Détection de rupture dans la copule d'observations multivariées avec possibilité de changement dans les lois marginales. Illustration et simulations de Monte Carlo.

–Séminaire LMAC de Compiègne–

Tom Rohmer

with collaboration from:
Ivan Kojadinovic (Université de Pau) & Jean-Franois Quessy (Université de Trois-Rivières)
and
Axel Bücher (Université d'Heidelberg) & Johan Segers (Université catholique de Louvain)

19 December 2017

Summary

Summary

Test for break detection

Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be d-dimensional random vectors, where for $i = 1, \ldots, n$, $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{id})$. We aim at testing the null hypothesis

$$\mathcal{H}_0 : \exists\, F \text{ such that } \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \text{ have c.d.f. } F.$$

Test for break detection

Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be d-dimensional random vectors, where for $i = 1, \ldots, n$, $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{id})$. We aim at testing the null hypothesis

$$\mathcal{H}_0 : \exists\, F \text{ such that } \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \text{ have c.d.f. } F.$$

Example of alternative hypotheses :

- An abrupt change at the instant $k^\star \in [\![1, n-1]\!]$:

$$\exists F^{(1)}, F^{(2)} \text{ such that } \boldsymbol{X}_1, \ldots, \boldsymbol{X}_{k^\star} \sim F^{(1)}, \quad \boldsymbol{X}_{k^\star+1}, \ldots, \boldsymbol{X}_n \sim F^{(2)}$$

Test for break detection

Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be d-dimensional random vectors, where for $i = 1, \ldots, n$, $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{id})$. We aim at testing the null hypothesis

$$\mathcal{H}_0 : \exists\, F \text{ such that } \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \text{ have c.d.f. } F.$$

Example of alternative hypotheses :

- A gradual change : $\exists 1 \leq k_1 \leq k_2 \leq n-1$
  $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{k_1} \sim F^{(1)}$
  $\boldsymbol{X}_{k_2}, \ldots, \boldsymbol{X}_n \sim F^{(2)}$
  For $i = k_1 + 1, \ldots, k_2$ the law of $\boldsymbol{X}_i$ will gradually go from $F^{(1)}$ to $F^{(2)}$.

Cumulative Sum test for $\mathcal{H}_0$, serially independent data I

Example 1 : Test for change in the mean, d=1

$$T_n^{\mu} = \max_{k=1,\dots,n-1} \frac{k(n-k)}{n^{3/2}} \left| \left\{ \frac{1}{k}\sum_{i=1}^{k} X_i - \frac{1}{n-k}\sum_{i=k+1}^{n} X_i \right\} \right|$$

$$= \sup_{s\in[0,1]} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} (X_i - \bar{X}_n) \right| .$$

- Under $\mathcal{H}_0$, as soon as $X_1, \dots, X_n$ are i.i.d.,

$$T_n^{\mu} \rightsquigarrow \sigma \sup_{s\in[0,1]} |\mathbb{U}(s)|,$$

where σ^2 is the unknown variance of X_i and $\mathbb{U}$ is a standard Brownian bridge, i.e. a centered Gaussian process with covariance function :

$$\mathrm{cov}\{\mathbb{U}(s), \mathbb{U}(t)\} = min(s,t) - st, \qquad s,t \in [0,1].$$

Cumulative Sum test for $\mathcal{H}_0$, serially independent data II

Example 2 : test à la [Csörgő et Horváth(1997)]

$$T_n^{\#} = \max_{k=1,\dots,n-1} \frac{k(n-k)}{n^{3/2}} \sup_{\boldsymbol{x} \in \mathbb{R}^d} |F_{1:k}(\boldsymbol{x}) - F_{k+1:n}(\boldsymbol{x})|$$
$$= \sup_{s \in [0,1]} \sup_{\boldsymbol{x} \in \mathbb{R}^d} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} \{\mathbf{1}(\boldsymbol{X}_i \leq \boldsymbol{x}) - F_{1:n}(\boldsymbol{x})\} \right|,$$

where for $1 \leq k \leq n$, $F_{1:k}$ (resp. $F_{k+1:n}$) is the empirical c.d.f. of the subsample $\boldsymbol{X}_1, \dots, \boldsymbol{X}_k$ (resp. $\boldsymbol{X}_{k+1}, \dots, \boldsymbol{X}_n$) :

$$F_{1:k}(\boldsymbol{x}) = \frac{1}{k} \sum_{i=1}^{k} \mathbf{1}(\boldsymbol{X}_i \leq \boldsymbol{x}), \qquad F_{k+1:n}(\boldsymbol{x}) = \frac{1}{n-k} \sum_{i=k+1}^{n} \mathbf{1}(\boldsymbol{X}_i \leq \boldsymbol{x}),$$

and

$$F_{1:n}(\boldsymbol{x}) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(\boldsymbol{X}_i \leq \boldsymbol{x}).$$

Strong mixing conditions

- In this work, we do not necessarily assume the observations to be serially independent. The asymptotic validity of techniques is proved for strongly mixing observations :

