

Testing for independence in arbitrary distributions

By C. GENEST, J. G. NEŠLEHOVÁ

*Department of Mathematics and Statistics, McGill University,
805, rue Sherbrooke ouest, Montréal, Québec, Canada H3A 0B9*
christian.genest@mcgill.ca johanna.neslehova@mcgill.ca

B. RÉMILLARD

*Department of Decision Sciences, HEC Montréal, 3000, chemin de la
Côte-Sainte-Catherine, Montréal, Québec, Canada H3T 2A4*
bruno.remillard@hec.ca

AND O. A. MURPHY

*Department of Mathematics and Statistics, McGill University,
805, rue Sherbrooke ouest, Montréal, Québec, Canada H3A 0B9*
orla.murphy@mail.mcgill.ca

SUMMARY

Statistics are proposed for testing the hypothesis that arbitrary random variables are mutually independent. The tests are consistent and well behaved for any marginal distributions; they can be used, for example, for contingency tables which are sparse or whose dimension depends on the sample size, as well as for mixed data. No regularity conditions, data jittering, or binning mechanisms are required. The statistics are rank-based functionals of Cramér–von Mises type whose asymptotic behaviour derives from the empirical multilinear copula process. Approximate p -values are computed using a wild bootstrap. The procedures are simple to implement and computationally efficient, and maintain their level well in moderate to large samples. Simulations suggest that the tests are robust with respect to the number of ties in the data, can easily detect a broad range of alternatives, and outperform existing procedures in many settings. Additional insight into their performance is provided through asymptotic local power calculations under contiguous alternatives. The procedures are illustrated on traumatic brain injury data.

Some key words: Independence test; Mixed data; Multilinear copula; Sparse contingency table; Wild bootstrap.

1. INTRODUCTION

The problem of testing the null hypothesis of mutual independence between the components of a random vector is old but important. Investigating relationships between variables is central to many scientific studies, and tests of independence typically precede any attempt at modelling association. The relevant literature is too vast to review here, but we observe that consistent tests are usually designed for cases where the variables are all of the same type.

Independence between categorical outcomes is often tested by classical tests such as Pearson's χ^2 or the likelihood ratio statistic (Agresti, 2013). When all variables are continuous, consistent tests can be based on empirical distribution functions (Blum et al., 1961; Deheuvels, 1980; Genest & Rémillard, 2004) or characteristic functions (Feuerverger, 1993; Meintanis & Iliopoulos, 2008; Fan et al., 2017); see, e.g., Rémillard (2011) for an overview and additional references. In practice, however, multivariate data often consist of a mix of discrete and continuous variables, and ties can occur even among the latter, for example when the variables are measured with limited precision. Often the data are divided into response variables and covariates, but before a model is developed, the issue of which variables are dependent or not should be investigated.

As a case in point, Gauthier et al. (2018) describe a study of cognitive communication impairment after traumatic brain injury in which various language tests were administered, before hospital discharge, to adults with this condition. One of their objectives was to explore connections between these tests in relation to the patient's age, years of education, and severity of injury on the Glasgow Coma Scale. Another goal was to assess the tests' ability to predict the patient's degree of disability upon discharge. The data, described in greater detail in § 3.3, are a mix of ordinal, discrete and continuous variables. Before attempting to model relations between the variables, it is of interest to explore potential associations using an omnibus procedure that can handle various types of data simultaneously and can be adjusted for multiple testing.

In this paper, we propose model-free tests of mutual independence that are applicable to any collection of variables measured at least on an ordinal scale. Some of the variables could be discrete while others are continuous, and still others might have distributions that are mixtures of the two types. The tests, which are consistent and powerful, do not involve any data binning or jittering. They remain valid even when the data are categorical and lead to sparse contingency tables or to tables whose dimension grows with the sample size.

2. MOTIVATION AND DESCRIPTION OF THE BASIC TEST STATISTIC

2.1. Motivational background and notation

Let $(X_{11}, \dots, X_{1d}), \dots, (X_{n1}, \dots, X_{nd})$ be a random sample from a vector (X_1, \dots, X_d) with distribution function H , and for each $k \in \{1, \dots, d\}$ let F_k denote the distribution function of X_k . By Sklar's theorem (Nelsen, 2006), there exists at least one distribution function C with standard uniform margins, i.e., a copula, such that

$$H(x_1, \dots, x_d) = C\{F_1(x_1), \dots, F_d(x_d)\}, \quad x_1, \dots, x_d \in \mathbb{R}. \quad (1)$$

When F_1, \dots, F_d are continuous, the joint distribution function of $F_1(X_1), \dots, F_d(X_d)$ is the only choice for C . Thus the null hypothesis $\mathcal{H}_0 : X_1 \perp\!\!\!\perp \dots \perp\!\!\!\perp X_d$ holds if and only if C is the independence copula given by $\Pi(u_1, \dots, u_d) = u_1 \times \dots \times u_d$ for $u_1, \dots, u_d \in [0, 1]$. A simple yet powerful rank-based test of \mathcal{H}_0 whose null asymptotic distribution is margin-free can then be based on the statistic

$$\|n^{1/2}(C_n - \Pi)\|_2^2 = \int_{[0,1]^d} n\{C_n(u_1, \dots, u_d) - \Pi(u_1, \dots, u_d)\}^2 d\Pi(u_1, \dots, u_d), \quad (2)$$

where $\|\cdot\|_2$ denotes the L_2 -norm and C_n is the empirical copula given by

$$C_n(u_1, \dots, u_d) = n^{-1} \sum_{i=1}^n 1\{F_{n1}(X_{i1}) \leq u_1, \dots, F_{nd}(X_{id}) \leq u_d\}, \quad u_1, \dots, u_d \in [0, 1],$$

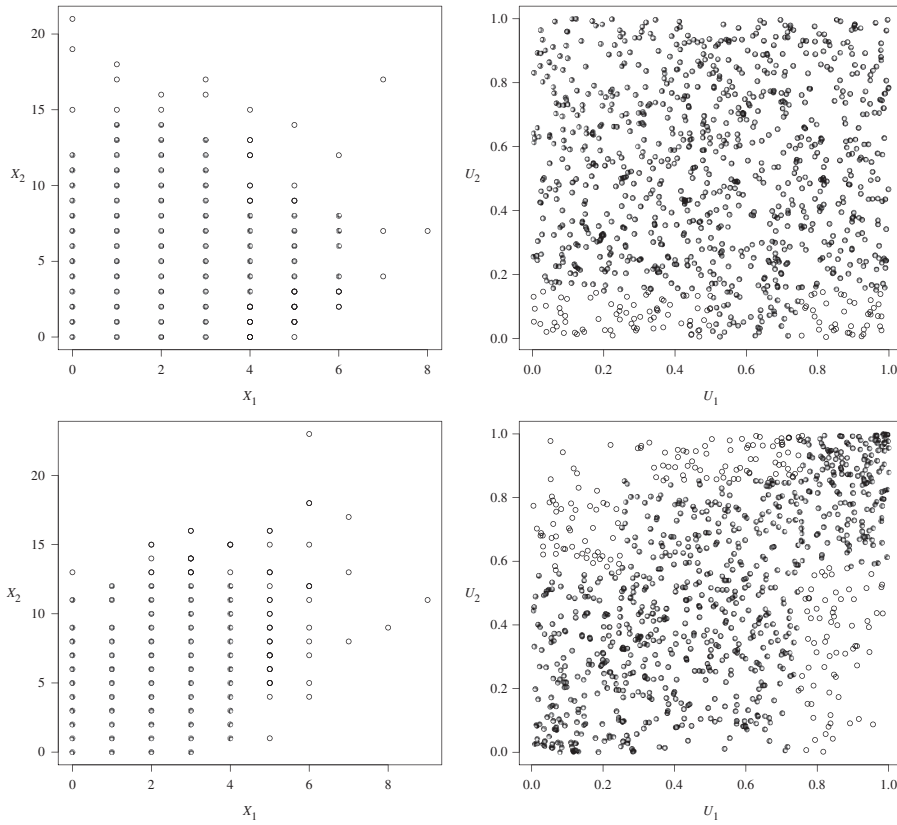


Fig. 1. Samples of size 1000 from (X_1, X_2) with $X_1 \sim \text{Po}(2)$ and $X_2 \sim \text{NeBi}(5, 0.5)$ (left column) when X_1 and X_2 are independent (top) or dependent (bottom); the pairs (U_1, U_2) from the checkerboard copula $C^{\mathbf{X}}$ are plotted in the right column.

in terms of the componentwise empirical distribution functions F_{n1}, \dots, F_{nd} . Here and throughout this paper, $1(B)$ stands for the indicator function of the set B . The statistic (2) was originally proposed by Deheuvels (1980); for details on the power, the asymptotic efficiency, and refinements of this test, see Genest & Rémillard (2004) and Genest et al. (2006, 2007), respectively.

When at least one of F_1, \dots, F_d is discontinuous, the copula C in Sklar's decomposition is unique only on $\text{ran}(F_1) \times \dots \times \text{ran}(F_d)$, but there exists an extension $C^{\mathbf{X}}$ on $[0, 1]^d$ that coincides with Π if and only if \mathcal{H}_0 holds (Genest & Nešlehová, 2007; Nešlehová, 2007). This checkerboard or multilinear extension copula, which can be traced back to Schweizer & Sklar (1974) and Moore & Spruill (1975), may be defined as follows based on Proposition 1 in Genest et al. (2017). In what follows, the notation $F(x_-)$ refers to the left limit of a univariate distribution function F at $x \in \mathbb{R}$, and $\Delta_F(x) = F(x) - F(x_-)$ with the convention that $F(-\infty_-) = 0$.

DEFINITION 1. Let (X_1, \dots, X_d) be a random vector with joint distribution H and univariate margins F_1, \dots, F_d . Also, let $(\Lambda_1, \dots, \Lambda_d)$ be a random vector with joint distribution Π independent of (X_1, \dots, X_d) . The checkerboard copula $C^{\mathbf{X}}$ of H is the distribution function of the vector (U_1, \dots, U_d) where, for $j \in \{1, \dots, d\}$, $U_j = F_j(X_{j-}) + \Lambda_j \Delta_{F_j}(X_j)$.

Figure 1 illustrates why $C^{\mathbf{X}}$ is useful as a tool for detecting dependence. The left column shows samples of size 1000 from (X_1, X_2) when X_1 is Poisson with mean 2 and X_2 is negative binomial with size 5 and $p = 0.5$; in the top panel the variables are independent, while in the

bottom panel they are linked via a Gaussian copula with correlation 0.5. The relationship between X_1 and X_2 is hard to assess. The right column of Fig. 1 displays samples from $C^{\mathbf{x}}$ obtained by transforming the original pairs (X_1, X_2) as described in Definition 1, using the known forms of F_1 and F_2 and additional independent samples of the same size from the randomizers Λ_1 and Λ_2 . The dependence, if any, is then manifest.

Of course, data jittering cannot be recommended for testing independence per se, as the added noise could decrease efficiency and will not lead to reproducible results in any case. Fortunately, an estimator $C_n^{\mathbf{x}}$ of $C^{\mathbf{x}}$ can be derived from data without involving randomization, and a global test of independence can be based on an analogue of (2) defined by

$$S_n = \|\mathbb{G}_n\|_2^2 = n \int_{[0,1]^d} \{C_n^{\mathbf{x}}(u_1, \dots, u_d) - \Pi(u_1, \dots, u_d)\}^2 d\Pi(u_1, \dots, u_d),$$

where $\mathbb{G}_n = n^{1/2}(C_n^{\mathbf{x}} - \Pi)$. We explain below how the empirical checkerboard copula $C_n^{\mathbf{x}}$ is constructed and can be used to carry out this test or design even better tests of independence.

