Centre de Recherche en Économie et Statistique

CREST

Série des Documents de Travail

n° 2013-25

A Asymptotic Total Variation Test for Copulas

J.-D. FERMANIAN[1]
D. RADULOVIĆ[2]
M. WEGKAMP[3]

2013

[1] CREST (ENSAE). Paris, France. Email : jean-david.fermanian@ensae.fr

[2] Department of Mathematics, Florida Atlantic University, USA. Email : radulovi@fau.edu

[3] Cornell University, USA. Email : marten.wegkam@cornell.edu

An Asymptotic Total Variation Test for Copulas

JEAN-DAVID FERMANIAN, DRAGAN RADULOVIĆ AND MARTEN WEGKAMP*

Jean-David Fermanian
Crest-Ensae
3 av. Pierre Larousse
92245 Malakoff cedex, France
E-mail: jean-david.fermanian@ensae.fr

Dragan Radulović
Department of Mathematics
Florida Atlantic University, USA
E-mail: radulovi@fau.edu

Marten Wegkamp
Department of Mathematics & Department of Statistical Science
Cornell University, USA
E-mail: marten.wegkamp@cornell.edu

We propose a new goodness-of-fit test for copulas, based on empirical copula processes and nonparametric bootstrap counterparts. The standard Kolmogorov-Smirnov type test for copulas that takes the supremum of the empirical copula process indexed by orthants is extended by test statistics based on the supremum of the empirical copula process indexed by families of L_n disjoint boxes, with L_n slowly tending to infinity. Although the underlying empirical process does not converge, the critical values of our new test statistic can be consistently estimated by nonparametric bootstrap techniques, under simple or composite null assumptions. Simulations confirm that the power of the new procedure is oftentimes higher than the power of the standard Kolmogorov-Smirnov or the Cramér-von Mises tests for copulas.

AMS 2000 subject classifications: Primary 60F17, 60K35; secondary 60G99.
Keywords: Bootstrap, copula, empirical copula process, goodness-of-fit Test, weak Convergence.

1. Introduction

This paper introduces a new powerful goodness-of-fit (GOF) test for copulas in $[0,1]^d$, $d \geqslant 2$, based on the empirical copula process

$$\mathbb{Z}_n(\mathbf{u}) = \sqrt{n}(\mathbb{C}_n - C)(\mathbf{u}), \qquad \mathbf{u} = (u_1, \ldots, u_d) \in [0,1]^d, \tag{1.1}$$

given a sample of n independent random vectors $\mathbf{X}_i = (X_{i1}, \ldots, X_{id}) \in \mathbb{R}^d$, $i = 1, \ldots, n$, from a common distribution function H. Let C be the associated copula function C, as given by Sklar's Theorem (Sklar, 1959). Here $\mathbb{C}_n$ is the usual empirical copula, as introduced by Deheuvels (1979):

*The research of Wegkamp is supported in part by NSF DMS 1007444 and NSF DMS 1310119 grants.

denoting by $\mathbb{H}_n$ the joint cdf of the sample $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$, $\mathbb{F}_{n,j}$ the j-th empirical cdf associated to $(X_{1j}, \ldots, X_{nj})$, $j = 1, \ldots, d$, and $\mathbb{F}^-_{n,j}$ its empirical quantile function, we have

$$\mathbb{C}_n(\mathbf{u}) = \mathbb{H}_n(\mathbb{F}^-_{n,1}(u_1), \ldots, \mathbb{F}^-_{n,d}(u_d))$$

by definition, for every $\mathbf{u} = (u_1, \ldots, u_d) \in [0,1]^d$. The Kolmogorov-Smirnov (KS) test statistic for testing of the null hypothesis $\mathrm{H}_0 : C = C_0$ is

$$\mathrm{KS}_n = \sup_{\mathbf{u} \in [0,1]^d} |\sqrt{n}(\mathbb{C}_n - C_0)(\mathbf{u})|. \tag{1.2}$$

The Cramér-von Mises statistic (CvM) is

$$\mathrm{CM}_n = \int \{\sqrt{n}(\mathbb{C}_n - C_0)(\mathbf{u})\}^2 \, \mathrm{d}\mathbb{C}_n(\mathbf{u}). \tag{1.3}$$

It is well-known, see, for instance, Fermanian et al. (2004), that $\mathbb{Z}_n$ and its bootstrap counterpart $\mathbb{Z}^*_n$, defined in (2.4) below, both converge weakly to the same tight Gaussian process in $\ell^\infty([0,1]^d)$ under the null hypothesis. Therefore, we can compute the α-upper points of KS_n and CM_n via the bootstrap. To the best of our knowledge, all the proposed GOF tests rely on simulation-based procedures to calculate their corresponding p-values, with the notable exception of the distribution-free test statistics of Fermanian (2005). The latter idea has been further developed by Scaillet (2007) and Fermanian and Wegkamp (2012). A parametric bootstrap has been proposed (Genest and Rémillard, 2008) to tackle composite null hypotheses, while Rémillard and Scaillet (2009) advocate the use of the multiplier central limit theorem to build an alternative bootstrap empirical copula process. Bücher and Dette (2010) give a survey and a comparison of various bootstrap methods.

The goal of this paper is to develop a more powerful test than the KS test (1.2) and CvM test (1.3) for simple and composite null hypotheses. In the case of a null simple hypothesis $\mathrm{H}_0 : C = C_0$, we propose the following test that rejects H_0 for large values of the test statistic

$$\mathbb{T}_n := \sup_{B_1, \ldots, B_{L_n}} \sum_{k=1}^{L_n} |\mathbb{Z}_n(B_k)|. \tag{1.4}$$

The supremum is taken over all disjoint boxes $B_1, \ldots, B_{L_n} \subset [0,1]^d$ of the form $\prod_{j=1}^d (a_j, b_j]$, using the convention

$$\mathbb{Z}_n((a_1, b_1] \times \cdots \times (a_d, b_d]) = \Delta^1_{a_1,b_1} \Delta^2_{a_2,b_2} \cdots \Delta^d_{a_d,b_d} \mathbb{Z}_n(\mathbf{u}), \tag{1.5}$$

for any arbitrary point $\mathbf{u} \in [0,1]^d$ and for all $0 \leqslant a_j < b_j \leqslant 1$, $j = 1, \ldots, d$. Here, we have used the usual operators Δ^j defined for every function f by

$$\left(\Delta^j_{a,b} f\right)(\mathbf{u}) = f(u_1, \ldots, u_{j-1}, b, u_{j+1}, \ldots, u_d) - f(u_1, \ldots, u_{j-1}, a, u_{j+1}, \ldots, u_d),$$

for all $\mathbf{u} \in [0,1]^d$, and all real numbers a and b.

In the empirical part, we will also consider the statistics $\widetilde{\mathbb{T}}_n$ defined in a similar way as $\mathbb{T}_n$:

$$\widetilde{\mathbb{T}}_n = \max_{B_1, \ldots, B_{L_n}} \sum_{i=1}^{L_n} |\mathbb{Z}_n(B_i)| \ , \tag{1.6}$$

but here the maximum is taken over all disjoint rectangles $B_1, ..., B_{L_n}$ of the form $B = \Pi_{j=1}^d (a_j, b_j]$ with a_j, b_j belonging to a grid $\{\frac{1}{n^{1/d}}, \frac{2}{n^{1/d}}, \ldots, \frac{\lfloor n^{1/d} \rfloor}{n^{1/d}}\}$. Asymptotically, $\widetilde{\mathbb{T}}_n$ and $\mathbb{T}_n$ behave identically (since $|\widetilde{\mathbb{T}}_n - \mathbb{T}_n| = o_p(1)$), but $\widetilde{\mathbb{T}}_n$ is computationally more tractable.

Now, if $L_n = L$ for all n, the collection of boxes is sufficiently small that we can still appeal to the weak convergence of $\mathbb{Z}_n$ and $\mathbb{Z}_n^*$ in conjunction with the continuous mapping theorem, to obtain α-upper points of the test statistic $\mathbb{T}_n$ via the bootstrap. Taking $L_n = +\infty$ for all n, or equivalently, if we consider all families of disjoint boxes in $[0,1]^d$ (possibly partitions), the statistic $\mathbb{T}_n$ is equal to the total variation distance $TV(\mathbb{Z}_n)$ of $\mathbb{Z}_n$. The resulting test is not statistically meaningful as $TV(\mathbb{Z}_n)$ is maximal, to wit, $TV(\mathbb{Z}_n) = n^{1/2} \to +\infty$. The problem is to find a rich collection that quickly detects departure from the null, but still yields a consistent test. The main novelty of our approach is the fact that we let L_n, the number of boxes, slowly tend to ∞ in that $L_n \sim (\log n)^\gamma$, $0 < \gamma < 1$. While in this case the process $\mathbb{Z}_n$ no longer converges, Theorem 1 in Section 2 states that we can still consistently estimate the distribution of the process $\mathbb{Z}_n$ by the bootstrap. We refer to our procedure as the Asymptotic Total Variation (ATV) test. The considered families of boxes are finer and finer, presumably improving the power of the test, while for each n large enough, we still have a consistent test in that we control the type 1 error. A key observation is that under the null hypothesis $\mathrm{H}_0 : C = C_0$, we have $\mathbb{T}_n \leqslant L_n \sup_B |\mathbb{Z}_n(B)| = O_p(L_n)$, while under the alternative $\mathrm{H}_A : C = C_1$ for some fixed $C_1 \neq C_0$, $\mathbb{T}_n$ is much larger since the bias is at least of order $O(n^{1/2})$.

Theorem 1 extends the surprising result obtained by Radulović (2012) for empirical processes indexed by sums of indicator functions of VC-graph classes (see Theorem 13 in the appendix). We require very mild conditions on the copula function C. This is one of the few notable exceptions known to us in the literature where the bootstrap "works", that is, the conditional bootstrap distribution consistently estimates the distribution of the test statistic, while the distribution of the statistic itself does not converge. For other instances of this phenomenon, we refer to Bickel and Freedman (1983) or Radulović (1998, 2012, 2013), more recently.

Section 3 considers the more general hypothesis that the underlying copula C belongs to some parametric copula family $\{C_\theta, \theta \in \Theta \subset \mathbb{R}^p\}$. Given a sufficiently regular estimator $\widehat{\theta}$ and its bootstrap counterpart $\widehat{\theta}^*$, we adjust our statistic (1.4) and its non-parametric bootstrap counterpart to obtain a consistent level α test (Theorem 4). Again, the result is established under very mild regularity conditions on the copula C_θ and the estimators $\widehat{\theta}$ and $\widehat{\theta}^*$. Incidentally, we introduce a new bootstrap procedure under composite null hypotheses, an alternative to the usual parametric bootstrap or the multiplier CLT.

Section 4 then reports a small numerical study where we show that, in complex but realistic situations, our test (1.4) is superior to the Kolmogorov-Smirnov and the Cramér-von Mises tests. We also comment on a possible inadequacy in the way the copula GOF tests are commonly evaluated. Finally, the proofs are collected in Section 5. The appendix contains some technical results from Segers (2012) and Radulović (2012) and a description of the implementation of the proposed tests.

2. The Asymptotic Total Variation Test

NOTATIONS. Let H be the distribution function of the random vector $\mathbf{X}$ with marginals $F_1, \ldots, F_d$. We will assume throughout the paper that H is continuous. Let $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ be independent copies of $\mathbf{X}$. We denote the generalized inverse of a distribution function F by F^-. For instance, $F_j^-(u) = \inf\{x \mid F_j(x) \geqslant u\}$. The empirical counterparts of H and any F_j are, respectively,

$$\begin{aligned}\mathbb{H}_n(\mathbf{x}) &= \frac{1}{n}\sum_{i=1}^n \mathbf{1}\{\mathbf{X}_i \leqslant \mathbf{x}\},\ \mathbf{x} \in \mathbb{R}^d \\ \mathbb{F}_{n,j}(x) &= \frac{1}{n}\sum_{i=1}^n \mathbf{1}\{X_{i,j} \leqslant x\},\ x \in \mathbb{R},\ \ j = 1,\ldots,d.\end{aligned}$$

The copula function of $\mathbf{X}$ is $C(\mathbf{u}) = H(F_1^-(u_1),\ldots,F_d^-(u_d))$, $\mathbf{u} = (u_1,\ldots,u_d) \in [0,1]^d$, and its empirical estimate is $\mathbb{C}_n(\mathbf{u}) = \mathbb{H}_n(\mathbb{F}_{n,1}^-(u_1),\ldots,\mathbb{F}_{n,d}^-(u_d))$. The empirical copula process $\mathbb{Z}_n(\mathbf{u}) = \sqrt{n}(\mathbb{C}_n - C)(\mathbf{u})$ is already defined in (1.1). We define $\mathcal{F}_n$ as the class of functions

$$f(\mathbf{x}) = \sum_{k=1}^{L_n} c_k \mathbf{1}\{\mathbf{x} \in B_k\}, \tag{2.1}$$

with $c_k \in \{-1,+1\}$ and disjoint boxes B_k of the form $\prod_{j=1}^d (a_j, b_j]$ in the unit cube $[0,1]^d$, for all $1 \leqslant k \leqslant L_n$. We let

$$\mathbb{Z}_n(f) = \sum_{k=1}^{L_n} c_k \mathbb{Z}_n(B_k),$$

and observe that

$$\mathbb{T}_n = \sup_{f \in \mathcal{F}_n} |\mathbb{Z}_n(f)| = \sup_{B_1,\ldots,B_{L_n}} \sum_{k=1}^{L_n} |\mathbb{Z}_n(B_k)|,$$

where the supremum is taken over all disjoint boxes $B_1,\ldots,B_{L_n}$ of the unit square $[0,1]^d$.

If $L_n = L$ for all n, then $\mathcal{F}_n = \mathcal{F}$ and $\mathbb{Z}_n$ converges in $\ell^\infty(\mathcal{F})$ to a Gaussian process under regularity conditions on C, see, for instance, Fermanian et al. (2004) and Segers (2012). As a consequence of the continuous mapping theorem, $\mathbb{T}_n$ trivially converges weakly as well. However, if $L_n \to \infty$, as $n \to \infty$, this is no longer true as the process $\mathbb{Z}_n$ does not converge weakly.