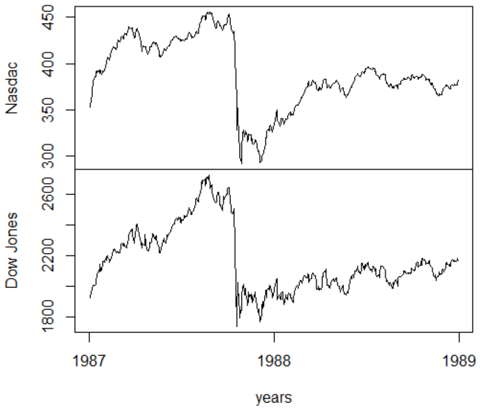
[Rosenblatt(1956)]

Consider a sequence of d-dimensional random vectors $(\boldsymbol{Y}_i)_{i\in\mathbb{Z}}$. For $s,t \in \mathbb{Z}\cup\{-\infty,+\infty\}$, let $\mathcal{F}_s^t$ the σ-algebra generated by $\boldsymbol{Y}_i$, $s \leq i \leq t$, i.e., $\mathcal{F}_s^t = \sigma(\boldsymbol{Y}_i, s \leq i \leq t)$. The strong mixing coefficient is defined by

$$\alpha_r = \sup_{t\in\mathbb{Z}} \sup_{A\in\mathcal{F}_{-\infty}^t, B\in\mathcal{F}_{t+r}^{+\infty}} |P(A\cap B) - P(A)P(B)|.$$

The sequence $(\boldsymbol{Y}_i)_{i\in\mathbb{Z}}$ is said α-mixing if $\alpha_r \to 0$ as $r \to +\infty$.

Nasdaq, Dow Jones and the "black Monday" (1987-10-19)



$\triangleright$: $\hat{p}_{val}^{T_n^{\#},\alpha=0.05} = 0.0004$: we conclude in favour of $\neg\mathcal{H}_0$.

Summary

The Sklar's theorem

Theorem of [Sklar(1959)]

Let $\boldsymbol{X} = (X_1, \ldots, X_d)$ be a d-dimensional random vector with c.d.f. $\boldsymbol{F}$ and let $F_1, \ldots F_d$ be the marginal c.d.f. of $\boldsymbol{X}$ assuming continuous. Then it exists a unique function $C : [0,1]^d \to [0,1]$ such that :

$$\boldsymbol{F}(\boldsymbol{x}) = C\{F_1(x_1), \ldots, F_d(x_d)\}, \qquad \boldsymbol{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d.$$

- The copula C characterizes the dependence structure of vector $\boldsymbol{X}$.
- The copula C can be expressed as follows :

$$C(\boldsymbol{u}) = \boldsymbol{F}\{F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)\}, \qquad \boldsymbol{u} = (u_1, \ldots, u_d) \in [0,1]^d.$$

A. Sklar.
Fonctions de répartition à n dimensions et leurs marges.
Publications de l'Institut de Statistique de l'Université de Paris, 8 : 229–231, 1959.

Classical copulas

- Independence copula :

$$C^{\Pi}(\boldsymbol{u}) = \prod_{j=1}^{d} u_j;$$

- Normal copulas

$$C^{N}_{\Sigma}(\boldsymbol{u}) = \Phi_{d,\Sigma}\{\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)\};$$

- Gumbel–Hougaard copulas :

$$C^{GH}_{\theta}(\boldsymbol{u}) = \exp\left(-\left[\sum_{j=1}^{d}\{-\log(u_j)\}^{\theta}\right]^{1/\theta}\right), \quad \theta \geq 1;$$

- Clayton copulas :

$$C^{Cl}_{\theta}(\boldsymbol{u}) = \left(\sum_{j=1}^{d} u_j^{-\theta} - d + 1\right)^{-1/\theta}, \quad \theta > 0.$$

→My Copulas here

The role of copulas to test for breaks detection

$$\mathcal{H}_0 : \exists\, F \text{ such that } \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \text{ have c.d.f. } F.$$

Sklar's theorem allows to rewrite $\mathcal{H}_0$ as $\mathcal{H}_{0,m} \cap \mathcal{H}_{0,c}$ where

$\mathcal{H}_{0,m} \cap \mathcal{H}_{0,c}$:

$$\begin{array}{ll} \mathcal{H}_{0,c} : & \exists C, \text{ such that } \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \text{ have copula } C \\ \mathcal{H}_{0,m} : & \exists F_1, \ldots, F_d \text{ such that } \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \text{ have m.c.d.f. } F_1, \ldots, F_d \,. \end{array}$$

- Construction of a test for $\mathcal{H}_0$ more powerful than its predecessors against alternatives involving a change in the copula, based on the CUSUM approach.
- F, $F_1, \ldots, F_d$ and C are unknown.

Summary

Empirical copula I

- Let $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ be d-dimensional random vectors with continuous m.c.d.f. $F_1, \dots, F_d$ and copula C.
- For $i = 1, \dots, n$, the random vectors $\boldsymbol{U}_i = (F_1(X_{i1}), \dots, F_d(X_{id})) \sim C$.
- When $F_1, \dots, F_d$ are supposed known, a natural estimator of C is given by the empirical c.d.f. of $\boldsymbol{U}_1, \dots, \boldsymbol{U}_n$.
- Here $F_1, \dots, F_d$ are unknown, an estimator of C is given by the empirical c.d.f. of *pseudo-observations* of copula C :

For $j = 1, \dots, d$ let $F_{1:n,j}$ be the empirical c.d.f. of sample $X_{1j}, \dots, X_{nj}$. For $i = 1, \dots, n$, consider the vectors

$$\hat{\boldsymbol{U}}_i^{1:n} = (F_{1:n,1}(X_{i1}), \dots, F_{1:n,d}(X_{id})) = \frac{1}{n}(R_{i1}^{1:n}, \dots, R_{id}^{1:n}),$$

where for $j = 1, \dots, d$, $R_{ij}^{1:n}$ is the maximal rank of X_{ij} among $X_{1j}, \dots, X_{nj}$.