2.2. Empirical checkerboard copula

To construct an estimator $C_n^{\mathbf{x}}$ of $C^{\mathbf{x}}$ that relies neither on any knowledge of the margins nor on any form of randomization or data jittering, Genest et al. (2014, 2017) used the fact that the empirical analogue H_n of H based on a random sample from (X_1, \dots, X_d) is a bona fide distribution function with discrete margins. The estimator $C_n^{\mathbf{x}}$ is then simply the checkerboard copula of H_n introduced in Definition 1. From Proposition 2 of Genest et al. (2017),

$$C_n^{\mathbf{x}}(u_1, \dots, u_d) = n^{-1} \sum_{i=1}^n \prod_{j=1}^d V_{F_{nj}}(X_{ij}, u_j), \quad u_1, \dots, u_d \in [0, 1],$$

where, for any $x \in \mathbb{R}$, $u \in [0, 1]$ and univariate distribution function F ,

$$V_F(x, u) = \lambda_F(u) 1\{x \leq F^{-1}(u)\} + \{1 - \lambda_F(u)\} 1\{x < F^{-1}(u)\}$$

with $F^{-1}(u) = \inf\{t \in \mathbb{R} : F(t) \geq u\}$ and

$$\lambda_F(u) = \begin{cases} [u - F\{F^{-1}(u)_-\}]/\Delta_F\{F^{-1}(u)\}, & \Delta_F\{F^{-1}(u)\} > 0, \\ 1, & \text{otherwise.} \end{cases} \quad (3)$$

As shown in Proposition 3 of Genest et al. (2017), $C_n^{\mathbf{x}}$ admits a density on $(0, 1)^d$ which is given below in terms of the probability mass function h_n of H_n . For any fixed $j \in \{1, \dots, d\}$, let $y_{1j} < \dots < y_{n_jj}$ be the ordered distinct values of $\{X_{1j}, \dots, X_{n_jj}\}$, with $y_{0j} = y_{1j} - 1$, and for each $i \in \{1, \dots, n_j\}$ set $f_{ij} = \Delta_{F_{nj}}(y_{ij})$. The Lebesgue density of $C_n^{\mathbf{x}}$ is then

$$c_n^{\mathbf{x}}(u_1, \dots, u_d) = \frac{h_n(y_{k_11}, \dots, y_{k_d d})}{f_{k_11} \times \dots \times f_{k_d d}} \quad (4)$$

whenever, for $j \in \{1, \dots, d\}$, $F_{nj}(y_{k_j-1j}) < u_j \leq F_{nj}(y_{k_jj})$ for some $k_j \in \{1, \dots, n_j\}$. Moreover, note that when H has continuous margins, $C_n^{\mathbf{x}}$ is asymptotically equivalent to the classical empirical copula C_n , as explained in Remark 2 of Genest et al. (2017).

2.3. Computation of S_n and link with classical statistics

An explicit expression for S_n can be derived from (4). By equation (A.1) in [Genest et al. \(2017\)](#) one has, for all $k \in \{1, \dots, d\}$ and $u \in [0, 1]$, $n^{-1}\{V_{F_{nk}}(X_{1k}, u) + \dots + V_{F_{nk}}(X_{nk}, u)\} = u$. For $k \in \{1, \dots, d\}$, also let I_k be an $n \times n$ symmetric matrix with (i, j) entry

$$\begin{aligned} I_{ijk} &= \int_0^1 V_{F_{nk}}(X_{ik}, u) V_{F_{nk}}(X_{jk}, u) du \\ &= \frac{1}{6} \sum_{\ell=1}^{n_k} \left\{ 2 \times 1(X_{ik} \leq y_{\ell k}) 1(X_{jk} \leq y_{\ell k}) + 1(X_{ik} \leq y_{\ell k}) 1(X_{jk} < y_{\ell k}) \right. \\ &\quad \left. + 1(X_{ik} < y_{\ell k}) 1(X_{jk} \leq y_{\ell k}) + 2 \times 1(X_{ik} < y_{\ell k}) 1(X_{jk} < y_{\ell k}) \right\} f_{\ell k}, \end{aligned} \quad (5)$$

whose i th row, and hence also the i th column average, is given by

$$I_{i \cdot k} = n^{-1} \sum_{j=1}^n I_{ijk} = \int_0^1 V_{F_{nk}}(X_{ik}, u) u du. \quad (6)$$

Then

$$S_n = n^{-1} \sum_{i=1}^n \sum_{j=1}^n \left(\prod_{k=1}^d I_{ijk} - \prod_{k=1}^d I_{i \cdot k} - \prod_{k=1}^d I_{j \cdot k} + 1/3^d \right).$$

In the special case of $d = 2$, this formula reduces to

$$S_n = n^{-1} \sum_{i=1}^n \sum_{j=1}^n \prod_{k=1}^2 (I_{ijk} - I_{i \cdot k} - I_{j \cdot k} + 1/3),$$

because $(I_{1 \cdot k} + \dots + I_{n \cdot k})/n = 1/3$. This can be contrasted with the standard χ^2 and likelihood ratio statistics for an $n_1 \times n_2$ contingency table with observed relative frequency $f_{k_1 k_2} = h_n(y_{k_1 1}, y_{k_2 2})$ in cell (k_1, k_2) , which, as noted by [Genest et al. \(2014\)](#), are given by

$$\begin{aligned} \chi_n^2 &= n \sum_{k_1=1}^{n_1} \sum_{k_2=1}^{n_2} \frac{(f_{k_1 k_2} - f_{k_1 \cdot} f_{\cdot k_2})^2}{f_{k_1 \cdot} f_{\cdot k_2}} = n \int_0^1 \int_0^1 \{c_n^{\mathfrak{F}}(u_1, u_2) - 1\}^2 du_2 du_1, \\ L_n &= 2n \sum_{k_1=1}^{n_1} \sum_{k_2=1}^{n_2} f_{k_1 k_2} \log \left(\frac{f_{k_1 k_2}}{f_{k_1 \cdot} f_{\cdot k_2}} \right) = 2n \int_0^1 \int_0^1 \log \{c_n^{\mathfrak{F}}(u_1, u_2)\} c_n^{\mathfrak{F}}(u_1, u_2) du_2 du_1. \end{aligned}$$

Thus, while χ_n^2 and L_n are based on $c_n^{\mathfrak{F}}$, which consists of cell-by-cell ratios of observed to expected frequencies, S_n and its variants described below are functionals of the cumulative distribution function $C_n^{\mathfrak{F}}$, which also exploits the natural ordering in the values of the variables. As will be seen in § 4, this approach leads to a more powerful test in cases that the χ_n^2 and L_n statistics were designed to handle. More importantly still, the asymptotic null distribution of $C_n^{\mathfrak{F}}$ exists whatever the marginal distributions of X_1, \dots, X_d may be. The new procedures can therefore be used to test for independence between any set of discrete, continuous or mixed variables, even in cases where the traditional statistics are known to perform poorly ([Agresti, 2013](#)).

2.4. Limiting null distribution of S_n and computation of critical values

It follows from the work of [Genest et al. \(2017\)](#) that a test of \mathcal{H}_0 based on S_n is consistent and that the limiting null distribution of S_n exists. This is stated below and proved in the Appendix.

THEOREM 1. *If $C^{\mathfrak{X}} \neq \Pi$, then for any $M > 0$, $\text{pr}(S_n > M) \rightarrow 1$ as $n \rightarrow \infty$. If $C^{\mathfrak{X}} = \Pi$, then as $n \rightarrow \infty$, $S_n \rightarrow S = \|\mathbb{G}\|_2^2$ in law, where \mathbb{G} is a centred Gaussian process given in (A2).*

As the limiting process \mathbb{G} is intricate, the distribution of S is unwieldy. More importantly, it depends on the unknown margins F_1, \dots, F_d unless they are continuous. To carry out the test based on S_n , one must therefore resort to resampling. The wild bootstrap approximation of the limiting process \mathbb{G} proposed by [Genest et al. \(2017\)](#) can be used to this end. It extends the work of [Rémillard & Scaillet \(2009\)](#) and [Bücher & Dette \(2010\)](#) beyond the case of continuous margins.

Algorithm 1. Fix a large integer B , and for each $b \in \{1, \dots, B\}$ proceed as follows:

Step 1. Generate a random sample $\xi_{1b}, \dots, \xi_{nb}$ from some distribution with mean 0, variance 1, and fourth moment at most 3, e.g., $\mathcal{N}(0, 1)$. Then set $\bar{\xi}_b = (\xi_{1b} + \dots + \xi_{nb})/n$.

Step 2. For any $u_1, \dots, u_d \in [0, 1]$ and $j \in \{1, \dots, d\}$, let

$$\mathbb{G}_{nb}(u_1, \dots, u_d) = n^{-1/2} \sum_{i=1}^n (\xi_{ib} - \bar{\xi}_b) \left\{ \prod_{j=1}^d V_{F_{nj}}(X_{ij}, u_j) - \sum_{j=1}^d V_{F_{nj}}(X_{ij}, u_j) \prod_{k \neq j} u_k \right\}.$$

Step 3. Set $S_{nb} = \|\mathbb{G}_{nb}\|_2^2$, which can be explicitly computed as

$$S_{nb} = n^{-1} \sum_{i=1}^n \sum_{j=1}^n (\xi_{ib} - \bar{\xi}_b)(\xi_{jb} - \bar{\xi}_b) J_{ij},$$

where J_{ij} , which does not depend on b , is given by

$$J_{ij} = \prod_{k=1}^d I_{ijk} - \sum_{\alpha=1}^d I_{ij\alpha} \prod_{k \neq \alpha} I_{i \cdot k} - \sum_{\alpha=1}^d I_{ij\alpha} \prod_{k \neq \alpha} I_{j \cdot k} + \frac{1}{3^{d-1}} \sum_{\alpha=1}^d I_{ij\alpha} + \frac{1}{3^{d-2}} \sum_{\alpha \neq \beta} I_{i \cdot \alpha} I_{j \cdot \beta}.$$

Once the B bootstrap replicates have been computed,

$$p_{nB} = B^{-1} \sum_{b=1}^B 1(S_{nb} > S_n)$$

is an approximate p -value of the test of \mathcal{H}_0 based on S_n .

The validity of the above wild bootstrap approximation relies on Proposition 8 of [Genest et al. \(2017\)](#), as detailed in the Appendix, where the following result is also proved.

PROPOSITION 1. *If $C^{\mathfrak{X}} = \Pi$, then for any $\alpha \in (0, 1)$, $\text{pr}(p_{nB} < \alpha) \rightarrow (\lceil \alpha B \rceil - 1)/(B + 1)$ as $n \rightarrow \infty$, which in turn converges to α as $B \rightarrow \infty$. Here, $\lceil x \rceil$ denotes the smallest integer greater than or equal to x . If $C^{\mathfrak{X}} \neq \Pi$, then for all $B \in \mathbb{N}$, $\text{pr}(p_{nB} = 0) \rightarrow 1$ as $n \rightarrow \infty$.*

3. ALTERNATIVE PROCEDURES BASED ON THE MÖBIUS DECOMPOSITION

3.1. Möbius decomposition

Based on a decomposition of the empirical copula process \mathbb{C}_n derived from the Möbius combinatorial formula (Spitzer, 1974), Deheuvels (1981) and Genest & Rémillard (2004) proposed and studied alternative tests of \mathcal{H}_0 which are, in the continuous case, more powerful than the test based on S_n (Genest et al., 2007). A similar approach is developed below.

For each set $A \in \mathcal{A}_d = \{A \subseteq \{1, \dots, d\} : \text{card}(A) = |A| > 1\}$, define the process

$$\mathbb{G}_{An}(u_1, \dots, u_d) = n^{-1/2} \sum_{i=1}^n \prod_{k \in A} \{V_{F_{nk}}(X_{ik}, u_k) - u_k\},$$

which depends only on u_k with $k \in A$. When $C^{\mathfrak{A}} = \Pi$, one has

$$\mathbb{G}_n(u_1, \dots, u_d) = \sum_{A \in \mathcal{A}_d} \mathbb{G}_{An}(u_1, \dots, u_d) \prod_{k \in A^c} u_k,$$

which is called the Möbius decomposition of the process \mathbb{G}_n . The usefulness of the latter for testing independence can be established as in Ghoudi et al. (2001). To see how, consider a vector (U_1, \dots, U_d) with distribution $C^{\mathfrak{A}}$. For arbitrary $A \in \mathcal{A}_d$, set

$$\mu_A(u_1, \dots, u_d) = E \left[\prod_{k \in A} \{1(U_k \leq u_k) - u_k\} \right]. \quad (7)$$

When $d = 2$, \mathcal{A}_d contains only the set $A = \{1, 2\}$ and $\mu_A \equiv 0$ is equivalent to independence between X_1 and X_2 . When $d > 2$, mutual independence between the variables $\{X_k : k \in A\}$ for a given set A continues to imply that $\mu_A \equiv 0$, but not the other way around. However, Proposition 2.1 of Ghoudi et al. (2001) states that $C^{\mathfrak{A}} = \Pi$ if and only if $\mu_A \equiv 0$ for all $A \in \mathcal{A}_d$.

