Tests of independence among continuous random vectors based on Cramér-von Mises functionals of the empirical copula process

Ivan Kojadinovic, Mark Holmes

Department of Statistics, The University of Auckland, Private Bag 92019, Auckland 1142, New Zealand.

Abstract

A decomposition of the independence empirical copula process into a finite number of asymptotically independent sub-processes was studied by Deheuvels. Starting from this decomposition, Genest and Rémillard recently investigated tests of independence among random variables based on Cramér-von Mises statistics derived from the sub-processes. A generalization of Deheuvels' decomposition to the case where independence is to be tested among *continuous random vectors* is presented. The asymptotic behavior of the resulting collection of Cramér-von Mises statistics is derived. It is shown that they are not distribution-free. One way of carrying out the resulting tests of independence then involves using the bootstrap or the permutation methodology. The former is shown to behave consistently, while the latter is employed in practice. Finally, simulations are used to study the finite-sample behavior of the tests.

Key words: Empirical process; Möbius decomposition; Cramér-von Mises statistic; Bootstrap; Permutation.

1 Introduction

Inspired by the work, among others, of Blum, Kiefer and Rosenblatt [1], Dugué [2] and Deheuvels [3], Genest and Rémillard [4] recently studied a test of multivariate independence based on a *Möbius decomposition* of the *empirical copula process*. Given $d \geq 2$ continuous random variables $X_1, \ldots, X_d$ with marginal cumulative distribution functions (c.d.f.s) $F_1, \ldots, F_d$ respectively, it

Email addresses: ivan@stat.auckland.ac.nz (Ivan Kojadinovic), mholmes@stat.auckland.ac.nz (Mark Holmes).

is well-known that their joint c.d.f. F can be uniquely represented by means of a function $C : [0,1]^d \to [0,1]$, called a *copula*, such that

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

This representation, due to Sklar [5], has become of central importance for the study of the notion of dependence among variables. Indeed, essentially all nonparametric measures of dependence can be expressed in terms of the function C; see e.g. [6,7,8] for a comprehensive introduction to copulas. Noticing that independence occurs when $C(u_1,\ldots,u_d) = \prod_{k=1}^d u_k$, $u \in [0,1]^d$, it appears natural to consider, as statistics for testing the mutual independence of $X_1,\ldots,X_d$, Kolmogorov-Smirnov or Cramér-von Mises functionals derived from the process

$$\sqrt{n}\left[C_n(u) - \prod_{k=1}^d u_k\right], \qquad u \in [0,1]^d, \tag{1}$$

where C_n, known as the *empirical copula*, is an estimate of the unique copula C based on a random sample $(X_{11},\ldots,X_{1d})$, $\ldots$, $(X_{n1},\ldots,X_{nd})$ from c.d.f. F. Initially studied in [9], it is usually defined by

$$C_n(u) = \frac{1}{n}\sum_{i=1}^n \prod_{j=1}^d 1[F_{j,n}(X_{ij}) \le u_j], \qquad u \in [0,1]^d,$$

where, for any $j \in \{1,\ldots,d\}$,

$$F_{j,n}(x) = \frac{1}{n}\sum_{i=1}^n 1[X_{ij} \le x], \qquad x \in \mathbb{R},$$

is the empirical c.d.f. of X_j. This amounts to working on the ranks $(R_{11},\ldots,R_{1d})$, $\ldots$, $(R_{n1},\ldots,R_{nd})$ associated with the random sample as, for any $i \in \{1,\ldots,n\}$, and any $j \in \{1,\ldots,d\}$, $R_{ij} = nF_{j,n}(X_{ij})$. In this context, it is also convenient to define the *pseudo-observations* $\hat{U}_{ij}$, depending on the ranks and the sample size, by

$$\hat{U}_{ij} = F_{j,n}(X_{ij}) = \frac{R_{ij}}{n}, \qquad \forall i \in \{1,\ldots,n\}, \forall j \in \{1,\ldots,d\}. \tag{2}$$

The empirical copula is then merely the empirical c.d.f. calculated from the pseudo-observations:

$$C_n(u) = \frac{1}{n}\sum_{i=1}^n \prod_{j=1}^d 1[\hat{U}_{ij} \le u_j], \qquad u \in [0,1]^d. \tag{3}$$

The mathematical beauty of the test suggested by Deheuvels [3] and studied in [4,10] comes from the fact that, under the mutual independence of $X_1,\ldots,X_d$, the empirical process (1) can be decomposed, using the *Möbius* *transform* [11], into $2^d - d - 1$ sub-processes $\sqrt{n}\mathcal{M}_A(C_n)$, $A \subseteq \{1, \ldots, d\}$, $|A| > 1$, that converge jointly to tight centered mutually independent Gaussian processes. One fundamental property of this decomposition, whose form will be precisely defined in Section 2.4, is that mutual independence among $X_1, \ldots, X_d$ is equivalent to having $\mathcal{M}_A(C)(u) = 0$, for all $u \in [0,1]^d$ and all $A \subseteq \{1, \ldots, d\}$ such that $|A| > 1$. Instead of one test statistic based on (1), this leads one to consider $2^d - d - 1$ test statistics of the form

$$\int_{[0,1]^d} [\sqrt{n}\mathcal{M}_A(C_n)(u)]^2 \mathrm{d}u \qquad \text{or} \qquad \sup_{u \in [0,1]^d} |\sqrt{n}\mathcal{M}_A(C_n)(u)|,$$

where $A \subseteq \{1, \ldots, d\}$, $|A| > 1$, that are asymptotically mutually independent under the null hypothesis of independence. Working with the above Cramér-von Mises versions of the test statistics, Genest and Rémillard [4] showed how to compute quantiles from their asymptotic and small-sample distributions. Furthermore, they investigated how these $2^d - d - 1$ statistics could be combined to obtain a global statistic for testing independence, thereby leading to a potentially more powerful test. More recently, Genest, Quessy and Rémillard [10] compared the asymptotic power of the Cramér-von Mises test derived from the copula process (1) with tests involving different combinations of the $2^d - d - 1$ Cramér-von Mises statistics derived from the Möbius decomposition of (1).

The main theoretical aim of this paper is to extend the Möbius decomposition proposed by Deheuvels [3] to the situation where one wants to test the mutual independence of p continuous random vectors using the empirical copula process. A more general objective was recently pursued by Beran, Bilodeau and Lafaye de Micheaux [12] using a characterization of mutual independence defined from probabilities of half-spaces. Being based on the empirical probability distribution, their test of independence between random vectors can be applied in a wide variety of situations (purely discrete, purely continuous or mixed setting). However, the characterization of mutual independence employed in their test results in a very high computational cost [12, §6]. The approach considered in this work is less ambitious and merely leads to an extension of the empirical copula-based procedures studied in [3] and [4]. As a consequence, the resulting rank-based procedures are only applicable to the situation where mutual independence among *continuous* random vectors is to be tested. One important advantage however, that follows from the rank-based nature of the studied tests, is their speed, which will allow us to investigate their finite-sample properties. The extension of this work to the continuous multivariate time series setting can be found in [13].

The paper is organized as follows. The second section is devoted to the empirical copula process for testing independence among random vectors and to its Möbius decomposition. The resulting processes, unlike in the "univariate" case studied by Deheuvels, are shown to be distribution-dependent. In the third section, we give the expressions of the Cramér-von Mises statistics derived from these processes in terms of the pseudo-observations and show that the bootstrap methodology, which can be used to practically carry out the tests, behaves consistently. The last subsection is devoted to a straightforward adaptation of the solutions proposed in [4] to practically implement the tests. Finally, simulations are presented in the last section.

Note that all the empirical copula-based tests studied in this paper are implemented, along those studied in [4], in the R package `copula` [14] available on the Comprehensive R Archive Network (`http://cran.r-project.org`).

2 The independence empirical copula process and its Möbius decomposition

2.1 Notation and setting

We want to test the mutual independence of p continuous random vectors $\mathbb{X}_1$, ..., $\mathbb{X}_p$ of dimensions $d_1,\dots,d_p$ respectively. Let $S = \{1,\dots,p\}$ and let $d = d_1+\cdots+d_p$ be the dimension of the random vector $(\mathbb{X}_1,\dots,\mathbb{X}_p)$. Furthermore, define the integers $b_1,\dots,b_p$ as

$$b_j = \sum_{k=1}^{j} d_k, \qquad \forall j \in S,$$

with the convention that $b_0 = 0$. Clearly, $b_j = b_{j-1} + d_j$ for all $j \in S$. These integers will be used to name the components of the random vectors $\mathbb{X}_1$, ..., $\mathbb{X}_p$: for any $k \in S$, the d_k components of the random vector $\mathbb{X}_k$ will be denoted by $X_{b_{k-1}+1}, X_{b_{k-1}+2}, \dots, X_{b_k}$ respectively.

The copula of the random vector $(\mathbb{X}_1,\dots,\mathbb{X}_p) = (X_1,\dots,X_d)$ will be denoted by C. Moreover, given a vector $u \in [0,1]^d$ and a subset B of S, the vector $u^B \in [0,1]^d$ is defined, for any $i \in \{1,\dots,d\}$, by

$$u_i^B = \begin{cases} u_i, \text{ if } i \in \bigcup_{j\in B}\{b_{j-1}+1,\dots,b_j\}, \\ 1, \ \text{ otherwise.} \end{cases}$$

For any $k \in S$, the marginal copula of $\mathbb{X}_k$ is then given by $C(u^{\{k\}})$, $u \in [0,1]^d$, and mutual independence among $\mathbb{X}_1$, ..., $\mathbb{X}_p$ occurs when

$$C(u) = \prod_{k=1}^{p} C(u^{\{k\}}), \qquad \forall u \in [0,1]^d.$$

As we continue, we shall assume that we have at hand n independent copies of the random vector $(\mathbb{X}_1, \ldots, \mathbb{X}_p) = (X_1, \ldots, X_d)$ that are denoted by $(X_{11}, \ldots, X_{1d})$, $\ldots$, $(X_{n1}, \ldots, X_{nd})$ with associated pseudo-observations defined by (2). A natural extension of the independence copula process (1) based on this random sample is then

$$\sqrt{n}\left[C_n(u) - \prod_{k=1}^{p} C_n(u^{\{k\}})\right], \qquad u \in [0,1]^d, \tag{4}$$

where the empirical copula C_n is defined as in (3).

Before studying the asymptotic behavior of the above process, let us recall recent results on the estimation of the empirical copula.

2.2 Asymptotic behavior of the empirical copula

Let $\ell^\infty([0,1]^d)$ be the space of all bounded real-valued functions on $[0,1]^d$ equipped with the uniform metric. The asymptotic behavior of the empirical copula defined by (3) is classically studied through the *empirical copula process*. The most modern treatment seems to be due to Fermanian, Radulovic and Wegkamp [15] and Tsukahara [16] whose results are summarized in the following theorem.

