Let (X, Y) be a bivariate random vector with joint distribution function $H(x, y)$ and continuous marginal distribution function $F(x)$ and $G(y)$. Its associated copula C is defined by $H(x, y) = C\{F(x), G(y)\}$. Since F and G are continuous, the copula C is unique and we can write $C(u, v) = H(F^{\leftarrow}(u), G^{\leftarrow}(v))$ for $0 \leq u, v \leq 1$ and where $F^{\leftarrow}(u) = inf\{v \in \mathbb{R} | F(v) \geq u$ and $G \leftarrow (u) = inf\{v \in \mathbb{R} | G(v) \geq u\}$ are the generalized inverse functions of F and G respectively. At each $t \in \{1, \ldots, T\}$, we suppose that one of both entries may be missing. The observations consist of a sample of independent, identically distributed quadruples

$$(I_t, J_t, I_t X_t, J_t Y_t) \quad t \in 1, \ldots, T$$

The indicator variables I_t (respectively J_t) is equal to 1 or 0 according to wheter X_t or Y_t is observed or not. We suppose that the indicator functions I_t and J_t are independent. The probability of observing a realisation partially or completely is denoted by $p_X = \mathbb{P}(I_t = 1) > 0$ and $p_Y = \mathbb{P}(J_t = 1) > 0$.

We define :

$$C(u, v) = H(F^{\leftarrow}(u), G^{\leftarrow}(v)) = \varphi(H)(u, v)$$

and,

$$Z_T(u, v) = \sqrt{T}\{\hat{C}_T(u, v) - C(u, v)\} \tag{1}$$

where $\hat{H}_T$ corresponds to the empirical distribution function of the sample $(X_1, Y_1), \ldots, (X_T, Y_T)$

$$\hat{H}_T(u, v) = \frac{\sum_{t=1}^{T} 1_{\{X_t \leq u, Y_t \leq v\}} I_t J_t}{\sum_{t=1}^{T} I_t J_t}$$

We define also the corresponding empirical distribution functions in the case of missing data :

$$\hat{F}_T(u) = \frac{\sum_{t=1}^{T} 1_{\{X_t \leq u\}} I_t}{\sum_{t=1}^{T} I_t}$$

$$\hat{G}_T(v) = \frac{\sum_{t=1}^{T} 1_{\{Y_t \leq v\}} J_t}{\sum_{t=1}^{T} J_t}$$

Hypothèse 1. *We suppose for all $t \in \{1, \ldots, T\}$:*

$$\mathbb{E}[X_t Y_t I_t J_t] = \mathbb{E}[X_t Y_t]\mathbb{E}[I_t J_t]$$

Furthermore, we suppose that there exist at least one $t \in \{1, \ldots, T\}$ such that $I_t J_t \neq 0$.

Proposition 1. *Under hypothesis 1, $\hat{H}_T, \hat{F}_T, \hat{G}_T$ are consistant estimators of H, F, G.*

Démonstration. We check the consistency for $\hat{H}_T$. By independence, we have

$$\mathbb{E}[T^{-1} \sum_{t=1}^{T} I_t J_t] = T^{-1} \sum_{t=1}^{T} \mathbb{E}[I_t]\mathbb{E}[J_t] = p_X p_Y$$

So, by applying the law of large numbers, we have :

$$T^{-1}\sum_{t=1}^{T} I_t J_t \longrightarrow p_X p_Y \quad a.s. \quad as \quad T\to\infty$$

Then, we now use the first hypothesis to get :

$$T^{-1}\sum_{t=1}^{T}\mathbb{E}[1_{\{X_t\le u, Y_t\le v\}} I_t J_t] = T^{-1}\sum_{t=1}^{T}\mathbb{E}[1_{\{X_t\le u, Y_t\le v\}}]\mathbb{E}[I_t J_t] = H(x,y)p_X p_Y$$

By applying again the law of large numbers, we derive :

$$\sum_{t=1}^{T} 1_{\{X_t\le u, Y_t\le v\}} I_t J_t \longrightarrow H(x,y)p_X p_Y \quad a.s. \quad as \quad T\to\infty$$

We can now apply the continuous mapping theorem to the function $f : (x,y)\mapsto \frac{x}{y}$ which are continuous on $\mathbb{R}_+\times\mathbb{R}_+\setminus 0$ to conclude that :

$$\hat{H}_T(x,y)\longrightarrow H(x,y) \quad a.s. \quad as \quad T\to\infty$$

□

Hypothèse 2. *1. The bivariate distribution function H has continuous margins F, G and copula C.*

2. The first order partial derivatives $\dot{C}_1(u,v)=\frac{\partial C}{\partial u}(u,v)$ and $\dot{C}_2(u,v)=\frac{\partial C}{\partial v}(u,v)$ exists and is continuous on the set $\{(u,v)\in[0,1]^2, 0<u,v<1\}$

Hypothèse 3. *There exists $\gamma_t>0$ and $r_t>0$ such that $r_t\longrightarrow\infty$ as $t\to\infty$ such that in the space $l^\infty(\mathbb{R}^2)\otimes(l^\infty(\mathbb{R}), l^\infty(\mathbb{R}))$ equipped with the topology of uniform convergence, we have the joint weak convergence*

$$(r_t(\hat{H}_T-H); r_t(\hat{F}_t-F), r_t(\hat{G}_t-G))\rightsquigarrow(\alpha\circ(F,G), \beta_1\circ F, \beta_2\circ G)$$

The stochastic processes α and β_j take values in $l^\infty([0,1]^2)$ and $l^\infty([0,1])$ respectively, and are such that $\alpha\circ F$ and $\beta_j\circ F_j$ have continuous trajectories on $[-\infty,\infty]^2$ and $[-\infty,\infty]$ almost surely.