3.2. Computation of test statistics and an asymptotic approximation of their critical values

For any $A \in \mathcal{A}_d$, a test of the null hypothesis $\mathcal{H}_A : \mu_A \equiv 0$ can be based on the statistic

$$S_{An} = \|\mathbb{G}_{An}\|_2^2 = n^{-1} \sum_{i=1}^n \sum_{j=1}^n \prod_{k \in A} (I_{ijk} - I_{i \cdot k} - I_{jk} + 1/3),$$

where the same notation as in (5) and (6) is used. The following result, whose proof is given in the Appendix, explores the asymptotic properties of the statistics S_{An} .

THEOREM 2. *If $C^{\mathfrak{A}} \neq \Pi$, then there exists $A \in \mathcal{A}_d$ such that $\text{pr}(S_{An} > M) \rightarrow 1$ as $n \rightarrow \infty$. If $C^{\mathfrak{A}} = \Pi$, then as $n \rightarrow \infty$, the vector $(S_{An} : A \in \mathcal{A}_d)$ converges in law to a vector $(S_A : A \in \mathcal{A}_d)$ of mutually independent random variables, where for each $A \in \mathcal{A}_d$, $S_A = \|\mathcal{M}_A(\mathbb{G})\|_2^2$ with \mathbb{G} being the centred Gaussian process defined in (A2) and \mathcal{M}_A the operator defined in (A6).*

As was the case for S , the distribution of S_A is unwieldy and depends on the unknown margins F_1, \dots, F_d unless all of them are continuous. To carry out a test based on the statistics S_{An} , therefore, one must again rely on resampling, as described below.

Algorithm 2. Fix a large integer B , and for each $b \in \{1, \dots, B\}$ proceed as follows:

Step 1. Generate a random sample $\xi_{1b}, \dots, \xi_{nb}$ from some distribution with mean 0, variance 1, and fourth moment at most 3, e.g., $\mathcal{N}(0, 1)$. Then set $\bar{\xi}_b = (\xi_{1b} + \dots + \xi_{nb})/n$.

Step 2. For every $A \in \mathcal{A}_d$, let

$$\mathbb{G}_{Anb}(u) = n^{-1/2} \sum_{i=1}^n (\xi_{ib} - \bar{\xi}_b) \prod_{k \in A} \{V_{F_{nk}}(X_{ik}, u_k) - u_k\}.$$

Step 3. For every $A \in \mathcal{A}_d$, set $S_{Anb} = \|\mathbb{G}_{Anb}\|_2^2$, which can be explicitly computed as

$$S_{Anb} = n^{-1} \sum_{i=1}^n \sum_{j=1}^n (\xi_{ib} - \bar{\xi}_b)(\xi_{jb} - \bar{\xi}_b) \prod_{k \in A} (I_{ijk} - I_{i \cdot k} - I_{j \cdot k} + 1/3).$$

Once the B bootstrap replicates have been computed, one has that for any $A \in \mathcal{A}_d$,

$$p_{AnB} = B^{-1} \sum_{b=1}^B 1(S_{Anb} > S_{An})$$

is an approximate p -value of the test of $\mathcal{H}_A : \mu_A \equiv 0$ based on S_{An} .

The validity of this algorithm is proved in the Appendix, along with the following result.

PROPOSITION 2. *If $C^{\boxtimes} = \Pi$, then for all $B \in \mathbb{N}$, $\{p_{AnB} : A \in \mathcal{A}_d\}$ converge jointly in law, as $n \rightarrow \infty$, to mutually independent copies of a uniform random variable on $\{0, 1/B, \dots, B/B\}$. If $C^{\boxtimes} \neq \Pi$, then for at least one $A \in \mathcal{A}_d$ and all $B \in \mathbb{N}$, $\text{pr}(p_{AnB} = 0) \rightarrow 1$ as $n \rightarrow \infty$.*

This result can be exploited to derive tests of \mathcal{H}_0 that are both easier to compute than S_n and, as we shall see, more powerful. Following [Genest et al. \(2007\)](#), we propose the Fisher statistic

$$T_n = -2 \sum_{A \in \mathcal{A}_d} \log p_{AnB}.$$

The resulting test is then consistent by Proposition 2. From the work of [Littell & Folks \(1973\)](#), T_n is asymptotically Bahadur-optimal and one can reject \mathcal{H}_0 at asymptotic level α if T_n exceeds the upper α th quantile of the χ^2 distribution with $2(2^d - d - 1)$ degrees of freedom.

As the computation of T_n may become prohibitive for large d , one could also compute

$$T_{nK} = -2 \sum_{A \in \mathcal{A}_d, |A| \leq K} \log p_{AnB}$$

for some suitable $K \in \{2, \dots, d\}$. Under \mathcal{H}_0 , T_{nK} is approximately χ^2 with

$$2\{d/2\} + \dots + 2\{d/K\}$$

degrees of freedom. The test based on T_{nK} is not consistent for \mathcal{H}_0 unless $K = d$. However, it is consistent for the hypothesis of pairwise independence for all $K \in \{2, \dots, d\}$.

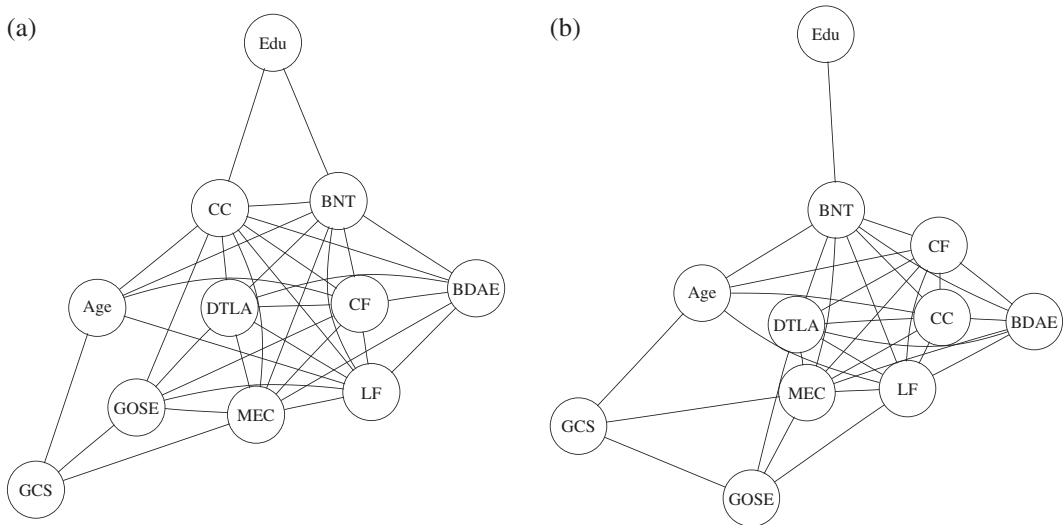


Fig. 2. Graphs of pairwise associations between patient's age, Age, education, Edu, severity of injury, GCS, level of disability at discharge, GOSE, and results on seven language tests: the Boston Diagnostic Aphasia Examination, BDAE, the Boston Naming Test, BNT, the Chapman–Cook Speed of Reading Test, CC, the Categorical Fluency test, CF, the Detroit Test of Learning Aptitude, DTLA, the Protocole Montréal d'évaluation de la communication, MEC, and the Literal Fluency test, LF. The two graphs were obtained (a) without and (b) with the Benjamini–Hochberg adjustment for false discovery rate; edges indicate tests of pairwise independence that were significant at the 5% level.

When the hypothesis of mutual independence is rejected based on S_n , T_n or T_{nK} , it may be of interest to determine which subsets of the variables led to this conclusion. By analogy with the correlogram, one could construct and plot a dependogram, as proposed by [Genest & Rémillard \(2004\)](#). In it the p -values p_{An} are displayed together with a horizontal line at level α . This way, one can identify the subsets of variables that are significantly dependent at that level. However, as discussed in § 3.1, the fact that $p_{An} > \alpha$ for some A does not imply that all variables in that subset are independent unless $|A| = 2$; all subsets of A of size 2 and above must also be checked.

3.3. Data illustration

[Gauthier et al. \(2018\)](#) investigated communication impairment following a traumatic brain injury. Of interest is the impact of the severity of a patient's injury on performance in a number of language tests administered during the hospital stay, as well as the results of these tests as an indicator of the level of disability upon discharge, as measured on the Glasgow Outcome Scale-Extended. The severity of the injury was measured on the 13-point Glasgow Coma Scale. The language tests considered here are listed in the caption of Fig. 2 and detailed in the Supplementary Material. The clinicians suspect that not all tests are equally informative; also of interest to them is the effect of the patient's age and years of education. Complete data were available for 135 adult patients with traumatic brain injury.

Based on the statistics S_n and T_{n3} with $B = 1000$ multiplier bootstrap replicates, the null hypothesis of mutual independence between these 11 variables was rejected at the 0.1% level. This result is neither surprising nor instructive. Because $d = 11$, \mathcal{A}_d contains 2036 sets and the dependogram is unmanageable. However, the statistics S_{nA} with $|A| = 2$ and their p -values can be used to explore pairwise dependence. The results can then be visualized with a graph, as in Fig. 2, where each node represents a variable. An edge joins two nodes if and only if the association

between the two corresponding variables is significant at the 5% level. In panel (a), the p -values are raw, i.e., unadjusted for multiple testing, while in panel (b), the adjustment of [Benjamini & Hochberg \(1995\)](#) was used to control the false discovery rate.

Focusing on Fig. 2(b), one can see that the seven language tests are mutually dependent. However, they are not equally indicative of disability status at discharge; nor are they influenced by age, education and severity of injury in the same way. Age affects the performance in confrontation naming as measured by the Boston Naming Test, literal and categorial fluency, and speed of reading. This is in accordance with the conclusion of [Gauthier et al. \(2018\)](#) and references therein that older patients perform these tasks with greater difficulty. Age is also related to injury severity, which may be explained by the cause of injury. The fact that education is related only to some of the tests is also documented but not yet elucidated from a clinical point of view; here, it is associated only with the Boston Naming Test after adjusting for the false discovery rate. More importantly, only the Protocole Montréal is associated with the severity of the injury, while the degree of disability at discharge is related to performance in the Protocole Montréal, the Detroit Test of Learning Aptitude, and the Literal Fluency test. These findings vindicate the use of the Protocole Montréal, which is easily administered at bedside, and may help clinicians to make a selection among the above tests, which was one of the study objectives.

4. NUMERICAL EXPERIMENTS

4.1. *Experimental design*

The good performance of the tests based on S_n and T_n has already been documented for copula models with continuous margins ([Genest & Rémillard, 2004](#); [Genest et al., 2006, 2007](#)). The case of two-way contingency tables was discussed in [Genest et al. \(2014\)](#), where the test based on S_n was often found to be significantly more powerful than the χ^2 test, even when the critical values of the latter are obtained by Monte Carlo to compensate for its well-known liberality. The Supplementary Material includes additional simulation results that show the good behaviour of S_n as a test of independence on contingency tables, when compared to the χ^2 , G^2 and Zelterman statistics, as well as to the test of [Bakirov et al. \(2006\)](#).

In this section, we study the tests based on S_n and T_n when some margins are discrete and others are continuous or mixed. In all cases, $N = 1000$ simulation runs were conducted and the nominal level of the tests was $\alpha = 5\%$. The number of bootstrap replicates was $B = 1000$ throughout.