Theorem 1 *Suppose that C has continuous partial derivatives. Then, the empirical copula process*

$$\sqrt{n}[C_n(u) - C(u)], \qquad u \in [0,1]^d,$$

converges weakly in $\ell^\infty([0,1]^d)$ to the tight centered Gaussian process

$$\mathcal{G}(u) = \mathcal{B}(u) - \sum_{i=1}^{d} \partial_i C(u) \mathcal{B}(1, \ldots, 1, u_i, 1, \ldots, 1), \qquad u \in [0,1]^d,$$

where $\partial_i C$ denotes the i-th partial derivative of C and $\mathcal{B}$ is a tight centered Gaussian process on $[0,1]^d$ with covariance function

$$E[\mathcal{B}(u)\mathcal{B}(u')] = C(u \wedge u') - C(u)C(u'),$$

i.e., $\mathcal{B}$ is a multivariate tied-down Brownian bridge.

2.3 *Independence empirical copula process*

In order to study the asymptotic behavior of the empirical process (4), we consider the map $\mathcal{I} : \ell^\infty([0,1]^d) \to \ell^\infty([0,1]^d)$ defined by

$$\mathcal{I}(f)(x) = f(x) - \prod_{k=1}^{p} f(x^{\{k\}}), \qquad x \in [0,1]^d. \tag{5}$$

Lemma 2 *The map $\mathcal{I}$ is Hadamard differentiable tangentially to $\ell^\infty([0,1]^d)$ and its derivative (a continuous linear map from $\ell^\infty([0,1]^d)$ to $\ell^\infty([0,1]^d)$) at $f \in \ell^\infty([0,1]^d)$ is*

$$\mathcal{I}'_f(a)(x) = a(x) - \sum_{i=1}^{p} a(x^{\{i\}}) \prod_{\substack{j=1\\ j\neq i}}^{p} f(x^{\{j\}}), \qquad x \in [0,1]^d.$$

Proof. Let $f \in \ell^\infty([0,1]^d)$ and let t_n be a sequence of reals converging to 0. Let $a_n \in \ell^\infty([0,1]^d)$ be a sequence of functions converging to $a \in \ell^\infty([0,1]^d)$ such that $f + t_n a_n \in \ell^\infty([0,1]^d)$ for every n. Then, uniformly in $x \in [0,1]^d$,

$$\begin{aligned} &\frac{\mathcal{I}(f + t_n a_n)(x) - \mathcal{I}(f)(x)}{t_n} \\ &= \frac{(f + t_n a_n)(x) - \prod_{k=1}^{p} (f + t_n a_n)(x^{\{k\}}) - f(x) + \prod_{k=1}^{p} f(x^{\{k\}})}{t_n} \\ &\to a(x) - \sum_{k=1}^{p} a(x^{\{k\}}) \prod_{\substack{j=1\\ j\neq k}}^{p} f(x^{\{j\}}). \end{aligned}$$

□

The following result extends that presented in [3, §2] and establishes, under independence, the asymptotic behavior of the independence empirical copula process (4).

Theorem 3 *Suppose that C has continuous partial derivatives. Then, when*

$$\mathcal{I}(C)(u) = C(u) - \prod_{k=1}^{p} C(u^{\{k\}}) = 0, \qquad u \in [0,1]^d,$$

i.e., when $\mathbb{X}_1, \ldots, \mathbb{X}_p$ are mutually independent, the empirical process $\sqrt{n}\mathcal{I}(C_n)(u)$, $u \in [0,1]^d$, converges weakly in $\ell^\infty([0,1]^d)$ to the tight centered Gaussian process

$$\mathcal{I}'_C(\mathcal{G})(u) = \mathcal{B}(u) - \sum_{k=1}^{p} \mathcal{B}(u^{\{k\}}) \prod_{\substack{j=1\\ j\neq k}}^{p} C(u^{\{j\}}), \qquad u \in [0,1]^d.$$

For $u,v \in [0,1]^d$, the covariance function $E[\mathcal{I}'_C(\mathcal{G})(u)\mathcal{I}'_C(\mathcal{G})(v)]$ is given by

$$\prod_{k=1}^p C(u^{\{k\}}\wedge v^{\{k\}}) - \sum_{k=1}^p C(u^{\{k\}}\wedge v^{\{k\}}) \prod_{\substack{j=1\\j\neq k}}^p C(u^{\{j\}})C(v^{\{j\}}) + (p-1)\prod_{k=1}^p C(u^{\{k\}})C(v^{\{k\}}).$$

Proof. From Theorem 1 and the application of the functional delta method [17, Theorem 3.9.4] with the Hadamard differentiable map $\mathcal{I}$ (see Lemma 2), we have that the empirical process

$$\sqrt{n}\left[\mathcal{I}(C_n)(u) - \mathcal{I}(C)(u)\right], \qquad u \in [0,1]^d,$$

converges weakly in $\ell^\infty([0,1]^d)$ to the tight centered Gaussian process

$$\mathcal{I}'_C(\mathcal{G})(u) = \mathcal{G}(u) - \sum_{k=1}^p \mathcal{G}(u^{\{k\}}) \prod_{\substack{j=1\\j\neq k}}^p C(u^{\{j\}}), \qquad u \in [0,1]^d.$$

The first claim then follows from the fact that, under independence, for any $k \in S$ and any $j \in \{1,\dots,d_k\}$,

$$\partial_{b_{k-1}+j}C(u) = \partial_{b_{k-1}+j}\left[\prod_{l=1}^p C(u^{\{l\}})\right] = \prod_{\substack{l=1\\l\neq k}}^p C(u^{\{l\}})\partial_{b_{k-1}+j}C(u^{\{k\}}), \qquad u \in [0,1]^d.$$

For the second claim, fix $u,v \in [0,1]^d$. Then $E[\mathcal{I}'_C(\mathcal{G})(u)\mathcal{I}'_C(\mathcal{G})(v)]$ is equal to

$$\begin{aligned} &E[\mathcal{B}(u)\mathcal{B}(v)] - \sum_{l=1}^p E[\mathcal{B}(u)\mathcal{B}(v^{\{l\}})] \prod_{\substack{m=1\\m\neq l}}^p C(v^{\{m\}}) \\ &- \sum_{k=1}^p E[\mathcal{B}(v)\mathcal{B}(u^{\{k\}})] \prod_{\substack{j=1\\j\neq k}}^p C(u^{\{j\}}) + \sum_{k=1}^p\sum_{l=1}^p E[\mathcal{B}(u^{\{k\}})\mathcal{B}(v^{\{l\}})] \prod_{\substack{j=1\\j\neq k}}^p C(u^{\{j\}}) \prod_{\substack{m=1\\m\neq l}}^p C(v^{\{m\}}). \end{aligned}$$

Using the expression of the covariance function of the process $\mathcal{B}$ given in Theorem 1 and the fact that mutual independence among $\mathbb{X}_1,\dots,\mathbb{X}_p$ is assumed, this is equal to

$$\begin{aligned} C(u\wedge v) - C(u)C(v) - 2\sum_{k=1}^p [C(u^{\{k\}}\wedge v^{\{k\}}) - C(u^{\{k\}})C(v^{\{k\}})] \prod_{\substack{j=1\\j\neq k}}^p C(u^{\{j\}})C(v^{\{j\}}) \\ + \sum_{k=1}^p\sum_{l=1}^p [C(u^{\{k\}}\wedge v^{\{l\}}) - C(u^{\{k\}})C(v^{\{l\}})] \prod_{\substack{j=1\\j\neq k}}^p C(u^{\{j\}}) \prod_{\substack{m=1\\m\neq l}}^p C(v^{\{m\}}). \end{aligned}$$

The last term is clearly zero if $k \neq l$ in the summand. The desired result is then immediately obtained after simplification. □

In the general situation where at least one of the random vectors $\mathbb{X}_1, \ldots, \mathbb{X}_p$ is of dimension two or more, it follows from the previous result that, unlike in the "univariate" case considered in [4], any test statistic derived from the independence empirical copula process (4) will not be distribution-free.

2.4 Möbius decomposition of the independence process

The aim of this subsection is to generalize the decomposition presented in [3, §2] and used in [4] and [10]. This will eventually enable us to consider $2^p - p - 1$ statistics that are asymptotically mutually independent under the assumption of independence among $\mathbb{X}_1, \ldots, \mathbb{X}_p$, instead of one statistic derived from (4).

Let $\mathcal{P}_S = \{B \subseteq S : |B| > 1\}$. Since $|S| = p$, $\mathcal{P}_S$ contains $2^p - p - 1$ elements. Let $A \subseteq S$ and consider the map $\mathcal{M}_A : \ell^\infty([0,1]^d) \to \ell^\infty([0,1]^d)$ defined by

$$\mathcal{M}_A(f)(x) = \sum_{B \subseteq A} (-1)^{|A|-|B|} f(x^B) \prod_{k \in A \setminus B} f(x^{\{k\}}), \qquad x \in [0,1]^d. \tag{6}$$

Lemma 4 *For any $A \in \mathcal{P}_S$, the map $\mathcal{M}_A$ is Hadamard differentiable tangentially to $\ell^\infty([0,1]^d)$ and its derivative (a continuous linear map from $\ell^\infty([0,1]^d)$ to $\ell^\infty([0,1]^d)$) at $f \in \ell^\infty([0,1]^d)$ is*

$$\mathcal{M}'_{A,f}(a)(x) = \sum_{B \subseteq A} (-1)^{|A|-|B|} \left[f(x^B) \sum_{k \in A \setminus B} a(x^{\{k\}}) \prod_{\substack{i \in A \setminus B \\ i \neq k}} f(x^{\{i\}}) + a(x^B) \prod_{k \in A \setminus B} f(x^{\{k\}}) \right],$$

where $x \in [0,1]^d$.

Proof. Fix $A \in \mathcal{P}_S$ and $f \in \ell^\infty([0,1]^d)$, and let t_n be a sequence of reals converging to 0. Let $a_n \in \ell^\infty([0,1]^d)$ be a sequence of functions converging to $a \in \ell^\infty([0,1]^d)$ such that $f + t_n a_n \in \ell^\infty([0,1]^d)$ for every n. Then, uniformly in $x \in [0,1]^d$,

$$\begin{aligned}
&\frac{1}{t_n} \left[\mathcal{M}_A(f + t_n a_n)(x) - \mathcal{M}_A(f)(x) \right] \\
&\quad = \sum_{B \subseteq A} \frac{(-1)^{|A|-|B|}}{t_n} \left[(f + t_n a_n)(x^B) \prod_{k \in A \setminus B} (f + t_n a_n)(x^{\{k\}}) - f(x^B) \prod_{k \in A \setminus B} f(x^{\{k\}}) \right] \\
&\quad \to \sum_{B \subseteq A} (-1)^{|A|-|B|} \left[f(x^B) \sum_{k \in A \setminus B} a(x^{\{k\}}) \prod_{\substack{i \in A \setminus B \\ i \neq k}} f(x^{\{i\}}) + a(x^B) \prod_{k \in A \setminus B} f(x^{\{k\}}) \right].
\end{aligned}$$

□

The $2^p - p - 1$ functions $\{\mathcal{M}_A(C) : A \in \mathcal{P}_S\}$ (resp. empirical processes $\{\mathcal{M}_A(C_n) : A \in \mathcal{P}_S\}$) are known as the *Möbius decomposition* of $\mathcal{I}(C)$ (resp. $\mathcal{I}(C_n)$).