The main point of this paper is to show that, provided $L_n = (\log n)^\gamma$ for some $0 < \gamma < 1$, the distribution of $\mathbb{T}_n$ can be estimated by the bootstrap. The bootstrap counterparts of the above processes are defined as follows. Let the bootstrap sample $(\mathbf{X}_1^*,\cdots,\mathbf{X}_n^*)$ be obtained by sampling with replacement from $\mathbf{X}_1,\cdots,\mathbf{X}_n$. We write

$$\mathbb{H}_n^*(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}\{\mathbf{X}_i^* \leqslant \mathbf{x}\},\ \mathbf{x} \in \mathbb{R}^d, \tag{2.2}$$

for the empirical cdf based on the bootstrap, with marginals

$$\mathbb{F}_{n,j}^*(x) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}\{X_{i,j}^* \leqslant x\},\ x \in \mathbb{R},\ j = 1,\ldots,d. \tag{2.3}$$

We denote its associated empirical copula function by $\mathbb{C}_n^*$. The bootstrap empirical copula process is

$$\mathbb{Z}_n^* = \sqrt{n}(\mathbb{C}_n^* - \mathbb{C}_n) = \sqrt{n}\left\{\mathbb{H}_n^*(\mathbb{F}_{n,1}^{*-},\ldots,\mathbb{F}_{n,d}^{*-}) - \mathbb{H}_n(\mathbb{F}_{n,1}^-,\ldots,\mathbb{F}_{n,d}^-)\right\}. \tag{2.4}$$

Assumptions. We will assume the following set of assumptions:

(C1) For any $j = 1, \dots, d$, for all $\mathbf{u} \in [0,1]^d$ with $0 < u_j < 1$, the first-order partial derivative $C_j(\mathbf{u}) = \partial C(\mathbf{u})/\partial u_j$ exists and is of bounded variation (Hildebrandt, 1963, e.g.). Moreover, it satisfies, for some $r > 0$, $\beta \geqslant 0$ and $K < \infty$,

$$|C_j(\mathbf{u}) - C_j(\mathbf{v})| \leqslant K \left(u_j^{-\beta}(1-u_j)^{-\beta} + v_j^{-\beta}(1-v_j)^{-\beta} \right) \sum_{l=1}^{d} |u_l - v_l|^r,$$

for all $\mathbf{u}, \mathbf{v} \in [0,1]^d$, $0 < u_j, v_j < 1$. As in Segers (2013), we extend the domain of each C_j to the whole $[0,1]^d$ by setting

$$C_j(\mathbf{u}) := \begin{cases} \limsup_{h \downarrow 0} \frac{C(\mathbf{u}+h\mathbf{e}_j)}{h} & \text{if } \mathbf{u} \in [0,1]^d, u_j = 0; \\ \limsup_{h \downarrow 0} \frac{C(\mathbf{u})-C(\mathbf{u}-h\mathbf{e}_j)}{h} & \text{if } \mathbf{u} \in [0,1]^d, u_j = 1. \end{cases}$$

Here $\mathbf{e}_j$ is the jth coordinate vector in $\mathbb{R}^d$.

(C2) The number L_n is of order $(\log n)^\gamma$ for some $0 < \gamma < 1$.

Remark. We know that continuity of the partial derivatives of C on $(0,1)^d$ is required for weak convergence, see Fermanian et al. (2004) and Segers (2012). The requirement that the partial derivatives are of bounded variation is natural since we compute the supremum of $\mathbb{Z}_n$ over increasingly finer families of boxes in $[0,1]^d$. The process $\mathbb{Z}_n(\mathbf{u})$ is asymptotically equivalent to $\alpha_n(\mathbf{u}) - \sum_{j=1}^d C_j(\mathbf{u})\alpha_{n,j}(u_j)$ with $\alpha_n(\mathbf{u}) = \sqrt{n}(\mathbb{H}_n - H)(\mathbf{u})$ and $\alpha_{n,j}(u_j) = \sqrt{n}(\mathbb{F}_{n,j} - F)(u_j)$ (see Proposition 10).

Remark. The additional requirement (C1) is weaker than imposing a Hölder condition on the derivatives. Segers (2012) imposes a slightly stronger condition on the second-order partial derivatives of C (corresponding to $r = 1$) to obtain an almost sure representation of the empirical copula process.

Indeed, consider the bivariate Archimedean copula C whose generator is given by $\psi : (0,1] \to \mathbb{R}^+$, $\psi(t) := \exp(t^{-\theta}) - e$ for some $\theta > 0$. This copula, numbered (4.2.20) in Nelsen (2006), is

$$C(u_1, u_2) = \left[\ln \left(\exp(u_1^{-\theta}) + \exp(u_2^{-\theta}) - e \right) \right]^{-1/\theta},$$

for any $\mathbf{u} \in [0,1]^2$. It can be checked easily that, when $u \to 0$, the copula density

$$C_{12}(u,u) \sim \frac{\theta^2}{4} u^{-\theta-1}.$$

Therefore, C cannot fulfill Condition 4.1 in Segers (2012). Nonetheless, by the mean value theorem and simple calculations, we can prove that

$$|C_1(\mathbf{u}) - C_1(\mathbf{v})| \leqslant K(\min(u_1,v_1))^{-2\theta-2}|u_1 - v_1| + K(\min(u_1,v_1))^{-\theta-1}|u_2 - v_2|.$$

Since the same reasoning can be done with C_2, our condition (C1) is fulfilled.

The second assumption (C2) allows for sub-logarithmic rate in the sample size for the number of boxes considered. In practice, even this fairly slow rate yields much better tests, see our simulations in Section 4. And we have not observed any significant differences empirically between choosing $\gamma = 1$ and γ closed to one.

Our first result states that the processes $\mathbb{Z}_n$ and $\mathbb{Z}_n^*$ are close in the bounded Lipschitz distance that characterizes weak convergence. Formally, we show that

$$\mathbb{E}\left[\sup_h |\mathbb{E}[h(\mathbb{Z}_n)] - \mathbb{E}^*[h(\mathbb{Z}_n^*)]|\right] \tag{2.5}$$

is asymptotically negligible. Here $\mathbb{E}^*$ is the conditional expectation with respect to the bootstrap sample and the supremum in (2.5) is taken over $BL_1 = BL_1(\ell^\infty(\mathcal{F}_n))$, the class of all uniformly bounded, Lipschitz functionals $h : \ell^\infty(\mathcal{F}_n) \to \mathbb{R}$ with Lipschitz constant 1, that is,

$$\sup_{x \in \ell^\infty(\mathcal{F}_n)} |h(x)| \leqslant 1 \tag{2.6}$$

and, for all $x, y \in \ell^\infty(\mathcal{F}_n)$,

$$|h(x) - h(y)| \leqslant \sup_{f \in \mathcal{F}_n} |x(f) - y(f)|. \tag{2.7}$$

Theorem 1. *Let $\mathbb{Z}_n = \{\mathbb{Z}_n(f),\ f \in \mathcal{F}_n\}$ and $\mathbb{Z}_n^* = \{\mathbb{Z}_n^*(f),\ f \in \mathcal{F}_n\}$ with $\mathcal{F}_n$ as defined in (2.1) above. Under conditions (C1) and (C2), we have*

$$\lim_{n\to\infty} \mathbb{E}\left[\sup_{h \in BL_1} |\mathbb{E}[h(\mathbb{Z}_n)] - \mathbb{E}^*[h(\mathbb{Z}_n^*)]|\right] = 0. \tag{2.8}$$

Corollary 2. *Under conditions (C1) and (C2) and for any Lipschitz functional $\phi : \ell^\infty(\mathcal{F}_n) \to \mathbb{R}$, we have*

$$\lim_{n\to\infty} \mathbb{E}\left[\sup_g |\mathbb{E}[g(\phi(\mathbb{Z}_n))] - E^*[g(\phi(\mathbb{Z}_n^*))]|\right] = 0.$$

The supremum is taken over all uniformly bounded Lipschitz functions $g : \mathbb{R} \to \mathbb{R}$ with $\sup_x |g(x)| \leqslant 1$ and $|g(x) - g(y)| \leqslant |x - y|$.

Corollary 2 follows directly from Theorem 1 since the composition $g(\phi(\cdot))$ is bounded Lipschitz as long as ϕ is Lipschitz. In particular, since the mapping $\phi(X) = \sup_{f \in \mathcal{F}_n} |\phi(f)|$ is Lipschitz, Corollary 2 implies that we can approximate the distribution of the statistic $\mathbb{T}_n$ by the conditional (bootstrap) distribution of

$$\mathbb{T}_n^* = \sup_{f \in \mathcal{F}_n} |\mathbb{Z}_n^*(f)| = \sup_{B_1,\ldots,B_{L_n}} \sum_{k=1}^{L_n} |\mathbb{Z}_n^*(B_k)| : \tag{2.9}$$

Corollary 3. *Under conditions (C1) and (C2), we have*

$$\lim_{n\to\infty} \mathbb{E}\left[\sup_g |\mathbb{E}\left[g\left(\mathbb{T}_n\right)\right] - \mathbb{E}^*\left[g\left(\mathbb{T}_n^*\right)\right]|\right] = 0. \tag{2.10}$$

The supremum is taken over all uniformly bounded Lipschitz functions $g : \mathbb{R} \to \mathbb{R}$ with $\sup_x |g(x)| \leqslant 1$ and $|g(x) - g(y)| \leqslant |x - y|$ for all $x, y \in \mathbb{R}$.

Actually, $\mathbb{T}_n$ is just one of many potentially useful asymptotic variation type statistics. We mention two other possible statistics:

- **Generalized χ^2 statistics**. Form an equidistant grid i/p, $i = 0, \ldots, p = \left\lfloor L_n^{1/d} \right\rfloor + 1$ on each axis of $[0,1]^d$, and use the $(p+1)^d$ points of the resulting equidistant grid on $[0,1]^d$ as the corners of p^d disjoint boxes B_i. We define the statistic $\sum_i |\mathbb{Z}_n(B_i)|^2$, which, for fixed L_n, reduces to a non-normalized χ^2 statistics, in the same spirit as in Dobrić and Schmid (2005). Here, since the statistic as a function of $\mathbb{Z}_n$ is Lipschitz on $\ell^\infty(\mathcal{F}_n)$, $L_n \to \infty$ is allowed. However, we suspect that the full power of Theorem 1 is not needed, since Radulović (2013) proved a result similar to Theorem 1 via a more direct approach, in the non-copula, i.i.d. setting under a weaker restriction on the partition size.

- **Generalized Kuiper statistics.** We start with the usual Kuiper statistics
$$K_1 = \mathbb{Z}_n(B_1) = \sup_B |\mathbb{Z}_n(B)|,$$
where supremum is taken over all boxes $B \subseteq [0,1]^d$, and achieved at B_1. Then we define recursively, given boxes $B_1, ..., B_m$ with $m < L_n$,
$$K_{m+1} = \mathbb{Z}_n(B_{m+1}) = \sup_{B \cap B_j = \varnothing,\, j=1,\ldots,m} |\mathbb{Z}_n(B)|.$$
The supremum is taken over all boxes B that are disjoint with $B_1, \ldots, B_m$, and we denote by B_{m+1} for the box at which supremum is achieved. The resulting sum $\sum_{j=1}^{L_n} K_j$ of statistics K_j, based on disjoint boxes B_j, is a Lipschitz functional of $\mathbb{Z}_n$ and Corollary 2 applies to this statistics (in lieu of $\mathbb{T}_n$) as well.

The performance and the actual implementation of these additional statistics will not be discussed here, but we will report on them elsewhere. This paper offers a numerical study only as a proof of principle and for this purpose we used the straightforward statistic $\widetilde{\mathbb{T}}_n$ and optimization scheme (pure random search) to demonstrate the applicability of Theorem 1. Nevertheless, even this conservative approach resulted in a superior performance.

Remark. While exact computation of $\mathbb{T}_n$ is impossible, we found that a simple random search algorithm performed very well in our simulation studies: see Appendix C.

JDF : I think we should remove the latter remark.

Remark. We may approximate the α-upper point of the statistic $\mathbb{T}_n$ by that of the bootstrap counterpart $\mathbb{T}_n^*$. Unlike the classical bootstrap situation that assumes a continuous limiting distribution function[1], the bootstrap quantile approximation can be used as follows. Let $\varepsilon > 0$ be

[1]In order to formally transition from weak convergence results to uniform distribution approximations, the standard approach follows the steps described above and yields
$$\sup_t |\mathbb{P}\{\mathbb{T}_n \leqslant t\} - \mathbb{P}^*\{\mathbb{T}_n^* \leqslant t\}| \leqslant \frac{\delta_n}{\varepsilon} + \min(\Delta_{n,\varepsilon}, \Delta_{n,\varepsilon}^*)$$
for $\Delta_{n,\varepsilon} = \sup_t \mathbb{P}(t - \varepsilon \leqslant \mathbb{T}_n \leqslant t + \varepsilon)$ and $\Delta_{n,\varepsilon}^* = \sup_t \mathbb{P}^*(t - \varepsilon \leqslant \mathbb{T}_n^* \leqslant t + \varepsilon)$. The weak limit result implies $\delta_n \to 0$, and the classical approach, in order to argue $\Delta_{n,\varepsilon} \to 0$, assumes that (a) pointwise limit of $\Phi_n(t) := \mathbb{P}(\mathbb{T}_n \leqslant t)$ exists

arbitrary (independent of n) and define the Lipschitz function

$$g_{t,\varepsilon}(x) = \mathbf{1}\{x \leqslant t\} + \frac{t+\varepsilon-x}{\varepsilon}\mathbf{1}\{t < x \leqslant t+\varepsilon\}.$$

We have, for $\delta_n := \sup_h |\mathbb{E}[h(\mathbb{T}_n)] - \mathbb{E}^*[h(\mathbb{T}_n^*)]|$ with the supremum taken over all $h \in BL_1$, uniformly in $t \in \mathbb{R}$,

$$\begin{aligned}\mathbb{P}\{\mathbb{T}_n \leqslant t\} &= \mathbb{E}^*\left[g_{t,\varepsilon}(\mathbb{T}_n^*)\right] + \mathbb{E}\left[g_{t,\varepsilon}(\mathbb{T}_n)\right] - \mathbb{E}^*\left[g_{t,\varepsilon}(\mathbb{T}_n^*)\right] \\ &\leqslant \mathbb{P}^*\{\mathbb{T}_n^* \leqslant t+\varepsilon\} + \delta_n/\varepsilon,\end{aligned}$$

since $g_{t,\varepsilon}$ has Lipschitz constant $1/\varepsilon$. A similar computation shows that $\mathbb{P}^*\{\mathbb{T}_n^* \leqslant t-\varepsilon\} - \delta_n/\varepsilon \leqslant \mathbb{P}\{\mathbb{T}_n \leqslant t\}$, so that, uniformly in t, and each $\varepsilon > 0$

$$\mathbb{P}^*\{\mathbb{T}_n^* \leqslant t-\varepsilon\} - \delta_n/\varepsilon \leqslant \mathbb{P}\{\mathbb{T}_n \leqslant t\} \leqslant \mathbb{P}^*\{\mathbb{T}_n^* \leqslant t+\varepsilon\} + \delta_n/\varepsilon \tag{2.11}$$

and in the same way we may prove

$$\mathbb{P}\{\mathbb{T}_n \leqslant t-\varepsilon\} - \delta_n/\varepsilon \leqslant \mathbb{P}^*\{\mathbb{T}_n^* \leqslant t\} \leqslant \mathbb{P}\{\mathbb{T}_n \leqslant t+\varepsilon\} + \delta_n/\varepsilon, \tag{2.12}$$

uniformly in t, and each $\varepsilon > 0$. For instance, if t^* is the bootstrap 95% critical value of $\mathbb{T}_n^*$, it is prudent to reject the null for values of $\mathbb{T}_n$ larger than $t^* + \varepsilon$.