4.2. *The $d = 2$ case*

When $d = 2$, $S_n = T_n$. Although the statistic is rank-based, its limiting distribution in Theorem 1 depends on the marginal distributions if the latter are discontinuous ([Genest et al., 2017](#)). The behaviour of the test based on S_n is therefore expected to depend on the marginals, the sample size n , and the nature of dependence. Five marginals were considered in the study: a Poisson with mean 1, which produces a large number of ties; a Poisson with mean 20, which is discrete but produces fewer ties than a Poisson with mean 1; a rounded Pareto with survival function $1/(k+1)^{1/3}$ for $k \in \{0, 1, \dots\}$, which is discrete and heavy-tailed and has no expectation; a standard Cauchy, which is continuous and heavy-tailed and also has no expectation; and a Student t with three degrees of freedom, having an atom at 0 with mass 0.05, which is a mixture of a continuous and a discrete distribution; its mean is zero but its variance does not exist.

We first investigate the level of the test based on S_n when the margins are identical. Table 1 reports the estimated size of the test and the system time when run on a Thinkpad W530 using Matlab parallel processing with six processors. The nominal level is well maintained when $n >$

Table 1. *Simulated results under the null hypothesis: the reported values are average rejection rates (%) for the test statistic $S_n = T_n$ when $d = 2$ and all margins are equal, based on 1000 multiplier samples and 1000 permutation samples*

n	20		50		100	250
Margins	M	P	M	P	M	M
P1	7.7	6.5	6.6	5.5	5.2	5.1
P20	6.1	4.5	5.7	4.1	5.0	5.0
RP	6.2	4.3	5.7	4.3	5.2	4.9
CA	6.1	4.0	5.7	4.2	5.2	4.7
SA	6.0	4.1	5.5	4.2	4.9	4.7
Time (s)	452	3465	514	5394	691	1561

M, multiplier samples; P, permutation samples; P1, Poisson with mean 1; P20, Poisson with mean 20; RP, rounded Pareto; CA, standard Cauchy; SA, mixture of Student t_3 and an atom at 0.

50. The test is slightly liberal in small samples, but this issue can be resolved by computing the p -values using the permutation approach. However, the solution becomes computationally prohibitive as n increases. This motivates the use of the wild bootstrap proposed here.

In the bivariate case, the performance of S_n could be compared with many alternative procedures. We limited ourselves to tests that can handle any type of marginal distributions and therefore considered: a test of [Hoeffding \(1948\)](#) as implemented in the `hoeffd` procedure of the R ([R Development Core Team, 2019](#)) package `Hmisc`; a test of [Székely et al. \(2007\)](#) as implemented in the `indep.test` procedure with `method = 'dcov'` in the R package `energy`; a test based on an extension of Spearman's ρ due to [Genest et al. \(2013\)](#); and a test of [Genest & Rémillard \(2004\)](#) applied after breaking the tied ranks at random. The test of [Heller et al. \(2016\)](#) was found to be too slow to run for a large-scale simulation study.

As reported in the Supplementary Material, all tests hold their nominal level reasonably well for all combinations of margins when $n \in \{100, 250\}$, except for the test of [Hoeffding \(1948\)](#), which is conservative in the presence of many ties. This issue with Hoeffding's test persists even when $n = 250$. As $N = 1000$, the margin of error is about 2%.

For power comparisons, we first considered all combinations of margins and various copulas in (1) with different degrees of dependence as measured by Kendall's τ of the underlying copula. This way, the effect of the marginal distributions and the overall association can be studied separately. As expected, the power of all tests increases with n and τ . In Table 2, we therefore only report the results for $n = 100$ and $\tau = 0.1$ when the copula in (1) is Clayton, Gaussian, Cauchy or Student t_3 ; see, e.g., [McNeil et al. \(2015\)](#) for definitions of these families of copulas. Results for other choices of n , τ , and copula alternatives are in the Supplementary Material.

The power of the test based on S_n depends on the copula, but only slightly. More interestingly, the power is fairly stable across all choices of margins, regardless of whether they are discrete, continuous, mixed, or heavy-tailed. It is mildly negatively affected by the large number of ties in at least one margin. Data jittering cannot be recommended because it systematically reduces power and is too costly if repeated over a large number of jitters to ensure reproducibility.

Compared to S_n , Hoeffding's test is more dependent on the choice of margins and performs poorly when there are many ties in at least one margin. When there are few or no ties, it performs similarly to the test based on S_n , and does better when the copula is Cauchy.

Table 2. Power (%) for four dependence structures with $\tau = 0.1$ and $n = 100$

Copula		Clayton					Gaussian				
F_1	F_2 Test	Sn	JSn	H	S	R	Sn	JSn	H	S	R
P1	P1	22.9	18.0	4.2	20.1	22.0	25.3	19.6	4.7	25.1	25.3
	P20	27.2	24.2	10.9	24.8	27.0	27.6	24.2	12.6	26.5	27.6
	RP	26.9	23.7	11.2	7.0	26.3	27.6	24.0	13.3	13.1	27.4
	CA	26.1	24.0	12.8	15.8	26.6	27.3	23.5	14.0	18.3	28.2
	SA	26.6	24.0	12.5	25.8	26.4	27.2	24.1	14.1	27.1	28.2
P20	P20	30.1	29.0	24.9	32.1	33.3	29.0	28.1	24.7	30.1	31.3
	RP	30.0	28.4	24.8	5.2	31.9	29.2	28.1	24.3	10.1	30.3
	CA	29.3	28.3	27.7	28.2	32.7	28.9	29.3	27.3	20.9	31.3
	SA	29.5	28.7	27.3	35.7	32.9	28.8	29.0	27.2	29.0	31.1
RP	RP	28.5	28.2	25.8	5.7	31.5	28.2	28.0	24.8	12.0	30.8
	CA	29.0	28.3	27.5	4.2	32.5	28.9	27.9	26.6	8.5	31.3
	SA	29.0	28.2	27.4	4.0	32.4	28.8	27.9	26.7	8.9	31.4
CA	CA	28.7	29.0	30.3	20.2	33.2	29.1	28.8	28.9	9.5	31.4
	SA	29.3	28.8	30.0	26.1	33.2	28.8	28.3	28.8	15.9	31.6
SA	SA	29.5	28.7	29.9	38.4	33.4	29.2	28.4	28.8	29.6	31.4
Copula		Cauchy					Student t_3				
F_1	F_2 Test	Sn	JSn	H	S	R	Sn	JSn	H	S	R
P1	P1	22.4	21.0	9.6	48.0	25.7	24.4	19.5	6.3	31.6	26.6
	P20	24.1	24.1	18.3	56.1	24.7	25.5	22.3	14.4	34.8	29.2
	RP	24.2	24.2	18.6	48.9	25.3	25.6	21.8	14.3	33.8	29.5
	CA	24.5	24.0	19.4	72.9	25.1	26.1	22.5	15.4	43.0	29.4
	SA	24.3	24.2	19.4	71.6	24.9	26.0	21.7	15.6	40.6	29.4
P20	P20	29.7	30.9	39.4	75.4	28.5	28.0	27.4	29.4	40.9	31.3
	RP	30.0	30.1	37.5	78.2	28.0	28.6	28.1	29.8	40.6	30.6
	CA	30.0	30.6	42.0	98.6	28.6	29.3	28.0	32.1	64.7	30.7
	SA	29.9	30.7	41.4	92.2	28.6	29.3	28.1	32.1	51.2	30.8
RP	RP	29.3	29.4	37.4	52.2	28.4	28.8	28.3	29.7	34.9	30.2
	CA	29.7	29.8	40.7	77.2	28.6	29.1	28.4	31.5	38.2	30.6
	SA	29.6	29.9	40.3	81.9	28.7	29.1	28.5	31.2	40.6	30.6
CA	CA	30.2	30.2	45.9	99.6	28.8	29.5	28.2	33.2	63.9	31.3
	SA	30.2	30.1	45.6	99.5	28.9	29.3	28.4	33.4	67.6	31.2
SA	SA	30.2	30.4	45.2	98.1	29.0	29.2	28.9	33.8	64.2	31.4

P1, Poisson with mean 1; P20, Poisson with mean 20; RP, rounded Pareto; CA, standard Cauchy; SA, mixture of Student t_3 and an atom at 0; Sn, test based on S_n ; JSn, test of [Genest & Rémillard \(2004\)](#) based on jittered data; H, test of [Hoeffding \(1948\)](#); S, test of [Székely et al. \(2007\)](#); R, test of [Genest et al. \(2013\)](#) based on Spearman's ρ .

The test of [Genest et al. \(2013\)](#) based on Spearman's ρ has stable power across margins. It is often slightly more powerful than the test based on S_n for the copulas in Table 2, but it is not consistent and will perform poorly if Spearman's ρ of the joint distribution is zero or close to it. An example is the tent map copula for which $\rho = 0$, as detailed in the Supplementary Material.

The test of [Székely et al. \(2007\)](#) is much more sensitive to the choice of copula and margins, regardless of the proportion of ties. When the copula is Clayton or Gaussian, the lack of existence of the mean for Cauchy or rounded Pareto margins leads to a considerable loss of power. When the copula is Cauchy or Student t_3 , however, this test does much better than that based on S_n , and the nonexistence of the mean appears not to be much of an issue. This is possibly because the Cauchy and Student t_3 margins do not reduce to Π when $\tau = 0$. Instead, when $\tau = 0$, samples

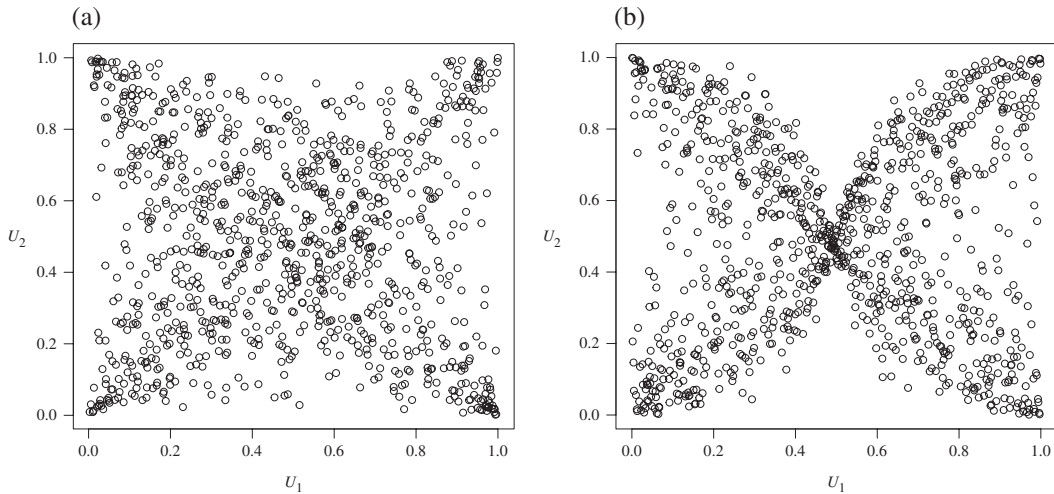


Fig. 3. Samples of size 1000 from (a) the Cauchy copula with $\tau = 0$ and (b) the copula of (X_1, ZX_1) where X_1 and Z are $\mathcal{N}(0, 1)$.

Table 3. Power (%) for regression alternatives with $n = 100$

Factor X test	Normal					Binomial				
	Sn	JSn	H	S	R	Sn	JSn	H	S	R
P1	100.0	100.0	100.0	100.0	11.0	100.0	99.8	88.2	100.0	9.2
P20	16.7	16.3	14.8	31.2	5.4	35.3	32.7	74.3	37.4	5.2
RP	100.0	100.0	100.0	100.0	12.5	100.0	100.0	100.0	91.9	7.9
CA	31.4	31.8	85.9	100.0	12.0	17.7	17.6	65.0	97.2	11.2
SA	26.7	27.1	71.9	100.0	11.5	17.5	16.9	56.0	99.6	11.1
GA	20.1	20.2	47.7	96.7	10.9	16.4	16.0	50.0	97.4	11.3

P1, Poisson with mean 1; P20, Poisson with mean 20; RP, rounded Pareto; CA, standard Cauchy; SA, mixture of Student t_3 and an atom at 0; GA, normal with mean 0 and variance 1; Sn, test based on S_n ; JSn, test of [Genest & Rémillard \(2004\)](#) based on jittered data; H, test of [Hoeffding \(1948\)](#); S, test of [Székely et al. \(2007\)](#); R, test of [Genest et al. \(2013\)](#) based on Spearman's ρ .

from these copulas are often clustered along both diagonals of the unit square. This phenomenon, sometimes referred to as arachnitude, is illustrated in Fig. 3(a).