The two following lemmas are immediate extensions of known results. We include the proofs for completeness.

Lemma 5 *Let H be a c.d.f. on $[0,1]^d$ and let $A \in \mathcal{P}_S$. Then,*

$$\mathcal{M}_A(H)(u) = \sum_{B \subseteq A} (-1)^{|A|-|B|} \mathcal{I}(H)(u^B) \prod_{k \in A \setminus B} H(u^{\{k\}}), \qquad u \in [0,1]^d.$$

Proof. Since $\sum_{B \subseteq A} (-1)^{|A|-|B|} = 0$, for any $u \in [0,1]^d$, we can write

$$\begin{aligned}
\mathcal{M}_A(H)(u) &= \sum_{B \subseteq A} (-1)^{|A|-|B|} H(u^B) \prod_{k \in A \setminus B} H(u^{\{k\}}) - \prod_{k \in A} H(u^{\{k\}}) \sum_{B \subseteq A} (-1)^{|A|-|B|} \\
&= \sum_{B \subseteq A} (-1)^{|A|-|B|} \left[H(u^B) - \prod_{k \in B} H(u^{\{k\}})(u^{\{k\}}) \right] \prod_{k \in A \setminus B} H(u^{\{k\}}).
\end{aligned}$$

□

Lemma 6 *Let H be a c.d.f. on $[0,1]^d$. Then,*

$$\sum_{A \in \mathcal{P}_S} \mathcal{M}_A(H)(u) \prod_{k \in S \setminus A} H(u^{\{k\}}) = \mathcal{I}(H)(u), \qquad u \in [0,1]^d.$$

Proof. For any $u \in [0,1]^d$, Lemma 5 yields

$$\sum_{A \in \mathcal{P}_S} \mathcal{M}_A(H)(u) \prod_{k \in S \setminus A} H(u^{\{k\}}) = \sum_{A \in \mathcal{P}_S} \sum_{B \subseteq A} (-1)^{|A|-|B|} \mathcal{I}(H)(u^B) \prod_{k \in S \setminus B} H(u^{\{k\}}).$$

Interchanging the sums, we obtain,

$$\begin{aligned}
\sum_{A \in \mathcal{P}_S} \mathcal{M}_A(H)(u) \prod_{k \in S \setminus A} H(u^{\{k\}}) &= \sum_{B \subseteq S} \sum_{\substack{A \in \mathcal{P}_S \\ A \supseteq B}} (-1)^{|A|-|B|} \mathcal{I}(H)(u^B) \prod_{k \in S \setminus B} H(u^{\{k\}}) \\
&= \sum_{B \subseteq S} \mathcal{I}(H)(u^B) \prod_{k \in S \setminus B} H(u^{\{k\}}) \sum_{\substack{A \in \mathcal{P}_S \\ A \supseteq B}} (-1)^{|A|-|B|} \\
&= \sum_{B \subsetneq S} \mathcal{I}(H)(u^B) \prod_{k \in S \setminus B} H(u^{\{k\}}) \sum_{\substack{A \in \mathcal{P}_S \\ A \supseteq B}} (-1)^{|A|-|B|} + \mathcal{I}(H)(u).
\end{aligned}$$

Using the fact that $\sum_{\substack{A \in \mathcal{P}_S \\ A \supseteq B}} (-1)^{|A|-|B|} = 0$ when $B \subsetneq S$, we obtain the desired result. □

The following proposition, again an immediate generalization of a known result (see e.g. [18]), is of central importance as it will be used to provide an alternative characterization of mutual independence among $\mathbb{X}_1, \dots, \mathbb{X}_p$.

Proposition 7 *Let H be a c.d.f. on $[0,1]^d$. The two following statements are equivalent:*

(i) $\mathcal{I}(H)(u) = 0$ *for all* $u \in [0,1]^d$,

(ii) $\mathcal{M}_A(H)(u) = 0$ *for every* $A \in \mathcal{P}_S$ *and* $u \in [0,1]^d$.

Proof. The fact that (ii) implies (i) follows from the previous lemma, while the fact that (i) implies (ii) follows from Lemma 5. □

The following result generalizes Theorem 1 in [3]. It shows that, under independence, the empirical processes arising from the Möbius decomposition of the independence empirical copula process (4) are asymptotically mutually independent.

Theorem 8 *Suppose that C has continuous partial derivatives. Then, under mutual independence of $\mathbb{X}_1, \dots, \mathbb{X}_p$, the vector of $2^p - p - 1$ empirical processes $\{\sqrt{n}\mathcal{M}_A(C_n)(u), u \in [0,1]^d : A \in \mathcal{P}_S\}$ converges weakly in $\ell^\infty([0,1]^d)$ to the corresponding vector of tight centered Gaussian processes $\{\mathcal{M}'_{A,C}(\mathcal{G})(u), u \in [0,1]^d : A \in \mathcal{P}_S\}$, where*

$$\mathcal{M}'_{A,C}(\mathcal{G})(u) = \sum_{B \subseteq A} (-1)^{|A|-|B|} \mathcal{B}(u^B) \prod_{k \in A \setminus B} C(u^{\{k\}}), \qquad u \in [0,1]^d.$$

For $u, v \in [0,1]^d$ and $A, A' \in \mathcal{P}_S$, the cross-covariance function is given by

$$E[\mathcal{M}'_{A,C}(\mathcal{G})(u)\mathcal{M}'_{A',C}(\mathcal{G})(v)] = 1(A = A') \prod_{k \in A} [C(u^{\{k\}} \wedge v^{\{k\}}) - C(u^{\{k\}})C(v^{\{k\}})].$$

Proof. Let $\vec{\mathcal{M}}_S : \ell^\infty([0,1]^d) \to (\ell^\infty([0,1]^d))^{2^p-p-1}$ denote the (Hadamard differentiable) map whose $2^p - p - 1$ components are the maps $\mathcal{M}_A$, $A \in \mathcal{P}_S$. From Theorem 1 and the application of the functional delta method with the Hadamard differentiable map $\vec{\mathcal{M}}_S$, we obtain that $\sqrt{n}\vec{\mathcal{M}}_S(C_n)(u)$, $u \in [0,1]^d$ converges weakly in $(\ell^\infty([0,1]^d))^{2^p-p-1}$ to $\vec{\mathcal{M}}'_{S,C}(\mathcal{G})(u)$, whose corresponding components are defined by

$$\mathcal{M}'_{A,C}(\mathcal{G})(u) = \sum_{B\subseteq A} (-1)^{|A|-|B|} \left[C(u^B) \sum_{k\in A\setminus B} \mathcal{G}(u^{\{k\}}) \prod_{\substack{i\in A\setminus B\\ i\neq k}} C(u^{\{i\}}) \right.$$
$$\left. +\mathcal{G}(u^B) \prod_{k\in A\setminus B} C(u^{\{k\}}) \right].$$

Using the expression of the process $\mathcal{G}$ given in Theorem 1, for any $u \in [0,1]^d$, $\mathcal{M}'_{A,C}(\mathcal{G})(u)$ is given by

$$\sum_{B\subseteq A} (-1)^{|A|-|B|} \left[C(u^B) \sum_{k\in A\setminus B} \left\{ \mathcal{B}(u^{\{k\}}) - \sum_{i=b_{k-1}+1}^{b_k} \partial_i C(u^{\{k\}}) \mathcal{B}(1, u_i^{\{k\}}, 1) \right\} \prod_{\substack{i\in A\setminus B\\ i\neq k}} C(u^{\{i\}}) \right.$$
$$\left. + \left\{ \mathcal{B}(u^B) - \sum_{k\in B} \sum_{i=b_{k-1}+1}^{b_k} \partial_i C(u^B) \mathcal{B}(1, u_i^B, 1) \right\} \prod_{k\in A\setminus B} C(u^{\{k\}}) \right].$$

Using the fact that $\mathbb{X}_1, \ldots, \mathbb{X}_p$ are mutually independent, we get that $\mathcal{M}'_{A,C}(\mathcal{G})(u)$ is equal to

$$\sum_{B\subseteq A} (-1)^{|A|-|B|} \left[\sum_{k\in A\setminus B} \left\{ \mathcal{B}(u^{\{k\}}) - \sum_{i=b_{k-1}+1}^{b_k} \partial_i C(u^{\{k\}}) \mathcal{B}(1, u_i^{\{k\}}, 1) \right\} \prod_{\substack{i\in A\\ i\neq k}} C(u^{\{i\}}) \right.$$
$$\left. +\mathcal{B}(u^B) \prod_{k\in A\setminus B} C(u^{\{k\}}) - \sum_{k\in B} \sum_{i=b_{k-1}+1}^{b_k} \partial_i C(u^{\{k\}}) \mathcal{B}(1, u_i^{\{k\}}, 1) \prod_{\substack{i\in A\\ i\neq k}} C(u^{\{i\}}) \right].$$

The first claim then follows from the fact that

$$\sum_{B\subseteq A} (-1)^{|A|-|B|} \sum_{k\in A} \sum_{i=b_{k-1}+1}^{b_k} \partial_i C(u^{\{k\}}) \mathcal{B}(1, u_i^{\{k\}}, 1) \prod_{\substack{i\in A\\ i\neq k}} C(u^{\{i\}}) = 0,$$

as the main summand does not depend on B, and from the fact that

$$\sum_{B\subseteq A} \sum_{k\in A\setminus B} (-1)^{|A|-|B|} \mathcal{B}(u^{\{k\}}) \prod_{\substack{i\in A\\ i\neq k}} C(u^{\{i\}}) = \sum_{k\in A} \sum_{B\subseteq A\setminus\{k\}} (-1)^{|A|-|B|} \mathcal{B}(u^{\{k\}}) \prod_{\substack{i\in A\\ i\neq k}} C(u^{\{i\}})$$
$$= \sum_{k\in A} \mathcal{B}(u^{\{k\}}) \prod_{\substack{i\in A\\ i\neq k}} C(u^{\{i\}}) \sum_{B\subseteq A\setminus\{k\}} (-1)^{|A|-|B|} = 0.$$

Fix $A, A' \in \mathcal{P}_S$ and $u, v \in [0,1]^d$. Using Theorem 1, $E[\mathcal{M}'_{A,C}(\mathcal{G})(u)\mathcal{M}'_{A',C}(\mathcal{G})(v)]$

is equal to

$$E\left[\left(\sum_{B\subseteq A}(-1)^{|A|-|B|}\mathcal{B}(u^B)\prod_{k\in A\setminus B}C(u^{\{k\}})\right)\left(\sum_{B'\subseteq A'}(-1)^{|A'|-|B'|}\mathcal{B}(v^{B'})\prod_{k'\in A'\setminus B'}C(v^{\{k'\}})\right)\right]$$
$$=\sum_{B\subseteq A}\sum_{B'\subseteq A'}(-1)^{|A|-|B|+|A'|-|B'|}[C(u^B\wedge v^{B'})-C(u^B)C(v^{B'})]\prod_{k\in A\setminus B}C(u^{\{k\}})\prod_{k\in A'\setminus B'}C(v^{\{k\}}).$$

