TEST NON-PARAMÉTRIQUE DE DÉTECTION DE RUPTURE DANS LA DÉPENDANCE D'OBSERVATIONS MULTIVARIÉES

Tom Rohmer [1]

[1] *Université du Maine Département de Mathématiques Avenue Olivier Messiaen 72085 LE MANS Cedex 9*
tom.rohmer@univ-lemans.fr

Résumé. Dans la littérature, on retrouvera un certain nombre de tests non-paramétriques de détection de rupture dans la distribution d'observations multivariées. Cependant, ces tests se révèlent souvent très peu sensibles à un changement dans la dépendance entre les composantes des vecteurs aléatoires (v.a.), lorsque les lois marginales sont inchangées. Si l'on considère un v.a. dont les marges sont continues, le théorème de Sklar affirme qu'il existe une unique fonction appelée copule du v.a. telle que la donnée de la copule et des fonctions de répartitions marginales, caractérise la loi du v.a. De plus, la copule quant-à-elle caractérise la structure de dépendance entre les composantes du v.a. Dans des contributions récentes, des tests basés sur le processus de copule empirique séquentiel, sensibles à un changement dans la copule lorsque les lois marginales sont inchangées ont été mis en place. Ces tests ne permettent pas de conclure en une rupture dans la copule dès lors que les lois marginales ne sont pas constantes. Lorsque des changements dans les lois marginales surviennent, il est néanmoins possible d'adapter les dernières approches mentionnées pour prendre en compte ces potentiels instants de ruptures.

Mots-clés. Tests non-paramétriques, processus empiriques séquentiels, copule, détection de rupture.

Abstract. Many non-parametric tests for detecting distributional changes in multivariate observations based on empirical processes are available. These tests are often not very sensitive to detecting a change in the dependence between the components of multivariate data that leaves the marginal cumulative distribution functions (m.c.d.f.s) unchanged. For a random vector with continuous marginal cumulative distributions, Sklar's Theorem stipulates that there is a unique function called copula, characterizing the dependence between the components of the random vector. The cumulative distribution function of the random vector can then be rewritten using only the copula and the m.c.d.f.s. In recent publications, non-parametric tests for break detection that are sensitive to changes in the copula of observations and based on the two-sided sequential empirical copula are considered. These tests do not make it possible to conclude in favour of change in the copula if the m.c.d.f.s are not constant. An alternative test for detecting changes in the dependence structure of random vectors that is sensitive to changes in copula of observations and adapted in the case of alternative hypotheses involving abrupt changes in the m.c.d.f.s is considered.

Keywords. Non parametric tests, sequential empirical processes, copulas, break-detection.

1 Introduction

Considérons $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$, n vecteurs aléatoires (v.a.) de dimension $d \geq 2$ quelconque, dont les fonctions de répartitions marginales (f.d.r.m.) sont continues. Intéressons nous à l'hypothèse $\mathcal{H}_0$ suivante :

$$\mathcal{H}_0 : \exists F \text{ telle que } F \text{ est la f.d.r. des vecteurs } \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n.$$

De nombreux auteurs se sont intéressés à la mise en place de tests non-paramétriques pour $\neg\mathcal{H}_0$, basés sur des processus empiriques séquentiels, (à titre d'exemple on pourra voir les articles de Bai, 1994; Csörgő and Horváth, 1997; Inoue, 2001; Bücher et al., 2014) et à l'estimation du ou des instants de ruptures. On parlera couramment de 'change-point detection'.

Cependant ces tests mis en place pourront trouver des faiblesses en terme de puissance face à certaines alternatives considérées. Si l'on se place sous l'hypothèse $\mathcal{H}_0$, le théorème de Sklar (1959) stipule qu'il existe une unique fonction $C : [0,1]^d \mapsto \mathbb{R}$ appelée copule du vecteur aléatoire, telle que

$$F(\boldsymbol{x}) = C\{F_1(x_1), \ldots, F_d(x_d)\}, \qquad \boldsymbol{x} \in \mathbb{R}^d. \tag{1}$$

Il est particulièrement intéressant de noter que la copule C caractérise la structure de dépendance entre les composantes des vecteurs aléatoires $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$. La décomposition de Sklar va notamment permettre de réécrire $\mathcal{H}_0$ de la façon suivante : $\mathcal{H}_0 = \mathcal{H}_{0,m} \cap \mathcal{H}_{0,c}$ avec

$$\mathcal{H}_{0,m} : \exists F_1, \ldots F_d \text{ telles que } F_1, \ldots F_d \text{ sont les f.d.r.m. des v.a. } \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \tag{2}$$

$$\mathcal{H}_{0,c} : \exists C \text{ telle que } C \text{ est la copule des v.a. } \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n. \tag{3}$$

Hormis l'article de Bücher et al. (2014), les tests des articles précédemment cités se révèlent peu puissants pour des échantillons de tailles modérées face à des alternatives de type $\mathcal{H}_{0,m} \cap (\neg\mathcal{H}_{0,c})$ (voir p.ex. les simulations de la Section 4 dans Holmes et al., 2013). Concernant les tests de Bücher et al. (2014), malgré ses meilleurs résultats de puissances à tailles d'échantillons modérés sous $\mathcal{H}_{0,m} \cap (\neg\mathcal{H}_{0,c})$, les tests présentés ont le désavantage de ne pas permettre de conclure en un rejet de $\mathcal{H}_{0,c}$ en cas de rejet de $\mathcal{H}_0$: la rupture peut avoir lieu dans une des lois marginales des vecteurs aléatoires. Une adaptation du test de Bücher et al. (2014) que l'on retrouvera dans Rohmer (2016) permet de tester $\neg\mathcal{H}_{0,c}$ en présence d'au plus $p \geq 1$ ruptures dans une (ou plusieurs) des lois marginales, les p instants de ruptures étant supposés connus.