Finally, following [Bakirov et al. \(2006\)](#), we considered regression alternatives of the form $X_2 = X_1Z$ with X_1 either being $\mathcal{N}(0, 1)$ or having one of the five previously mentioned margins. The error Z was taken to be either $\mathcal{N}(0, 1)$ or centred binomial, namely $Z + 5 \sim \text{Bi}(10, 0.5)$. The law of X_2 and its dependence on X_1 are induced by the distributions of X_1 and Z ; samples from (X_1, X_2) are shown in the Supplementary Material. From Table 3 it can be seen that the test of [Székely et al. \(2007\)](#) does much better than the test based on S_n when $n = 100$ and X_1 is normal, Cauchy, or t_3 with an atom at 0. As illustrated in Fig. 3(b), the underlying dependence again exhibits arachnitude. These regression alternatives are also examples of situations where the test of [Genest et al. \(2013\)](#) performs very poorly because $\rho(X_1, X_2) = 0$.

4.3. The $d > 2$ case

When $d > 2$, the tests based on S_n and T_n differ. As it is computationally infeasible to consider all possible combinations of the margins described in § 4.2, all margins were taken to be identical.

Table 4. *Estimated sizes (%) for test statistics T_n and S_n when $d = 5$ and all margins are equal, based on 1000 multiplier samples and 1000 permutation samples*

n	20				50				75				100		250	
	M		P		M		P		M		P		M		M	
Margins	Tn	Sn	Tn	Sn	Tn	Sn	Tn	Sn	Tn	Sn	Tn	Sn	Tn	Sn	Tn	Sn
P1	15.0	7.7	8.6	5.2	6.7	7.2	8.0	4.9	6.9	5.7	8.7	5.6	4.9	5.7	5.8	4.5
P20	9.4	4.3	8.6	5.2	5.2	5.6	7.5	5.6	4.2	3.9	8.0	4.9	3.5	4.9	5.5	4.6
RP	10.1	5.0	8.1	5.6	5.4	4.3	6.0	6.5	4.9	4.2	7.5	3.9	4.1	4.2	5.3	4.5
CA	8.8	4.2	8.9	5.3	4.0	5.0	8.1	5.3	4.1	3.5	8.3	4.8	3.0	5.1	5.1	4.0
SA	8.3	4.1	8.6	5.1	4.4	5.1	7.5	5.3	4.4	3.5	7.5	5.0	3.2	5.0	4.9	4.0
Time (h)	1.59		2.37		1.74		7.22		2.14		8.90		2.39		5.29	

M, multiplier samples; P, permutation samples; P1, Poisson with mean 1; P20, Poisson with mean 20; RP, rounded Pareto; CA, standard Cauchy; SA, mixture of Student t_3 and an atom at 0; Sn, test based on S_n ; Tn, test based on T_n .

The first observation to be made is that the tests based on S_n and T_n are again liberal in small samples; this issue worsens as d increases. Table 4 reports the levels of both tests when $d = 5$. As in the bivariate case, one can use the permutation approach when n is small at the cost of increased computational burden; the system times reported in Table 4 were obtained using the same personal laptop computer as in § 4.2. The overall recommendation based on this table is that when $n > 75$, the multiplier method performs well and is preferred.

In the power study, we considered the tests based on S_n and T_n as well as the test based on T_n when the cardinality of the subsets is bounded above by $K = 2$. Furthermore, we ran these tests on a jittered sample obtained by breaking the tied ranks at random. As competitors, we used the tests based on ρ_{n1}^* and W_n from Genest et al. (2013), which are multivariate generalizations of the tests of bivariate independence based on Spearman's ρ . The statistic W_n is constructed in a similar manner to T_n by combining p -values of tied-corrected statistics stemming from a decomposition of Spearman's ρ using the Möbius combinatorial formula.

The observed power when $n = 100$ is summarized in Table 5 for $d = 5$ and four exchangeable copulas whose pairwise value of Kendall's τ is 0.1. These copulas are extensions of those considered in Table 2. Results for other values of n and τ , $d = 3$, and the Frank, Gumbel, Farlie–Gumbel–Morgenstern and tent map copulas are in the Supplementary Material.

As in the $d = 2$ case, the tests based on S_n and T_n perform similarly across margins; their power can be slightly affected when there are many ties. Nor is jittering to be recommended when $d > 2$ either. For the alternatives in Table 5, the power of the tests increases with dimension d because all subsets of variables of size at least 2 contribute evidence against \mathcal{H}_0 . In contrast, the power drops with d for Farlie–Gumbel–Morgenstern copulas, as only the full set exhibits dependence.

Which one of the tests based on S_n and T_n is superior depends on the copula. For example, the test based on S_n is preferable when the copula is Clayton, Gaussian or Student t_3 , whereas with few exceptions the test based on T_n has higher power for Cauchy, Frank or Gumbel alternatives.

Truncating the statistic T_n by considering subsets of size $K = 2$ generally leads to a rise in power. However, this test performs poorly when the variables are pairwise but not mutually independent; the Farlie–Gumbel–Morgenstern alternatives are again a case in point.

The tests of Genest et al. (2013) based on Spearman's ρ are generally, but not always, more powerful than those based on S_n and T_n . However, this is not the case when $\rho \approx 0$. An illustration is again provided by the tent map copula, for which both tests of Genest et al. (2013) perform very poorly, in stark contrast to those based on S_n and T_n .

Table 5. Power (%) when $d = 5$ for four dependence structures with $\tau = 0.1$ and $n = 100$

Copula	Margins Test	Sn	Tn	TnT	JSn	JTn	JTnT	R	Wn
Clayton	P1	68.8	54.4	69.5	56.5	42.3	58.5	84.4	47.1
	P20	71.5	67.9	83.6	71.2	66.8	83.9	93.5	75.3
	RP	72.8	67.4	81.4	69.9	65.0	81.4	93.0	71.2
	CA	70.4	67.9	83.3	70.3	68.2	83.3	93.8	75.2
	SA	70.2	67.5	83.4	69.9	68.4	83.5	93.8	75.5
Cauchy	P1	82.1	62.0	80.1	68.7	50.3	69.1	84.1	93.4
	P20	78.6	83.0	98.0	78.5	81.9	97.4	83.2	99.2
	RP	82.9	81.0	96.9	80.9	80.2	96.3	83.0	98.6
	CA	78.6	82.9	97.7	79.4	83.5	97.5	83.1	99.2
	SA	78.7	83.0	97.9	79.4	83.8	97.5	83.2	99.2
Student t_3	P1	81.7	58.8	72.9	68.3	47.9	66.1	88.0	81.8
	P20	77.6	66.2	81.7	76.8	65.7	81.4	90.2	92.9
	RP	80.0	66.6	81.3	78.0	64.5	81.4	89.6	91.9
	CA	78.3	66.7	82.2	77.8	66.8	82.9	90.6	93.4
	SA	77.9	67.0	82.3	77.9	66.6	82.4	90.7	93.6
Gaussian	P1	82.5	61.6	75.1	68.6	49.4	65.9	90.3	60.3
	P20	82.4	67.4	81.7	80.5	66.2	81.6	93.3	71.6
	RP	83.6	68.5	82.5	81.7	66.8	82.1	92.9	71.7
	CA	82.4	67.8	82.5	81.7	67.9	82.6	93.0	72.3
	SA	82.4	68.1	82.2	81.7	68.2	82.5	93.1	72.3

P1, Poisson with mean 1; P20, Poisson with mean 20; RP, rounded Pareto; CA, standard Cauchy; SA, mixture of Student t_3 and an atom at 0; Sn, test based on S_n ; Tn, test based on T_n ; TnT, test based on T_n and subsets of cardinality at most 2; JSn, JTn and JTnT, tests based on jittered data; R, test of [Genest et al. \(2013\)](#) based on Spearman's ρ ; Wn, test based on W_n of [Genest et al. \(2013\)](#).

5. CONTIGUOUS ALTERNATIVES

As a complement to the numerical results presented in § 4, we now look at the limiting distribution of the stochastic processes \mathbb{G}_n and \mathbb{G}_{An} for all $A \in \mathcal{A}_d$, under a sequence of contiguous alternatives. To define the latter, we will proceed as in [Genest et al. \(2007\)](#). Let $\Theta \subset \mathbb{R}$ be an open interval, and let $\{C_\theta : \theta \in \Theta\}$ be a family of copulas such that for a unique $\theta_0 \in \Theta$, $C_{\theta_0} = \Pi$. Next, assume that the following regularity conditions hold.

Condition 1. C_θ is absolutely continuous and its density c_θ has a square-integrable right derivative $\dot{c}_\theta = \partial c_\theta(u)/\partial \theta$ at $\theta = \theta_0$ for each $(u_1, \dots, u_d) \in (0, 1)^d$, with the latter satisfying

$$\lim_{n \rightarrow \infty} \int_{(0,1)^d} \left(n^{1/2} [\{c_{\theta_n}(u_1, \dots, u_d)\}^{1/2} - 1] - \delta \dot{c}_{\theta_0}(u_1, \dots, u_d)/2 \right)^2 du_d \cdots du_1 = 0.$$

Condition 2. For every $(u_1, \dots, u_d) \in (0, 1)^d$,

$$\dot{C}_{\theta_0}(u_1, \dots, u_d) = \lim_{\theta \rightarrow \theta_0} \frac{\partial}{\partial \theta} C_\theta(u_1, \dots, u_d) = \int_0^{u_1} \cdots \int_0^{u_d} \dot{c}_{\theta_0}(v_1, \dots, v_d) dv_d \cdots dv_1.$$

Let $\delta_n \rightarrow \delta \in \mathbb{R}$ be such that $\theta_n = \theta_0 + n^{-1/2}\delta_n \in \Theta$ for n sufficiently large. Define Q_n as the joint distribution of the random sample $(U_{11n}, \dots, U_{1dn}), \dots, (U_{n1n}, \dots, U_{ndn})$ of

size n from C_{θ_n} , and let C_n^{\boxtimes} be the empirical checkerboard copula corresponding to the sample $(F_1^{-1}(U_{11n}), \dots, F_d^{-1}(U_{1dn}), \dots, (F_1^{-1}(U_{n1n}), \dots, F_d^{-1}(U_{ndn}))$. Also, let P_n be the joint distribution of the random sample of the same size from the d -variate independence copula Π .

When Conditions 1 and 2 hold, Q_n is contiguous with respect to P_n . The regularity conditions are the same as in Genest et al. (2007) and do not concern the marginal distributions. As detailed in the Appendix, the following result can then be derived from LeCam's Third Lemma (van der Vaart & Wellner, 1996). In what follows, $C([0, 1]^d)$ denotes the space of continuous real-valued functions defined on $[0, 1]^d$ endowed with the supremum norm $\|\cdot\|$.