Let $A\cap A'=R\neq\emptyset$. Then, this is equal to

$$\sum_{K\subseteq A\setminus R}\sum_{L\subseteq R}\sum_{K'\subseteq A'\setminus R}\sum_{L'\subseteq R}(-1)^{|A|-|K|-|L|+|A'|-|K'|-|L'|}$$
$$\times[C(u^{K\cup L}\wedge v^{K'\cup L'})-C(u^{K\cup L})C(v^{K'\cup L'})]\prod_{k\in A\setminus(K\cup L)}C(u^{\{k\}})\prod_{k'\in A'\setminus(K'\cup L')}C(v^{\{k'\}}).$$

By construction, $K\cap(L\cup L'\cup K')=\emptyset$ and similarly for K'. The assumption of mutual independence then gives

$$E[\mathcal{M}'_{A,C}(\mathcal{G})(u)\mathcal{M}'_{A',C}(\mathcal{G})(v)]=\sum_{L\subseteq R}\sum_{L'\subseteq R}(-1)^{|L|+|L'|}[C(u^L\wedge v^{L'})-C(u^L)C(v^{L'})]$$
$$\times\prod_{k\in A\setminus L}C(u^{\{k\}})\prod_{k'\in A'\setminus L'}C(v^{\{k'\}})\sum_{K\subseteq A\setminus R}(-1)^{|A|-|K|}\sum_{K'\subseteq A'\setminus R}(-1)^{|A'|-|K'|},$$

whence the covariance is zero unless $A=R=A'$. In this case, using $(-1)^m=(-1)^{-m}$ and $(-1)^{2m}=1$, we see that $E[\mathcal{M}'_{A,C}(\mathcal{G})(u)\mathcal{M}'_{A,C}(\mathcal{G})(v)]$ is equal to

$$\sum_{B\subseteq A}\sum_{K\subseteq A\setminus B}\sum_{L\subseteq B}(-1)^{|B|+|K|+|L|}[C(u^B\wedge v^{K\cup L})-C(u^B)C(v^{K\cup L})]\prod_{k\in A\setminus B}C(u^{\{k\}})\prod_{k\in A\setminus(K\cup L)}C(v^{\{k\}})$$
$$=\sum_{B\subseteq A}\sum_{L\subseteq B}(-1)^{|B|+|L|}[C(u^B\wedge v^L)-C(u^B)C(v^L)]\prod_{k\in A\setminus B}C(u^{\{k\}})\prod_{k\in A\setminus L}C(v^{\{k\}})\sum_{K\subseteq A\setminus B}(-1)^{|K|},$$

the summand being zero when $B\neq A$. Thus, $E[\mathcal{M}'_{A,C}(\mathcal{G})(u)\mathcal{M}'_{A,C}(\mathcal{G})(v)]$ is equal to

$$\sum_{L\subseteq A}(-1)^{|A|+|L|}[C(u^A\wedge v^L)-C(u^A)C(v^L)]\prod_{k\in A\setminus L}C(v^{\{k\}})$$
$$=\sum_{L\subseteq A}(-1)^{|A|+|L|}[C(u^L\wedge v^L)-C(u^L)C(v^L)]\prod_{k\in A\setminus L}C(u^{\{k\}})C(v^{\{k\}})$$
$$=\sum_{L\subseteq A}(-1)^{|A|-|L|}\left[\prod_{j\in L}C(u^{\{j\}}\wedge v^{\{j\}})-\prod_{j\in L}C(u^{\{j\}})C(v^{\{j\}})\right]\prod_{k\in A\setminus L}C(u^{\{k\}})C(v^{\{k\}}).$$

This is the difference of two terms, the second of which is zero since $\prod_{k\in A}C(u^{\{k\}})C(v^{\{k\}})$ is independent of L and $\sum_{L\subseteq A}(-1)^{|A|-|L|}=0$. Using the multinomial formula on the first term, we obtain the desired result. □

3 Tests for independence

A natural next step consists of considering, as measures of departure from independence, Kolmogorov-Smirnov or Cramér-von Mises statistics derived from the previously studied processes. The first two subsections give the expressions of the corresponding Cramér-von Mises statistics in terms of the pseudo-observations as well as some simple convergence results. As these statistics are not distribution-free, the bootstrap or the permutation methodology can be used to obtain approximate p-values and critical values. The former approach is shown to be consistent in the third subsection, while the latter is used in practice. The last subsection is devoted to the practical implementation of the tests, which mainly consists of transposing the solutions proposed in [4] to the current context.

3.1 Statistic derived from the independence empirical copula process

The Cramér-von Mises statistic derived from the empirical process (4) is given by

$$I_n = n \int_{[0,1]^d} \left[C_n(u) - \prod_{k=1}^{p} C_n(u^{\{k\}}) \right]^2 \mathrm{d}u = n \int_{[0,1]^d} \mathcal{I}(C_n)(u)^2 \mathrm{d}u. \quad (7)$$

The following result is an immediate consequence of Theorem 3 and the continuous mapping theorem.

Corollary 9 *Suppose that C has continuous partial derivatives. Then, under mutual independence of $\mathbb{X}_1, \ldots, \mathbb{X}_p$, the random variable I_n converges in distribution to*

$$\int_{[0,1]^d} \mathcal{I}'_C(\mathcal{G})(u)^2 \mathrm{d}u.$$

We now give the expression of the statistic in terms of the pseudo-observations.

Proposition 10 *We have*

$$I_n = \frac{1}{n} \sum_{i=1}^{n} \sum_{l=1}^{n} \prod_{j=1}^{d} [1 - \hat{U}_{ij} \vee \hat{U}_{lj}] - \frac{2}{n^p} \sum_{i=1}^{n} \prod_{k=1}^{p} \sum_{l=1}^{n} \prod_{j=b_{k-1}+1}^{b_k} [1 - \hat{U}_{ij} \vee \hat{U}_{lj}]$$
$$+ \frac{1}{n^{2p-1}} \prod_{k=1}^{p} \sum_{i=1}^{n} \sum_{l=1}^{n} \prod_{j=b_{k-1}+1}^{b_k} [1 - \hat{U}_{ij} \vee \hat{U}_{lj}].$$

Proof. For any $u \in [0,1]^d$, we have

$$\left[C_n(u) - \prod_{k=1}^{p} C_n(u^{\{k\}})\right]^2 = C_n(u)^2 - 2C_n(u)\prod_{k=1}^{p} C_n(u^{\{k\}}) + \prod_{k=1}^{p} C_n(u^{\{k\}})^2.$$

Integrating the first term over $[0,1]^d$, we obtain

$$\begin{aligned}\int_{[0,1]^d} C_n(u)^2 \mathrm{d}u &= \int_{[0,1]^d} \left[\frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{d} 1[\hat{U}_{ij} \le u_j]\right]\left[\frac{1}{n}\sum_{l=1}^{n}\prod_{j=1}^{d} 1[\hat{U}_{lj} \le u_j]\right] \mathrm{d}u \\ &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{l=1}^{n}\prod_{j=1}^{d}\int_{[0,1]} 1[\hat{U}_{ij} \le u_j]1[\hat{U}_{lj} \le u_j]\mathrm{d}u_j = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{l=1}^{n}\prod_{j=1}^{d}[1 - \hat{U}_{ij} \vee \hat{U}_{lj}].\end{aligned}$$

Similarly, for the last term, it is easy to verify that

$$\int_{[0,1]^d} \prod_{k=1}^{p} C_n(u^{\{k\}})^2 \mathrm{d}u = \frac{1}{n^{2p}}\prod_{k=1}^{p}\sum_{i=1}^{n}\sum_{l=1}^{n}\prod_{j=b_{k-1}+1}^{b_k}[1 - \hat{U}_{ij} \vee \hat{U}_{lj}].$$

Finally, for the second term, we have

$$\begin{aligned}\int_{[0,1]^d} C_n(u)\prod_{k=1}^{p} C_n(u^{\{k\}})\mathrm{d}u &= \int_{[0,1]^d}\left[\frac{1}{n}\sum_{i=1}^{n}\prod_{k=1}^{p}\prod_{j=b_{k-1}+1}^{b_k} 1[\hat{U}_{ij} \le u_j]\right]\prod_{k=1}^{p} C_n(u^{\{k\}})\mathrm{d}u \\ &= \int_{[0,1]^d}\frac{1}{n}\sum_{i=1}^{n}\prod_{k=1}^{p}\left[C_n(u^{\{k\}})\prod_{j=b_{k-1}+1}^{b_k} 1[\hat{U}_{ij} \le u_j]\right]\mathrm{d}u \\ &= \int_{[0,1]^d}\frac{1}{n}\sum_{i=1}^{n}\prod_{k=1}^{p}\left[\frac{1}{n}\sum_{l=1}^{n}\prod_{j=b_{k-1}+1}^{b_k} 1[\hat{U}_{ij} \le u_j]1[\hat{U}_{lj} \le u_j]\right]\mathrm{d}u \\ &= \frac{1}{n^{p+1}}\sum_{i=1}^{n}\prod_{k=1}^{p}\sum_{l=1}^{n}\prod_{j=b_{k-1}+1}^{b_k}[1 - \hat{U}_{ij} \vee \hat{U}_{lj}].\end{aligned}$$

□

3.2 Statistics derived from the Möbius decomposition of the independence process

The $2^p - p - 1$ Cramér-von Mises statistics obtained from the Möbius decomposition of the independence empirical copula process are given by

$$M_{A,n} = n\int_{[0,1]^d} \mathcal{M}_A(C_n)(u)^2 \mathrm{d}u, \qquad A \in \mathcal{P}_S.$$

The following result is an immediate consequence of Theorem 8 and the continuous mapping theorem.

Corollary 11 *Suppose that C has continuous partial derivatives. Then, under mutual independence of $\mathbb{X}_1,\ldots,\mathbb{X}_p$, the random vector $\{M_{A,n} : A \in \mathcal{P}_S\}$ converges in distribution to the random vector*

$$\left\{ \int_{[0,1]^d} \mathcal{M}'_{A,C}(\mathcal{G})(u)^2 \mathrm{d}u : A \in \mathcal{P}_S \right\},$$

whose components are mutually independent.

The following lemma will be useful to establish the expression of the statistics in terms of the pseudo-observations. It is a simple adaptation of a known result; see e.g. [18].

Lemma 12 *Let $A \in \mathcal{P}_S$. Then, for any $u \in [0,1]^d$, we have*

$$\mathcal{M}_A(C_n)(u) = \frac{1}{n}\sum_{i=1}^{n}\prod_{k\in A}\left[\prod_{j=b_{k-1}+1}^{b_k} 1[\hat{U}_{ij} \le u_j] - C_n(u^{\{k\}})\right].$$

Proof. For any $u \in [0,1]^d$, starting from (6), we can write

$$\begin{aligned}\mathcal{M}_A(C_n)(u) &= \sum_{B\subseteq A}\left[\frac{1}{n}\sum_{i=1}^{n}\prod_{k\in B}\prod_{j=b_{k-1}+1}^{b_k} 1[\hat{U}_{ij} \le u_j]\right]\prod_{l\in A\setminus B}(-1)^{|A|-|B|}C_n(u^{\{l\}}) \\ &= \frac{1}{n}\sum_{i=1}^{n}\prod_{k\in A}\left[\prod_{j=b_{k-1}+1}^{b_k} 1[\hat{U}_{ij} \le u_j] - C_n(u^{\{k\}})\right],\end{aligned}$$

where the last equality follows from the multinomial formula. □

In terms of the pseudo-observations, the statistics can be expressed as follows.

