Statistics and Probability Letters 119 (2016) 45–54

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Some results on change-point detection in cross-sectional dependence of multivariate data with changes in marginal distributions

Tom Rohmer *

Université de Bordeaux, ISPED, 146 rue Léo Saignat, 33000 Bordeaux, France
Inserm, Centre Inserm U-1219, 146 rue Léo Saignat, 33000 Bordeaux, France

ARTICLE INFO

Article history:
Received 22 January 2016
Received in revised form 26 May 2016
Accepted 29 June 2016
Available online 14 July 2016

Keywords:
Non-parametric tests
Sequential empirical copula process
Monte Carlo experiments

ABSTRACT

A non-parametric test is proposed for detecting changes in the dependence between the components of multivariate data, when changes in marginal distributions occur at known instants. Monte Carlo simulations have been carried out to illustrate the performance of the procedure.

1. Introduction

Let $\boldsymbol{X}$ be a d-dimensional random vector ($d \geq 2$), with cumulative distribution function (c.d.f.) F and marginal cumulative distribution functions (m.c.d.f.s) $F_1, \dots, F_d$. When the m.c.d.f.s $F_1, \dots, F_d$ are continuous, Sklar's Theorem (see Sklar, 1959) stipulates that there is a unique function C called copula, characterizing the dependence of the random vector $\boldsymbol{X}$, such that F can be written as:

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

Let $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ be d-dimensional observations. The purpose of change-points detection is to test the hypothesis

$$H_0 : \exists F \text{ such as } \boldsymbol{X}_1, \dots, \boldsymbol{X}_n \text{ have c.d.f. } F,$$

against $\neg H_0$. In agreement with Sklar's Theorem, H_0 can be rewritten as $H_0 = H_{0,m} \cap H_{0,c}$, with

$$H_{0,m} : \exists F_1, \dots, F_d \text{ such as } \boldsymbol{X}_1, \dots, \boldsymbol{X}_n \text{ have m.c.d.f.s } F_1, \dots, F_d, \quad (1)$$

$$H_{0,c} : \exists C \text{ such as } \boldsymbol{X}_1, \dots, \boldsymbol{X}_n \text{ have copula } C. \quad (2)$$

A change either in the copula of random vectors or in one of the m.c.d.f.s implies the rejection of the null hypothesis H_0. Many non-parametric tests for H_0 based on empirical processes are available; see for example Bai (1994), Csörgő and Horváth (1997) and Inoue (2001). These tests are not very sensitive to detecting a change in the copula that leaves the m.c.d.f.s unchanged. This conclusion is highlighted in Holmes et al. (2013, Section 4) through Monte Carlo simulations.

* Correspondence to: Université de Bordeaux, ISPED, 146 rue Léo Saignat, 33000 Bordeaux, France.
E-mail address: tom.rohmer@isped.u-bordeaux2.fr.

http://dx.doi.org/10.1016/j.spl.2016.06.026
0167-7152/

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 process are considered in Bücher et al. (2014). These tests do not make it possible to conclude in favour of $\neg H_{0,c}$ if the m.c.d.f.s are not constant. In many situations, see for example Section 2 of the supplementary material (see Appendix B), a specific event can lead to changes in the marginal cumulative distributions. The question then becomes whether the specific event changes the copula or not. The aim of this paper is to propose a test for detecting a change 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.

The paper is organized as follows. The procedure to test the null hypothesis of a break in c.d.f. when a change in the m.c.d.f.s occurs is presented in Section 2. An adaptation of results of Section 2 when multiple changes in m.c.d.f.s occur is described in Section 3. Finally, Section 4 contains the results of Monte Carlo simulations. The supplementary materiel (see Appendix B) discusses the case of α-mixing observations and presents an illustration with a specific situation.

2. Break detection in the copula when a break time in the m.c.d.f.s is known

In the sequel, the weak convergence, denoted by $\rightsquigarrow$, must be understood as being the weak convergence in the sense of Definition 1.3.3 in van der Vaart and Wellner (2000). For a set T, $\ell^\infty(T)$ denotes the space of bounded real-valued functions on T equipped with the uniform metric.