THEOREM 3. *Let $C = \{C_\theta : \theta \in \Theta\}$ be a family of copulas such that $C_{\theta_0} = \Pi$ for a unique $\theta_0 \in \Theta$, and suppose that Conditions 1 and 2 hold. Then under Q_n , as $n \rightarrow \infty$, the empirical process \mathbb{G}_n converges weakly in $C([0, 1]^d)$ to $\mathbb{G} + \delta\mathbb{M}_F(\dot{C}_{\theta_0})$, where \mathbb{M}_F is as defined in (A1). Furthermore, under Q_n , the processes \mathbb{G}_{A_n} with $A \in \mathcal{A}_d$ jointly converge weakly in $C([0, 1]^d)$ to $\mathbb{G}_A + \delta\mathbb{M}_F(q_A)$, where $q_A = \mathcal{M}_A(\dot{C}_{\theta_0})$ and \mathcal{M}_A is as defined in (A6).*

Formulas for q_A are given in § 3 of Genest et al. (2007) for various families of multivariate copulas. In the exchangeable or equicorrelated Gaussian case, for instance, one has

$$q_A(u_1, \dots, u_d) = 1(|A| = 2) \prod_{j \in A} \varphi\{\Phi^{-1}(u_j)\}, \quad (u_1, \dots, u_d) \in (0, 1)^d,$$

where Φ denotes the $\mathcal{N}(0, 1)$ cumulative distribution function and φ the corresponding density. Thus, within that class of copulas, only the pairwise interactions are significant in a neighbourhood of independence. The same phenomenon occurs in the Clayton family of copulas, which plays a key role in survival analysis (Clayton, 1978). For this model, one has

$$q_A(u_1, \dots, u_d) = 1(|A| = 2) \prod_{j \in A} u_j \log(u_j), \quad (u_1, \dots, u_d) \in (0, 1)^d.$$

This explains why in Table 5, the tests based on the truncated statistics are significantly more powerful than the corresponding tests based on the nontruncated versions of these two classes of alternatives. In contrast, for the Frank and Gumbel families of Archimedean copulas, one has that for all $(u_1, \dots, u_d) \in (0, 1)^d$,

$$q_A(u_1, \dots, u_d) = \frac{1}{2} \prod_{j \in A} (u_j^2 - u_j),$$

$$q_A(u_1, \dots, u_d) = -\left(\prod_{j \in A} u_j\right) \sum_{B \subset A} (-1)^{|A \setminus B|} \left(\sum_{j \in B} \log u_j\right) \log\left(-\sum_{j \in B} \log u_j\right),$$

respectively. In these models, therefore, it is essential to look at all possible subsets of variables. As $|A|$ increases, the terms in the expression for q_A decrease in absolute value for the Frank family but not for the Gumbel family. This explains why, as reported in the Supplementary Material, the tests based on the truncated statistics have comparable, and sometimes even lower, power than the nontruncated versions of the Frank and Gumbel alternatives.

Finally, for the Farlie–Gumbel–Morgenstern family of copulas, one finds that

$$q_A(u_1, \dots, u_d) = 1(|A| = d) (-1)^d \prod_{j=1}^d u_j (1 - u_j), \quad (u_1, \dots, u_d) \in (0, 1)^d.$$

So unless the full set is included in the statistic, one cannot expect any test to have significant power for these alternatives, as already noted in § 4.3.

Under the conditions of Theorem 3, it easily follows that under Q_n , as $n \rightarrow \infty$, S_n and the collection of $(S_{An} : A \in \mathcal{A}_d)$ converge jointly in law to $\|\mathbb{G} + \delta \mathbb{M}_F(\dot{C}_{\theta_0})\|_2^2$ and $(\|\mathbb{G}_A + \delta \mathbb{M}_F(q_A)\|_2^2 : A \in \mathcal{A}_d)$, respectively. In principle this result can be used to compute the asymptotic local power of the tests considered here. To do this in full generality is a much more complicated endeavour than in Genest et al. (2006, 2007), however, because the limiting distributions depend on the marginals if the latter are discontinuous. Therefore, here we only consider the local power of the test based on S_n when $d = 2$ with binomial margins. The binomial distribution $\text{Bi}(p, N)$ converges to a continuous distribution as $N \rightarrow \infty$ while having a finite support, which makes calculations easier.

Example 1. Suppose $d = 2$ and that for each $j \in \{1, 2\}$, F_j is binomial $\text{Bi}(p_j, N_j)$. From (A4), note that under Q_n , $\mathbb{G}_n \rightarrow \mathbb{M}_F(\mathbb{B}_\delta)$ weakly, where $\mathbb{B}_\delta = \mathbb{B} + \delta \dot{C}_{\theta_0}$ with \mathbb{B} given by (A3). Consequently, under Q_n , $S_n \rightarrow \|\mathbb{M}_F(\mathbb{B}_\delta)\|_2^2$ in law. The limit can be evaluated explicitly,

$$\begin{aligned} \|\mathbb{M}_F(\mathbb{B}_\delta)\|_2^2 &= \frac{1}{9} \sum_{k_1=0}^{N_1} \sum_{k_2=0}^{N_2} \Delta_{F_1}(k_1) \Delta_{F_2}(k_2) \times \left[\Omega_{k_1, k_2}^2 + \Omega_{k_1-1, k_2}^2 + \Omega_{k_1, k_2-1}^2 + \Omega_{k_1-1, k_2-1}^2 \right. \\ &\quad + \Omega_{k_1, k_2} \Omega_{k_1-1, k_2} + \Omega_{k_1, k_2} \Omega_{k_1, k_2-1} + \Omega_{k_1-1, k_2} \Omega_{k_1-1, k_2-1} \\ &\quad \left. + \Omega_{k_1, k_2-1} \Omega_{k_1-1, k_2-1} + \Omega_{k_1, k_2} \Omega_{k_1-1, k_2-1} / 2 + \Omega_{k_1-1, k_2} \Omega_{k_1, k_2-1} / 2 \right], \end{aligned}$$

where, for any $k_1, k_2 \in \mathbb{N}$, $\Omega_{k_1, k_2} = \mathbb{B}_\delta\{F_1(k_1), F_2(k_2)\}$. Furthermore, as is well known, the process \mathbb{B} admits the Karhunen–Loève expansion

$$\mathbb{B}(u_1, u_2) = \sum_{\ell_1=1}^{\infty} \sum_{\ell_2=1}^{\infty} \frac{1}{\ell_1 \ell_2 \pi^2} \sin(\ell_1 \pi u_1) \sin(\ell_2 \pi u_2) Z_{\ell_1, \ell_2}, \quad (8)$$

where $\{Z_{\ell_1, \ell_2} : \ell_1, \ell_2 \in \mathbb{N}\}$ are mutually independent $\mathcal{N}(0, 1)$ random variables (Shorack & Wellner, 1984, p. 213). Using this expansion, the asymptotic critical values $c_F(\alpha)$ and the asymptotic local power $\pi(\delta) = \text{pr}\{\|\mathbb{M}_F(\mathbb{B}_\delta)\|_2^2 > c_F(\alpha)\}$ can be determined by simulation.

Table 6 shows the results for 1000 Monte Carlo repetitions when ℓ_1 and ℓ_2 are truncated at 50 in (8) and the marginal parameters are $p_1 = p_2 = 0.5$ and $N_1, N_2 \in \{2, 5, 50\}$. As alternatives, we chose the Clayton, Gaussian and Farlie–Gumbel–Morgenstern copula families parameterized by Kendall’s τ to ensure comparability. The results largely confirm the findings from the simulation study in § 4.2 reported in Table 2. The power rises with δ , as expected; more importantly, it is fairly homogeneous across the three alternatives. Also visible is the phenomenon of the asymptotic power decreasing with the degree of discreteness; it is lowest when $N_1 = N_2 = 2$. The loss of power is, however, rather small. Therefore, we can conclude that the test based on S_n is very robust with respect to the choice of margins and successfully corrects for ties in the data.

ACKNOWLEDGEMENT

The authors thank Dr Elaine de Guise for permission to use the brain injury data and for insightful discussions. Thanks are also due to the editor, associate editor and referees for comments that have improved the manuscript. This work was partially supported by the Canada Research

Table 6. *Asymptotic local power of the test based on S_n when $d = 2$ for three local alternatives; the results are based on 1000 Monte Carlo samples and $\ell_1, \ell_2 \leq 50$ in (8)*

Copula	N_1	$N_2 \mid \delta$	0.0	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0
Clayton	2	2	4.5	8.3	22.6	44.9	70.0	87.4	96.7	99.4	99.9
		5	5.6	10.5	27.3	53.6	78.0	93.2	98.5	99.9	100.0
		50	6.0	11.5	27.1	54.5	79.9	93.3	99.0	99.6	100.0
	5	5	6.2	10.6	28.6	55.6	80.4	93.9	99.0	99.8	100.0
		50	4.5	10.1	27.4	54.7	80.2	94.6	99.1	99.8	99.9
	50	50	5.8	11.9	30.5	59.0	83.2	96.4	99.5	99.9	100.0
	Gaussian	2	4.5	8.6	21.8	43.2	67.8	85.1	95.8	98.9	100.0
		5	5.6	10.5	27.0	52.3	75.1	90.4	97.6	99.8	100.0
		50	6.0	11.3	25.8	51.8	76.8	91.6	97.9	99.9	100.0
FGM	2	2	4.5	8.6	21.8	42.9	67.7	85.0	95.6	98.9	100.0
		5	5.6	10.8	27.3	53.1	75.7	90.8	97.8	99.8	100.0
		50	6.0	11.4	26.2	52.7	78.2	92.4	98.2	99.9	100.0
	5	5	6.2	10.6	28.9	54.8	78.7	92.0	98.5	99.8	100.0
		50	4.5	10.8	28.1	54.9	79.3	93.2	98.9	99.8	100.0
	50	50	5.8	12.5	31.1	59.8	82.5	95.4	99.4	99.9	100.0
	FGM	2	4.5	8.6	21.8	42.9	67.7	85.0	95.6	98.9	100.0
		5	5.6	10.8	27.3	53.1	75.7	90.8	97.8	99.8	100.0
		50	6.0	11.4	26.2	52.7	78.2	92.4	98.2	99.9	100.0

FGM, Farlie–Gumbel–Morgenstern copula family.

Chairs Program, the Natural Sciences and Engineering Research Council of Canada, the Canadian Institute of Statistical Sciences, and the Fonds de recherche du Québec-Nature et technologies.

SUPPLEMENTARY MATERIAL

Supplementary material available at *Biometrika* online includes results for two-way contingency tables, additional simulation results, and material for the data application.

APPENDIX

Multilinear interpolation

Given univariate distribution functions F_1, \dots, F_d , set $F = (F_1, \dots, F_d)$. For each $j \in \{1, \dots, d\}$, define the range of F_j by $\mathcal{R}_{F_j} = \{F_j(y) : y \in \mathbb{R}\} \subseteq [0, 1]$. For arbitrary $u = (u_1, \dots, u_d) \in [0, 1]^d$ and $S \subseteq \{1, \dots, d\}$, let $u_{F,S}$ refer to the $d \times 1$ vector whose j th component $(u_{F,S})_j$ is given, for $j \in \{1, \dots, d\}$, by $F_j\{F_j^{-1}(u_j)\}$ if $j \in S$ and $F_j\{F_j^{-1}(u_j)\}_-$ otherwise. Note that $u_{F,S}$ is an element in the closure $\bar{\mathcal{R}}_F$ of $\mathcal{R}_F = \mathcal{R}_{F_1} \times \dots \times \mathcal{R}_{F_d}$, and that if F_j is continuous at $F_j^{-1}(u_j)$, then $(u_{F,S})_j = u_j$ for any S , given that $F_j\{F_j^{-1}(u_j)\}_- = F_j\{F_j^{-1}(u_j)\} = u_j$. Further, let $\ell^\infty(K)$ be the collection of bounded real-valued functions on $K \subseteq [0, 1]^d$. Finally, recalling the definition of λ_F given in (3), one can define the multilinear interpolation operator \mathbb{M}_F , for $f \in \ell^\infty(\bar{\mathcal{R}}_F)$ and $u_1, \dots, u_d \in [0, 1]$, by

$$\mathbb{M}_F(f)(u_1, \dots, u_d) = \sum_{S \subseteq \{1, \dots, d\}} f(u_{F,S}) \left\{ \prod_{j \in S} \lambda_{F_j}(u_j) \right\} \left[\prod_{j \notin S} \{1 - \lambda_{F_j}(u_j)\} \right]. \quad (\text{A1})$$

Asymptotic behaviour of the empirical multilinear process

For convenience, we provide a combined restatement of Theorem 1 and Corollary 1 from [Genest et al. \(2017\)](#) concerning the limiting behaviour of $\mathbb{C}_n = n^{1/2}(C_n^\mathfrak{X} - C^\mathfrak{X})$. Let \mathbb{B}^* be a Brownian bridge, i.e., a continuous centred Gaussian process on $[0, 1]^d$ such that for all $u_1, \dots, u_d, v_1, \dots, v_d \in [0, 1]$,

$$\text{cov}\{\mathbb{B}^*(u_1, \dots, u_d), \mathbb{B}^*(v_1, \dots, v_d)\} = \prod_{j=1}^d \min(u_j, v_j) - \prod_{j=1}^d u_j v_j.$$

Further, set $\mathbb{C}^* = \mathbb{M}_F(\mathbb{B}^*)$ where \mathbb{M}_F is as in (A1). Then \mathbb{C}^* is a continuous centred Gaussian process. For any $v \in \mathbb{R}$, let $\iota_j(v)$ be the vector in \mathbb{R}^d whose j th component is v and whose other components all equal 1. Also let $C([0, 1]^d)$ be the space of continuous functions $f : [0, 1]^d \rightarrow \mathbb{R}$ endowed with the supremum norm $\|\cdot\|$.