Proposition 13 *For any $A \in \mathcal{P}_S$, we have*

$$\begin{aligned}M_{A,n} = \frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{n}\prod_{k\in A}\Bigg[&\prod_{j=b_{k-1}+1}^{b_k}[1-\hat{U}_{ij}\vee\hat{U}_{lj}] - \frac{1}{n}\sum_{m=1}^{n}\prod_{j=b_{k-1}+1}^{b_k}[1-\hat{U}_{ij}\vee\hat{U}_{mj}] \\ &-\frac{1}{n}\sum_{m=1}^{n}\prod_{j=b_{k-1}+1}^{b_k}[1-\hat{U}_{lj}\vee\hat{U}_{mj}] + \frac{1}{n^2}\sum_{r=1}^{n}\sum_{s=1}^{n}\prod_{j=b_{k-1}+1}^{b_k}[1-\hat{U}_{rj}\vee\hat{U}_{sj}]\Bigg].\end{aligned}$$

Proof. Starting from Lemma 12, for any $u \in [0,1]^d$, we can write

$$\begin{aligned}&\mathcal{M}_A(C_n)(u)^2 \\ &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{l=1}^{n}\prod_{k\in A}\left[\prod_{j=b_{k-1}+1}^{b_k} 1[\hat{U}_{ij} \le u_j] - C_n(u^{\{k\}})\right]\left[\prod_{j=b_{k-1}+1}^{b_k} 1[\hat{U}_{lj} \le u_j] - C_n(u^{\{k\}})\right],\end{aligned}$$

which is equivalent to

$$\mathcal{M}_A(C_n)(u)^2 = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{l=1}^{n}\prod_{k\in A}\left[\prod_{j=b_{k-1}+1}^{b_k} 1[\hat{U}_{ij}\le u_j]1[\hat{U}_{lj}\le u_j]\right.$$
$$\left. - \prod_{j=b_{k-1}+1}^{b_k} 1[\hat{U}_{ij}\le u_j]C_n(u^{\{k\}}) - \prod_{j=b_{k-1}+1}^{b_k} 1[\hat{U}_{lj}\le u_j]C_n(u^{\{k\}}) + C_n(u^{\{k\}})^2\right].$$

Using the expression for the empirical copula, the term between brackets becomes

$$\prod_{j=b_{k-1}+1}^{b_k} 1[\hat{U}_{ij}\le u_j]1[\hat{U}_{lj}\le u_j] - \frac{1}{n}\sum_{m=1}^{n}\prod_{j=b_{k-1}+1}^{b_k} 1[\hat{U}_{ij}\le u_j]1[\hat{U}_{mj}\le u_j]$$
$$-\frac{1}{n}\sum_{m=1}^{n}\prod_{j=b_{k-1}+1}^{b_k} 1[\hat{U}_{lj}\le u_j]1[\hat{U}_{mj}\le u_j]+\frac{1}{n^2}\sum_{r=1}^{n}\sum_{s=1}^{n}\prod_{j=b_{k-1}+1}^{b_k} 1[\hat{U}_{rj}\le u_j]1[\hat{U}_{sj}\le u_j].$$

Integrating over $[0,1]^{d_k}$, we obtain

$$\prod_{j=b_{k-1}+1}^{b_k}[1-\hat{U}_{ij}\vee\hat{U}_{lj}] - \frac{1}{n}\sum_{m=1}^{n}\prod_{j=b_{k-1}+1}^{b_k}[1-\hat{U}_{ij}\vee\hat{U}_{mj}]$$
$$-\frac{1}{n}\sum_{m=1}^{n}\prod_{j=b_{k-1}+1}^{b_k}[1-\hat{U}_{lj}\vee\hat{U}_{mj}] + \frac{1}{n^2}\sum_{r=1}^{n}\sum_{s=1}^{n}\prod_{j=b_{k-1}+1}^{b_k}[1-\hat{U}_{rj}\vee\hat{U}_{sj}].$$

□

3.3 Practical computation of the statistics

In order to increase the speed of the calculations of the statistics $M_{A,n}$, $A\in\mathcal{P}_S$, it is convenient to first compute the quantity

$$J(\hat{U},i,l,k) = \prod_{j=b_{k-1}+1}^{b_k}[1-\hat{U}_{ij}\vee\hat{U}_{lj}], \qquad i,l\in\{1,\dots,n\}, k\in S,$$

that depends on the pseudo-observations, and then, for any $i\in\{1,\dots,n\}$ and any $k\in S$, the quantities

$$K(\hat{U},i,k) = \frac{1}{n}\sum_{l=1}^{n}\prod_{j=b_{k-1}+1}^{b_k}[1-\hat{U}_{ij}\vee\hat{U}_{lj}] = \frac{1}{n}\sum_{l=1}^{n}J(\hat{U},i,l,k),$$

and

$$L(\hat{U},k) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{l=1}^{n}\prod_{j=b_{k-1}+1}^{b_k}[1-\hat{U}_{ij}\vee\hat{U}_{lj}] = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{l=1}^{n}J(\hat{U},i,l,k) = \frac{1}{n}\sum_{i=1}^{n}K(\hat{U},i,k).$$

The statistics are then given by

$$I_n = \frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{n}\prod_{k=1}^{p} J(\hat{U},i,l,k) - 2\sum_{i=1}^{n}\prod_{k=1}^{p} K(\hat{U},i,k) + n\prod_{k=1}^{p} L(\hat{U},k),$$

and by

$$M_{A,n} = \frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{n}\prod_{k\in A}\left[J(\hat{U},i,l,k) - K(\hat{U},i,k) - K(\hat{U},l,k) + L(\hat{U},k)\right], \qquad A \in \mathcal{P}_S.$$

3.4 Bootstrap of the test statistics

As already noted, in the current multivariate setting, the Cramér-von Mises test statistics under consideration are not distribution-free (see Theorems 3 and 8). In such a situation, a sensible way of obtaining critical values and p-values involves using the bootstrap methodology.

Under the assumption of independence among $\mathbb{X}_1,\ldots,\mathbb{X}_p$, a natural estimate of the copula C is given by $\prod_{k=1}^{p} C_n(u^{\{k\}})$, $u \in [0,1]^d$. It is then natural to consider that the bootstrap sample is constructed by sampling independently from the empirical marginal c.d.f. of each vector $\mathbb{X}_i$, $i \in S$; see e.g. [12,19]. Let F_n denote the joint empirical c.d.f. of $\mathbb{X}_1,\ldots,\mathbb{X}_p$ computed from the available data. The bootstrap sample is therefore a random sample drawn from the empirical c.d.f. $\prod_{k=1}^{p} F_n(x^{\{k\}})$, $x \in \mathbb{R}^d$, where, for any $k \in \{1,\ldots,p\}$, $x^{\{k\}}$ is a d-dimensional vector defined, for any $i \in \{1,\ldots,d\}$, by

$$x_i^{\{k\}} = \begin{cases} x_i, \text{ if } i \in \{b_{k-1}+1,\ldots,b_k\}, \\ \infty, \text{ otherwise.} \end{cases}$$

The aim of this subsection is to verify that the bootstrap distributions converge appropriately. Before presenting the main results, we introduce some definitions and notation, and state two lemmas.

For any c.d.f. $G : \mathbb{R} \to [0,1]$, define its *generalized inverse* by

$$G^{-}(u) = \inf\{x \in \mathbb{R} : G(x) \geq u\}.$$

Also, let $D([0,1]^d)$ (resp. $\mathcal{C}([0,1]^d)$) be the space of *càdlàg* (resp. continuous) functions on $[0,1]^d$ equipped with the Skorohod (resp. uniform) topology.

The following result is due to Fermanian, Radulovic and Wegkamp [15].

Lemma 14 *Let G be a c.d.f. with compact support $[0,1]^d$ and marginal c.d.f.s $G_1,\ldots,G_d$ that are continuously differentiable on $[0,1]$ with strictly positive* *densities. Furthermore, assume that G is continuously differentiable on $[0,1]^d$. Then, the map $\phi : D([0,1]^d) \to \ell^\infty([0,1]^d)$ defined by*

$$\phi(G)(u) = G(G_1^-(u_1), \dots, G_d^-(u_d)), \qquad u = (u_1, \dots, u_d) \in [0,1]^d, \qquad (8)$$

is Hadamard differentiable tangentially to $\mathcal{C}([0,1]^d)$.

Let F denote the c.d.f. of $(\mathbb{X}_1, \dots, \mathbb{X}_p) = (X_1, \dots, X_d)$, and, for any $i \in \{1, \dots, d\}$, let F_i denote the marginal c.d.f. of the random variable X_i. Define the random vector $(U_1, \dots, U_d)$ by $U_j = F_j(X_j)$, $j \in \{1, \dots, d\}$, and let H be its c.d.f. Notice that $(U_1, \dots, U_d)$ has uniform marginals. Similarly, let $(U_{11}, \dots, U_{1d})$,...,$(U_{n1}, \dots, U_{nd})$ be the n independent copies of $(U_1, \dots, U_d)$ obtained from the available sample $(X_{11}, \dots, X_{1d})$, ..., $(X_{n1}, \dots, X_{nd})$. The empirical c.d.f. obtained from the $(U_{i1}, \dots, U_{id})$, $i \in \{1, \dots, n\}$, is denoted by H_n. Finally, let the bootstrap sample drawn from the empirical c.d.f. $\prod_{k=1}^p F_n(x^{\{k\}})$ be denoted by $(X_{i1}^*, \dots, X_{id}^*)$, $i \in \{1, \dots, n\}$, let $(U_{i1}^*, \dots, U_{id}^*)$, $i \in \{1, \dots, n\}$, be its corresponding probability-transformed version and let H_n^* be the empirical c.d.f. obtained from this last sample.

The following lemma is a known result; see e.g. [17, Theorem 3.8.3].

Lemma 15 *Under mutual independence among $\mathbb{X}_1, \dots, \mathbb{X}_p$, the conditional distribution of the process*

$$\sqrt{n}\left[H_n^*(u) - \prod_{k=1}^p H_n(u^{\{k\}})\right], \qquad u \in [0,1]^d,$$

given the data, converges to the same limiting distribution as that of

$$\sqrt{n}\left[H_n(u) - \prod_{k=1}^p H(u^{\{k\}})\right], \qquad u \in [0,1]^d,$$

in $\ell^\infty([0,1]^d)$ almost surely.

Denote by C_n^* the empirical copula obtained from the bootstrap sample $(X_{i1}^*, \dots, X_{id}^*)$, $i \in \{1, \dots, n\}$.