Remark. The test for $\mathrm{H}_0 : C = C_0$ based on the critical regions $\{\mathbb{T}_n > c\}$ is consistent. Indeed, under the null, since $\mathbb{T}_n \leqslant L_n \sup_B |\mathbb{Z}_n(B)|$, we have $L_n^{-1}\mathbb{T}_n$ is bounded in probability, while under the alternative hypothesis, $\mathrm{H}_A : C = C_1$ for a fixed $C_1 \neq C_0$, we have that $\mathbb{T}_n \geqslant \sqrt{n}|C_0(B) - C_1(B)| - |\mathbb{Z}_n(B)|$, so that $n^{-1/2}\mathbb{T}_n \geqslant \frac{1}{2}|C_0(B) - C_1(B)|$, with probability tending to one, for any box B where C_0 and C_1 differ. Such a box exists under the alternative and the increasing sequence $\mathcal{F}_n$ likely contains at least one such box for relatively small n. The improved power of our test statistic is confirmed in our simulation study.

3. Parametric hypothesis

In this section we consider the problem of testing if the underlying copula C belongs to a parametric family $\mathcal{C} := \{C_\theta,\ \theta \in \Theta\}$. That is, the null hypothesis states that $C = C_{\theta_0}$ for some $\theta_0 \in \Theta$. Here $\Theta \subset \mathbb{R}^p$, equipped with the Euclidean norm $\|\cdot\|_2$. Suppose that we have a consistent estimator $\widehat{\theta} = \widehat{\theta}(\mathbb{H}_n)$ of θ_0.

Replacing C_0 by $C_{\widehat{\theta}}$ in the definition of the test statistic $\mathbb{T}_n$, we consider the process

$$\mathbb{Y}_n = \sqrt{n}(\mathbb{C}_n - C_{\widehat{\theta}}) = \mathbb{Z}_n - \sqrt{n}(C_{\widehat{\theta}} - C), \tag{3.1}$$

and its bootstrap version

$$\mathbb{Y}_n^* = \mathbb{Z}_n^* - \sqrt{n}(C_{\widehat{\theta}*} - C_{\widehat{\theta}}), \tag{3.2}$$

and (b) the limiting distribution function is continuous. While assumption (a) does not hold in our case, we only need that $\mathbb{E}[\min(\Delta_{n,\varepsilon}, \Delta_{n,\varepsilon}^*)] \to 0$. The quantity $\Delta_{n,\varepsilon}^*$ is computable, which allows for numerical verification (I.e., for a fixed ε and $n_1 < n_2 ...$, compute $\Delta_{n_k,\varepsilon}^*$ and verify if it decreases towards zero).

based on the *non-parametric* bootstrap estimate $\widehat{\theta}^* = \widehat{\theta}(\mathbb{H}_n^*)$, obtained after resampling with replacement from the original sample. Note that

$$\mathbb{Y}_n^* = \sqrt{n}(\mathbb{C}_n^* - C_{\widehat{\theta}*}) - \sqrt{n}(\mathbb{C}_n - C_{\widehat{\theta}}) \neq \sqrt{n}(\mathbb{C}_n^* - C_{\widehat{\theta}*}). \tag{3.3}$$

Indeed, the process $\sqrt{n}(\mathbb{C}_n^* - C_{\widehat{\theta}*})$, while perhaps a natural candidate, does not yield a consistent estimate of the distribution of $\mathbb{Y}_n$. Indeed, the "distance" between $\mathbb{Y}_n$ and the latter process will be of the order of $\mathbb{Z}_n^*$, thus asymptotically tight. On the other hand, the distance between $\mathbb{Y}_n$ and $\mathbb{Y}_n^*$ will be of the same order of magnitude as the distance between $\mathbb{Z}_n$ and $\mathbb{Z}_n^*$, that tends to zero (see the proof of Theorem 1).

We stress that our approach does not involve the *parametric* bootstrap, as studied by Genest and Rémillard (2008), to estimate the limiting law of copula-based statistics. In other words, we calculate $\widehat{\theta}^*$ after resampling from the empirical distribution $\mathbb{H}_n$, and not from the law given by the parametric copula $C_{\widehat{\theta}}$.

We impose some regularity on our parameter estimate $\widehat{\theta}$.

(C3) There exists a $\psi : \mathbb{R}^d \mapsto \mathbb{R}^p$ with $\int \|\psi\|_2^4 \,\mathrm{d}H < \infty$ such that

$$\widehat{\theta} - \theta_0 = \int \psi \, d(\mathbb{H}_n - H) + \varepsilon_n, \text{ and } \widehat{\theta}^* - \widehat{\theta} = \int \psi \, d(\mathbb{H}_n^* - \mathbb{H}_n) + \varepsilon_n^*,$$

under the null hypothesis, with $\|\varepsilon_n\|_2 = o_p(n^{-1/2}/L_n)$ and $\|\varepsilon_n^*\|_2 = o_{p*}(n^{-1/2}/L_n)$ in probability.

Note that the estimators satisfying (C3) are closely related to the estimators in the class $\mathcal{R}$ of regular estimators, as defined by Genest and Rémillard (2008).

Example (Estimators based on the inversion of Kendall's tau). As an example, we verify condition (C3) for estimators based on the inversion of Kendall's tau in the bivariate case ($d = 2$). Let $\theta = g(\tau)$ for some twice differentiable function g and Kendall's $\tau := 4\mathbb{E}[C_\theta(U, V)] - 1$, with the expectation taken over $(U, V) \sim C_\theta$. Kendall's τ is estimated empirically by

$$\widehat{\tau}_n := \frac{4}{n(n-1)} \sum_{i=1}^{n} \sum_{j=i+1}^{n} \mathbf{1}\left\{(Y_j - Y_i)(X_j - X_i) > 0\right\} - 1.$$

Then $U_n := \widehat{\tau}_n + 1$ is a U-statistic of order 2 for the kernel

$$h\left((x_1, y_1); (x_2, y_2)\right) = 2 \cdot \mathbf{1}\{(y_2 - y_1)(x_2 - x_1) > 0\}.$$

The projection of $U_n - \mathbb{E}[U_n]$ onto the space of all statistics of the form $\sum_{i=1}^n g_i(X_i, Y_i)$, for arbitrary measurable functions g_i with $\mathbb{E}[g_i^2(X, Y)] < \infty$, is

$$\hat{U}_n = \sum_{i=1}^{n} \mathbb{E}[U_n - \mathbb{E}[U_n] \,|\, X_i, Y_i] = \frac{2}{n} \sum_{i=1}^{n} \{\psi(X_i, Y_i) - \mathbb{E}[\psi(X_i, Y_i)]\}$$

with

$$\psi(x, y) = P(X < x, Y < y) + P(X > x, Y > y).$$

By Hájek's projection principle,

$$\mathrm{Var}(U_n - \widehat{U}_n) = \mathrm{Var}(U_n) - \mathrm{Var}(\widehat{U}_n).$$

From the proof of Theorem 12.3 in Van der Vaart (1998), due to Hoeffding (1948),

$$\mathrm{Var}(U_n) - \mathrm{Var}(\widehat{U}_n) = \frac{4(n-2)}{n(n-1)}\zeta_1 + \frac{2}{n(n-1)}\zeta_2 - \frac{4}{n}\zeta_1 = \frac{2\zeta_2 - 4\zeta_1}{n(n-1)}$$

with $\zeta_1 = \mathrm{Cov}(h(X,Y_1), h(X,Y_2))$ for X independent of Y_1 and Y_2, and with the same distribution as X_1, and $\zeta_2 = \mathrm{Var}(h(X_1,Y_1))$. Thus the difference is $\mathrm{Var}(U_n) - \mathrm{Var}(\widehat{U}_n)$ is of order $O(1/n^2)$. Consequently, $U_n - \mathbb{E}[U_n] = \widehat{U}_n + R_n$ with $R_n = O_p(1/n)$ so that

$$\widehat{\tau}_n - \tau = U_n - \mathbb{E}[U_n] = \widehat{U}_n + R_n = \frac{2}{n}\sum_{i=1}^n \{\psi(X_i,Y_i) - \mathbb{E}[\psi(X_i,Y_i)]\} + O_p(1/n).$$

Hence, if g is twice continuously differentiable in the neighborhood of τ, a limited expansion ensures that $\hat{\theta} := g(\widehat{\tau}_n)$ satisfies the first part of (C3). The second (bootstrap) part of (C3) follows from the same reasoning: We set $\widehat{\tau}_n^* := U_n^* - 1$ with

$$U_n^* = \frac{4}{n(n-1)}\sum_{i=1}^n \sum_{j=i+1}^n \mathbf{1}\left\{(Y_j^* - Y_i^*)(X_j^* - X_i^*) > 0\right\}$$

and for

$$\widehat{U}_n^* = \sum_{i=1}^n \mathbb{E}^*[U_n^* - \mathbb{E}^*[U_n^*] \,|\, X_i^*, Y_i^*]$$

we can show that

$$\mathrm{Var}^*(U_n^* - \widehat{U}_n^*) = \mathrm{Var}^*(U_n^*) - \mathrm{Var}^*(\widehat{U}_n^*)$$

is of order $O(1/n)$ almost surely, using the same arguments as above, keeping in mind that the empirical counterparts of ζ_1 and ζ_2 are bounded everywhere. Moreover, for

$$\psi_n(x,y) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}\{X_i < x, Y_i < y\} + \frac{1}{n}\sum_{i=1}^n \mathbf{1}\{X_i > x, Y_i > y\},$$

we find

$$\begin{aligned}
\widehat{U}_n^* &= \sum_{i=1}^n \mathbb{E}^*[U_n^* - \mathbb{E}^*[U_n^*] \,|\, X_i^*, Y_i^*] \\
&= \frac{2}{n}\sum_{i=1}^n \psi_n(X_i^*, Y_i^*) - \mathbb{E}^*[\psi_n(X_i^*, Y_i^*)] \\
&= \frac{2}{n}\sum_{i=1}^n \{\psi(X_i^*, Y_i^*) - \mathbb{E}^*[\psi(X_i^*, Y_i^*)]\} + \frac{2}{n}\sum_{i=1}^n \{(\psi_n - \psi)(X_i^*, Y_i^*) - \mathbb{E}^*[(\psi_n - \psi)(X_i^*, Y_i^*)]\}.
\end{aligned}$$

The second term on the right is of order $O_{p*}(1/n)$ as its variance equals

$$\frac{4}{n}\mathrm{Var}^*\left((\psi_n - \psi)(X_1^*, Y_1^*)\right) \leqslant \frac{4}{n}\sum_{i=1}^n (\psi_n - \psi)^2(X_i, Y_i) = O_{p*}(1/n^2),$$

by the reasoning in Bickel and Freedman (1981, p.1202). This implies

$$\widehat{\tau}_n^* - \widehat{\tau}_n = U_n^* - \mathbb{E}^*[U_n^*] = \frac{2}{n}\sum_{i=1}^n \{\psi(X_i^*, Y_i^*) - \mathbb{E}^*[\psi(X_i^*, Y_i^*)]\} + O_{p^*}(1/n).$$

Again, for a g that is twice continuously differentiable in the neighborhood of τ, a limited expansion ensures that $\hat{\theta}^*$ satisfies the second part of (C3).

Moreover, we need more regularity concerning $\theta \mapsto C_\theta$ itself.

(C4) For every $(s,t) \in [0,1]^d$, the function $\theta \mapsto C_\theta(\mathbf{u})$ has continuous partial derivatives $\dot{C}_\theta(\mathbf{u}) = (\partial/\partial\theta)C_\theta(\mathbf{u})$ that satisfy a Hölder condition with Hölder exponent $\nu > 0$ locally: there exists a constant $K < \infty$ such that

$$\sup_{\mathbf{u}} \|\dot{C}_\theta(\mathbf{u}) - \dot{C}_{\theta_0}(\mathbf{u})\|_2 \leqslant K\|\theta - \theta_0\|_2^\nu,$$

for every θ in a neighborhood of θ_0. Moreover, $\dot{C}_{\theta_0}$ is of bounded variation.

The regularity condition (C4) is satisfied for most of standard copula families. Simple calculations show that it is the case for the Gaussian-, Clayton- and the Frank-copula families in particular. Although copula partial derivatives with respect to their arguments often exhibit discontinuities or non-existence near their boundaries, justifying conditions such as (C1) (see Segers, 2012), the derivatives $\partial C_\theta(x,y)/\partial\theta$ with respect to the copula parameter θ behave a lot more regularly.

