$.

La fonction $x \longmapsto -\ln(1-x)$ est croissante donc :

$\min[-\ln(1-u), -\ln(1-\sqrt{v})] = -\ln(1 - \min(u, \sqrt{v}))$.

Par suite :

$$C(u,v) = (1 + e^{2\ln(1-\sqrt{v})}) - \left[\frac{1}{1-\min(u,\sqrt{v})}(1 - \sqrt{v})^2 + 1 - \min(u, \sqrt{v})\right].$$

$$\boxed{C(u,v) = (1-\sqrt{v})^2 + \min(u, \sqrt{v}) - \frac{(1-\sqrt{v})^2}{1 - \min(u, \sqrt{v})}}$$

2) Corrélations.

Si on utilise le coefficient de corrélation linéaire et le taux de Kendall et qu'on trouve des différences, cela veut dire qu'il y a des phénomènes non linéaires dans notre jeu de données $\Rightarrow$ il faut utiliser du carré, parce que tout ce qu'on a vu de côté linéaire est inutile.

Calculons le taux de Kendall de 3 manières :

$(X_1, X_2) \perp\!\!\!\perp (X_1', X_2')$; on a X_1, X_1' iid et X_2, X_2' iid

i) A l'aide de la définition $\tau = P[(X_1 - X_1')(X_2 - X_2') > 0]$
$- P[(X_1 - X_1')(X_2 - X_2') < 0]$

Posons $A = (X_1 - X_1')(X_2 - X_2')$.

On a :

$$\tau = P[A > 0] - P[A < 0] = 1 - 2P[A < 0]$$
$$= P[A > 0] - 1 + P[A > 0] = 2P[A > 0] - 1$$

On calcule $P[A > 0]$ car c'est plus simple, étant donné que c'est un produit :

$$P[A > 0] = P[(X_1 - X_1') > 0 \cap (X_2 - X_2') > 0] +$$
$$P[(X_1 - X_1') < 0 \cap (X_2 - X_2') < 0].$$

les 2 ont la m̂ probabilité car iid

$X_1 > X_1'$ et $X_2 > X_2'$

$X_1 < X_1'$ et $X_2 < X_2'$

les 2 ont la m̂ proba car iid.

$$\text{Donc } P[A > 0] = 2 \underbrace{P[X_1 < X_1', X_2 < X_2']}_{B}$$

$$B = \mathbb{P}[X_1 < X_1', X_2 < X_2']$$

$$= \mathbb{P}[X_1 < X_1', X_1 + X_3 < X_1' + X_3']$$

$$= \int_0^{\infty} \int_0^{x_1'} \mathbb{P}[X_1 + X_3 < X_1' + X_3' \mid X_1 = x_1, X_1' = x_1'] f(x_1) f(x_1') \, dx_1 \, dx_1'$$

$$= \int_0^{\infty} \int_0^{x_1'} \mathbb{P}[X_3 < X_3' - (x_1 - x_1')] f(x_1) f(x_1') \, dx_1 \, dx_1'$$

$$= \int_0^{+\infty} \int_0^{x_1'} \int_{x_1 - x_1'}^{+\infty} \mathbb{P}(X_3 < x_3' - (x_1 - x_1')) f(x_1) f(x_1') f(x_3') \, dx_3' \, dx_1 \, dx_1'$$

$$= \int_0^{+\infty} \int_0^{x_1'} \int_{x_1 - x_1'}^{+\infty} (1 - e^{-2(x_3' - (x_1 - x_1'))}) e^{-x_1 - x_1' - 2x_3'} \, dx_3' \, dx_1 \, dx_1'.$$

$$B = \frac{5}{12}$$

$$P[A > 0] = 2 \times \frac{5}{12} = \frac{5}{6}$$

$$\varepsilon = 2P[A > 0] - 1 = \frac{5}{3} - 1 = \frac{5}{3} - \frac{3}{3}$$

$$\boxed{\varepsilon = \frac{2}{3}}$$