Theorem 16 *Suppose that C has continuous partial derivatives. Under mutual independence among $\mathbb{X}_1, \dots, \mathbb{X}_p$, the conditional distribution of the process*

$$\sqrt{n}\left[C_n^*(u) - \prod_{k=1}^p C_n(u^{\{k\}})\right], \qquad u \in [0,1]^d,$$

given the data, converges to the same limiting distribution as that of

$$\sqrt{n}\left[C_n(u) - \prod_{k=1}^p C(u^{\{k\}})\right], \qquad u \in [0,1]^d,$$

in $\ell^\infty([0,1]^d)$ in probability.

Proof. Observe that H satisfies the conditions of Lemma 14. Invoke the functional delta method and the functional delta method for the bootstrap [17, Theorems 3.9.4 and 3.9.11] with the Hadamard differentiable map (8) applied to Lemma 15. Then, under mutual independence, the conditional distribution of the process

$$\sqrt{n}\left[\phi(H_n^*)(u) - \prod_{k=1}^{p} \phi(H_n)(u^{\{k\}})\right], \qquad u \in [0,1]^d,$$

given the data, converges to the same limiting distribution as that of

$$\sqrt{n}\left[\phi(H_n)(u) - \prod_{k=1}^{p} \phi(H)(u^{\{k\}})\right], \qquad u \in [0,1]^d,$$

in $\ell^\infty([0,1]^d)$ in probability. Now, it is well-known that

$$\phi(H)(u) = H(H_1^-(u_1), \ldots, H_d^-(u_d)) = H(u) = C(u), \qquad u \in [0,1]^d.$$

The result follows since, as shown for instance in [15, Lemma 1 and p 854], almost surely,

$$\sup_{u\in[0,1]^d} |\phi(H_n)(u) - C_n(u)| \le O(n^{-1}).$$

The same clearly holds for $\phi(H_n^*)$ and C_n^*. $\square$

Let I_n^* denote the version of I_n computed from the bootstrap sample, and $M_{A,n}^*$, $A \in \mathcal{P}_S$, be the version of $M_{A,n}$, $A \in \mathcal{P}_S$, respectively, computed from the bootstrap sample.

Proposition 17 *Suppose that C has continuous partial derivatives. Then, under mutual independence, the conditional distribution of the random variable I_n^* given the data, converges to the same limiting distribution as that of I_n in probability.*

Proof. Starting from Theorem 16, apply the functional delta method and the functional delta method for the bootstrap, both with the Hadamard differentiable map (5) (see Lemma 2). Conclude that under mutual independence among $\mathbb{X}_1, \ldots, \mathbb{X}_p$, the conditional distribution of the process $\sqrt{n}\mathcal{I}(C_n^*)(u)$, $u \in [0,1]^d$ given the data, converges to the same limiting distribution as that of $\sqrt{n}\mathcal{I}(C_n)(u)$, $u \in [0,1]^d$, in $\ell^\infty([0,1]^d)$ in probability. Finally, recall the characterization of convergence in probability in terms of every subsequence having a further subsequence that converges almost surely. For almost every data set, we may apply the continuous mapping theorem for the Cramér-von Mises functional to the appropriate subsequences to complete the proof. □

We have a similar result for the statistics obtained from the Möbius decomposition.

Proposition 18 *Suppose that C has continuous partial derivatives. Then, under mutual independence, the conditional distribution of the vector $\{M^*_{A,n} : A \in \mathcal{P}_S\}$ given the data, converges to the same limiting distribution as that of $\{M_{A,n} : A \in \mathcal{P}_S\}$ in probability.*

Proof. Starting from Theorem 16, apply the functional delta method and the functional delta method for the bootstrap, both with the Hadamard differentiable map $\vec{\mathcal{M}}_S$ defined at the beginning of the proof of Theorem 8, and proceed as above. □

3.5 Practical implementation of the tests

As mentioned in the previous subsection, in the studied context, bootstrap samples should be formed by sampling from the empirical c.d.f. $\prod_{k=1}^{p} F_n(x^{\{k\}})$, $x \in \mathbb{R}^d$. Once a bootstrap sample is obtained, the next step would involve computing the pseudo-observations from the ranks and then the various statistics under consideration. The practical inconvenience of this approach is that the ranks would not be unambiguously defined as ties can occur in the bootstrap sample. A simple way to resolve this issue consists of sampling independently but *without replacement* from the empirical marginal c.d.f. of each vector $\mathbb{X}_i$, $i \in S$, which amounts to adopting a *permutation* approach. From a theoretical perspective, as discussed in [17, p 371], although a proof of the analogue of Lemma 15 for the resulting *permutation independence process* appears to be unavailable at this point, it is likely that it has the same asymptotic behavior as the *bootstrap independence process* considered therein.

With the exception of the aforementioned permutation approach, the rest of this subsection presents straightforward adaptations of the practical solutions adopted in [4] to implement the "univariate" versions of the tests under consideration.

Approximate p-values for the test statistics: Denote by $(\mathbb{X}_{1,1}, \ldots, \mathbb{X}_{1,p})$, $\ldots$, $(\mathbb{X}_{n,1}, \ldots, \mathbb{X}_{n,p})$ the n available independent copies of the random vector $(\mathbb{X}_1, \ldots, \mathbb{X}_p)$. Let Q_n stand for I_n or $M_{A,n}$, $A \in \mathcal{P}_S$. An approximate p-value for Q_n can be obtained as follows:

(1) Let $Q_{n,0}$ be the value of Q_n computed from the original sample.
(2) Generate $N \times p$ random permutations $\sigma_{i,j}$, $i \in \{1, \ldots, N\}$, $j \in \{1, \ldots, p\}$, on $\{1, \ldots, n\}$. For any $i \in \{1, \ldots, N\}$, let $Q_{n,i}$ be the value of Q_n obtained from the sample $(\mathbb{X}_{\sigma_{i,1}(1),1}, \ldots, \mathbb{X}_{\sigma_{i,p}(1),p})$, $\ldots$, $(\mathbb{X}_{\sigma_{i,1}(n),1}, \ldots, \mathbb{X}_{\sigma_{i,p}(n),p})$.
(3) An approximate p-value for the test statistic is then

$$\frac{1}{N+1}\left\{\frac{1}{2} + \sum_{i=1}^{N} 1[Q_{n,i} \geq Q_{n,0}]\right\}.$$

Note that when all the random vectors under consideration have dimension one, i.e. when $p = d$, the permutation approach presented above is equivalent to the procedure used in [4, §4.4] for simulating under the null hypothesis.

Rejection region for the test based on the Möbius decomposition: For the test based on the Möbius decomposition of the independence empirical copula process, a rejection region is constructed as

$$\bigcup_{A \in \mathcal{P}_S} \{M_{A,n} \geq m_A\},$$

where m_A are critical values chosen to achieve an asymptotic global significance level α. As discussed in [4], it is convenient to choose these critical values such that, under independence,

$$P\left[\int_{[0,1]^d} \mathcal{M}'_{A,C}(\mathcal{G})(u)^2 \mathrm{d}u \geq m_A\right] = 1 - \beta, \qquad A \in \mathcal{P}_S,$$

where $\beta = (1-\alpha)^{1/(2^p - p - 1)}$. Approximate critical values can be obtained from the values of the test statistics computed from the randomized samples.

Dependogram: In the case of the test based on the Möbius decomposition of the independence empirical copula process, Genest and Rémillard [4] proposed a graphical representation of the values of the observed test statistics: for each subset $A \in \mathcal{P}_S$, a vertical bar is drawn whose height is proportional to the value of $M_{A,n}$. The approximate critical values m_A, $A \in \mathcal{P}_S$, are represented on the bars by black bullets. Subsets such that the bar exceeds the critical value can be considered as being composed of dependent vectors.

To illustrate the practical interest of a dependogram, we extend the example considered in [4, §4.2].

Let $\mathbb{Z} = (Z_1, Z_2)$ be a two-dimensional normal random vector whose components are standard normals such that $\mathrm{cor}(Z_1, Z_2) = 0.5$, and let $\mathbb{Z}' = (Z'_1, Z'_2)$ and $\mathbb{Z}'' = (Z''_1, Z''_2)$ be two independent copies of $\mathbb{Z} = (Z_1, Z_2)$. Let $\mathbb{Y} = (Y_1, Y_2, Y_3)$ be a three-dimensional normal random vector, whose components are standard normals such that $\mathrm{cor}(Y_i, Y_j) = 0.3$, $i \neq j$, and let $\mathbb{Y}' = (Y'_1, Y'_2, Y'_3)$ be an independent copy of $\mathbb{Y}$, where $\mathbb{Y}$ and $\mathbb{Y}'$ are

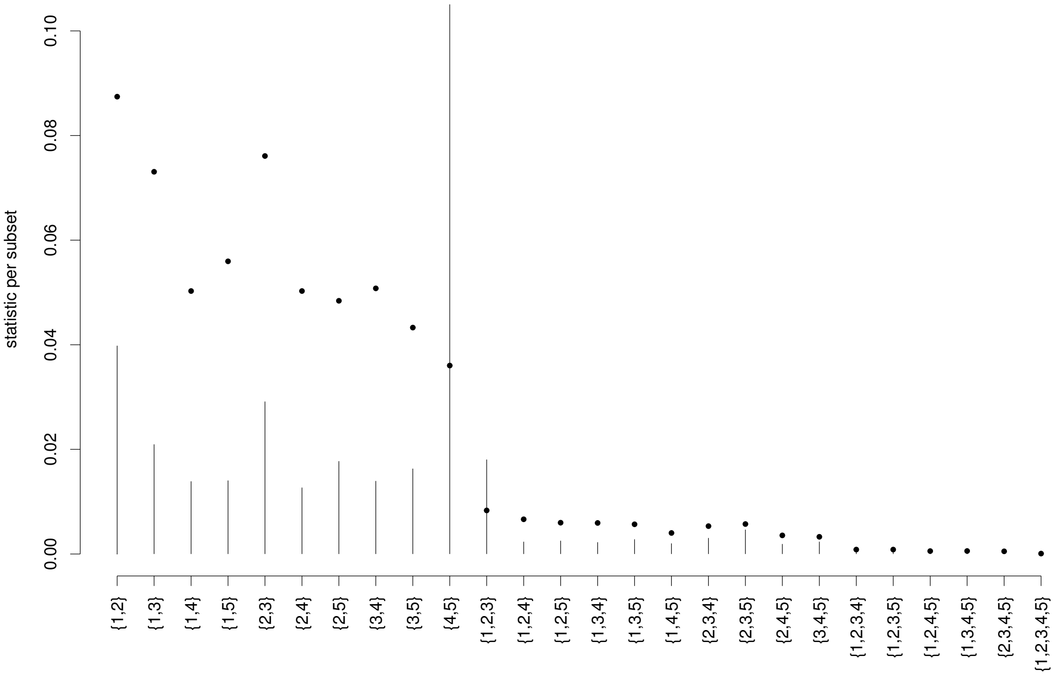

Fig. 1. Dependogram of asymptotic global level $\alpha = 5\%$ constructed from a sample of size $n = 100$ from the 12-dimensional random vector defined in the last but one paragraph of Subsection 3.5.

both independent of the $\mathbb{Z}$s. Now, define the two-dimensional random vector $\mathbb{X} = (X_1, X_2)$ by

$$X_i = |Z_i|\text{sign}(Z_1' Z_1''), \qquad \text{for } i = 1, 2,$$

and the three-dimensional random vector $\mathbb{T} = (T_1, T_2, T_3)$ by

$$T_i = Y_i + Y_i', \qquad \text{for } i = 1, 2, 3.$$

The random vectors $\mathbb{Y}$ and $\mathbb{T}$ are clearly not independent. Following [20], X_1, Z_1', Z_1'' are pairwise but not jointly independent. The same holds for X_2, Z_1', Z_1''. In fact, it can be shown that the random vectors $\mathbb{X}$, $\mathbb{Z}'$ and $\mathbb{Z}''$ are pairwise (but not jointly) independent [1].