Théorème 1 (Theorem 2.3 and example 3.5 in [Seg14])**.** *If conditions 2.1 and 2.2 holds, then uniformly in $u\in[0,1]^2$,*

$$r_T\{\hat{C}_T(u,v)-C(u,v)\} = r_T\{\hat{H}_T((F,G)^{\leftarrow}(u,v)-C(u,v)\} \tag{2}$$

$$-\dot{C}_1(u,v)r_T\{\hat{F}_T(F^{\leftarrow}(u))-u\}1_{(0,1)}(u) \tag{3}$$

$$-\dot{C}_2(u,v)r_T\{\hat{G}_T(G^{\leftarrow}(v))-v\}1_{(0,1)}(v)+\circ_{\mathbb{P}}(1) \tag{4}$$

as $T\to\infty$. Hence in $l^\infty([0,1]^2)$ equipped with the supremum norm, as $T\to\infty$,

$$(r_T\{\hat{C}_T(u,v)-C(u,v)\})_{u,v\in[0,1]^2}\rightsquigarrow(\alpha(u,v)-\dot{C}_1(u,v)\beta_1(u)-\dot{C}_2(u,v)\beta_2(v))_{u,v\in[0,1]^2}$$

We denote by $N_C(u,v)$ the process defined on the right-hand side in the weak convergence from above.

In our specific case with missing data, (ref) shows that $r_T\{\hat{C}_T(u,v)-C(u,v)\}$ is weakly convergent toward $\alpha(u,v)-\dot{C}_1(u,v)\beta_1(u)-\dot{C}_2(u,v)\beta_2(v)$ where $\beta_1(u)=p_X^{-1}\mathbb{G}(1_{X\leq F^{\leftarrow}(u)}-u1_{I=1})$, $\beta_2(v)=p_Y^{-1}\mathbb{G}(1_{Y\leq G^{\leftarrow}(v)}-v1_{J=1})$ and $\alpha(u,v)=(p_Xp_Y)^{-1}\mathbb{G}(1_{X\leq F^{\leftarrow}(u)}1_{Y\leq G^{\leftarrow}(v)}-C(u,v)1_{I=1}1_{J=1})$. Furthermore, by these expressions, we can detail the structure of the covariance matrix between the three processes.

Definition 1. *Let $(X_1,Y_1),\dots,(X_T,Y_T)$ a T bivariate random vectors with unknown margins F and G. A λ-FMadogram is the quantity defined by :*

$$\nu(\lambda)=\frac{1}{2}\mathbb{E}[|F^{\lambda}(X)-G^{1-\lambda}(Y)|] \tag{5}$$

We estimative the λ-FMadogram with the following quantity :

$$\hat{\nu}_T(\lambda)=\frac{1}{2\sum_{t=1}^T I_tJ_t}\sum_{t=1}^T|\hat{F}_T^{\lambda}(X_t)-\hat{G}_T^{1-\lambda}(Y_t)|I_tJ_t \tag{6}$$

Proposition 2 (Proposition 3 of [NGCD09])**.** *Suppose that the assumption of theorem (1) holds. Let J be a function of bounded variation, continuous from above and with discontinuities of the first kind. Then*

$$(\sum_{t=1}^T I_tJ_t)^{-1/2}\sum_{t=1}^T(J(\hat{F}_T(X_t),\hat{G}_T(Y_t))I_tJ_t-\mathbb{E}[J(F(X),G(Y))])$$

converges in distribution to $\int_{[0,1]}N_C(u,v)dJ(u,v)$ where $N_C(u,v)$ is defined in theorem (1) and the integral is well defined as a Lebesgue-Stieltjes integral (see section 5.2 of [Bre20]). The special case, $J(x,y)=\frac{1}{2}|x^{\lambda}-y^{1-\lambda}|$ provide the weak of convergence of the λ-Madogram estimator defined by (6) :

$$(\sum_{t=1}^T I_tJ_t)^{1/2}\{\hat{\nu}_T(\lambda)-\frac{1}{2}\mathbb{E}[|F^{\lambda}(X)-G^{1-\lambda}(Y)|]\}$$

converge in distribution to $\int_{[0,1]}N_C(u,v)dJ(u,v)$ where the latter integral satisfies :

$$\int_{[0,1]^2}f(x,y)dJ(x,y)=\frac{1}{2}\int_0^1 f(0,y^{1/(1-\lambda)})dy+\frac{1}{2}\int_0^1 f(x^{1/\lambda},0)dx-\int_0^1 f(x^{1/\lambda},x^{1/(1-\lambda)})dx$$

for all bounded functions f.

Références

[Bre20] Jean-Christophe Breton. Processus stochastiques, 2020.

[NGCD09] Philippe Naveau, Armelle Guillou, Dan Cooley, and Jean Diebolt. Modeling pairwise dependence of maxima in space. *Biometrika*, 96(1) :1–17, 2009.

[Seg14] Johan Segers. Hybrid copula estimators, 2014.