THEOREM A1. *The sequence $\|\mathbb{C}_n\|$ is tight. If in addition $C^\mathfrak{X} = \Pi$, then as $n \rightarrow \infty$, $\mathbb{G}_n = \mathbb{C}_n \rightarrow \mathbb{G}$ weakly in $C([0, 1]^d)$, where for any $u_1, \dots, u_d \in [0, 1]$,*

$$\mathbb{G}(u_1, \dots, u_d) = \mathbb{C}^*(u_1, \dots, u_d) - \sum_{j=1}^d \mathbb{C}^*\{\iota_j(u_j)\} \prod_{k \neq j} u_k. \quad (\text{A2})$$

The process \mathbb{G} in (A2) can also be conveniently expressed in terms of the process \mathbb{B} given by

$$\mathbb{B}(u_1, \dots, u_d) = \mathbb{B}^*(u_1, \dots, u_d) - \sum_{j=1}^d \mathbb{B}^*\{\iota_j(u_j)\} \prod_{k \neq j} u_k, \quad u_1, \dots, u_d \in [0, 1], \quad (\text{A3})$$

which is the limiting process \mathbb{G} in Theorem A1 when the margins F_1, \dots, F_d are continuous. To this end, note that for each $j \in \{1, \dots, d\}$ and $u_j \in [0, 1]$, $u_j = \lambda_{F_j}(u_j)F_j\{F_j^{-1}(u_j)\} + \{1 - \lambda_{F_j}(u_j)\}F_j\{F_j^{-1}(u_j)_-\}$, while $\mathbb{C}^*\{\iota_j(u_j)\} = \mathbb{M}_F(\mathbb{B}^*)\{\iota_j(u_j)\}$ and hence $\mathbb{C}^*\{\iota_j(u_j)\} = \lambda_{F_j}(u_j)\mathbb{B}^*(\iota_j[F_j\{F_j^{-1}(u_j)\}]) + \{1 - \lambda_{F_j}(u_j)\}\mathbb{B}^*(\iota_j[F_j\{F_j^{-1}(u_j)_-\}])$. Substituting in (A2) and using the multinomial formula and the fact that \mathbb{M}_F is a linear operator, one obtains

$$\mathbb{G}(u_1, \dots, u_d) = \mathbb{M}_F(\mathbb{B})(u_1, \dots, u_d), \quad u_1, \dots, u_d \in [0, 1]. \quad (\text{A4})$$

Proof of Theorem 1

If $\mathbb{C}_n = n^{1/2}(C_n^\mathfrak{X} - C^\mathfrak{X})$, then $\mathbb{G}_n = \mathbb{C}_n + n^{1/2}(C^\mathfrak{X} - \Pi)$ and

$$S_n = \|\mathbb{G}_n\|_2^2 \geq n\|C^\mathfrak{X} - \Pi\|_2^2 - \|\mathbb{C}_n\|_2^2 \geq n\|C^\mathfrak{X} - \Pi\|_2^2 - \|\mathbb{C}_n\|^2. \quad (\text{A5})$$

Given that $\|\mathbb{C}_n\|$ is tight by Theorem A1, it follows that for any $M > 0$, $\text{pr}(S_n > M) \rightarrow 1$ as $n \rightarrow \infty$ whenever $C^\mathfrak{X} \neq \Pi$. Therefore, the test based on S_n is consistent.

Suppose now that $C^\mathfrak{X} = \Pi$. From Theorem A1, as $n \rightarrow \infty$, $\mathbb{G}_n \rightarrow \mathbb{G}$ weakly in $C([0, 1]^d)$. The asymptotic distribution of S_n then follows because the map $f \mapsto \|f\|_2^2$ is continuous on $C([0, 1]^d)$.

Proof of Proposition 1

Under \mathcal{H}_0 , $C^\mathfrak{X} = \Pi$ and, for $j \in \{1, \dots, d\}$ and $u_1, \dots, u_d \in (0, 1)$, $\partial \Pi(u_1, \dots, u_d) / \partial u_j = (u_1 \cdots u_d) / u_j$. Proposition 8 in [Genest et al. \(2017\)](#) then implies that $\mathbb{G}_n, \mathbb{G}_{n1}, \dots, \mathbb{G}_{nB}$ jointly converge weakly in $C([0, 1]^d)$ to mutually independent copies of \mathbb{G} . By the continuous mapping theorem, the statistics $S_n, S_{n1}, \dots, S_{nB}$ then converge jointly in law to mutually independent copies S, S_1, \dots, S_B of $\|\mathbb{G}\|_2^2$. Next, p_{nB} takes values in $\{0, 1/B, \dots, B/B\}$ and that for each $k \in \{0, \dots, B\}$,

$$\text{pr}\left(p_{nB} = \frac{k}{B}\right) = \binom{B}{k} \text{pr}(S_{n1} > S_n, \dots, S_{nk} > S_n, S_{n(k+1)} \leq S_n, \dots, S_{nB} \leq S_n).$$

Given that S, S_1, \dots, S_B are continuous variables, the continuous mapping theorem now implies that as $n \rightarrow \infty$, the right-hand side converges to

$$\binom{B}{k} \text{pr}(S_1 > S, \dots, S_k > S, S_{k+1} \leq S, \dots, S_B \leq S) = \frac{1}{B+1}.$$

Consequently, as $n \rightarrow \infty$, $p_n \rightarrow p_B = \sum_{b=1}^B 1(S_b > S)/B$ in law, where Bp_B is uniform on $\{0, \dots, B\}$. Thus, for any $\alpha \in (0, 1)$, $\text{pr}(p_{nB} < \alpha) \rightarrow \text{pr}(p_B < \alpha) = (\lceil \alpha B - 1 \rceil + 1)/(B+1)$ as $n \rightarrow \infty$. The convergence of $\text{pr}(p_B < \alpha)$ to α as $B \rightarrow \infty$ is immediate.

If $C^\sharp \neq \Pi$, Proposition 8 of [Genest et al. \(2017\)](#) implies that $\mathbb{G}_{n1}, \dots, \mathbb{G}_{nB}$ jointly converge weakly in $\ell^\infty([0, 1]^d)$ to mutually independent copies of the process given, for $u_1, \dots, u_d \in [0, 1]$, by

$$\mathbb{M}_F(\mathbb{B}_{C^\sharp}^*)(u_1, \dots, u_d) = \sum_{j=1}^d \mathbb{M}_F(\mathbb{B}_{C^\sharp}^*)\{u_j(u_j)\} \times \prod_{k \neq j} u_k,$$

where $\mathbb{B}_{C^\sharp}^*$ is the C^\sharp -Brownian bridge; see [Genest et al. \(2017\)](#) for details. Consequently, the sequences $\|\mathbb{C}_n\|_2^2, \|\mathbb{G}_{n1}\|_2^2, \dots, \|\mathbb{G}_{nB}\|_2^2$ are tight. Using (A5), one further has that

$$\text{pr}(p_{nB} = 0) \geq 1 - \sum_{b=1}^B \text{pr}(\|\mathbb{C}_n\|_2^2 + \|\mathbb{G}_{nb}\|_2^2 > n\|C^\sharp - \Pi\|_2^2).$$

The claim thus follows at once from the fact that $n\|C^\sharp - \Pi\|_2^2 \rightarrow \infty$ as $n \rightarrow \infty$.

Proof of Theorem 2

Following [Genest & Rémillard \(2004\)](#), first define, for $A \in \mathcal{A}_d$ and $f \in C([0, 1]^d)$,

$$\mathcal{M}_A f(u_1, \dots, u_d) = \sum_{B \subset A} (-1)^{|A \setminus B|} f(u_1^B, \dots, u_d^B) \prod_{j \in A \setminus B} u_j, \quad u_1, \dots, u_d \in [0, 1], \quad (\text{A6})$$

where $u_j^B = u_j$ if $j \in B$ and 1 otherwise. It follows from the multinomial formula that if $f(u_1, \dots, u_d) = f_1(u_1) \cdots f_d(u_d)$ for all $u_1, \dots, u_d \in [0, 1]$, then

$$\mathcal{M}_A f(u_1, \dots, u_d) = \prod_{j \in A} \{f_j(u_j) - u_j\}. \quad (\text{A7})$$

Consequently, for any $A \in \mathcal{A}_d$, one has $\mathcal{M}_A(\Pi) \equiv 0$ and $\mu_A = \mathcal{M}_A(C^\sharp)$ as defined in (7). Thus, $\mathbb{G}_n = \mathbb{C}_n + n^{1/2}(C^\sharp - \Pi)$ and $\mathbb{G}_{An} = \mathcal{M}_A(\mathbb{G}_n) = \mathcal{M}_A(\mathbb{C}_n) + n^{1/2}\mu_A$. For any $A \in \mathcal{A}_d$, (A6) implies that $\|\mathcal{M}_A(f)\|_2 \leq \|\mathcal{M}_A(f)\| \leq 2^{|A|}\|f\|$ for all $f \in C([0, 1]^d)$. Hence

$$S_{An} = \|\mathbb{G}_{An}\|_2^2 \geq n\|\mu_A\|^2 - \|\mathcal{M}_A(\mathbb{C}_n)\|_2^2 \geq n\|\mu_A\|^2 - 4^{|A|}\|\mathbb{C}_n\|^2. \quad (\text{A8})$$

If $\mu_A \neq 0$, then for any $M > 0$, $\text{pr}(S_{An} > M) \rightarrow 1$ as $n \rightarrow \infty$, because $\|\mathbb{C}_n\|$ is tight by Theorem A1.

Now suppose that $C^\sharp = \Pi$. Then $\mathbb{G}_{An} = \mathcal{M}_A(\mathbb{C}_n)$ and, since the map $f \mapsto \mathcal{M}_A(f)$ is continuous on $C([0, 1]^d)$, it follows from Theorem A1 that the processes $(\mathbb{G}_{An} : A \in \mathcal{A}_d)$ jointly converge weakly to $\mathbb{G}_A = \mathcal{M}_A(\mathbb{G}) = \mathcal{M}_A(\mathbb{C}^*)$, using (A7) and representation (A2). Next, since $\mathbb{C}^* = \mathbb{M}_F(\mathbb{B}^*)$, it follows easily that $\mathbb{G}_A = \mathbb{M}_F(\mathbb{B}_A^*)$, where $\mathbb{B}_A^* = \mathcal{M}_A(\mathbb{B}^*)$. Now it is well known that $(\mathbb{B}_A^* : A \in \mathcal{A}_d)$ is a collection of mutually independent continuous centred Gaussian processes; see, e.g., [Ghoubi et al. \(2001\)](#). Their covariance function is given, for $A, A' \in \mathcal{A}_d$ and $u_1, \dots, u_d, v_1, \dots, v_d \in [0, 1]$, by

$$\text{cov}\{\mathbb{B}_A^*(u_1, \dots, u_d), \mathbb{B}_{A'}^*(v_1, \dots, v_d)\} = 1(A = A') \prod_{j \in A} \{\min(u_j, v_j) - u_j v_j\}.$$

The processes $(\mathbb{G}_A : A \in \mathcal{A}_d)$ are thus mutually independent; their covariance is given in the Supplementary Material. The claim now follows from the continuity of the map $f \mapsto \|f\|_2^2$ on $C([0, 1]^d)$.