To illustrate a test of independence among the random vectors $\mathbb{X}$, $\mathbb{Z}'$, $\mathbb{Z}''$, $\mathbb{Y}$

[1] To verify the equality $P[\mathbb{X} \in A, \mathbb{Z}' \in B] = P[\mathbb{X} \in A]P[\mathbb{Z}' \in B]$, write

$$A = \dot{\bigcup_{i,j \in \{-1,0,1\}}} A^{i,j}, \quad \text{and} \quad B = \dot{\bigcup_{k,l \in \{-1,0,1\}}} B^{k,l}$$

where $A^{i,j} = \{(x_1, x_2) \in A : \text{sign}(x_1) = i, \text{sign}(x_2) = j\}$ and similarly for B. Then, it suffices to prove the equality for each $A^{i,i}$ and $B^{k,l}$ with $i, k, l \neq 0$. Proceed by conditioning on the signs of Z_1' and Z_2'.

and $\mathbb{T}$, $n = 100$ realizations of the random vector $(\mathbb{X}, \mathbb{Z}', \mathbb{Z}'', \mathbb{Y}, \mathbb{T})$ were generated. The dependogram obtained from the observations is represented in Figure 1. As one can see, the pair $\{\mathbb{Y}, \mathbb{T}\}$ (denoted by $\{4, 5\}$ on the dependogram) exhibits a clear dependence. The same holds for the triplet $\{\mathbb{X}, \mathbb{Z}', \mathbb{Z}''\}$ (denoted by $\{1, 2, 3\}$ on the dependogram) which highlights the joint dependence among $\mathbb{X}$, $\mathbb{Z}'$, $\mathbb{Z}''$. Notice that the pairwise statistics (denoted by $\{1, 2\}$, $\{1, 3\}$ and $\{2, 3\}$ on the dependogram) are not significant. The approximate critical values were computed on the basis of $N = 1000$ randomized samples. The computation took approximately 10 seconds on a Pentium M 2.2 GHz processor.

Combining p-values: As discussed in [4], under independence, the p-values obtained from the statistics $M_{A,n}$, $A \in \mathcal{P}_S$, are approximately uniform on $[0, 1]$. From Corollary 11, it follows that, under independence, these p-values are also asymptotically mutually independent. This led Genest and Rémillard [4] to consider a global test of independence based on Fisher's p-value combination method. Additional combination rules were studied in [10]. In the rest of the paper, we restrict ourselves to the approaches proposed by Fisher [21] and Tippett [22]. From a practical perspective, the corresponding global p-values are obtained as follows:

(1) Let $M_{A,n,0}$, $A \in \mathcal{P}_S$, be the statistics computed from the original data.
(2) Generate N randomized samples from the original data and let $M_{A,n,i}$, $A \in \mathcal{P}_S$, be the statistics computed from the i-th sample.
(3) An approximate p-value for the statistic $M_{A,n,j}$, $A \in \mathcal{P}_S$, is then

$$\psi(M_{A,n,j}) = \frac{1}{N+1}\left\{\frac{1}{2} + \sum_{i=1}^{N} 1[M_{A,n,i} \geq M_{A,n,j}]\right\}, \qquad j \in \{0, 1, \ldots, N\}.$$

Next, for all $i \in \{0, 1, \ldots, N\}$, compute

$$W_{n,i} = -2 \sum_{A \in \mathcal{P}_S} \log\left[\psi(M_{A,n,i})\right] \qquad \text{and} \qquad T_{n,i} = \min_{A \in \mathcal{P}_S}\left[\psi(M_{A,n,i})\right].$$

(4) An approximate p-value for the global test *à la* Fisher (resp. *à la* Tippett) is then given by

$$\frac{1}{N}\sum_{i=1}^{N} 1[W_{n,i} \geq W_{n,0}] \qquad \left(\text{resp. } \frac{1}{N}\sum_{i=1}^{N} 1[T_{n,i} \leq T_{n,0}]\right).$$

4 Simulations

In order to investigate the finite-sample behavior of the different tests studied thus far, simulations were performed. More precisely, the dependence between three continuous r-dimensional random vectors $\mathbb{X} = (X_1, \dots, X_r)$, $\mathbb{Y} = (Y_1, \dots, Y_r)$, $\mathbb{Z} = (Z_1, \dots, Z_r)$ was investigated. To that end, the random vector $(X_1, \dots, X_r, Y_1, \dots, Y_r, Z_1, \dots, Z_r)$ of dimension $d = 3r$ was assumed to have a copula of the elliptical type, either the normal copula, or the t-copula with 2 degrees of freedom. These two copulas can be defined respectively as follows:

- The d-dimensional, $d \geq 2$, normal copula with $d \times d$ correlation matrix $\Sigma = (\rho_{ij})$ is defined by $F_\Sigma(F_N^{-1}(u_1), \dots, F_N^{-1}(u_d))$, $u \in [0,1]^d$, where F_N is the standard normal c.d.f. and

$$F_\Sigma(x) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_d} \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2} y^t \Sigma^{-1} y\right) \mathrm{d}y_1 \dots \mathrm{d}y_d, \qquad x \in \mathbb{R}^d.$$

- The d-dimensional, $d \geq 2$, t-copula with ν degrees of freedom and $d \times d$ correlation matrix $\Sigma = (\rho_{ij})$ is defined as $F_{\Sigma,\nu}(F_{t,\nu}^{-1}(u_1), \dots, F_{t,\nu}^{-1}(u_d))$, $u \in [0,1]^d$, where $F_{t,\nu}$ is the standard univariate t_ν c.d.f. and

$$F_{\Sigma,\nu}(x) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_d} \frac{\Gamma\left(\frac{\nu+d}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)(\pi\nu)^{d/2}|\Sigma|^{1/2}} \left(1 + \frac{y^t \Sigma^{-1} y}{\nu}\right)^{-\frac{\nu+d}{2}} \mathrm{d}y_1 \dots \mathrm{d}y_d.$$

For both copulas, the $d \times d$ correlation matrices Σ were structured as follows:

	X_1	$\dots$	X_r	Y_1	$\dots$	Y_r	Z_1	$\dots$	Z_r
X_1	1		ρ_{intra}						
$\vdots$					ρ_{inter}			ρ_{inter}	
X_r	ρ_{intra}		1						
Y_1				1		ρ_{intra}			
$\vdots$		ρ_{inter}						ρ_{inter}	
Y_r				ρ_{intra}		1			
Z_1							1		ρ_{intra}
$\vdots$		ρ_{inter}			ρ_{inter}				
Z_r							ρ_{intra}		1

where ρ_{inter} (resp. ρ_{intra}) controls the amount of dependence among (resp. within) the random vectors.

Both for the normal copula and the t copula with 2 degrees of freedom, we considered the values $0, 0.1, 0.2, 0.3, 0.4$ for ρ_{inter}, and 0 and 0.5 for ρ_{intra}. For each copula family, and each combination of ρ_{inter} and ρ_{intra}, we generated $R = 1000$ samples composed of $n = 100$ independent realizations of

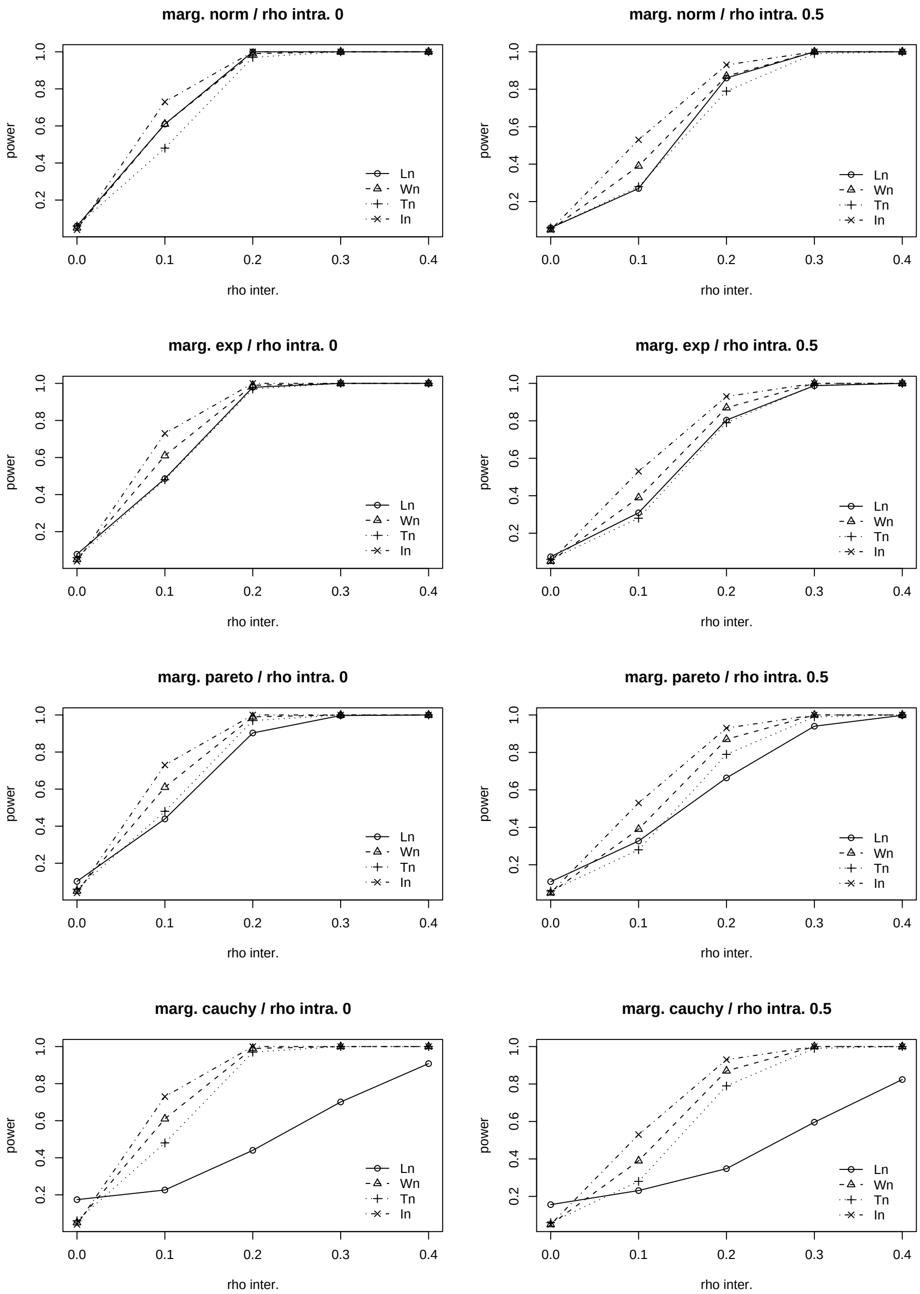

Fig. 2. Proportion of times that the different tests rejected independence at the 5 % significance level for the normal copula with $r = 2$ (i.e., $d = 6$). The statistic I_n is defined in (7), W_n is the test statistic *à la* Fisher, T_n is the test statistic *à la* Tippett, and L_n is the likelihood ratio test statistic.