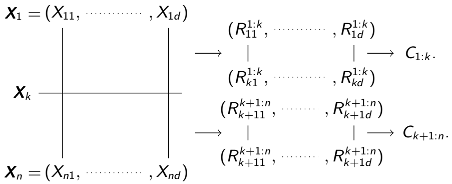
Empirical copula II

[Rüschendorf(1976)], [Deheuvels(1979)]

$$C_{1:n}(\boldsymbol{u}) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(\hat{\boldsymbol{U}}_i^{1:n} \leq \boldsymbol{u}), \qquad \boldsymbol{u} \in [0,1]^d.$$

Let $C_{1:k}$ (resp. $C_{k+1:n}$) the empirical copula evaluated on $\boldsymbol{X}_1, \dots, \boldsymbol{X}_k$ (resp. $\boldsymbol{X}_{k+1}, \dots, \boldsymbol{X}_n$) :

$$C_{1:k}(\boldsymbol{u}) = \frac{1}{k}\sum_{i=1}^{k} \mathbf{1}(\hat{\boldsymbol{U}}_i^{1:k} \leq \boldsymbol{u}), \qquad C_{k+1:n}(\boldsymbol{u}) = \frac{1}{n-k}\sum_{i=k+1}^{n} \mathbf{1}(\hat{\boldsymbol{U}}_i^{k+1:n} \leq \boldsymbol{u}) \quad \boldsymbol{u} \in [0,1]^d.$$



Break detection in copula

We consider the process

$$\mathbb{D}_n(s,\boldsymbol{u}) = \sqrt{n}\frac{\lfloor ns \rfloor}{n}\frac{(n-\lfloor ns \rfloor)}{n}\{C_{1:\lfloor ns \rfloor}(\boldsymbol{u}) - C_{\lfloor ns \rfloor+1:n}(\boldsymbol{u})\}, \qquad (s,\boldsymbol{u}) \in [0,1]^{d+1},$$

A Cramér-von Mises statistic :

$$S_{n,k} = \int_{[0,1]^d} \mathbb{D}_n^2(k/n,\boldsymbol{u})\mathrm{d}C_{1:n}(\boldsymbol{u}) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{D}_n^2(k/n,\hat{\boldsymbol{U}}_i^{1:n}),$$

and

$$S_n = \max_{k\in\{1,\ldots,n-1\}} S_{n,k}.$$

- Under $\mathcal{H}_0$, for all $k \in \{1,\ldots,n-1\}$, $C_{1:k}$ and $C_{k+1:n}$ estimate the same copula C, thus S_n tends to be relatively weak.
- For an abrupt change in copula, the unknown break time can be estimate by the integer k which maximise $S_{n,k}$.

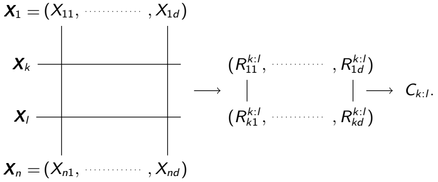
The sequential empirical copula process

Let for $s \le t \in [0,1]$, $\lambda_n(s,t) = (\lfloor nt \rfloor - \lfloor ns \rfloor)/n$.

Sequential empirical copula process

$$\mathbb{C}_n(s,t,\boldsymbol{u}) = \sqrt{n}\lambda_n(s,t)\{C_{\lfloor ns \rfloor+1:\lfloor nt \rfloor}(\boldsymbol{u}) - C(\boldsymbol{u})\}$$
$$= \frac{1}{\sqrt{n}} \sum_{i=\lfloor ns \rfloor+1}^{\lfloor nt \rfloor} \left\{ \mathbf{1}(\hat{\boldsymbol{U}}_i^{\lfloor ns \rfloor+1:\lfloor nt \rfloor} \le \boldsymbol{u}) - C(\boldsymbol{u}) \right\}.$$

$$C_{k:l}(\boldsymbol{u}) = \frac{1}{l-k+1} \sum_{i=k}^{l} \mathbf{1}(\hat{\boldsymbol{U}}_i^{k:l} \le \boldsymbol{u}) \quad \boldsymbol{u} \in [0,1]^d, \quad 1 \le k \le l \le n.$$



Asymptotic behaviour of the sequential empirical copula process

Condition : [Segers(2012)]

For any $j \in \{1, \ldots, d\}$, the partial derivatives $\dot{C}_j = \partial C / \partial u_j$ exist and are continuous on $V_j = \{\boldsymbol{u} \in [0,1]^d, u_j \in (0,1)\}$.