Theorem 4. *Let* $\mathbb{Y}_n = \{\mathbb{Y}_n(f),\ f \in \mathcal{F}_n\}$ *and* $\mathbb{Y}_n^* = \{\mathbb{Y}_n^*(f),\ f \in \mathcal{F}_n\}$ *with* $\mathcal{F}_n$ *in (2.1) as defined above. Assume that conditions (C1), (C2), (C3) and (C4) hold. Then, under the null hypothesis* $\mathrm{H}_0 : C = C_\theta,\ \theta \in \Theta$, *we have*

$$\lim_{n\to\infty} \mathbb{E}\left[\sup_{h\in BL_1} |\mathbb{E}[h(\mathbb{Y}_n)] - \mathbb{E}^*[h(\mathbb{Y}_n^*)]|\right] = 0. \tag{3.4}$$

This result implies that the distribution of the test statistic

$$\widehat{\mathbb{T}}_n = \sup_{f\in\mathcal{F}_n} |\mathbb{Y}_n(f)| = \sup_{B_1,\dots,B_{L_n}} \sum_{k=1}^{L_n} |\mathbb{Y}_n(B_k)| \tag{3.5}$$

can be "bootstrapped" by the distribution of

$$\widehat{\mathbb{T}}_n^* = \sup_{f\in\mathcal{F}_n} |\mathbb{Y}_n^*(f)| = \sup_{B_1,\dots,B_{L_n}} \sum_{k=1}^{L_n} |\mathbb{Y}_n^*(B_k)|. \tag{3.6}$$

Corollary 5. *Assume that conditions (C1), (C2), (C3) and (C4) hold. Then, under the null hypothesis* $\mathrm{H}_0 : C = C_\theta$, $\theta \in \Theta$,

$$\lim_{n\to\infty} \mathbb{E}\left[\sup_g \left|\mathbb{E}[g(\widehat{\mathbb{T}}_n)] - \mathbb{E}^*[g(\widehat{\mathbb{T}}_n^*)]\right|\right] = 0, \tag{3.7}$$

with the supremum taken over all Lipschitz functions $g : \mathbb{R} \to [-1,1]$ *with Lipschitz constant 1.*

Often, (C3) can be replaced by

(C3') There exists a $\psi : \mathbb{R}^d \mapsto \mathbb{R}^p$ with $\int \|\psi\|_2^4 \, dC < \infty$ such that

$$\widehat{\theta} - \theta_0 = \frac{1}{n}\sum_{i=1}^n \{\psi(\mathbb{F}_{n,1}(X_{i,1}), \dots, \mathbb{F}_{n,d}(X_{i,d})) - \mathbb{E}[\psi(F_1(X_{i,1}), \dots, F_d(X_{i,d}))]\} + \varepsilon_n,$$

$$\widehat{\theta}^* - \widehat{\theta} = \frac{1}{n}\sum_{i=1}^n \{\psi(\mathbb{F}^*_{n,1}(X^*_{i,1}), \dots, \mathbb{F}^*_{n,d}(X^*_{i,d})) - \psi(\mathbb{F}_{n,1}(X_{i,1}), \dots, \mathbb{F}_{n,d}(X_{i,d}))\} + \varepsilon^*_n,$$

under the null hypothesis, with $\|\varepsilon_n\|_2 = o_p(n^{-1/2}/L_n)$ and $\|\varepsilon^*_n\|_2 = o_{p*}(n^{-1/2}/L_n)$ in probability.

This is a consequence of the following result.

Proposition 6. *Assume (C1) holds. Any estimator $\widehat{\theta}$ satisfying (C3'), satisfies (C3).*

Copula parameters are typically estimated through pseudo-observations or ranks, without any assumption on the marginal distributions. For this reason the copula estimators that satisfy (C3') are relevant. They are very closely related to the estimators in the class $\mathcal{R}_1$ of Genest and Rémillard (2008). In particular, the maximum pseudo-likelihood estimator, that maximizes the pseudo log-likelihood function $\int \log c_\theta \, d\mathbb{C}_n$ over $\theta \in \Theta$, see, for instance, Genest et al. (1995) or Shih and Louis (1995), satisfies (C3') under suitable regularity conditions on the copula density c_θ.

Since the bootstrapped copula process $\mathbb{Y}^*_n$ is new, it is noteworthy to stress that it provides a valuable alternative to the usual parametric bootstrap. Now, assume $L_n = L$ is a constant, to retrieve the standard framework.

Corollary 7. *Assume that conditions (C1), (C3) and (C4) hold. Then, the process $\{\mathbb{Y}_n(\mathbf{u}), \mathbf{u} \in [0,1]^d\}$ tends weakly towards a Gaussian process in $\ell^\infty([0,1]^d)$. Moreover, the bootstrapped process $\{\mathbb{Y}^*_n(\mathbf{u}), \mathbf{u} \in [0,1]^d\}$ converges weakly to the same Gaussian process in probability in $\ell^\infty([0,1]^d)$.*

4. Applications and Numerical Studies

We present a limited numerical study, serving as a proof of principle rather than the final word on this subject. The evaluation of GOF tests in copula settings is a complex problem and only partial answers can be found in literature: see the surveys of Berg (2009), Genest et al. (2009) and, more recently, Fermanian (2012). Here, we restrict ourselves to the bivariate case. A full-scale numerical analysis is beyond the scope of this paper.

As we said in the introduction, we have implemented $\widetilde{\mathbb{T}}_n$, a computationally simpler version of $\mathbb{T}_n$: see Equation (1.6). Obviously, in the case of a composite zero assumption, we have implemented a simplified version of $\hat{\mathbb{T}}_n$ similarly. For convenience, it will be denoted by $\widetilde{\mathbb{T}}_n$ too. Since the distance (in probability) between $\mathbb{T}_n$ and $\widetilde{\mathbb{T}}_n$ tends to zero (see Appendix C), the weak convergence results are valid with $\widetilde{\mathbb{T}}_n$ instead of $\mathbb{T}_n$ or $\hat{\mathbb{T}}_n$. And the reasoning to approximate p-values by bootstrap still applies.

4.1. Heuristics

For two copula densities c_0 and c_1, we define the *difference* sets A^+ and A^- as

$$A^+ = \{(s,t) : c_0(s,t) > c_1(s,t)\}, \text{ and } A^- = \{(s,t) : c_0(s,t) < c_1(s,t)\}.$$

The proposed statistic $\mathbb{T}_n$ (or $\widetilde{\mathbb{T}}_n$) is designed to sample L_n boxes in order to maximize the difference between the "true" and postulated copulas. In situations where the geometry of the difference sets A^+ and A^- is complex, $\mathbb{T}_n$ can "pick out" disjoint subregions of A^+ and A^-, and one could expect superior performance consequently. However, sometimes just a single well placed box can pick essentially all the mass of sets A^+ or A^-, while the remaining $L_n - 1$ boxes are just collecting noise and consequently diminish the power of the statistic $\mathbb{T}_n$.

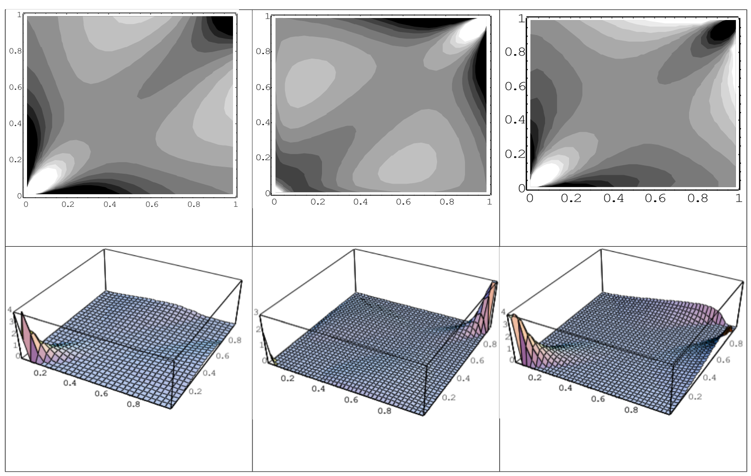

Fig 1. **Common comparisons**. *Copula density differences, through contour plots and 3D plots of synthetic data: Clayton - Frank (left), Gumbel - Frank (center), Clayton - Gumbel (right). Their Kendall's tau is* 0.4.

Most common scenarios encountered in the literature compare Frank, Clayton, Gumbel, and Gauss copulas with each other, after controlling for some dependence indicator (typically Kendall's tau): see, for instance, Berg (2009), Genest and Rémillard (2008) and Genest et al. (2009). However, all these pairings produce trivial difference sets A^+ and A^-, as revealed in the contour plots and 3D plots of $c_0 - c_1$ of Figure 1. We see that nearly all the mass difference between copula densities c_0 and c_1 is concentrated in a single spot, located in either the lower left or upper right corner. Here Kendall's $\tau = 0.4$, but we observed similar plots for different values of τ. Therefore, these common simulation scenarios are tailored towards many standard GOF tests such as KS and CvM tests. We are not aware of any argument that justifies such specific types of pairing, except for analytical tractability. Figures 2 and 3, however, paint a very different scenario with more elaborate difference sets A^+ and A^- that appear in real life situations. How often and to what extent this complex situation is encountered in reality is largely an open issue.

In this study, the copula densities c_1 were estimated by kernel density estimators based on the following data:

- The bivariate *ARCH*-like process $(X_1, Y_1), \dots, (X_n, Y_n)$, with $n = 10^6$, was generated as follows: First, we created independent $Z_i \sim N(0,1)$ and $W_i = Z_i(1 + 0.6W_{i-1}^2)^{1/2}$, with $W_0 = 0$. Second, we set $(X_i, Y_i) := (W_{100i}, W_{100i+1})$, creating nearly independent couples (of strongly dependent observations). Such models are commonly used in empirical finance, for instance.
- The *Mixture Copula* data $(X_1, Y_1), \dots, (X_n, Y_n)$, with $n = 10^6$, are generated from the mixture $c_1(s,t) = \frac{1}{2}c_F(s,t) + \frac{1}{2}c_F(1-s,t)$ for the Frank copula c_F with Kendall's $\tau = 0.4$. Therefore, this copula has asymmetrical features, contrary to most copulas that are tested in the literature. Obviously, other asymmetrical copulas could be built, following Liebscher (2008) for instance.
- The *Euro-Dollar* data $(X_1, Y_1), \dots, (X_n, Y_n)$, with $n = 1800$, are quoted currency exchange values. X is the daily percentage change of the Euro against the US dollar, while Y corresponds to the daily change of the Canadian dollar against the US dollar.
- The *Silver-Gold* data $(X_1, Y_1), \dots, (X_n, Y_n)$, with $n = 5000$, presents the log ratio of the average daily price of silver and gold futures respectively. For instance, $X_i = \log(S_{i+1}/S_i)$ based on the average price S_i of silver in US dollars on day i.

We compared *Mixture copula* and *ARCH* with the independence copula, for which $c_0(s,t) = 1$. In the case of real data (*Euro-Dollar* and *Silver-Gold*), we choose the Frank copula density with parameters $\tau = 2.6$ and $\tau = 3.4$, respectively, for c_0. The latter parameters were chosen after minimizing the (estimated) L_1-distance between c_0 and c_1. The difference sets are easily depicted by dark and bright sections of the contour plots, and the 3D plots clearly indicate that the mass difference between copula densities c_0 and c_1 is not concentrated in a single spot.

4.2. GOF tests in practice

We generated the data sets *ARCH* and *Mixture Copula* as described above. For each data set, we run two sets of simulations:

- (ARCH-S and Mixture-S) Test the simple null hypothesis $C_0(s,t) = st$ using the methodology of Section 2.
- (ARCH-C and Mixture-C) Test the composite null hypothesis that C_0 is a Frank copula using the procedure described in Section 3.

In both cases, the null hypothesis is wrong and should be rejected.

In our simulations, the statistics $\mathbb{T}_n$ and $\widehat{\mathbb{T}}_n$ are approximated (by $\widetilde{\mathbb{T}}_n$ in the case of $\mathbb{T}_n$: see Equation (1.6)). The number of boxes is $L_n = \lfloor \ln^{0.95}(n) \rfloor - 2$. We approximated the p-values of our statistics via the bootstrap procedures introduced in sections 2 and 3. For each approximation, we used 1,000 bootstrap samples. For the second set of simulations (ARCH-C and Mixture-C), we computed the parameters $\widehat{\theta}$ and $\widehat{\theta}^*$ by the usual pseudo-maximum likelihood procedure. Each procedure is repeated 100 times. We report the percentage of times that the computed p-value is

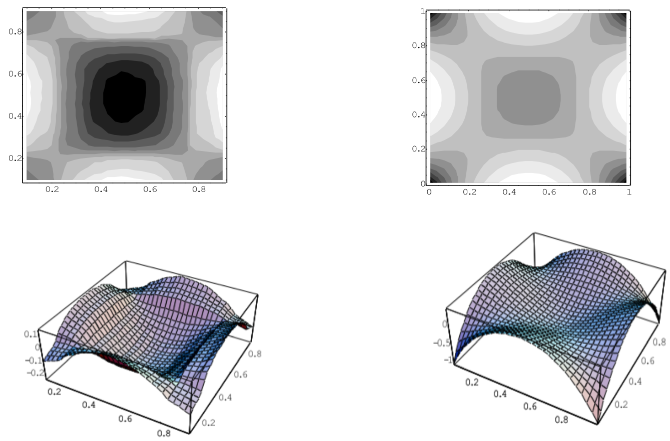

FIG 2. **Complex relation (synthetic data)**. *Copula density differences, through contour plots and 3D plots: ARCH (left) and mixture copula (right), compared to the independence copula.*

type	n	ARCH-S	ARCH-C	Mixture-S	Mixture-C
ATV	400	**75**%	**80**%	**41**%	**25**%
KS	400	6%	4%	8%	12%
CvM	400	25%	50%	6%	15%
ATV	800	**100**%	**99**%	**94**%	**98**%
KS	800	32%	50%	20%	25%
CvM	800	50%	92%	31%	84%

TABLE 1
Complex pairing, *related to Figure 2: relative frequencies of rejected null hypotheses under* $\alpha = 0.05$.

below $\alpha = 0.05$.

Our limited numerical study confirms the above assessment. Table 1 shows that the ATV test outperforms largely the KS and CvM tests in the case of complex pairing, while Table 2 confirms that the ATV test is inferior in case of the commonly used pairings of Figure 1. In Table 2, for each pair of copulas, say Clayton - Frank, we generated n observations from the first copula (Clayton), and we tested the null hypothesis that the second copula (Frank) is the true underlying copula. Table 3 shows that the significance level of the ATV test is below 0.05. The data were simulated from the null hypothesis. In all tables, Kendall's $\tau = 0.4$.

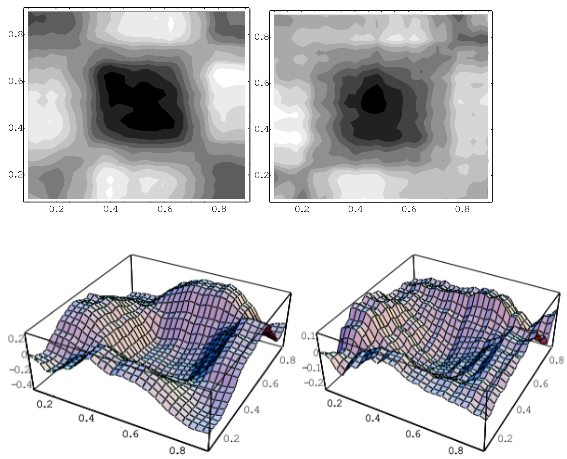

FIG 3. **Complex relation (actual data)**. *Copula density differences, through contour plots and 3D plots: Euro - Dollar (left) and Silver - Gold (right), compared to Frank copulas (with Kendall's tau equal to* 2.6 *and* 3.4 *respectively).*

5. Proofs

Throughout the proofs, we assume without loss of generality that $F_j = I$ for every $j = 1, \ldots, d$ (uniform marginal distributions). This implies that $H = C$. This is justified by the following lemma.