$(X_1, \ldots, X_r, Y_1, \ldots, Y_r, Z_1, \ldots, Z_r)$ using the R `copula` package. Note that, in all the simulations, the number of randomized samples was set to $N = 1000$.

For $r = 2$ (i.e., $d = 6$), the proportion of times that the different tests rejected independence at the 5 % significance level is represented in Figure 2. Each row of graphs in Figure 2 corresponds to a different marginal c.d.f. As in [4], this was done to enable the comparisons of the studied tests with the likelihood ratio test [23], which is the most common procedure for checking independence, and is asymptotically optimal under multivariate normality. In the current context, the likelihood ratio test statistic can be written as

$$L_n = -n \log \left(\frac{|S_n|}{|S_{n,\mathbb{X}}||S_{n,\mathbb{Y}}||S_{n,\mathbb{Z}}|} \right),$$

where S_n is the sample covariance matrix computed from the data and $S_{n,\mathbb{X}}$, $S_{n,\mathbb{Y}}$, and $S_{n,\mathbb{Z}}$ are the "within vector" sample covariance matrices computed from the available realizations of $\mathbb{X}$, $\mathbb{Y}$ and $\mathbb{Z}$ respectively. The latter covariance matrices are clearly sub-matrices of S_n. Under mutual independence of $\mathbb{X}$, $\mathbb{Y}$ and $\mathbb{Z}$ with finite fourth moments, L_n is known to converge in distribution to a chi-square random variable with $3r^2$ degrees of freedom; see e.g. [24,25]. As in [4], four marginal c.d.f.s were considered: the standard normal, the standard exponential, the standard Cauchy and the Pareto $x \mapsto 1-x^{-4}$, $x \geq 1$. Of course the studied tests, being rank-based, are margin-free, which explains why their corresponding curves (within each column in Figure 2) are not affected by the margins.

Except for the likelihood ratio test, similar results are presented in Figure 3 for the t-copula with $d = 6$ (first row of graphs), for the normal copula with $d = 12$ (second row of graphs), and for the t-copula with $d = 12$ (third row of graphs).

The following observations can be made from Figures 2 and 3:

- In the setting under consideration, the test based on I_n appears to be always more powerful than the likelihood ratio test, which, in the multivariate normal case (first row of graphs in Figure 2), comes as a surprise.
- Except in the multivariate normal case, the test based on W_n always seems to outperform the likelihood ratio test, which is consistent with the results obtained in the "univariate" case by Genest and Rémillard [4].
- The rank-based tests appear to always hold their nominal level (verified from the rejection percentages under independence) which is not the case of the likelihood ratio test as can be seen for instance from the last row of graphs in Figure 2. The fact that the rejection percentages are clearly above 5 % in the first and third rows of graphs in Figure 3 for $\rho_{inter} = 0$ is due to the fact that t copulas cannot model independence even in this case.
- Among the empirical copula-based tests, the test statistic I_n seems to lead to the best results for $d = 6$ (i.e. $r = 2$), whereas W_n seems slightly more powerful for $d = 12$ (i.e. $r = 4$). Additional simulations seem to suggest that this behavior perdures as the dimension increases, and as the number of random vectors increases.
- The power of the tests appears to globally decrease as $|\rho_{\text{intra}}|$ increases.

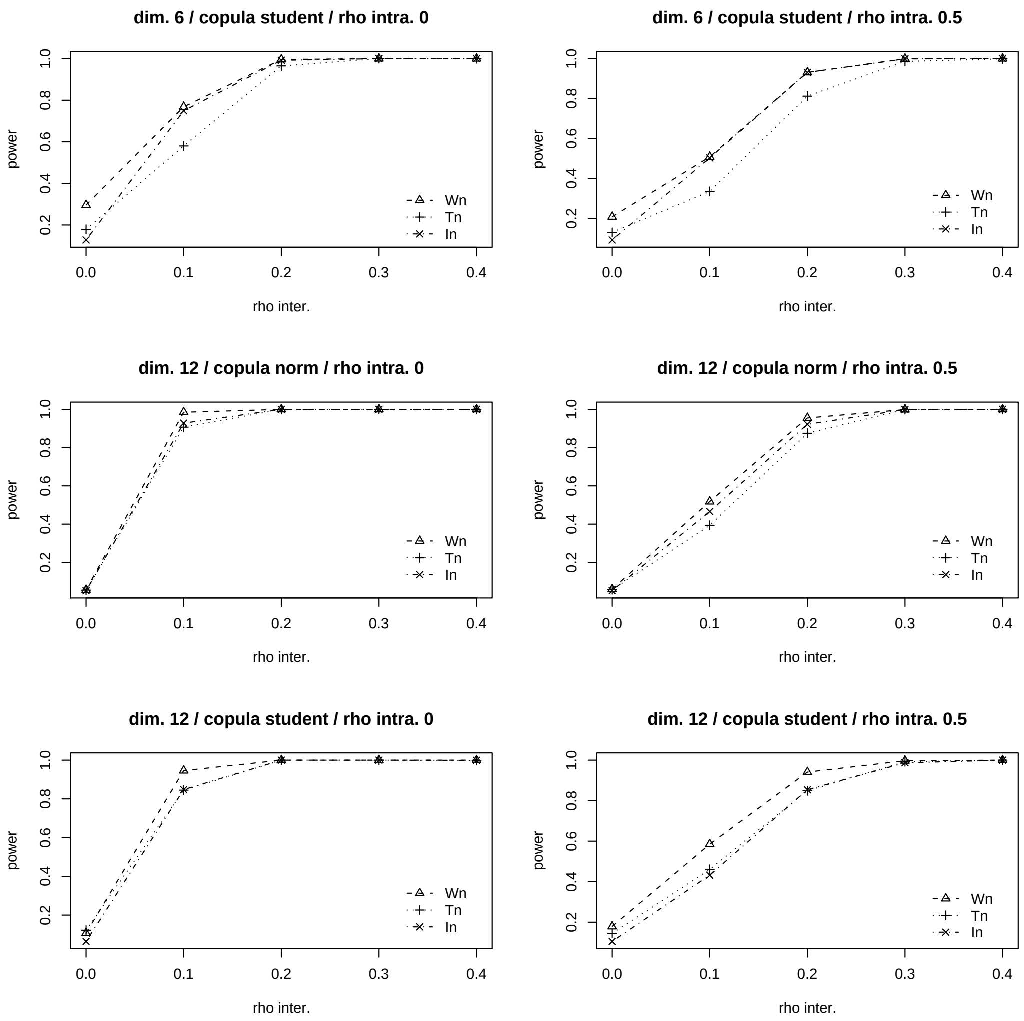

Fig. 3. Proportion of times that the different tests rejected independence at the 5 % significance level for the t-copula with $d = 6$ (first row of graphs), for the normal copula with $d = 12$ (second row of graphs), and for the t-copula with $d = 12$ (third row of graphs). The statistic I_n is defined in (7), W_n is the test statistic *à la* Fisher, and T_n is the test statistic *à la* Tippett.

Other usual copula families were not considered as most of them are not flexible enough to model simultaneously dependence *among* and *within* random vectors.

Acknowledgements

The authors are very grateful to the Associate Editor and three referees whose comments helped to improve the quality of this work.

References

[1] J. Blum, J. Kiefer, M. Rosenblatt, Distribution free tests of independence based on the sample distribution function, Annals of Mathematical Statistics 32 (1961) 485–498.

[2] D. Dugué, Sur les tests d'indépendance 'indépendants de la loi', Comptes rendus de l'Académie des Sciences de Paris, Série A 281 (1975) 1103–1104.

[3] P. Deheuvels, An asymptotic decomposition for multivariate distribution-free tests of independence, Journal of Multivariate Analysis 11 (1981) 102–113.

[4] C. Genest, B. Rémillard, Tests of independence and randomness based on the empirical copula process, Test 13 (2) (2004) 335–369.

[5] A. Sklar, Fonctions de répartition à n dimensions et leurs marges, Publ. Inst. Statist. Univ. Paris 8 (1959) 229–231.

[6] H. Joe, Multivariate models and dependence concepts, Chapman and Hall, London, 1997.

[7] D. Drouet-Mari, S. Kotz, Correlation and dependence, Imperial College Press, London, 2001.

[8] R. Nelsen, An introduction to copulas, Springer, New-York, 2006, second edition.

[9] 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 (1979) 274–292.

[10] C. Genest, J.-F. Quessy, B. Rémillard, Asymptotic local efficiency of Cramér-von Mises tests for multivariate independence, The Annals of Statistics 166–191 (2007) 35.

[11] G.-C. Rota, On the foundations of combinatorial theory. I. Theory of Möbius functions, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 2 (1964) 340–368 (1964).

[12] R. Beran, M. Bilodeau, P. L. de Micheaux, Nonparametric tests of independence between random vectors, Journal of Multivariate Analysis 98 (9) (2007) 1805–1824.

[13] I. Kojadinovic, J. Yan, Tests of serial independence for multivariate time series based on a Möbius decomposition of the independence empirical copula process, submitted.

[14] J. Yan, I. Kojadinovic, copula: Multivariate dependence with copulas, R package version 0.7-6 (2008).

[15] J.-D. Fermanian, D. Radulovic, M. Wegkamp, Weak convergence of empirical copula processes, Bernoulli 10 (5) (2004) 847–860.

[16] H. Tsukahara, Semiparametric estimation in copula models, The Canadian Journal of Statistics 33 (3) (2005) 357–375.

[17] A. van der Vaart, J. Wellner, Weak convergence and empirical processes, Springer, New York, 1996.

[18] K. Ghoudi, R. Kulperger, B. Rémillard, A nonparametric test of serial independence for times series and residuals, Journal of Multivariate Analysis 79 (2001) 191–218.

[19] P. Jing, L. Zhu, Some Blum-Kiefer-Rosenblatt type tests for the joint independence of variables, Communications in Statistics: Theory and Method 25 (9) (1996) 2127–2139.

[20] J. Romano, A. Siegel, Counterexamples in probability and statistics, Wadsworth, London, 1986.

[21] R. Fisher, Statistical methods for research workers, Olivier and Boyd, London, 1932.

[22] L. Tippett, The method of statistics, Williams and Norgate, London, 1931.

[23] S. Wilk, On the independence of k sets of normally distributed statistical variables, Econometrica 3 (1935) 309–326.

[24] J. Allaire, Y. Lepage, On a likelihood ratio test for independence, Statistis & Probability Letters 11 (1991) 449–452.

[25] S. Taskinen, H. Oja, R. Randles, Multivariate nonparametric tests of independence, Journal of the American Statistical Association 100 (471) (2005) 916–925.