Let $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ be d-dimensional random vectors ($d \geq 2$) and consider for $1 \leq k \leq l \leq n$ the empirical copula $C_{k:l}$ of the subsample $\boldsymbol{X}_k, \dots, \boldsymbol{X}_l$ as suggested in Deheuvels (1979):

$$C_{k:l}(\boldsymbol{u}) = \frac{1}{l-k+1} \sum_{i=k}^{l} \prod_{j=1}^{d} \mathbf{1}(F_{k:l,j}(X_{ij}) \leq u_j), \quad (u_1, \dots, u_d) \in [0,1]^d, \tag{3}$$

where for $j = 1, \dots, d$, $F_{k:l,j}$ is the empirical cumulative distribution function (e.c.d.f.) of sample $X_{kj}, \dots, X_{lj}$:

$$F_{k:l,j}(x) = \frac{1}{l-k+1} \sum_{i=k}^{l} \mathbf{1}(X_{ij} \leq x), \quad x \in \mathbb{R}. \tag{4}$$

In Bücher et al. (2014), the following Cramér–von Mises type statistic to test H_0 is suggested:

$$S_n = \sup_{s \in [0,1]} \sqrt{n} \lambda_n(s,1) \lambda_n(0,s) \int_{[0,1]^d} \{C_{1:\lfloor ns \rfloor}(\boldsymbol{u}) - C_{\lfloor ns \rfloor + 1:n}(\boldsymbol{u})\}^2 \mathrm{d}C_{1:n}(\boldsymbol{u}), \tag{5}$$

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

Monte Carlo simulations (see section 5 of Bücher et al., 2014) show that a strategy of bootstrapping with independent or dependent multipliers according to the observations (see Bücher and Kojadinovic, 2015; Bücher et al., 2014) of the statistic S_n leads to very good performances in term of powers for alternatives hypotheses that involve a change in copula that leave the m.c.d.f.s unchanged.

Let us suppose that there is a break time $m = \lfloor nb \rfloor$ in the continuous m.c.d.f.s, $b \in (0,1)$ known. We propose a test for $H_0^m = H_{0,c} \cap H_{1,m}$, where $H_{0,c}$ is defined in (2) and $H_{1,m}$ is defined by:

$$H_{1,m} : \exists F_1, \dots, F_d \text{ and } F'_1, \dots, F'_d \text{ such that } \begin{matrix} \boldsymbol{X}_1, \dots, \boldsymbol{X}_m \text{ have m.c.d.f. } F_1, \dots, F_d, \\ \boldsymbol{X}_{m+1}, \dots, \boldsymbol{X}_n \text{ have m.c.d.f. } F'_1, \dots, F'_d. \end{matrix} \tag{6}$$

We do not suppose that $F'_1, \dots, F'_d$ are necessarily different from $F_1, \dots, F_d$. In other words, we do not assume a change in the m.c.d.f.s. However we suppose that if there is a change in the m.c.d.f.s, it is a unique and abrupt change at time m.

Let us consider that the random sample $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ satisfies the hypothesis $H_0^m = H_{0,c} \cap H_{1,m}$ and that $C, F_1, \dots, F_d$, $F'_1, \dots, F'_d$ are unknown.

Note that the unobservable random vectors $\boldsymbol{U}_{i,m}$, $i \in \{1, \dots, n\}$, defined by

$$\boldsymbol{U}_{i,m} = \begin{cases} (F_1(X_{i1}), \dots, F_d(X_{id})) & i \in \{1, \dots, m\} \\ (F'_1(X_{i1}), \dots, F'_d(X_{id})) & i \in \{m+1, \dots, n\}, \end{cases} \tag{7}$$

have C for c.d.f. For $i \in \{1, \dots, n\}$, let $\hat{\boldsymbol{U}}_{i,m}^{1:n}$ defined by:

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

where for all $1 \leq k \leq l \leq n$ and $j = 1, \dots, d$ $F_{k:l,j}$ is the e.c.d.f. of $X_{kj}, \dots, X_{lj}$ as defined in (4). The vectors $\hat{\boldsymbol{U}}_{i,m}^{1:n}$, $i = 1, \dots, n$ can be seen to be pseudo-observations of the copula C. An estimator of C is given by the empirical distribution of $\hat{\boldsymbol{U}}_{1,m}^{1:n}, \dots, \hat{\boldsymbol{U}}_{n,m}^{1:n}$:

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