Statistical Methodology 21 (2014) 59–87

Contents lists available at ScienceDirect

Statistical Methodology

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

General tests of independence based on empirical processes indexed by functions

Salim Bouzebda *

Laboratoire de Mathématiques Appliquées de Compiègne, Université de Technologie de Compiégne, France

A R T I C L E   I N F O

Article history:
Received 29 July 2013
Received in revised form
4 March 2014
Accepted 15 March 2014

Keywords:
Empirical process
Exchangeability
Multivariate empirical copula processes
Tests of independence
Gaussian approximation
Contiguous alternatives
Möbius decomposition
Half-spaces
Cramér–von Mises statistic

A B S T R A C T

The present paper is mainly concerned with the statistical tests of the independence problem between random vectors. We develop an approach based on general empirical processes indexed by a particular class of functions. We prove two abstract approximation theorems that include some existing results as particular cases. Finally, we characterize the limiting behavior of the Möbius transformation of empirical processes indexed by functions under contiguous sequences of alternatives.

1. Introduction and main results

One of the classical and important problems in statistics is testing independence between two or more components of a random vector. The traditional approach is based on Pearson's correlation coefficient, but its lack of robustness to outliers and departures from normality eventually led researchers to consider alternative nonparametric procedures. To overcome such a problem, some rank tests of independence – those of Savage, Spearman and van der Waerden in particular – rely on linear rank statistics are proposed. Another way to test the independence is the use of functionals of empirical processes, which is the approach that we will develop in the present work.

We first set out some notation and the basic definitions which will be used throughout the paper. Consider a random sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$ of independent draws from a probability measure $\mathbb{P}$ on an

* Correspondence to: UTC Rue Roger Couttolenc CS 60319, 60203 Compiégne Cedex, France. Tel.: +33 0611224000.
E-mail addresses: bouzebda@gmail.com, salim.bouzebda@utc.fr.

http://dx.doi.org/10.1016/j.stamet.2014.03.001
1572-3127/

arbitrary sample space $\mathcal{X}$. We define the empirical measure to be

$$\mathbb{P}_n = n^{-1} \sum_{i=1}^{n} \delta_{\mathbf{X}_i},$$

where $\delta_{\mathbf{x}}$ is the measure that assigns mass 1 at $\mathbf{x}$ and zero elsewhere. Let $f : \mathcal{X} \mapsto \mathbb{R}$ be a measurable function. In the modern theory of the empirical processes it is customary to identify $\mathbb{P}$ and $\mathbb{P}_n$ with the mappings given by

$$f \to \mathbb{P}f = \int_{\mathcal{X}} f d\mathbb{P}, \quad \text{and} \quad f \to \mathbb{P}_n f = \int_{\mathcal{X}} f d\mathbb{P}_n = \frac{1}{n} \sum_{k=1}^{n} f(\mathbf{X}_i).$$

For any class $\mathcal{F}$ of measurable functions $f : \mathcal{X} \mapsto \mathbb{R}$, an empirical process

$$\{\mathbb{G}_n f = \sqrt{n}\,(\mathbb{P}_n f - \mathbb{P}f) : f \in \mathcal{F}\}$$

can be defined. Throughout this paper, it will be assumed that $\mathcal{F} \subset L_2(\mathcal{X}, d\mathbb{P})$, which in turn implies that the finite-dimensional distributions of the sequence of random functions $\{\mathbb{G}_n(f) : f \in \mathcal{F}\}$ converge weakly, as $n \to \infty$, to the finite-dimensional distributions of a mean zero Gaussian random function $\{\mathbb{B}(f) : f \in \mathcal{F}\}$ with the same matrix of covariance as $\{\mathbb{G}_n(f) : f \in \mathcal{F}\}$, that is,

$$\langle f, g\rangle = \mathrm{cov}(\mathbb{B}(f), \mathbb{B}(g)) = \mathbb{E}(f(\mathbf{X})g(\mathbf{X})) - \mathbb{E}(f(\mathbf{X}))\mathbb{E}(g(\mathbf{X})), \quad \text{for } f, g \in \mathcal{F}.$$

In the context of the present article the process $\{\mathbb{B}(f) : f \in \mathcal{F}\}$ will always admit a version which is almost surely bounded and continuous with respect to the intrinsic semi-metric

$$d_{\mathbb{P}}(f, g) = \sqrt{\mathbb{E}(f(\mathbf{X}) - g(\mathbf{X}))^2}, \quad \text{for } f, g \in \mathcal{F}.$$

We call the process $\{\mathbb{B}(f) : f \in \mathcal{F}\}$ a $\mathbb{P}$-Brownian bridge indexed by $\mathcal{F}$. We can assume without loss of generality that $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_p$. We will consider the following particular class of functions

$$\mathscr{F} = \left\{ f \in \mathcal{F} : f = \prod_{j=1}^{p} f_j \text{ such that } f_j : \mathcal{X}_j \mapsto \mathbb{R} \right\}. \tag{1.1}$$

Let us introduce, for $j = 1, \ldots, p$,

$$\mathscr{F}_j = \left\{ g : \mathcal{X}_j \mapsto \mathbb{R} : g \in L_2(\mathcal{X}_j, d\mathbb{P}_j) \right\}. \tag{1.2}$$

We will use the following notation

$$\| \cdot \|_{\mathscr{F}} = \sup_{f \in \mathscr{F}} | \cdot |.$$

Let $\mathbb{P}$ be the joint probability for $\mathbf{X} = (\mathbb{X}_1, \ldots, \mathbb{X}_p)$ and $\mathbb{P}^{(j)}$ be the marginal probability for $\mathbb{X}_j$. Throughout this paper, it will be assumed tacitly that $\mathscr{F}$ and $\mathscr{F}_j$, defined in (1.1) and (1.2), for $j = 1, \ldots, p$, respectively, are $\mathbb{P}$-Donsker classes of functions and

$$\|\mathbb{P}^{(j)}\|_{\mathscr{F}_j} < \infty, \quad \text{for } j = 1, \ldots, p,$$

see Remark 2.4 below for more details. Indeed, these assumptions allow us to apply Theorem 3.8.1 of van der Vaart and Wellner [68] to the process $\{\mathbb{A}_n(f) : f \in \mathscr{F}\}$ defined in (1.3) below. Notice that $\mathbb{X}_1, \ldots, \mathbb{X}_p$ are independent, if and only if,

$$\mathcal{H}_0 : \mathbb{P} = \prod_{j=1}^{p} \mathbb{P}^{(j)},$$

which implies that

$$\mathcal{H}_0 : \mathbb{P}f = \prod_{j=1}^{p} \mathbb{P}^{(j)} f_j, \quad \text{for all } f \in \mathscr{F}.$$