Theorem

Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be drawn from a strictly stationary sequence $(\boldsymbol{X}_i)_{i \in \mathbb{Z}}$ with continuous margins and whose the strong mixing coefficient satisfies $\alpha_r = O(r^{-a})$, $a > 1$. Under the previous condition,

$$\sup_{\boldsymbol{u} \in [0,1]^d} |\mathbb{C}_n(s,t,\boldsymbol{u}) - \widetilde{\mathbb{C}}_n(s,t,\boldsymbol{u})| \xrightarrow{\mathbb{P}} 0,$$

where for $\boldsymbol{u} \in [0,1]^d$, and $\boldsymbol{u}^{\{j\}} = (1, \ldots, 1, u_j, 1, \ldots, 1)$,

$$\widetilde{\mathbb{C}}_n(s,t,\boldsymbol{u}) = \mathbb{B}_n(s,t,\boldsymbol{u}) - \sum_{j=1}^{d} \dot{C}_j(\boldsymbol{u}) \mathbb{B}_n(s,t,\boldsymbol{u}^{\{j\}}).$$

A decomposition for the process $\mathbb{D}_n$

$$\mathbb{B}_n(s,t,\boldsymbol{u}) = \frac{1}{\sqrt{n}} \sum_{i=\lfloor ns \rfloor+1}^{\lfloor nt \rfloor} \{1(\boldsymbol{U}_i \leq \boldsymbol{u}) - C(\boldsymbol{u})\}, \quad (s,t,\boldsymbol{u}) \in [0,1]^{d+2}, s \leq t.$$

$$\rightsquigarrow \mathbb{Z}_C(t,\boldsymbol{u}) - \mathbb{Z}_C(s,\boldsymbol{u})$$

in $\ell^\infty(\{s \leq t \in [0,1]^2\} \times [0,1]^d)$, $\mathbb{Z}_C$ a centered Gaussian process (a C-Kiefer–Müller process).

The process $\mathbb{D}_n$ can be written as function of the sequential empirical copula process $\mathbb{C}_n$:

$$\mathbb{D}_n(s,\boldsymbol{u}) = \left(1 - \frac{\lfloor ns \rfloor}{n}\right) \mathbb{C}_n(0,s,\boldsymbol{u}) - \frac{\lfloor ns \rfloor}{n} \mathbb{C}_n(s,1,\boldsymbol{u}),$$

with $s \in [0,1]$ and $\boldsymbol{u} \in [0,1]^d$.

Asymptotic behaviour of S_n under $\mathcal{H}_0$

Proposition

Under $\mathcal{H}_0$

$$S_n = \max_{k \in \{1,\dots,n-1\}} \frac{1}{n} \sum_{i=1}^{n} \mathbb{D}_n(k/n, \hat{\boldsymbol{U}}_i^{1:n})^2 \rightsquigarrow S_C = \sup_{s \in [0,1]} \int_{[0,1]^d} \{\mathbb{D}_C(s, \boldsymbol{u})\}^2 \mathrm{d}C(\boldsymbol{u}).$$

$$\mathbb{D}_C(s, \boldsymbol{u}) = (1-s)\mathbb{C}_C(0, s, \boldsymbol{u}) - s\mathbb{C}_C(s, 1, \boldsymbol{u}), \quad (s, \boldsymbol{u}) \in [0,1]^{d+1}.$$

and for $s \leq t \in [0,1]^2$ and $\boldsymbol{u} \in [0,1]^d$

$$\mathbb{C}_C(s, t, \boldsymbol{u}) = \{\mathbb{Z}_C(t, \boldsymbol{u}) - \mathbb{Z}_C(s, \boldsymbol{u})\} - \sum_{j=1}^{d} \dot{C}_j(\boldsymbol{u})\{\mathbb{Z}_C(t, \boldsymbol{u}^{\{j\}}) - \mathbb{Z}_C(s, \boldsymbol{u}^{\{j\}})\},$$

with $\mathbb{Z}_C$ the centered Gaussian process and $\boldsymbol{u}^{\{j\}} = (1, \dots, 1, u_j, 1, \dots, 1)$.

Summary

A resampling scheme of the sequential empirical copula process

- From the previous theorem, the process $\mathbb{C}_n$ is asymptotically equivalent to process $\widetilde{\mathbb{C}}_n$

$$\widetilde{\mathbb{C}}_n(s, t, \boldsymbol{u}) = \{\mathbb{Z}_n(t, \boldsymbol{u}) - \mathbb{Z}_n(s, \boldsymbol{u})\} - \sum_{j=1}^{d} \dot{C}_j(\boldsymbol{u})\{\mathbb{Z}_n(t, \boldsymbol{u}^{\{j\}}) - \mathbb{Z}_n(s, \boldsymbol{u}^{\{j\}})\},$$

with

$$\mathbb{Z}_n(s, \boldsymbol{u}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} \{\mathbf{1}(\boldsymbol{U}_i \leq \boldsymbol{u}) - C(\boldsymbol{u})\} \quad (s, \boldsymbol{u}) \in [0, 1]^{d+1}.$$

- $\mathbb{Z}_n \rightsquigarrow \mathbb{Z}_C$ in $\ell^\infty([0, 1]^{d+1})$, $\mathbb{Z}_C$ the C-Kiefer-Müller process
- To resample $\mathbb{C}_n$, we construct a resampling of $\mathbb{Z}_n$ and we estimate the partial derivatives $\dot{C}_j$.

A resampling scheme for $\mathbb{Z}_n$, case of i.i.d. observations

i.i.d. multipliers [van der Vaart and Wellner(2000)]

A sequence of i.i.d. multipliers $(\xi_i)_{i\in\mathbb{Z}}$ satisfies the following conditions :

- For all $i \in \mathbb{Z}$, ξ_i are independent of observations $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$
- $\mathbb{E}(\xi_0) = 0$, $\mathrm{var}(\xi_0) = 1$ and $\int_0^\infty \{\mathrm{P}(|\xi_0| > x)\}^{1/2} \mathrm{d}x < \infty$.