LEMMA 8. *Let F_j, $j = 1, \ldots, d$ be continuous distribution functions. Denote by $\tilde{H}$ the cdf of $(F_1(X_1), \ldots, F_d(X_d))$ and by $\tilde{C}$ its associated copula. The empirical copula associated to the sample $(F_1(X_{i1}), \ldots, F_d(X_{id}))$, $i = 1, \ldots, n$, is denoted by $\tilde{\mathbb{C}}_n$. We have*

$$C(\mathbf{u}) = \tilde{C}(\mathbf{u}) = \tilde{H}(\mathbf{u}) \textit{ for all } \mathbf{u} \in [0,1]^d.$$

type	n	Clayton - Frank	Gumbel - Frank	Clayton - Gumbel
ATV	400	42%	26%	88%
KS	400	58%	25%	90%
CvM	400	**84**%	**47**%	**95**%
ATV	800	92%	58%	94%
KS	800	98%	53%	98%
CvM	800	**100**%	**73**%	**100**%

TABLE 2
Trivial pairing, *related to Figure 1: relative frequencies of rejected null hypotheses under* $\alpha = 0.05$.

type	n	Clayton - Clayton	Gumbel - Gumbel	Frank -Frank
ATV	400	3%	2%	2%
KS	400	4%	5%	4%
CvM	400	4%	5%	4%
ATV	800	2%	4%	3%
KS	800	3%	3%	5%
CvM	800	5%	3%	6%

TABLE 3
Errors of the first kind: *relative frequencies of rejected null hypotheses under* $\alpha = 0.05$.

Moreover,

$$\mathbb{C}_n\left(\frac{i_1}{n},\ldots,\frac{i_d}{n}\right) = \tilde{\mathbb{C}}_n\left(\frac{i_1}{n},\ldots,\frac{i_d}{n}\right) \text{ for } i_1,\ldots,i_d \in \{0,1,\ldots,n\}.$$

Proof. This is a straightforward extension of Lemma 1 in Fermanian et al. (2004). □

Since the letter C is reserved for the copula function, we use the letters K, K_0, K_1, etc. in the sequel to denote generic constants, and we write $\|\mathbf{s}\|_\infty = \max_{1\leqslant j\leqslant d}|s_j|$ of $\mathbf{s} = (s_1,\ldots,s_d) \in [0,1]^d$.

5.1. Proof of preliminary results

In general, note that, for each $f \in \mathcal{F}_n$ defined in (2.1), we can write

$$\mathbb{Z}_n(f) = \sum_{k=1}^{L_n} c_k \mathbb{Z}_n(B_k) = \sum_{l=1}^{2^d L_n} \sigma_l \mathbb{Z}_n(\mathbf{s}_l),$$

and

$$\mathbb{Z}_n^*(f) = \sum_{l=1}^{2^d L_n} \sigma_l \mathbb{Z}_n^*(\mathbf{s}_l),$$

for some $\sigma_l \in \{-1,+1\}$ and $\mathbf{s}_l \in [0,1]^d$, using formula (1.5). Let $\alpha_n(\mathbf{u}) := \sqrt{n}(\mathbb{H}_n - H)(\mathbf{u}) = \sqrt{n}(\mathbb{H}_n(\mathbf{u}) - \mathbf{u})$ be the ordinary uniform empirical process in $[0,1]^d$, and let its oscillation modulus be defined as

$$\mathbb{M}_n(\delta) := \sup\left\{\left|\alpha_n(\mathbf{s}) - \alpha_n(\mathbf{s}')\right| : \|\mathbf{s}-\mathbf{s}'\|_\infty \leqslant \delta;\, \mathbf{s},\mathbf{s}' \in [0,1]^d\right\}, \tag{5.1}$$

for any $\delta > 0$.

LEMMA 9. *Let $(\delta_n)_{n\geqslant 0}$ be a sequence of positive real numbers such that $n\delta_n/\log n \to \infty$. Then, we have*

$$\mathbb{M}_n(\delta_n) = O(\delta_n^{1/2}(\log n)^{1/2}) \qquad \text{almost surely.}$$

Proof. We apply Proposition 14 with $\lambda_n = K_0\delta_n^{1/2}(\log n)^{1/2}$ for some constant $K_0 > 0$. Since $n^{-1/2}\lambda_n/\delta_n = K_0(\log n/(n\delta_n))^{1/2}$ tends to zero, this inequality can be rewritten

$$\mathbb{P}\{\mathbb{M}_n(\delta_n) > \lambda_n\} \leqslant \frac{K_1}{\delta_n}\exp\left(-\frac{K_2\psi(1)\lambda_n^2}{\delta_n}\right) = K_1 n \exp\left(-K_2K_0^2\psi(1)\log n\right),$$

for some constants K_1, K_2 and n sufficiently large. When K_0 is sufficiently large, we check that

$$\mathbb{P}\left\{\mathbb{M}_n(\delta_n) > \lambda_n\right\} \leqslant \frac{K_3}{n^2},$$

for some constant K_3. Invoke the Borel-Cantelli Lemma to conclude the proof. □

In addition, let $\alpha_{n,j}(u) = \sqrt{n}(\mathbb{F}_{n,j} - F_j)(u) = \sqrt{n}(\mathbb{F}_{n,j}(u) - u)$ be the ordinary uniform (marginal) empirical process in $[0,1]$, and we define

$$\widetilde{\mathbb{Z}}_n(\mathbf{s}) = \alpha_n(\mathbf{s}) - \sum_{j=1}^{d} C_j(\mathbf{s})\alpha_{n,j}(s_j). \tag{5.2}$$

Proposition 10. *Under conditions (C1) and (C2), we have*

$$\lim_{n\to\infty} \sup_{h\in BL_1} \left|\mathbb{E}[h(\mathbb{Z}_n)] - \mathbb{E}[h(\widetilde{\mathbb{Z}}_n)]\right| = 0.$$

Proof. First, we observe that

$$\sup_{h\in BL_1} \left|\mathbb{E}[h(\mathbb{Z}_n) - h(\widetilde{\mathbb{Z}}_n)]\right| \quad \leqslant \quad \delta + 2\mathbb{P}\left\{\sup_{f\in\mathcal{F}_n} |\mathbb{Z}_n(f) - \widetilde{\mathbb{Z}}_n(f)| > \delta\right\}.$$

The latter inequality holds for any $\delta > 0$, and uses the fact that $|h|$ is bounded by 1 and has Lipschitz constant 1. It remains to show that

$$\sup_{f\in\mathcal{F}_n} |\mathbb{Z}_n(f) - \widetilde{\mathbb{Z}}_n(f)| \to 0,$$

in probability, as $n \to \infty$. The remainder of the proof generalizes Proposition 4.2 of Segers (2012). Now, we note that

$$\sup_{f\in\mathcal{F}_n} |\mathbb{Z}_n(f) - \widetilde{\mathbb{Z}}_n(f)| \leqslant 2^d L_n \sup_{\mathbf{s}\in[0,1]^d} |\mathbb{Z}_n(\mathbf{s}) - \widetilde{\mathbb{Z}}_n(\mathbf{s})| \leqslant 2^d L_n(I + II)$$

with

$$\begin{aligned} I &= \sup_{\mathbf{s}\in[0,1]^d} \left|\alpha_n(\mathbb{F}_{n,1}^- s_1, \ldots, \mathbb{F}_{n,d}^- s_d) - \alpha_n(\mathbf{s})\right|, \\ II &= \sup_{\mathbf{s}\in[0,1]^d} \left|\sqrt{n}\left[C(\mathbb{F}_{n,1}^- s_1, \ldots, \mathbb{F}_{n,d}^- s_d) - C(\mathbf{s})\right] + \sum_{j=1}^{d} C_j(\mathbf{s})\alpha_{n,j}(s_j)\right|. \end{aligned}$$

The first term, I, can be bounded as follows. Set $\beta_{n,j}(s) = \sqrt{n}(\mathbb{F}_{n,j}^- s - s)$, $j = 1, \ldots, d$. By the Chung-Smirnov LIL, we have

$$\max_{1\leqslant j\leqslant d} \sup_{0\leqslant s\leqslant 1} |\beta_{n,j}(s)| = O((\log\log n)^{1/2}) \qquad \text{almost surely.}$$

Using Lemma 9 with $\delta = n^{-1/2}(\log\log n)^{1/2}$, we get

$$\sup_{\|\mathbf{s}-\mathbf{s}'\|_\infty < \delta} \left|\alpha_n(\mathbf{s}) - \alpha_n(\mathbf{s}')\right| = O(n^{-1/4}(\log n)^{1/2}(\log\log n)^{1/4}),$$

almost surely. This implies that $I = O(n^{-1/4}(\log n)^{1/2}(\log\log n)^{1/4})$, almost surely.

For the second term, we get by the mean value theorem that

$$\begin{aligned} II &= \sup_{\mathbf{s}\in[0,1]^d} \left| \sqrt{n}[C(\mathbb{F}^-_{n,1}s_1,\dots,\mathbb{F}^-_{n,d}s_d) - C(\mathbf{s})] + \sum_{j=1}^d C_j(\mathbf{s})\alpha_{n,j}(s_j) \right| \\ &\leqslant \sup_{\mathbf{s}\in[0,1]^d} \left| \sum_{j=1}^d C_j(\mathbf{s}_n)\beta_{nj}(s_j) + \sum_{j=1}^d C_j(\mathbf{s})\alpha_{n,j}(s_j) \right|, \end{aligned}$$

where $\mathbf{s}_n$ is a vector in $[0,1]^d$ s.t. $\|\mathbf{s}_n - \mathbf{s}\|_\infty \leqslant n^{-1/2}\max_{1\leqslant j\leqslant d}|\beta_{n,j}(s_j)|$. Since $|C_j| \leqslant 1$ for every $j = 1,\dots,d$ (because copulas are Lipschitz with Lipschitz constant 1), we deduce

$$\begin{aligned} II &\leqslant \sup_{\mathbf{s}\in[0,1]^d} \sum_{j=1}^d |\beta_{nj}(s_j) + \alpha_{n,j}(s_j)| + \sup_{\mathbf{s}\in[0,1]^d} \sum_{j=1}^d |[C_j(\mathbf{s}_n) - C_j(\mathbf{s})]\alpha_{n,j}(s_j)| \\ &\leqslant IIa + IIb. \end{aligned}$$

The Bahadur-Kiefer theorem (Shorack and Wellner, 2009, p. 585) states that

$$\max_{1\leqslant j\leqslant d} \sup_{0\leqslant s\leqslant 1} |\beta_{n,j}(s) + \alpha_{n,j}(s_j)| = O(n^{-1/4}(\log n)^{1/2}(\log\log n)^{1/4}) \quad \text{almost surely.}$$

Then, $IIa = O(n^{-1/4}(\log n)^{1/2}(\log\log n)^{1/4})$ almost surely.

Concerning IIb, we consider a positive sequence (ε_n), $\varepsilon_n \to 0$, that will be specified later independently of any $\mathbf{s} = (s_1,\dots,s_d) \in [0,1]^d$. For any index $j = 1,\dots,d$ and any $\mathbf{s} \in [0,1]^d$, we will distinguish the two cases: $s_j \in [\varepsilon_n, 1-\varepsilon_n]$ and the opposite.

If $s_j \in [\varepsilon_n, 1-\varepsilon_n]$ then

$$s_{nj} = s_j\left(1 + \frac{s_{nj} - s_j}{s_j}\right) \geqslant s_j\left(1 - \frac{|s_{nj} - s_j|}{\varepsilon_n}\right) \geqslant \frac{s_j}{2},$$

and

$$1 - s_{nj} \geqslant (1 - s_j)\left(1 - \frac{|s_{nj} - s_j|}{\varepsilon_n}\right) \geqslant \frac{1 - s_j}{2},$$

almost surely and for n sufficiently large, for all $\varepsilon_n \to 0$ and $n\varepsilon_n^2/\log n \to \infty$. Corollary 2 in Mason (1981) implies that

$$\max_{1\leqslant j\leqslant d} \sup_{0\leqslant s_j\leqslant 1} |s_j^{-1/2}(1-s_j)^{-1/2}\alpha_{n,j}(s_j)| \leqslant K(\log n)^{1/2}\log\log n,$$

almost surely, for some constant $K > 0$.

In this case, using condition (C1), we deduce,

$$\begin{aligned} |C_j(\mathbf{s}_n) - C_j(\mathbf{s})|\,|\alpha_{n,j}(s_j)| &\leqslant K_0\|\mathbf{s}_n - \mathbf{s}\|^r \left\{ s_j^{-\beta}(1-s_j)^{-\beta} + s_{nj}^{-\beta}(1-s_{nj})^{-\beta} \right\} |\alpha_{n,j}(s_j)| \\ &\leqslant K_1\|\mathbf{s}_n - \mathbf{s}\|^r s_j^{1/2-\beta}(1-s_j)^{1/2-\beta}(\log n)^{1/2}\log\log n \\ &\leqslant K_2 n^{-r/2}(\log\log n)^{r/2}\max(\varepsilon_n^{1/2-\beta}, 1)(\log n)^{1/2}\log\log n, \end{aligned}$$

almost surely, for some constants $K_0, K_1, K_2 > 0$ and every j.

If $s_j \notin [\varepsilon_n, 1-\varepsilon_n]$ then

$$\begin{aligned} |C_j(\mathbf{s}_n) - C_j(\mathbf{s})|\,|\alpha_{n,j}(s_j)| &\leqslant 2|\alpha_{n,j}(s_j)| \\ &\leqslant 2\varepsilon_n^{1/2} s_j^{-1/2}(1-s_j)^{-1/2}|\alpha_{n,j}(s_j)| \\ &\leqslant K\varepsilon_n^{1/2}(\log n)^{1/2}\log\log n \quad \text{almost surely,} \end{aligned}$$

see Corollary 2 in Mason (1981).

Combining all these bounds entails then

$$IIb \leqslant K_3\left[n^{-r/2}(\log\log n)^{r/2}\max(\varepsilon_n^{1/2-\beta},1) + \varepsilon_n^{1/2}\right](\log n)^{1/2}\log\log n,$$

with $K_3 > 0$. We now specify the choice of $\varepsilon_n = n^{-p}$, with p depending on β and r only. If $2\beta > 2r+1$, we take $0 < p < r/(2\beta-1)$. If $\beta < 1/2$, set $p = 1/4$. Otherwise, take $p = \min(1/4, r/(4\beta-2))$, for instance. In each case, these choices ensure that $IIb = O(n^{-q})$ almost surely, for some $q > 0$.