2 Statistique de test

Plus précisément, considérons pour $j = 1, \ldots, p$, le vecteur $\boldsymbol{m} = (m_j)_{j=1,\ldots,p}$ composé des p instants potentiels de ruptures : $m_j = \lfloor nb_j \rfloor$, $0 \leq b_1 < \ldots < b_p \leq 1$ supposés connus. Il est proposé ici un test pour $\neg\mathcal{H}_0^{\boldsymbol{m}}$ avec $\mathcal{H}_0^{\boldsymbol{m}} = \mathcal{H}_{0,c} \cap \mathcal{H}_{1,\boldsymbol{m}}$ et

$$\begin{aligned}\mathcal{H}_{1,\boldsymbol{m}} : \text{pour } j = 1, \ldots, p+1 \ \exists F_{1j}, \ldots, F_{dj} \\ \text{telles que } F_{1j}, \ldots, F_{dj} \text{ sont les f.d.r.m. des v.a. } \boldsymbol{X}_{m_{j-1}+1}, \ldots, \boldsymbol{X}_{m_j},\end{aligned} \quad (4)$$

où $m_0 = 1$ et $m_{p+1} = n$.

Intéressons nous pour simplifier le problème, au cas d'au plus une rupture dans les lois marginales, soit $p = 1$ et $m = \lfloor nb \rfloor$, $b \in [0, 1]$ connu. Pour un sous-échantillon $\boldsymbol{X}_k, \ldots, \boldsymbol{X}_l$, $1 \leq k \leq l \leq n$, nous utilisons alors les pseudo-observations de la copule C suivantes :

$$\hat{\boldsymbol{U}}_{i,m}^{k:l} = \begin{cases} \begin{cases} (F_{k:m,1}(X_{i1}), \ldots, F_{k:m,d}(X_{i1})) & i \in \{k, \ldots, m\} \\ (F_{m+1:l,1}(X_{i1}), \ldots, F_{m+1:l,d}(X_{i1})) & i \in \{m+1, \ldots, l\} \end{cases} & m \in [k, l], \\ (F_{k:l,1}(X_{i1}), \ldots, F_{k:l,d}(X_{i1})) & m \notin [k, l], \end{cases}$$

où pour $j = 1, \ldots, d$, $F_{k:l,j}$ est la fonction de répartition empirique calculée à partir de l'échantillon $X_{kj}, \ldots, X_{lj}$. Nous prendrons comme estimation de la copule sur l'échantillon $\boldsymbol{X}_k, \ldots, \boldsymbol{X}_l$ la fonction de répartition empirique dénotée $C_{k:l,m}$ calculée à partir des vecteurs aléatoires $\hat{\boldsymbol{U}}_{k,m}^{k:l}, \ldots, \hat{\boldsymbol{U}}_{l,m}^{k:l}$.

La statistique utilisée est une statistique de type Cramér–von Mises construite à partir du processus

$$\mathbb{D}_{n,m}(s, \boldsymbol{u}) = \sqrt{n} \frac{\lfloor ns \rfloor}{n} \frac{n - \lfloor ns \rfloor}{n} \{C_{1:\lfloor ns \rfloor,m}(\boldsymbol{u}) - C_{\lfloor ns \rfloor+1:n,m}(\boldsymbol{u})\},$$

$(s, \boldsymbol{u}) \in [0, 1]^{d+1}$. Une technique de rééchantillonnage de la statistique à base de multiplicateurs (voir p. ex. van der Vaart and Wellner, 2000; Bücher and Kojadinovic, 2014) est mise en place pour l'estimation de la p-value du test. Des résultats de convergence sous l'hypothèse nulle considérée sont établis et des simulations de Monte Carlo à tailles d'échantillons modérées ont été réalisées.

La mise en place du test et l'estimation du point de rupture dans la copule (en cas de rupture abrupte) peut être effectuée à l'aide du package R *npCopTest*, disponible sur le CRAN.

Références

J. Bai. Weak convergence of the sequential empirical processes of residuals in ARMA models. *The Annals of Statistics*, 22(4) :2051–2061, 1994.

A. Bücher and I. Kojadinovic, A dependent multiplier bootstrap for the sequential empirical copula process under strong mixing. *Bernoulli*,2015

A. Bücher, I. Kojadinovic, T. Rohmer & J. Segers. Detecting changes in cross-sectional dependence in multivariate time series *Journal of Multivariate Analysis*, Volume 132, November 2014, Pages 111-128

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

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

A. Inoue. Testing for distributional change in time series. *Econometric Theory*, volume 17, number 1, pages 156–187, 2001

T. Rohmer. Some results on change-point detection in cross-sectional dependance of multivariate data with changes in marginal distributions *Statistics & Probability Letters*, Volume 119, December 2016, Pages 45-54, ISSN 0167-7152.

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*, Second edition, 2000