For $m = 1, \ldots, M$ consider the processes

$$\mathbb{Z}_n^{(m)}(s, \boldsymbol{u}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} \xi_i^{(m)} \{\mathbf{1}(\boldsymbol{U}_i \leq \boldsymbol{u}) - C(\boldsymbol{u})\}, \qquad (s, \boldsymbol{u}) \in [0,1]^{d+1}.$$

[Holmes, Kojadinovic et Quessy(2013)]

$$\left(\mathbb{Z}_n, \mathbb{Z}_n^{(1)}, \ldots, \mathbb{Z}_n^{(M)}\right) \rightsquigarrow \left(\mathbb{Z}_C, \mathbb{Z}_C^{(1)}, \ldots, \mathbb{Z}_C^{(M)}\right)$$

in $\{\ell^\infty([0,1]^{d+1})\}^{M+1}$, where $\mathbb{Z}_C$ is the C-Kiefer-Müller process and the processes $\mathbb{Z}_C^{(1)}, \ldots, \mathbb{Z}_C^{(M)}$ are independent copies of $\mathbb{Z}_C$.

A resampling scheme for $\mathbb{C}_n$, case of serially dependent observations I

dependent multipliers [Bühlmann(1993)]

A sequence $(\xi_{i,n})_{i\in\mathbb{Z}}$ of dependent multipliers satisfy the following conditions :

- $(\xi_{i,n})_{i\in\mathbb{Z}}$ is strictly stationary and for all $i \in \mathbb{Z}$, $\xi_{i,n}$ are independent of observations $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$;
- $\mathbb{E}(\xi_{0,n}) = 0$, $\mathrm{var}(\xi_{0,n}) = 1$ and $\sup_{n\geq 1} \mathbb{E}\left(|\xi_{0,n}|^{\nu}\right) < \infty$ for any $\nu \geq 1$;
- There exists a sequence $\ell_n \to \infty$ of strictly positive constants such that $\ell_n = o(n)$, and the sequence $(\xi_{i,n})_{i\in\mathbb{Z}}$ is ℓ_n-dependant ; i.e., such that $\xi_{i,n}$ is independent of $\xi_{i+h,n}$ for all $h > \ell_n$ and all $i \in \mathbb{N}$;
- There exists a function $\varphi : \mathbb{R} \to [0,1]$, symmetric around 0, continuous at 0 satisfying $\varphi(0) = 1$ and $\varphi(x) = 0$ for all $|x| > 1$, such that $\mathbb{E}(\xi_{0,n}\xi_{h,n}) = \varphi(h/\ell_n)$ for all $h \in \mathbb{Z}$.

A resampling scheme for $\mathbb{C}_n$, case of serially dependent observations II

- Using *dependent multipliers* $\xi_{i,n}$ "well-chosen", we can construct a resampling for $\mathbb{Z}_n$ (or $\mathbb{B}_n$) adapted to the case of dependent data.

$$\mathbb{B}_n^{(m)}(s,t,\boldsymbol{u}) = \frac{1}{\sqrt{n}} \sum_{i=\lfloor ns \rfloor+1}^{\lfloor nt \rfloor} \xi_{i,n}^{(m)} \{\mathbf{1}(\boldsymbol{U}_i \leq \boldsymbol{u}) - C(\boldsymbol{u})\}, \quad (s,t,\boldsymbol{u}) \in [0,1]^{d+2}, s \leq t.$$

- For $m = 1, \dots, M$, consider the estimated processes :

$$\hat{\mathbb{B}}_n^{(m)}(s,t,\boldsymbol{u}) = \frac{1}{\sqrt{n}} \sum_{i=\lfloor ns \rfloor+1}^{\lfloor nt \rfloor} \xi_{i,n}^{(m)} \{\mathbf{1}(\hat{\boldsymbol{U}}_i^{1:n} \leq \boldsymbol{u}) - C_{1:n}(\boldsymbol{u})\}, \quad (s,t,\boldsymbol{u}) \in [0,1]^{d+2}.$$

- We consider an estimator $\dot{C}_{j,1:n}$ of $\dot{C}_j$ constructed with finite differencing of the empirical copula at a bandwidth of h_n :

$$\dot{C}_j^{1:n}(\boldsymbol{u}) = \frac{C_{1:n}(\boldsymbol{u} + h_n \boldsymbol{e}_j) - C_{1:n}(\boldsymbol{u} - h_n \boldsymbol{e}_j)}{\min(u_j + h_n, 1) - \max(u_j - h_n, 0)}, \quad \boldsymbol{u} \in [0,1]^d,$$

$\boldsymbol{e}_j = (0, \dots, 0, 1, 0, \dots, 0)$ and $h_n = n^{-1/2}$.