Since $L_n = O(\log n)$ by assumption (C2), we obtain $L_n(I + II) \to 0$ almost surely, as $n \to \infty$, and the proof is complete. $\square$

Next, we turn our attention to the bootstrap counterparts. We define $\alpha_n^*(\mathbf{s}) = \sqrt{n}(\mathbb{H}_n^* - \mathbb{H}_n)(\mathbf{s})$ as the ordinary bootstrap empirical process in $[0,1]^d$. We prove the following exponential inequality for the oscillation modulus

$$\mathbb{M}_n^*(\delta) = \sup_{\|\mathbf{s}-\mathbf{s}'\|_\infty < \delta} |\alpha_n^*(\mathbf{s}) - \alpha_n^*(\mathbf{s}')|.$$

Lemma 11. *For all bounded sequences δ_n such that $n\delta_n/\log(n) \to \infty$ as $n \to \infty$,*

$$\mathbb{M}_n^*(\delta_n) = O(\delta_n^{1/2}(\log n)^{1/2}) \qquad \textit{almost surely.} \tag{5.3}$$

Note that the sequence (δ_n) may be constant.

Proof. Since α_n^* is a step function, we find that

$$\sup_{\|\mathbf{s}-\mathbf{s}'\|_\infty < \delta_n} |\alpha_n^*(\mathbf{s}) - \alpha_n^*(\mathbf{s}')| = \max|\alpha_n^*(X_{i_1,1},\ldots,X_{i_d,d}) - \alpha_n^*(X_{i'_1,1},\ldots,X_{i'_d,d})|,$$

with the maximum taken over all $|X_{i_j,j} - X_{i'_j,j}| < \delta_n$, $j = 1,\ldots,d$, $i_1, i'_1, \ldots, i_d, i'_d \in \{1,\ldots,n\}$. For any $\mathbf{i} := (i_1,\ldots,i_d)$ and $\mathbf{i}' = (i'_1,\ldots,i'_d)$ in $\{1,\ldots,n\}^d$, we rewrite

$$|\alpha_n^*(X_{i_1,1},\ldots,X_{i_d,d}) - \alpha_n^*(X_{i'_1,1},\ldots,X_{i'_d,d})| = n^{-1/2}\sum_{k=1}^n \{V_{k,\mathbf{i},\mathbf{i}'} - \mathbb{E}^*[V_{k,\mathbf{i},\mathbf{i}'}]\},$$

as a sum of bounded independent random variables with

$$V_{k,\mathbf{i},\mathbf{i}'} := \mathbf{1}\{X_{k,j}^* \leqslant X_{i_j,j},\, j=1,\ldots,d\} - \mathbf{1}\{X_{k,j}^* \leqslant X_{i'_j,j},\, j=1,\ldots,d\},$$

conditionally on the sample $(\mathbf{X}_1,\dots,\mathbf{X}_n)$. Moreover, a simple calculation and Lemma 9 yield

$$\begin{aligned}
\mathrm{Var}^*(V_{k,\mathbf{i},\mathbf{i}'}) &\leqslant \sum_{j=1}^d \mathbb{P}^*\left\{X_{i_j,j}\leqslant X^*_{k,j}\leqslant X_{i'_j,j}\right\}\\
&\leqslant \sum_{j=1}^d \sup_{s_j}[\mathbb{F}_{n,j}(s_j+\delta_n)-\mathbb{F}_{n,j}(s_j)]\\
&\leqslant d\delta_n + dn^{-1/2}\mathbb{M}_n(\delta_n)\\
&\leqslant d\max(\delta_n,\mathbb{M}_n(\delta_n)/\sqrt{n})\\
&\leqslant K\max(\delta_n,\sqrt{\delta_n\log n}/\sqrt{n}) = K\delta_n,
\end{aligned}$$

for n large enough, for almost all realizations (X_i,Y_i) and for some constant $K>0$. Hence, by the union bound and Bernstein's exponential inequality for bounded random variables, we have, for some constant K_0,

$$\begin{aligned}
&\mathbb{P}^*\left\{\max_{\substack{\mathbf{i},\mathbf{i}'\in\{1,\dots,n\}^d\\ |X_{i_j,j}-X_{i'_j,j}|<\delta_n,\,\forall j}} |\alpha_n^*(X_{i_1,1},\dots,X_{i_d,d})-\alpha_n^*(X_{i'_1,1},\dots,X_{i'_d,d})|>x\right\}\\
&\leqslant 2n^{2d}\exp\left(-K_0(\sqrt{n}x\wedge x^2\delta_n^{-1})\right),
\end{aligned}$$

for all samples $(\mathbf{X}_1,\dots,\mathbf{X}_n)$. By integrating the previous inequality over $\mathbb{P}$, we get the same inequality, but replacing $\mathbb{P}^*$ by $\mathbb{P}$. Set $x=K_1\delta_n^{1/2}(\log n)^{1/2}$ and take a constant K_1 sufficiently large to obtain

$$\sum_{n=1}^{+\infty}\mathbb{P}\left\{\mathbb{M}_n^*(\delta_n)>K_1\delta_n^{1/2}(\log n)^{1/2}\right\}<+\infty.$$

Apply the Borel-Cantelli lemma to conclude the proof. □

Analogous to the approximation of the process $\mathbb{Z}_n$ by $\widetilde{\mathbb{Z}}_n$ before, we introduce a simpler process $\widetilde{\mathbb{Z}}_n^*$ to approximate $\mathbb{Z}_n^*$. Set

$$\widetilde{\mathbb{Z}}_n^*(\mathbf{s})=\sqrt{n}(\mathbb{H}_n^*-\mathbb{H}_n)(\mathbf{s})-\sum_{j=1}^d C_j(\mathbf{s})\sqrt{n}(\mathbb{F}^*_{n,j}-\mathbb{F}_{n,j})(s_j). \tag{5.4}$$

Proposition 12. *Under conditions (C1) and (C2), we have*

$$\lim_{n\to\infty}\mathbb{E}\left[\sup_{h\in BL_1}\left|\mathbb{E}^*[h(\mathbb{Z}_n^*)-h(\widetilde{\mathbb{Z}}_n^*)]\right|\right]=0.$$

Proof. First, we notice that, for any $\eta>0$,

$$\begin{aligned}
\mathbb{E}\left[\sup_{h\in BL_1}\left|\mathbb{E}^*[h(\mathbb{Z}_n^*)-h(\widetilde{\mathbb{Z}}_n^*)]\right|\right] &\leqslant \eta+2\mathbb{E}\left[\mathbb{P}^*\left\{\sup_{f\in\mathcal{F}_n}\left|\mathbb{Z}_n^*(f)-\widetilde{\mathbb{Z}}_n^*(f)\right|\geqslant\eta\right\}\right]\\
&\leqslant \eta+2\mathbb{E}\left[\mathbb{P}^*\left\{\sup_{\mathbf{s}}2^dL_n\left|\mathbb{Z}_n^*(\mathbf{s})-\widetilde{\mathbb{Z}}_n^*(\mathbf{s})\right|\geqslant\eta\right\}\right].
\end{aligned}$$

Some straightforward adding and subtracting yields $\mathbb{Z}_n^*(\mathbf{s}) = \bar{\mathbb{Z}}_n^*(\mathbf{s}) + R_n^*(\mathbf{s})$ with

$$\bar{\mathbb{Z}}_n^*(\mathbf{s}) = \sqrt{n}\{\mathbb{H}_n^*(\mathbf{s}) - \mathbb{H}_n(\mathbf{s})\} - \sqrt{n}\{C(\mathbb{F}_{n,1}^* s_1, \ldots, \mathbb{F}_{n,d}^* s_d) - C(\mathbb{F}_{n,1} s_1, \ldots, \mathbb{F}_{n,d} s_d)\}$$

and $R_n^*(\mathbf{s}) = R_{n,1}^*(\mathbf{s}) + R_{n,2}^*(\mathbf{s}) + R_{n,3}^*(\mathbf{s}) + R_{n,4}^*(\mathbf{s})$ with

$$\begin{aligned}
R_{n,1}^*(\mathbf{s}) &= \alpha_n^*(\mathbb{F}_{n,1}^{*-} s_1, \ldots, \mathbb{F}_{n,d}^{*-} s_d) - \alpha_n^*(\mathbb{F}_{n,1}^{-} s_1, \ldots, \mathbb{F}_{n,d}^{-} s_d) \\
R_{n,2}^*(\mathbf{s}) &= \alpha_n^*(\mathbb{F}_{n,1}^{-} s_1, \ldots, \mathbb{F}_{n,d}^{-} s_d) - \alpha_n^*(\mathbf{s}) \\
R_{n,3}^*(\mathbf{s}) &= \alpha_n(\mathbb{F}_{n,1}^{*-} s_1, \ldots, \mathbb{F}_{n,d}^{*-} s_d) - \alpha_n(\mathbb{F}_{n,1}^{-} s_1, \ldots, \mathbb{F}_{n,d}^{-} s_d) \\
R_{n,4}^*(\mathbf{s}) &= \sqrt{n}\left\{C(\mathbb{F}_{n,1}^{*-} s_1, \ldots, \mathbb{F}_{n,d}^{*-} s_d) - C(\mathbb{F}_{n,1}^{-} s_1, \ldots, \mathbb{F}_{n,d}^{-} s_d)\right\} \\
&\quad + \sqrt{n}\left\{C(\mathbb{F}_{n,1}^{*} s_1, \ldots, \mathbb{F}_{n,d}^{*} s_d) - C(\mathbb{F}_{n,1} s_1, \ldots, \mathbb{F}_{n,d} s_d)\right\}.
\end{aligned}$$

Let $\alpha_{n,j}^*(s) = \sqrt{n}(\mathbb{F}_{n,j}^* - \mathbb{F}_{n,j})(s)$ and $\beta_{n,j}^*(s) = \sqrt{n}(\mathbb{F}_{n,j}^{-*} - \mathbb{F}_{n,j}^{-})(s)$ be the bootstrap versions of the empirical processes $\alpha_{n,j}(s)$ and $\beta_{n,j}(s)$, respectively. Both converge to the same weak limit as

$$\sup_{0 \leqslant s_j \leqslant 1} |\beta_{n,j}^*(s_j) + \alpha_{n,j}^*(s_j)| = O(n^{-1/4}(\log n)^{1/2}(\log\log n)^{1/4}) \quad \text{almost surely,}$$

see displays (2.10') and (2.12') in Theorem 2.1 of Csörgó and Mason (1989). It remains to show that $\mathbb{P}^*\{L_n \sup_{\mathbf{s}} |R_n^*(\mathbf{s})| > \eta\} \to 0$ for all $\eta > 0$, conditionally given all sequences $(\mathbf{X}_1, \ldots, \mathbf{X}_n) \in \Omega_n$ for some sequence of events $\Omega_n \subset \mathbb{R}^{d\times n}$ with $\lim_{n\to\infty} \mathbb{P}(\Omega_n) = 1$.

Let $\delta_n = n^{-1/4}$. (Other choices are possible as well.) We have

$$\limsup_{n\to\infty} \mathbb{P}^*\{L_n \|R_{n,1}^*\|_\infty \geqslant \eta\} \leqslant \limsup_{n\to\infty} \mathbb{P}^*\{L_n \mathbb{M}_n^*(\delta_n) \geqslant \eta\} + \limsup_{n\to\infty} \mathbb{P}^*\{\max_j \|\beta_{n,j}^*\|_\infty \geqslant \sqrt{n}\delta_n\} = 0,$$

by Lemma 11. Next, on the event $\max_j \|\beta_{n,j}\|_\infty \leqslant \sqrt{n}\delta_n$ (that holds almost surely by the law of iterated logarithm),

$$\limsup_{n\to\infty} \mathbb{P}^*\{L_n \|R_{n,2}^*\|_\infty \geqslant \eta\} \leqslant \limsup_{n\to\infty} \mathbb{P}^*\{L_n \mathbb{M}_n^*(\delta_n) \geqslant \eta\} = 0,$$

by Lemma 11. On the event $L_n \mathbb{M}_n(\delta_n) < \eta$ (that holds almost surely by Lemma 9), we have

$$\limsup_{n\to\infty} \mathbb{P}^*\{L_n \|R_{n,3}^*\|_\infty \geqslant \eta\} \leqslant \limsup_{n\to\infty} \mathbb{P}^*\{\max_j \|\beta_{n,j}^*\|_\infty > \sqrt{n}\delta_n\} = 0$$

by the weak convergence of $\beta_{n,j}^*$. Finally, for some s_j^* between $\mathbb{F}_{n,j}^{-*}(s_j)$ and $\mathbb{F}_{n,j}^{-}(s_j)$, and s_j^{**} between $\mathbb{F}_{n,j}^*(s_j)$ and $\mathbb{F}_{n,j}(s_j)$, we have

$$\begin{aligned}
|R_{n,4}^*(\mathbf{s})| &= \left|\sum_{j=1}^d \{C_j(s_j^*)\beta_{n,j}^*(s_j) + C_j(s_j^{**})\alpha_{n,j}^*(s_j)\right| \\
&\leqslant \sum_{j=1}^d |\beta_{n,j}^*(s_j) + \alpha_{n,j}^*(s_j)| + \sum_{j=1}^d |\alpha_{n,j}^*(s_j)||C_j(s_j^*) - C_j(s_j^{**})|.
\end{aligned}$$

The first term is of order $O(n^{-1/4}(\log n)^{1/2}(\log\log n)^{1/4})$, uniformly in s_j. For the second term, we argue as in the proof of Proposition 10. First, we observe that $|s_j^{**} - s_j^*| \leqslant |s_j^* - s_j| + |s_j^{**} - s_j|$ is of order $O_{p*}(n^{-1/2})$. Second, since the class $\mathbf{1}\{x \leqslant t\}t^{-b}(1-t)^{-b}$ is a P-Donsker class for the uniform probability measure P on $[0,1]$, for all $0 \leqslant b < 1/2$, see Van der Vaart and Wellner (1996), Example 2.11.15 (page 214), the weak convergence of the bootstrap empirical process [Van der Vaart and Wellner (1996, Theorem 3.6.1, page 347)] implies that

$$\sup_{0<s<1} |\alpha^*_{n,j}(s)|/(s^b(1-s)^b) = O_{p*}(1).$$

Consequently, as in the proof of Proposition 10, we find that, for some constant $K < \infty$,

$$\sup_{\varepsilon_n \leqslant s_j \leqslant 1-\varepsilon_n} |\alpha^*_{n,j}(s_j)||C_j(s^*_j) - C_j(s^{**}_j)| \leqslant K|s^{**}_j - s^*_j|^r s_j^{b-\beta}(1-s_j)^{b-\beta} \sup_{s_j} |\alpha^*_{n,j}(s_j)|/(s^b(1-s)^b)$$

which is of order $O_{p*}(1) \cdot \max(n^{-r/2} \max(1, \varepsilon_n^{b-\beta})$. On the other side,

$$\begin{aligned}\sup_{s_j \notin [\varepsilon, 1-\varepsilon_n]} |\alpha^*_{n,j}(s_j)||C_j(s^*_j) - C_j(s^{**}_j)| &\leqslant 2 \sup_{s_j \notin [\varepsilon, 1-\varepsilon_n]} |\alpha^*_{n,j}(s_j)| \\ &\leqslant 2\varepsilon_n^b \sup_{s_j} |\alpha^*_{n,j}(s_j)|/(s^b(1-s)^b),\end{aligned}$$

which is of order $O_{p*}(\varepsilon_n^b)$. Combining both bounds yields $\sup_{\mathbf{s}} |R_{n,4}(\mathbf{s})| = O_{p*}(\varepsilon_n^b + n^{-r/2} \max(1, \varepsilon_n^{b-\beta}))$. Taking $\varepsilon_n = n^{-p}$ with p depending on b, β and r, we get that $\lim_{n\to\infty} \mathbb{P}^*\{L_n \sup_{\mathbf{s}} |R^*_{n,4}(\mathbf{s})| \geqslant \eta\} = 0$ for all $\eta > 0$, conditionally on all sequences $(\mathbf{X}_1, \ldots, \mathbf{X}_n) \in \Omega_n$ for some sequence of events Ω_n with $\lim_n \mathbb{P}(\Omega_n) = 1$. This completes our proof. □