Proof of Proposition 2

For any $A \in \mathcal{A}_d$ and $b \in \{1, \dots, B\}$, $\mathbb{G}_{Anb} = \mathcal{M}_A(\mathbb{G}_{nb})$ with \mathcal{M}_A as defined in (A6). First suppose $C^\sharp = \Pi$. Given that $\mathbb{G}_n, \mathbb{G}_{n1}, \dots, \mathbb{G}_{nB}$ jointly converge weakly to mutually independent copies of \mathbb{G} , as shown in the proof of Proposition 1, the continuity of \mathcal{M}_A implies that as $n \rightarrow \infty$, $(\mathbb{G}_{An} : A \in \mathcal{A}_d), (\mathbb{G}_{An1} : A \in \mathcal{A}_d), \dots, (\mathbb{G}_{AnB} : A \in \mathcal{A}_d)$ jointly converge weakly to mutually independent copies of $(\mathbb{G}_A : A \in \mathcal{A}_d)$ in $C([0, 1]^d)$. As discussed in the proof of Theorem 2, the processes $\mathbb{G}_A, A \in \mathcal{A}_d$, are mutually independent. Hence, as $n \rightarrow \infty$, $\{(S_{An}, S_{An1}, \dots, S_{AnB}) : A \in \mathcal{A}_d\}$ jointly converge weakly to $\{(S_A, S_{A1}, \dots, S_{AB}) : A \in \mathcal{A}_d\}$, where for each $A \in \mathcal{A}_d$, $S_A, S_{A1}, \dots, S_{AB}$ are mutually independent copies of $\|\mathbb{G}_A\|_2^2$. Along the same lines as the proof of Proposition 1, one finds that $(p_{AnB} : A \in \mathcal{A}_d)$ converge jointly in law to mutually independent copies of the uniform distribution on $\{0, 1/B, \dots, B/B\}$.

If $C^\sharp \neq \Pi$, then $\mu_A \neq 0$ for some $A \in \mathcal{A}_d$. From the proof of Proposition 1, the continuity of \mathcal{M}_A and Theorem A1, $\|\mathbb{C}_n\|$ and S_{An1}, \dots, S_{AnB} are tight. Thus (A8) yields the conclusion, as in the proof of Proposition 1.

Proof of Theorem 3

Let $(U_{11n}, \dots, U_{1dn}), \dots, (U_{n1n}, \dots, U_{ndn})$ be a random sample from C_{θ_n} , and define the process

$$\mathbb{B}_n(u_1, \dots, u_d) = n^{-1/2} \sum_{i=1}^n \left\{ \prod_{j=1}^d 1(U_{ijn} \leq u_j) - \prod_{j=1}^d u_j \right\}, \quad (u_1, \dots, u_d) \in [0, 1]^d.$$

As shown in Proposition 2.1 of Genest et al. (2007) using Theorem 3.10.12 of van der Vaart & Wellner (1996), under Q_n , as $n \rightarrow \infty$, \mathbb{B}_n converges weakly in the Skorohod space of càdlàg functions $D([0, 1]^d)$ to $\mathbb{B}^* + \delta \dot{C}_{\theta_0}$, where \dot{C}_{θ_0} is as in Condition 2. Now write $\mathbb{G}_n = \tilde{\mathbb{C}}_n^\sharp + \tilde{\mathbb{D}}_n$, where the summands are as in equation (5) of Genest et al. (2017). From the proof of Proposition 4 and in light of Proposition 5 in the latter paper together with contiguity, it continues to be the case that under Q_n , $\|\tilde{\mathbb{C}}_n^\sharp - \mathbb{M}_F(\mathbb{B}_n)\| \rightarrow 0$ in probability. Because the operator \mathbb{M}_F is continuous and linear, the continuous mapping theorem implies that under Q_n , as $n \rightarrow \infty$, $\tilde{\mathbb{C}}_n^\sharp \rightarrow \mathbb{M}_F(\mathbb{B}^* + \delta \dot{C}_{\theta_0}) = \mathbb{C}^* + \delta \mathbb{M}_F(\dot{C}_{\theta_0})$ weakly in $D([0, 1]^d)$. Now observe that for each $j \in \{1, \dots, d\}$ and $u_j \in [0, 1]$, Condition 2 implies that $\dot{C}_{\theta_0}\{t_j(u_j)\} = 0$ and hence $\mathbb{M}_F(\dot{C}_{\theta_0})\{t_j(u_j)\} = 0$; see also Genest et al. (2007, p. 170). From Proposition 6 in Genest et al. (2017), together with Proposition 4.5 in Genest et al. (2014) and contiguity, one then has that under Q_n , as $n \rightarrow \infty$, $\tilde{\mathbb{D}}_n \rightarrow \tilde{\mathbb{D}}$ weakly, where for each $(u_1, \dots, u_d) \in [0, 1]^d$,

$$\tilde{\mathbb{D}}(u_1, \dots, u_d) = - \sum_{j=1}^d \prod_{k \neq j} u_k [\mathbb{C}^*\{t_j(u_j)\} + \delta \mathbb{M}_F(\dot{C}_{\theta_0})\{t_j(u_j)\}] = - \sum_{j=1}^d \prod_{k \neq j} u_k \mathbb{C}^*\{t_j(u_j)\}.$$

From (A2), under Q_n , as $n \rightarrow \infty$, $\mathbb{G}_n \rightarrow \mathbb{G} + \delta \mathbb{M}_F(\dot{C}_{\theta_0})$ weakly in $D([0, 1]^d)$. Because \mathbb{G}_n and the limit have continuous paths, the convergence is in $C([0, 1]^d)$. From the proof of Theorem 2, $\mathbb{G}_A = \mathcal{M}_A(\mathbb{G}_n)$, where \mathcal{M}_A is a continuous map; so the continuous mapping theorem implies that under Q_n , the processes \mathbb{G}_A with $A \in \mathcal{A}_d$ converge jointly to $\mathcal{M}_A\{\mathbb{G} + \delta \mathbb{M}_F(\dot{C}_{\theta_0})\}$. For each $A \in \mathcal{A}_d$, the map \mathcal{M}_A is linear, and, as shown in the proof of Theorem 2, $\mathcal{M}_A(\mathbb{G}) = \mathcal{M}_A(\mathbb{C}^*) = \mathbb{G}_A$. The limiting process can therefore be rewritten as $\mathcal{M}_A\{\mathbb{G} + \delta \mathbb{M}_F(\dot{C}_{\theta_0})\} = \mathbb{G}_A + \delta \mathbb{M}_F(q_A)$ where $q_A = \mathcal{M}_A(\dot{C}_{\theta_0})$, as claimed.

REFERENCES

- AGRESTI, A. (2013). *Categorical Data Analysis*. New York: Wiley, 3rd ed.
 BAKIROV, N. K., RIZZO, M. L. & SZÉKELY, G. J. (2006). A multivariate nonparametric test of independence. *J. Mult. Anal.* **97**, 1742–56.
 BENJAMINI, Y. & HOCHBERG, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Statist. Soc. B* **57**, 289–300.

- BLUM, J. R., KIEFER, J. & ROSENBLATT, M. (1961). Distribution free test of independence based on the sample distribution function. *Ann. Math. Statist.* **32**, 485–98.
- BÜCHER, A. & DETTE, H. (2010). A note on bootstrap approximations for the empirical copula process. *Statist. Prob. Lett.* **80**, 1925–32.
- CLAYTON, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* **65**, 141–51.
- DEHEUVELS, P. (1980). Non parametric tests of independence. In *Nonparametric Asymptotic Statistics (Proceedings of the Conference held in Rouen in 1979)*, J.-P. Raoult, ed. New York: Springer, pp. 95–107.
- DEHEUVELS, P. (1981). An asymptotic decomposition for multivariate distribution-free tests of independence. *J. Mult. Anal.* **11**, 102–13.
- FAN, Y., LAFAYE DE MICHEAUX, P., PENEV, S. & SALOPEK, D. (2017). Multivariate nonparametric test of independence. *J. Mult. Anal.* **153**, 189–210.
- FEUERVERGER, A. (1993). A consistent test for bivariate dependence. *Int. Statist. Rev.* **61**, 419–33.
- GAUTHIER, S., LEBLANC, J., SERESOVA, A., LABERGE-POIRIER, A., CORREA, J. A., ALTURKIE, A. Y., MARCOUX, J., MALEKI, M., FEYZ, M. & DE GUISE, E. (2018). Acute prediction of outcome and cognitive-communication impairments following traumatic brain injury: The influence of age, education and site of lesion. *J. Commun. Disord.* **73**, 77–90.
- GENEST, C. & NEŠLEHOVÁ, J. (2007). A primer on copulas for count data. *Astin Bull.* **37**, 475–515.
- GENEST, C., NEŠLEHOVÁ, J. G. & RÉMILLARD, B. (2013). On the estimation of Spearman's rho and related tests of independence for possibly discontinuous multivariate data. *J. Mult. Anal.* **117**, 214–28.
- GENEST, C., NEŠLEHOVÁ, J. G. & RÉMILLARD, B. (2014). On the empirical multilinear copula process for count data. *Bernoulli* **20**, 1344–71.
- GENEST, C., NEŠLEHOVÁ, J. G. & RÉMILLARD, B. (2017). Asymptotic behavior of the empirical multilinear copula process under broad conditions. *J. Mult. Anal.* **159**, 82–110.
- GENEST, C., QUESSY, J.-F. & RÉMILLARD, B. (2006). Local efficiency of a Cramér–von Mises test of independence. *J. Mult. Anal.* **97**, 274–94.
- GENEST, C., QUESSY, J.-F. & RÉMILLARD, B. (2007). Asymptotic local efficiency of Cramér–von Mises tests for multivariate independence. *Ann. Statist.* **35**, 166–91.
- GENEST, C. & RÉMILLARD, B. (2004). Tests of independence or randomness based on the empirical copula process. *Test* **13**, 335–69.
- GHOUDI, K., KULPERGER, R. J. & RÉMILLARD, B. (2001). A nonparametric test of serial independence for time series and residuals. *J. Mult. Anal.* **79**, 191–218.
- HELLER, R., HELLER, Y., KAUFMAN, S., BRILL, B. & GORFINE, M. (2016). Consistent distribution-free K -sample and independence tests for univariate random variables. *J. Mach. Learn. Res.* **17**, 1–54.
- HOEFFDING, W. (1948). A non-parametric test of independence. *Ann. Math. Statist.* **19**, 546–57.
- LITTELL, R. C. & FOLKS, J. L. (1973). Asymptotic optimality of Fisher's method of combining independent tests. II. *J. Am. Statist. Assoc.* **68**, 193–4.
- MCNEIL, A. J., FREY, R. & EMBRECHTS, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton, New Jersey: Princeton University Press, revised ed.
- MEINTANIS, S. G. & ILIOPOULOS, G. (2008). Fourier methods for testing multivariate independence. *Comp. Statist. Data Anal.* **52**, 1884–95.
- MOORE, D. S. & SPRUILL, M. C. (1975). Unified large-sample theory of general chi-squared statistics for tests of fit. *Ann. Statist.* **3**, 599–616.
- NELSEN, R. B. (2006). *An Introduction to Copulas*. New York: Springer, 2nd ed.
- NEŠLEHOVÁ, J. (2007). On rank correlation measures for non-continuous random variables. *J. Mult. Anal.* **98**, 544–67.
- R DEVELOPMENT CORE TEAM (2019). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0, <http://www.R-project.org>.
- RÉMILLARD, B. (2011). Tests of independence. In *International Encyclopedia of Statistical Science*, M. Lovric, ed. Berlin: Springer, pp. 1598–601.
- RÉMILLARD, B. & SCAILLET, O. (2009). Testing for equality between two copulas. *J. Mult. Anal.* **100**, 377–86.
- SCHWEIZER, B. & SKLAR, A. (1974). Operations on distribution functions not derivable from operations on random variables. *Studia Math.* **52**, 43–52.
- SHORACK, G. R. & WELLNER, J. A. (1984). *Empirical Processes with Applications to Statistics*. New York: Wiley.
- SPITZER, F. L. (1974). Introduction aux processus de Markov à paramètre dans Z_v . In *École d'été de probabilités de Saint-Flour, III–1973*. Berlin: Springer, pp. 114–89.
- SZÉKELY, G. J., RIZZO, M. L. & BAKIROV, N. K. (2007). Measuring and testing dependence by correlation of distances. *Ann. Statist.* **35**, 2769–94.
- VAN DER VAART, A. W. & WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. New York: Springer.

[Received on 10 July 2017. Editorial decision on 2 August 2018]