A resampling scheme for $\mathbb{C}_n$, case of serially dependent observations III

For the estimated processes

$$\hat{\mathbb{C}}_n^{(m)}(s,t,\boldsymbol{u}) = \hat{\mathbb{B}}_n^{(m)}(s,t,\boldsymbol{u}) - \sum_{j=1}^{d} \dot{C}_{j,1:n}(\boldsymbol{u})\,\hat{\mathbb{B}}_n^{(m)}(s,t,\boldsymbol{u}^{\{j\}}), \quad (s,t,\boldsymbol{u}) \in [0,1]^{d+2}, s \leq t,$$

we have the following result :

[Bücher, Kojadinovic, Rohmer et Segers(2014)]

Let $\boldsymbol{X}_1,\dots,\boldsymbol{X}_n$ be drawn from a strictly stationary sequence $(\boldsymbol{X}_i)_{i\in\mathbb{Z}}$ whose strong mixing coefficients satisfy $\alpha_r = O(r^{-a})$, $a > 3 + 3d/2$. Consider $l_n = O(n^{1/2-\gamma})$, $\gamma \in (0,1/2)$. Then :

$$\left(\mathbb{C}_n, \hat{\mathbb{C}}_n^{(1)}, \dots, \hat{\mathbb{C}}_n^{(M)}\right) \rightsquigarrow \left(\mathbb{C}_C, \mathbb{C}_C^{(1)}, \dots, \mathbb{C}_C^{(M)}\right),$$

in $\{\ell^\infty(\Delta \times [0,1]^d)\}^{M+1}$, where $\mathbb{C}_C^{(1)}, \dots, \mathbb{C}_C^{(M)}$ are independent copies of $\mathbb{C}_C$.

A resampling scheme for $\mathbb{C}_n$, case of serially dependent observations IV

For $(s,t,\boldsymbol{u}) \in [0,1]^{d+2}$, $s \leq t$, consider the processes

$$\check{\mathbb{B}}_n^{(m)}(s,t,\boldsymbol{u}) = \frac{1}{\sqrt{n}} \sum_{i=\lfloor ns \rfloor+1}^{\lfloor nt \rfloor} \xi_{i,n}^{(m)} \{\mathbf{1}(\hat{\boldsymbol{U}}_i^{\lfloor ns \rfloor+1:\lfloor nt \rfloor} \leq \boldsymbol{u}) - C_{\lfloor ns \rfloor+1:\lfloor nt \rfloor}(\boldsymbol{u})\}.$$

and

$$\check{\mathbb{C}}_n^{(m)}(s,t,\boldsymbol{u}) = \check{\mathbb{B}}_n^{(m)}(s,t,\boldsymbol{u}) - \sum_{j=1}^{d} \dot{C}_{j,\lfloor ns \rfloor+1:\lfloor nt \rfloor}(\boldsymbol{u})\, \check{\mathbb{B}}_n^{(m)}(s,t,\boldsymbol{u}^{\{j\}}).$$

[Bücher, Kojadinovic, Rohmer et Segers(2014)]

Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be drawn from a strictly stationary sequence $(\boldsymbol{X}_i)_{i \in \mathbb{Z}}$. Under the same mixing conditions :

$$\left(\mathbb{C}_n, \check{\mathbb{C}}_n^{(1)}, \ldots, \check{\mathbb{C}}_n^{(M)}\right) \leadsto \left(\mathbb{C}_C, \mathbb{C}_C^{(1)}, \ldots, \mathbb{C}_C^{(M)}\right),$$

in $\{\ell^\infty(\Delta \times [0,1]^d)\}^{M+1}$.

A resampling for S_n

- The process $\mathbb{D}_n$ can be rewritten as :

$$\mathbb{D}_n(s, \boldsymbol{u}) = \lambda_n(s,1)\mathbb{C}_n(0,s,\boldsymbol{u}) - \lambda_n(0,s)\mathbb{C}_n(s,1,\boldsymbol{u}).$$

- Two possibilities to resample $\mathbb{D}_n$:

$$\hat{\mathbb{D}}_n^{(m)}(s, \boldsymbol{u}) = \lambda_n(s,1)\hat{\mathbb{C}}_n^{(m)}(0,s,\boldsymbol{u}) - \lambda_n(0,s)\hat{\mathbb{C}}_n^{(m)}(s,1,\boldsymbol{u}),$$

$$\check{\mathbb{D}}_n^{(m)}(s, \boldsymbol{u}) = \lambda_n(s,1)\check{\mathbb{C}}_n^{(m)}(0,s,\boldsymbol{u}) - \lambda_n(0,s)\check{\mathbb{C}}_n^{(m)}(s,1,\boldsymbol{u}).$$

- ... and for S_n :

$$\hat{S}_n^{(m)} = \max_{k\in\{1,\dots,n-1\}} \sum_{i=1}^{n} \{\hat{\mathbb{D}}_n^{(m)}(k/n, \hat{\boldsymbol{U}}_i^{1:n})\}^2/n,$$

$$\check{S}_n^{(m)} = \max_{k\in\{1,\dots,n-1\}} \sum_{i=1}^{n} \{\check{\mathbb{D}}_n^{(m)}(k/n, \hat{\boldsymbol{U}}_i^{1:n})\}^2/n.$$

Asymptotic validity of the resampling scheme

[Bücher, Kojadinovic, Rohmer et Segers(2014)]

Under $\mathcal{H}_0$ and with the same mixing conditions and $\xi_{i,n}$ "well-chosen",

$$(S_n, \hat{S}_n^{(1)}, \dots, \hat{S}_n^{(M)}) \rightsquigarrow (S_C, S_C^{(1)}, \dots, S_C^{(M)})$$

$$(S_n, \check{S}_n^{(1)}, \dots, \check{S}_n^{(M)}) \rightsquigarrow (S_C, S_C^{(1)}, \dots, S_C^{(M)})$$

in $\mathbb{R}^{M+1}$, $S_C^{(1)}, \dots, S_C^{(M)}$ independent copies of S_C.