5.2. Proof of Theorem 1

By triangle inequality, we have,

$$\begin{aligned}\mathbb{E}\left[\sup_{h\in BL_1} |h(\mathbb{Z}_n)] - \mathbb{E}^*[h(\mathbb{Z}^*_n)]|\right] &\leqslant \sup_{h\in BL_1} \left|\mathbb{E}[h(\mathbb{Z}_n) - h(\widetilde{\mathbb{Z}}_n)]\right| \\ &+\mathbb{E}\left[\sup_{h\in BL_1} \left|\mathbb{E}[h(\widetilde{\mathbb{Z}}_n)] - \mathbb{E}^*[h(\widetilde{\mathbb{Z}}^*_n)]\right|\right] \\ &+\mathbb{E}\left[\sup_{h\in BL_1} \left|\mathbb{E}^*[h(\widetilde{\mathbb{Z}}^*_n) - h(\mathbb{Z}^*_n)]\right|\right].\end{aligned}$$

In view of Proposition 10 and Proposition 12, it remains to show that the second term on the right is asymptotically negligible. We recall that

$$\widetilde{\mathbb{Z}}_n(f) = \sum_{k=1}^{2^d L_n} \sigma_k \widetilde{\mathbb{Z}}_n(\mathbf{s}_k) = \sum_{k=1}^{2^d L_n} \sigma_k \int f_k(\mathbf{x})\, d\alpha_n(\mathbf{x}),$$

for

$$f_k(\mathbf{x}) = \mathbf{1}\{\mathbf{x} \leqslant \mathbf{s}_k\} - \sum_{j=1}^d C_j(\mathbf{s}_k)\mathbf{1}\{x_j \leqslant s_{k,j}\}.$$

Now, let $h_f(\mathbf{x}) = \sum_{k=1}^{2^d L_n} \sigma_k f_k(\mathbf{x})$ so that

$$\widetilde{\mathbb{Z}}_n(f) = \int h_f\, d\alpha_n, \tag{5.5}$$

and we can derive in the same way

$$\widetilde{\mathbb{Z}}_n^*(f) = \int h_f \, d\alpha_n^*. \tag{5.6}$$

We now apply Theorem 3 in Radulović (2012), stated as Theorem 13 in the appendix for convenience. We need to verify that

- the $d+1$ classes

$$\begin{aligned} \mathcal{G}_k^a &= \left\{ \mathbf{1}\{\mathbf{x} \leqslant \mathbf{s}_k\}, \ \mathbf{s}_k \in [0,1]^d \right\}, \\ \mathcal{G}_k^{(j)} &= \left\{ C_j(\mathbf{s}_k)\mathbf{1}\{x \leqslant s_{k,j}\}, \ \mathbf{s}_k \in [0,1]^d \right\}, \ j = 1, \dots, d, \end{aligned}$$

  have VC–indices V_k^a and $V_k^{(j)}$, respectively, with $\sum_{k=1}^{2^d L_n} (V_k^a + \sum_{j=1}^d V_k^{(j)}) \leqslant K(\log n)^\gamma$ for some finite constant K and some $0 < \gamma < 1$.
- the class $\mathcal{H}_n = \{h_f : \ f \in \mathcal{F}_n\}$ has an envelope $H(\mathbf{x})$ with $\mathbb{E}[H^4(\mathbf{X})] < \infty$.

First we verify the VC property. The class $\mathcal{G}_k^a$ is VC with VC-dimension $V_k^a = d+1$ (Van der Vaart and Wellner, 2000, page 135), while the class $\mathcal{G}_k^{(j)}$ is a subclass of the class of functions $c\mathbf{1}\{a \leqslant x \leqslant b\}$ with $a, b \in \mathbb{R}$ and $c > 0$. This class has a VC index 3 : see van der Vaart and Wellner (2000), Problem 20, page 153. Consequently

$$\sum_{k=1}^{2^d L_n} (V_k^a + \sum_{j=1}^d V_k^{(j)}) \leqslant (4d+1)2^d L_n \leqslant K(\log n)^\gamma$$

for some $K < \infty$.

It remains to verify the envelope condition. We will show that $h_f(\mathbf{x})$ has envelope $1 + d + \sum_{j=1}^d TV(C_j)$. Writing

$$g_{\mathbf{x}}(\mathbf{s}) \quad = \quad \mathbf{1}\{\mathbf{x} \leqslant \mathbf{s}\} - \sum_{j=1}^d C_j(\mathbf{s})\mathbf{1}\{x_j \leqslant s_j\},$$

we see that

$$h_f(\mathbf{x}) = \sum_{k=1}^{L_n} c_k g_{\mathbf{x}}(B_k)$$

for $c_k = \pm 1$ and the operation $\phi(B_k)$ defined in (1.5) for any function $\phi : \mathbb{R}^d \to \mathbb{R}$. Furthermore, writing

$$\gamma_{\mathbf{x}}(\mathbf{s}) = \mathbf{1}\{\mathbf{x} \leqslant \mathbf{s}\}, \ \ \zeta_x^{(j)}(\mathbf{s}) = C_j(\mathbf{s})\mathbf{1}\{x \leqslant s_j\}, \ j = 1, \dots, d,$$

we have

$$\begin{aligned} |h_f(\mathbf{x})| &\leqslant \sum_{k=1}^{L_n} |\gamma_{\mathbf{x}}(B_k)| + \sum_{j=1}^d \sum_{k=1}^{L_n} |\zeta_{x_j}^{(j)}(B_k)| \\ &\leqslant 1 + \sum_{j=1}^d \sum_{k=1}^{L_n} |\zeta_{x_j}^{(j)}(B_k)| \end{aligned}$$

since the boxes $B_1, \dots, B_{L_n}$ are disjoint. Since each B_k is of the form $\prod_{j=1}^d (s^1_{k,j}, s^2_{k,j}]$, there is a (fine enough) lattice partition Π of $[0,1]^d$ with the property that each B_k can be written as a union of (disjoint) elements A_{k_j}, with $A_{k_j} \in \Pi$. A little reflexion shows that, for each $1 \leqslant j \leqslant d$,

$$\sum_{k=1}^{L_n} |\zeta^{(j)}_{x_j}(B_k)| \quad \leqslant \quad \sum_{A \in \Pi} |\zeta^{(j)}_{x_j}(A)|$$

and, moreover, for $A_m = \prod_{j=1}^d (s^1_{m,j}, s^2_{m,j}] \in \Pi$, $A_{m,-j} = \prod_{l \neq j} (s^1_{m,l}, s^2_{m,l}]$ and

$$C_j(\mathbf{s}_{-j}|t) := C_j(s_1, \dots, s_{j-1}, t, s_{j+1}, \dots, s_d),$$

for every $\mathbf{s}_{-j} \in [0,1]^{d-1}$ and every $t \in [0,1]$, a little algebra gives the identity

$$\zeta^{(j)}_{x_j}(A_m) \quad = \quad \mathbf{1}\{x_j \leqslant s^2_{m,j}\} C_j(A_m) + \mathbf{1}\{s^1_{m,j} < x_j \leqslant s^2_{m,j}\} C_j(A_{m,-j}|s^1_{m,j}).$$

Since

$$C_j(\mathbf{s}_{-j}|s_j) = \mathbb{P}\{\mathbf{X}_{-j} \leqslant \mathbf{s}_{-j} \,|\, X_j = s_j\},$$

we obtain

$$\begin{aligned}
\sum_{k=1}^{L_n} |\zeta^{(j)}_{x}(B_k)| \quad &\leqslant \quad \sum_{A_m \in \Pi} |\zeta^{(j)}_{x}(A_m)| \\
&\leqslant \quad \sum_{A_m \in \Pi} |C_j(A_m)| + \sum_{A_m \in \Pi} \mathbf{1}\{s^1_{m,j} < x_j \leqslant s^2_{m,j}\} \mathbb{P}\{\mathbf{X}_{-j} \in A_{m,-j} \,|\, X_j = s^1_{m,j}\} \\
&\leqslant \quad \mathrm{TV}(C_j) + \sum_{A_m \in \Pi} \mathbf{1}\{s^1_{m,j} < x_j \leqslant s^2_{m,j}\} \mathbb{P}\{\mathbf{X}_{-j} \in A_{m,-j} \,|\, X_j = s^1_{m,j}\}.
\end{aligned}$$

Let $A_{\mathbf{x}} \in \Pi$ with $\mathbf{x} \in A_{\mathbf{x}}$ and $s^1_j < x \leqslant s^2_j$ with $(s^1_j, s^2_j]$ be the projection of $A_{\mathbf{x}}$ on the j-th axis of the lattice. Then, the last term on the right of the previous display can be bounded as follows:

$$\begin{aligned}
&\sum_{A_m \in \Pi} \mathbf{1}\{s^1_{m,j} < x_j \leqslant s^2_{m,j}\} \mathbb{P}\{\mathbf{X}_{-j} \in A_{m,-j} \,|\, X_j = s^1_{m,j}\} \\
&\leqslant \sum_{A_m \in \Pi,\ s^1_{m,j} = s^1_j,\ s^2_{m,j} = s^2_j} \mathbb{P}\{\mathbf{X}_{-j} \in A_{m,-j} \,|\, X_j = s^1_j\} \\
&\leqslant 1
\end{aligned}$$

since the boxes $A_m \in \Pi$ and therefore $A_{m,-j}$ are disjoint. We have shown that the class $\mathcal{H}_n$ has envelope $1 + d + \sum_{j=1}^d TV(C_j)$.
We can now apply Theorem 13 to conclude that

$$\lim_{n \to \infty} \mathbb{E}\left[\sup_{h \in BL_1} \left| \mathbb{E}[h(\widetilde{\mathbb{Z}}_n)] - \mathbb{E}^*[h(\widetilde{\mathbb{Z}}^*_n)] \right| \right] = 0,$$

and the proof is complete. □

5.3. Proof of Theorem 4

We proceed as in the proof of Theorem 1. We write $\widehat{C} = C_{\widehat{\theta}}$ and $\widehat{C}^* = C_{\widehat{\theta}*}$. Recall that

$$\mathbb{Y}_n = \mathbb{Z}_n - \sqrt{n}(\widehat{C} - C).$$

We may replace $\mathbb{Z}_n$ by $\widetilde{\mathbb{Z}}_n$ with impunity since

$$\begin{aligned}
& \sup_{h\in BL_1} \left| \mathbb{E}[h(\mathbb{Y}_n) - h(\widetilde{\mathbb{Z}}_n - \sqrt{n}(\widehat{C} - C))] \right| \\
\leqslant \quad & \delta + 2\mathbb{P}\left\{ \sup_{f\in\mathcal{F}_n} \left| \mathbb{Y}_n(f) - \widetilde{\mathbb{Z}}_n(f) + \sqrt{n}(\widehat{C} - C)(f) \right| \geqslant \delta \right\} \\
= \quad & \delta + 2\mathbb{P}\left\{ \sup_{f\in\mathcal{F}_n} \left| \mathbb{Z}_n(f) - \widetilde{\mathbb{Z}}_n(f) \right| \geqslant \delta \right\} \\
\to \quad & \delta \text{ as } n \to \infty,
\end{aligned}$$

for every $\delta > 0$, as in the proof of Proposition 10. Next, by the mean value theorem and assumptions (C3) and (C4), we have

$$\begin{aligned}
\sqrt{n}(\widehat{C} - C)(\mathbf{s}) \quad &= \quad \sqrt{n}(\widehat{\theta} - \theta_0)'\dot{C}_{\theta_0}(\mathbf{s}) + \sqrt{n}(\widehat{\theta} - \theta_0)'\{\dot{C}_{\tilde{\theta}}(\mathbf{s}) - \dot{C}_{\theta_0}(\mathbf{s})\} \\
& \qquad \text{for some } \tilde{\theta} \text{ between } \widehat{\theta} \text{ and } \theta_0 \\
&= \quad \left(\int \psi \, d\alpha_n + n^{1/2}\varepsilon_n \right)' \dot{C}_{\theta_0}(\mathbf{s}) + \sqrt{n}(\widehat{\theta} - \theta_0)'\{\dot{C}_{\tilde{\theta}}(\mathbf{s}) - \dot{C}_{\theta_0}(\mathbf{s})\} \\
&= \quad \left(\int \psi \, d\alpha_n \right)' \dot{C}_{\theta_0}(\mathbf{s}) + R_n(\mathbf{s})
\end{aligned}$$

for some remainder term R_n that satisfies

$$\begin{aligned}
|R_n(\mathbf{s})| \quad &\leqslant \quad n^{1/2}\|\varepsilon_n\|_2 \|\dot{C}_{\theta_0}(\mathbf{s})\|_2 + Kn^{1/2}\|\widehat{\theta} - \theta_0\|_2^{1+\nu} \\
&= \quad O_p(n^{1/2}\|\varepsilon_n\|_2 + n^{-\nu/2}) \\
&= \quad o_p(1/L_n).
\end{aligned}$$

This bound holds uniformly in $\mathbf{s}$. Consequently, for

$$\widetilde{\mathbb{Y}}_n(f) = \sum_{k=1}^{2^d L} \sigma_k \widetilde{\mathbb{Y}}_n(\mathbf{s}_k)$$

based on

$$\widetilde{\mathbb{Y}}_n(\mathbf{s}) \quad = \quad \widetilde{\mathbb{Z}}_n(\mathbf{s}) - \left(\int \psi \, d\alpha_n \right)' \dot{C}_{\theta_0}(\mathbf{s}),$$

we have

$$\sup_{h\in BL_1} \left| \mathbb{E}[h(\widetilde{\mathbb{Z}}_n - \sqrt{n}(\widehat{C} - C))] - \mathbb{E}[h(\widetilde{\mathbb{Y}}_n)] \right| \quad = \quad \sup_{h\in BL_1} \left| \mathbb{E}[h(\widetilde{\mathbb{Y}}_n - R_n)] - \mathbb{E}[h(\widetilde{\mathbb{Y}}_n)] \right|.$$

