Journal of Statistical Planning and Inference 140 (2010) 3058–3075

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

Applications and asymptotic power of marginal-free tests of stochastic vectorial independence

Jean-François Quessy

Département de mathématiques et d'informatique, Université du Québec à Trois-Rivières, C. P. 500, Trois-Rivières (Québec), Canada G9A 5H7

ARTICLE INFO

Article history:
Received 26 November 2008
Received in revised form
31 March 2010
Accepted 9 April 2010
Available online 14 April 2010

Keywords:
Contiguous alternatives
Copula
Empirical copula process
Ranks
Weighted bootstrap

ABSTRACT

Fully nonparametric tests for the independence between random vectors are studied in this paper. The test statistics are functionals of an empirical process defined as the difference between the joint empirical copula and the product of the empirical copulas associated to the vectors that are suspected to be independent. The validity of a weighted bootstrap procedure is established, which allows for a quick computation of p-values. A special attention is given to the asymptotic behavior of the tests under contiguous sequences of distributions. Finally, a characteristic of the copulas in the Archimedean class in terms of independence of vectors is exploited in order to propose a new goodness-of-fit procedure.

1. Introduction

Testing the independence between the univariate components of a random vector is an old and well-documented issue. In a bivariate normal setting, independence is characterized by a null correlation coefficient so that the use of tests based on the sample correlation coefficient is justified. This approach has a straightforward adaptation to the multivariate case, where likelihood ratio tests have been developed.

In many practical situations, however, the multivariate normality hypothesis is violated. Thus the procedures developed under this assumption are not recommended. This problem led many researchers to propose distribution-free dependence measures like those in the family of linear rank statistics. See Hájek et al. (1999) for details. An extension to the multivariate case has been proposed by Puri and Sen (1971), where an empirical association matrix whose entries are pairwise linear rank statistics is defined. This idea leads to nonparametric, rank-based versions of the test statistics developed under the Gaussian paradigm. Although the latter are robust against outliers and free of the margins, they are not consistent in general. This last problem is avoided by considering the general characterization of multivariate independence as the product of the marginal distributions, as was originally considered by Hoeffding (1948) and Blum et al. (1961). Although these procedures are omnibus and powerful under many dependence scenarios, their distributions under alternative hypotheses depend on the behavior of the univariate margins.

A fruitful approach of the dependence between random variables is that using the modern theory of copulas. Briefly stated, a copula is a function associated to a multivariate distribution that contains all the information about the dependence structure. In the context of testing multivariate independence, this idea was first exploited by Deheuvels (1979, 1980, 1981), where Cramér–von Mises statistics computed from an empirical copula were proposed. This method is

E-mail address: jean-francois.quessy@uqtr.ca

0378-3758/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2010.04.004

shown by Genest and Rémillard (2004) to yield powerful tests under many fixed alternatives and by Genest et al. (2006, 2007) to have good power under contiguous sequences of distributions.

For the more general problem of testing the independence between many random vectors, only few works considered an empirical process point of view, e.g. Bilodeau and Lafaye de Micheaux (2005) and Beran et al. (2007). Indeed, many of the approaches privileged so far are based on parametric or nonparametric estimations of a correlation matrix, as in Cléroux et al. (1995), Coxhead (1974), Cramer and Nicewander (1979), Stephens (1979), Dauxois and Nkiet (1998) and El Maâche and Lepage (1998). Unfortunately, these tests are not consistent under general alternatives. Other possibilities consist in using some extensions of the sign or quadrant tests, as in Gieser and Randles (1997), Um and Randles (2001) and Taskinen et al. (2003, 2005). Although these tests are affine invariant, their limiting distributions are derived under the assumption that the multivariate margins are elliptically symmetric.