- $\hat{S}_n^{(1)}, \dots, \hat{S}_n^{(M)}$ and $\check{S}_n^{(1)}, \dots, \check{S}_n^{(M)}$ can be interpreted as M 'almost' independent copies of S_n.
- We compute two approximate p-values for S_n as

$$\hat{p}_{M,n} = \frac{1}{M} \sum_{m=1}^{M} \mathbf{1}\left(\hat{S}_n^{(m)} \geq S_n\right) \text{ and } \check{p}_{M,n} = \frac{1}{M} \sum_{m=1}^{M} \mathbf{1}\left(\check{S}_n^{(m)} \geq S_n\right)$$

The pros and cons

- + : Most powerful test than predecessors for detect a change in copula
- + : Consistent tests
- + : Adapted in the case of strong mixing observations
- - : Less sensitive to detect a change in marginal distribution
- - : Without the hypothesis than the margins are constant ; we can't conclude in favour of a change in copula.

Summary

Break detection in the copula when there exists a change in marginal distribution at time $m = \lfloor nt \rfloor$, $t \in (0,1)$ known.

Consider the following null hypothesis

$\mathcal{H}_0^m = \mathcal{H}_{1,m} \cap \mathcal{H}_{0,c}$:

$$\begin{aligned} \mathcal{H}_{0,c} :& \exists C, \text{ such that } \boldsymbol{X}_1, \dots, \boldsymbol{X}_n \text{ have copula } C \\ \mathcal{H}_{1,m} :& \exists F_1, \dots F_d \text{ and } F_1', \dots F_d' \text{ such that } \boldsymbol{X}_1, \dots, \boldsymbol{X}_m \\ & \quad \text{have m.c.d.f. } F_1, \dots F_d \text{ and } \boldsymbol{X}_{m+1}, \dots, \boldsymbol{X}_n \text{ have m.c.d.f. } F_1', \dots F_d'. \end{aligned}$$

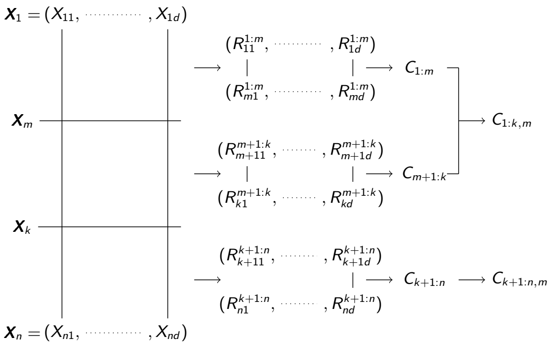
Taking the change at time m, an estimator of copula C can be built with the following pseudo-observations :

$$\hat{\boldsymbol{U}}_{i,m}^{1:n} = \begin{cases} (F_{1:m,1}(X_{i1}), \dots, F_{1:m,d}(X_{id})) & i \in \{1, \dots, m\} \\ (F_{m+1:n,1}(X_{i1}), \dots, F_{m+1:n,d}(X_{id})) & i \in \{m+1, \dots, n\}, \end{cases}$$

where in this case $j = 1, \dots, d, F_{1:m,j}$ (resp. $F_{m+1:n,j}$) is the empirical cumulative distribution function computed with $X_{1j}, \dots, X_{mj}$ (resp. $X_{m+1j}, \dots, X_{nj}$).

In the same way, we construct $C_{1:k,m}$ and $C_{k+1:n,m}$ from sub sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_k$ and $\boldsymbol{X}_{k+1}, \ldots, \boldsymbol{X}_n$ for k in $\{1, \ldots, n-1\}$.

FIGURE: Case of $m \leq k$:



The process

▷ The empirical c.d.f. $C_{1:k,m}$ and $C_{k+1:n,m}$ evaluate from pseudo-observations $\hat{\boldsymbol{U}}^{1:k}_{1,m}, \ldots, \hat{\boldsymbol{U}}^{1:k}_{k,m}$ and $\hat{\boldsymbol{U}}^{k+1:n}_{k+1,m}, \ldots, \hat{\boldsymbol{U}}^{k+1:n}_{n,m}$ are given by

$$C_{1:k,m}(\boldsymbol{u}) = \begin{cases} \frac{m}{k} C_{1:m}(\boldsymbol{u}) + \frac{k-m}{k} C_{m+1:k}(\boldsymbol{u}) & m \in [1, k] \\ C_{1:k}(\boldsymbol{u}) & m \notin [1, k], \end{cases}$$

and

$$C_{k+1:n,m}(\boldsymbol{u}) = \begin{cases} \frac{m-k+1}{n-k+1} C_{k+1:m}(\boldsymbol{u}) + \frac{n-m}{n-k+1} C_{m+1:n}(\boldsymbol{u}) & m \in [k+1, n] \\ C_{k+1:n}(\boldsymbol{u}) & m \notin [k+1, n]. \end{cases}$$