In this paper, an empirical process based on the characterization of stochastic vectorial independence in terms of copulas is defined as a base to build test statistics. The main characteristic of the proposed methodologies is that they are marginal-free in two ways: (i) the test statistics are defined in terms of the ranks of the observations, hence no assumption is required on the univariate margins and (ii) the multivariate margins associated to the individual vectors can be kept unknown, thanks to a weighted bootstrap procedure suitably adapted to the current situation. In the same context as in the present paper, Kojadinovic and Holmes (2009) recently considered test statistics based on the aforementioned empirical process. They obtained the asymptotic representation of the latter and studied the finite-sample behavior of test statistics rising from its so-called Möbius decomposition. It is thus important to clarify the difference between their work and this article.

First of all, the limiting representation of the empirical process already discovered by Kojadinovic and Holmes (2009) is obtained here using an alternate proof where less assumptions on the marginal distributions are necessary. Also, explicit expressions for new moment-based linear rank statistics are provided. Moreover, a quick re-sampling method based on the multiplier central limit theorem is introduced, its asymptotic validity is established and explicit formulas are provided for the test statistics under study. The most noteworthy contribution of the article is nevertheless the following. First, the asymptotic behavior of five test statistics under three types of local alternatives is obtained; the results are generalizations, in the vectorial context, of formulas obtained by Genest et al. (2007). Secondly, a new goodness-of-fit test statistic for copulas is introduced and its weak convergence is established as well as the validity of the re-sampling method that is employed to compute p-values.

The paper is organized as follows. In Section 2, the result of Kojadinovic and Holmes (2009) about the weak convergence of an empirical process underlying all test statistics in this work is recalled and proved in an alternate way enabling less assumptions on the multivariate margins. In Section 3, a weighted bootstrap procedure that aims at computing p-values is defined and its asymptotic validity is established. Sections 4 and 5 study the power of the tests under fixed and local alternatives, respectively. Section 6 is devoted to a new goodness-of-fit test for families of Archimedean copulas.

2. Properties of an empirical copula process

2.1. Description of the null and alternative hypotheses

For the problem of testing the stochastic independence between vectors, the approach based on copulas will prove very useful in order to build fully nonparametric procedures. To be specific, let $\mathbf{X} = (X_1, \ldots, X_d)$ be a random vector with joint distribution H and continuous margins $F_1, \ldots, F_d$. In that context, Sklar's Theorem ensures that there exists a unique copula $C : [0,1]^d \to [0,1]$ such that

$$H(x_1,\ldots,x_d) = C\{F_1(x_1),\ldots,F_d(x_d)\}. \tag{1}$$

This representation is no longer unique if one or many of the margins have discontinuities. Formula (1) emphasizes on the fact that a copula contains all the information about the dependence structure among the components of $\mathbf{X}$.

In order to develop a test for the null hypothesis of vectorial independence, let $\mathbf{X} = (\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(K)})$, where $\mathbf{X}^{(k)}$ has dimension $d_k \geq 1$ and $d_1 + \cdots + d_K = d$. Furthermore, define $S_k = \{i_{k,1},\ldots,i_{k,d_k}\}$ as the set of indices such that $\mathbf{X}^{(k)} = (X_{i_{k,1}},\ldots,X_{i_{k,d_k}})$. When $\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(K)}$ are independent,

$$H(\mathbf{x}_1,\ldots,\mathbf{x}_K) = \mathrm{P}(\mathbf{X}^{(1)} \leq \mathbf{x}_1,\ldots,\mathbf{X}^{(K)} \leq \mathbf{x}_K) = \prod_{k=1}^{K} H_k(\mathbf{x}_k),$$

where for $k \in \{1,\ldots,K\}$, the subvector $\mathbf{X}^{(k)}$ has joint distribution H_k and $\mathbf{x}_k = (x_{i_{k,1}},\ldots,x_{i_{k,d_k}})$. Hence, representation (1) allows to conclude that there exist copulas $C_1,\ldots,C_K$ and C such that

$$C\{F_1(x_1),\ldots,F_d(x_d)\} = \prod_{k=1}^{K} C_k\{F_{i_{k,1}}(x_{i_{k,1}}),\ldots,F_{i_{k,d_k}}(x_{i_{k,d_k}})\},$$