The test statistic is

$$S_{n,m} = \max_{k=1,\ldots,n} \frac{1}{n} \sum_{i=1}^{n} \mathbb{D}^2_{n,m}(k/n, \hat{\boldsymbol{U}}^{1:n}_{i,m})$$

where

$$\mathbb{D}_{n,m}(k/n, \boldsymbol{u}) = \frac{k(n-k)}{n^{3/2}} \left\{ C_{1:k,m}(\boldsymbol{u}) - C_{k+1:n,m}(\boldsymbol{u}) \right\}.$$

Asymptotic behaviour of $S_{n,m}$ under $\mathcal{H}_0$

Theorem [Rohmer(2016)]

Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ n independent (or strong mixing) random vectors, with copula C such that for a fixed integer m, $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m$ have marginal c.d.f. $F_1, \ldots, F_d$ and $X_{m+1}, \ldots, X_n$ have marginal c.d.f. $F'_1, \ldots, F'_d$. Then with similar conditions, we have :

$$S_{n,m} \rightsquigarrow S_C = \sup_{s \in [0,1]} \int_{[0,1]^d} \mathbb{D}_C^2(s, \boldsymbol{u}) \mathrm{d}C(\boldsymbol{u}).$$

- The asymptotic distribution does not depend on m !
- This result can be generalized at the case of multiple changes in marginal distributions.
- R package 'npCopTest'

Summary

Cramér-von Mises statistic, case of serially independent observations

Consider the alternative $\mathcal{H}_{1,c}$:

$\mathcal{H}_{1,c}$

$$H_{1,c} : \exists \text{ distinct } C_1 \text{ and } C_2, \text{ and } t \in (0,1) \text{ such that}$$
$$\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{\lfloor nt \rfloor} \text{ have copula } C_1 \text{ and } \boldsymbol{X}_{\lfloor nt \rfloor+1}, \ldots, \boldsymbol{X}_n \text{ have copula } C_2.$$

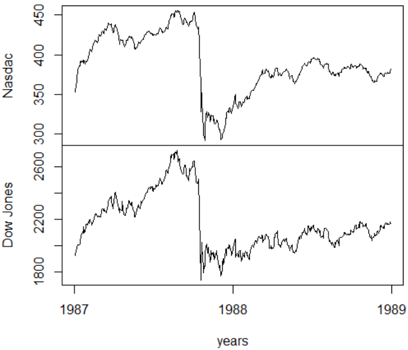
Percentages of rejection of hypothesis $\mathcal{H}_0$ computed using 1000 samples of size $n \in \{50, 100, 200\}$ generated under :

- $\mathcal{H}_0 = \mathcal{H}_{0,c} \cap \mathcal{H}_{0,m}$,
- $\mathcal{H}_{1,c} \cap \mathcal{H}_{0,m}$

With a change in marginal c.d.f. :

- $\mathcal{H}_0^m = \mathcal{H}_{0,c} \cap \mathcal{H}_{1,m}$
- $\mathcal{H}_{1,c} \cap \mathcal{H}_{1,m}$

Nasdaq, Dow Jones and the "black Monday" (1987-10-19)



Suppose there is at more a unique change in m.c.d.f. at time $m = 202$ (1987-10-19)

$\triangleright$ $\hat{p}_{val}^{S_{n,m}} = 0.201$: no evidence against $H_{0,c}$.

Bibliography I

A. Bücher, I. Kojadinovic, T. Rohmer et J. Segers.
Detecting changes in cross-sectional dependence in multivariate time series.
Journal of Multivariate Analysis, 132 : 111 ?128, 2014.

M. Csörgő et L. Horváth.
Limit theorems in change-point analysis.
Wiley Series in Probability and Statistics. John Wiley & Sons, Chichester, UK, 1997.

P. Bühlmann.
The blockwise bootstrap in time series and empirical processes.
PhD thesis, ETH Zürich, 1993.
Diss. ETH No. 10354.

P. Deheuvels.
La fonction de dépendance empirique et ses propriétés : un test non paramétrique d'indépendance.
Acad. Roy. Belg. Bull. Cl. Sci. 5th Ser., 65 :274–292, 1979.

Bibliography II

M. Holmes, I. Kojadinovic et J-F. Quessy.
Nonparametric tests for change-point detection à la Gombay and Horváth.
Journal of Multivariate Analysis, 115 :16–32, 2013.

I. Kojadinovic, J.-F. Quessy et T. Rohmer.
Testing the constancy of Spearman's rho in multivariate time series.
Annals of the Institute of Statistical Mathematics,2015, Pages 1-26.

T. Rohmer.
Some results on change-point detection in cross-sectional dependence of multivariate data with changes in marginal distributions.
Statistic and Probability letters, Volume 119, December 2016, Pages 45-54.

Murray Rosenblatt.
A central limit theorem and a strong mixing condition.
Proceedings of the National Academy of Sciences of the United States of America, 42(1) :43, 1956.

Bibliography III

L. Rüschendorf.
Asymptotic distributions of multivariate rank order statistics.
Annals of Statistics, 4 :912–923, 1976.

J. Segers.
Asymptotics of empirical copula processes under nonrestrictive smoothness assumptions.
Bernoulli, 18 :764–782, 2012.

A. Sklar.
Fonctions de répartition à n dimensions et leurs marges.
Publications de l'Institut de Statistique de l'Université de Paris, 8 : 229–231, 1959.

A.W. van der Vaart and J.A. Wellner.
Weak convergence and empirical processes.
Springer, New York, 2000.
Second edition.

Thank you for your attention !