Since

$$\sup_f |R_n(f)| \leqslant 2^d L_n \sup_{\mathbf{s}} |R_n(\mathbf{s})| \to 0$$

in probability, we get $\sup_h |\mathbb{E}[h(\widetilde{\mathbb{Z}}_n - \sqrt{n}(\widehat{C} - C))] - \mathbb{E}[h(\widetilde{\mathbb{Y}}_n)]| \to 0$, as $n \to \infty$. We conclude that

$$\limsup_{n\to\infty} \sup_{h\in BL_1} \left| \mathbb{E}[h(\mathbb{Y}_n)] - \mathbb{E}[h(\widetilde{\mathbb{Y}}_n)] \right| = 0.$$

For the bootstrap counterpart, we can argue in the same way. Using the expansion

$$\sqrt{n}(\widehat{C}^* - \widehat{C})(\mathbf{s}) \quad = \quad \left(\int \psi \, d\alpha_n^*\right)' \dot{C}_{\theta_0}(\mathbf{s}) + R_n^*(\mathbf{s})$$

for some remainder term R_n^* that satisfies

$$\sup_{\mathbf{s}} |R_n^*(\mathbf{s})| \quad \leqslant \quad K_0 n^{1/2} \|\varepsilon_n^*\|_2 + K_1 n^{1/2} \|\widehat{\theta} - \theta_0\|_2^{1+\nu} + K_2 n^{1/2} \|\widehat{\theta}^* - \widehat{\theta}\|_2^{1+\nu},$$

for some finite constants K_0, K_1 and K_2. We check that the processes $\mathbb{Y}_n^*$ and $\widetilde{\mathbb{Y}}_n^*$ are close with $\widetilde{\mathbb{Y}}_n^*$ based on

$$\widetilde{\mathbb{Y}}_n^*(\mathbf{s}) \quad = \quad \widetilde{\mathbb{Z}}_n^*(\mathbf{s}) - \left(\int \psi \, d\alpha_n^*\right)' \dot{C}_{\theta_0}(\mathbf{s}).$$

Note that $\widetilde{\mathbb{Y}}_n(f) = \sum_k \sigma_k \widetilde{\mathbb{Y}}_n(\mathbf{s}_k) = \int \left(\sum_k \sigma_k g_k\right) \, d\alpha_n$ with

$$g_k(\mathbf{x}) \quad = \quad \mathbf{1}\{\mathbf{x} \leqslant \mathbf{s}_k\} - \sum_{j=1}^d C_j(\mathbf{s}_k) \mathbf{1}\{x \leqslant s_{k,j}\} - (\psi(\mathbf{x}))' \dot{C}_{\theta_0}(\mathbf{s}_k).$$

As in the proof of Theorem 1, it remains to verify the two conditions of Theorem 13. Since the only difference with the proof of Theorem 1 is the addition of the term $(\psi(\mathbf{x}))' \dot{C}_{\theta_0}(\mathbf{s}_k)$, we concentrate on the class of functions $(\psi(\mathbf{x}))' \dot{C}_{\theta_0}(\mathbf{s}_k)$. Since it is a subclass of $c'\psi(\mathbf{x})$ with $c \in \mathbb{R}^p$, its VC dimension trivially is equal to p. Moreover, it is not hard to see from the proof of Theorem 1 that

$$\left| \sum_{k=1}^{2^d L_n} \sigma_k g_k(\mathbf{x}) \right| \quad \leqslant \quad 1 + d + \sum_{j=1}^d TV(C_j) + \|\psi(\mathbf{x})\| TV(\dot{C}_{\theta_0}).$$

Since $\mathbb{E}[\|\psi(\mathbf{X})\|_2^4] < \infty$, the conditions of Theorem 13 are met, and we conclude that

$$\mathbb{E}\left[\sup_{h\in BL_1} \left| \mathbb{E}[h(\widetilde{\mathbb{Y}}_n)] - \mathbb{E}^*[h(\widetilde{\mathbb{Y}}_n^*)] \right| \right] \to 0$$

as $n \to \infty$. □

5.4. Proof of Proposition 6

From the proofs of Proposition 10 and Proposition 12, we see that

$$\sup_{\mathbf{u}\in[0,1]^d} |\mathbb{Z}_n(\mathbf{u}) - \widetilde{\mathbb{Z}}_n(\mathbf{u})| = O_p\left(n^{-\mu}\right) \quad \text{and}$$

$$\sup_{\mathbf{u}\in[0,1]^d} |\mathbb{Z}_n^*(\mathbf{u}) - \widetilde{\mathbb{Z}}_n^*(\mathbf{u})| = O_{p*}\left(n^{-\mu}\right),$$

almost surely, for some $\mu > 0$. The result follows after integration by parts. □

5.5. Proof of Corollary 7

By the delta-method, $\{\widetilde{\mathbb{Y}}_n(\mathbf{s}), \mathbf{s} \in [0,1]^d\}$ converges towards a Gaussian process in $\ell^\infty([0,1]^d)$. The proof of Theorem 4 shows that $\limsup_{n\to\infty} \sup_{h\in BL_1} |\mathbb{E}[h(\mathbb{Y}_n) - h(\widetilde{\mathbb{Y}}_n)]| = 0$. Hence, the process $\mathbb{Y}_n$ converges weakly to the same weak limit as $\widetilde{\mathbb{Y}}_n$. This proves the first claim. The second part of the Corollary is a straightforward consequence of Theorem 4 and the triangle inequality. □

Appendix A

Let $X_1, \ldots, X_n$ be independent random variables with probability measure P. Let $\mathbb{P}_n$ be the empirical probability measure, putting mass $1/n$ at each observation, and let $\mathbb{P}_n^*$ be the nonparametric bootstrap measure based on n independent observations from $\mathbb{P}_n$. We index the empirical process $\sqrt{n}(\mathbb{P}_n - P)$ and its bootstrap counterpart $\sqrt{n}(\mathbb{P}_n^* - \mathbb{P}_n)$ by functions f that belong to a sequence of classes $\mathcal{F}_n$.

Theorem 13. *Let d_n be an integer sequence and, for each $1 \leqslant i \leqslant d_n$, let $\mathcal{G}_{i,n}$ be a VC class of functions with VC index $V_{i,n}$ and*

$$\sum_{i=1}^{d_n} V_{i,n} \leqslant K(\log n)^\gamma,$$

for some $K < \infty$ and $0 < \gamma < 1$. Set

$$\mathcal{F}_n = \left\{ f = \sum_{i=1}^{d_n} g_i : \ g_i \in \mathcal{G}_{i,n} \right\},$$

and suppose that there exists an envelope function $F \geqslant \sup_{f\in\mathcal{F}_n} |f|$, independent of n, with $\mathbb{E}[F^4(X)] < \infty$. Then,

$$\limsup_{n\to\infty} \mathbb{E}\left[\sup_{h\in BL_1} \left| \mathbb{E}[h(\sqrt{n}(\mathbb{P}_n - P))] - \mathbb{E}^*[h(\sqrt{n}(\mathbb{P}_n^* - \mathbb{P}_n))] \right| \right] = 0.$$

Proof. See Theorem 3 in Radulović (2012). □

Appendix B

Set $\mathbb{M}_n(\delta)$ as in (5.1) for $\delta \geqslant 0$, and define

$$\psi(x) = 2x^{-2}\{(1+x)\log(1+x) - x\}, \qquad x \in (-1,0) \cup (0,\infty)$$

and $\psi(-1) = 2$ and $\psi(0) = 1$. This function is continuous and decreasing.

Proposition 14. *There exist constants K_1 and K_2 such that*

$$\mathbb{P}\{\mathbb{M}_n(a) \geqslant \lambda\} \leqslant \frac{K_1}{a} \exp\left\{ -\frac{K_2 \lambda^2}{a} \psi\left(\frac{\lambda}{\sqrt{n}a}\right) \right\} \tag{B.1}$$

for all $a \in (0, 1/2]$ and all $\lambda \in [0,\infty)$.

Proof. See Proposition A.1 of Segers (2012). □

Appendix C

The computation of $\mathbb{T}_n$ is an optimization problem in $2dL_n$ variables. More precisely,

$$\mathbb{T}_n = \sup_{\mathbf{x}\in[0,1]^{2dL_n}} F(\mathbf{x}) = \sup_{x_i\in[0,1]} F(x_1, ..., x_{2dL_n}),$$

with

$$F(\mathbf{x}) := \begin{cases} \sum_{i=1}^{L_n} |\mathbb{Z}_n(B_i)| & \text{if } B_i := \prod_{j=1}^{d} \left(x_{2d(i-1)+2j-1}, x_{2d(i-1)+2j}\right],\ 1 \leqslant i \leqslant L_n \text{ are disjoint;} \\ 0 & \text{otherwise} \end{cases}$$

That is, for each $\mathbf{x} \in [0,1]^{2dL_n}$, we construct L_n boxes with $B_1 = (x_1, x_2] \times \cdots \times (x_{2d-1}, x_{2d}]$ formed by the first $2d$ coordinates, B_2 by the next $2d$ coordinates, and so on. If boxes overlap the function F is zero; otherwise it is $\sum_{i=1}^{L_n} |\mathbb{Z}_n(B_i)|$.

Typical situations ($d = 2, 3$, $n = 1000$ and $L_n = 4$) yield high dimensional domain spaces ($[0,1]^{16}$ and $[0,1]^{24}$ in this example). therefore, the optimization problem is not trivial due to the curse of dimension. That is why we have implemented the slightly "downgraded" test statistics $\widetilde{\mathbb{T}}_n$ where the corner of the boxes B_i are picked in a grid (see Equation (1.6)).

Figure 4 depicts one such an outcome with $d = 2$, $n = 800$, $L_n = 4$ and $M = 12 \times 10^4$. Clearly, these boxes are sampling the relevant regions. In general we do not have the access to contour plots such as Figure 4 and we cannot determine the shape and size of these relevant regions. But, as Figure 4 nicely demonstrates, sampling of quite a few boxes makes sense (more than 4 would probably do even better in this case).

A few details about numerical implementation

The computation of $\widetilde{\mathbb{T}}_n$ reduces to an optimization problem which needs to be done carefully. Here we present a stochastic optimization scheme that produces $\check{\mathbb{T}}_n$, which is near-optimal (i.e. close to $\widetilde{\mathbb{T}}_n$ with very high probability. Moreover, the algorithm is applicable in dimensions higher then 2 as well. In our simulation studies, we opted for this approach since it is very fast and easy to implement, and thus it enables easy verifications of our results. More sophisticated algorithms, with little extra computing time, could actually find the optimal value of $\widetilde{\mathbb{T}}_n$. In our experience this extra effort did not significantly change the performance of ATV statistics.

Step 1. Pre-compute $F(i_1, ..., i_d) := Z_n(\frac{i_1}{n^{1/d}}, ..., \frac{i_d}{n^{1/d}})$ for $i_j \in \{0, ..., \lfloor n^{1/d} \rfloor\}$.

Step 2. For all $B_i = \Pi_{j=1}^{d}(\frac{a_{i,j}}{n^{1/d}}, \frac{b_{i,j}}{n^{1/d}}]$, pre-compute $G(B_i) = \Delta^1_{a_{i,1},b_{i,1}} \Delta^2_{a_{i,2},b_{i,2}} ... \Delta^d_{a_{i,d},b_{i,d}} F$, where $a_{i,j}, b_{i,j} \in \{0, ..., \lfloor n^{1/d} \rfloor\}$ and $a_{i,j} < b_{i,j}$. Enumerate B_i's and choose m largest values (i.e. $G(B_1) \geqslant ... \geqslant G(B_m)$. In this experiment, we have chosen $m = n$.

Step 3. For a fixed i, sample randomly (without replacement) $A_{i,j} \in \{B_1, ..., B_n\}$, $j = 1, ..., L_n$. Repeat this for $i = 1, ..., K$, where $K = \max(10^4, 10n^L/(L!3^{2dL}))$ and compute

$$\mathcal{T}(\mathbf{A}_i) = \mathcal{T}(A_{i,1}, ..., A_{i,L}) = G(A_{i,1}) + G(A_{i,2})1_{A_{i,1}\cap A_{i,2}=\varnothing} + \ldots + G(A_{i,L})1_{A_{i,1}\cap\ldots\cap A_{i,L}=\varnothing}.$$

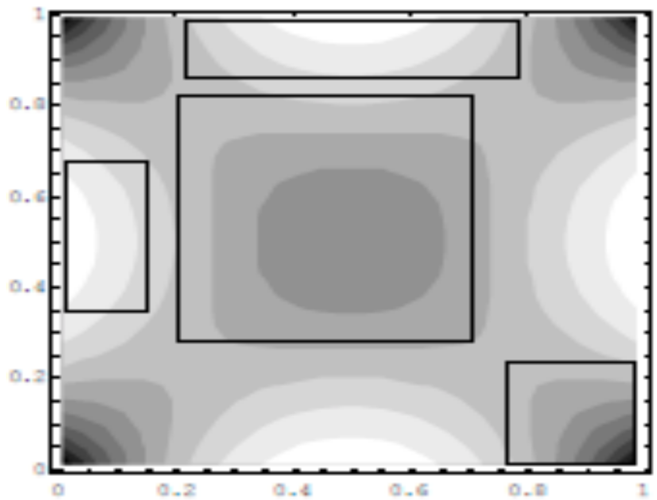

FIG 4. *Typical realization of the ATV statistic, $n = 800$, $L = 4$ for the Mixture-S synthetic data. The four rectangles capture the dark or bright regions which correspond to higher differences between the null and the alternative.*

Find $\mathbf{A}^o = (A_1^o, ..., A_L^o)$ such that $\mathcal{T}(\mathbf{A}^o) \geqslant \mathcal{T}(\mathbf{A}_i)$ for all $i \leqslant K$ and let

$$\check{\mathbb{T}}_n := \mathcal{T}(\mathbf{A}^o) \simeq \widetilde{\mathbb{T}}_n.$$

JDF : I have a problem here. Among the m (=n here) boxes we choose at step 2, we could have selected overlapping boxes (ie two arbitrarily chosen boxes among $B_1, \dots, B_m$ could overlap). In this case, $\mathcal{T}(\mathcal{A}_i) = G(A_{i,1})$ and $\mathcal{T}(\mathcal{A}^o) = \max_{i=1,\dots,m} G(B_i)$. Are we sure that $\mathcal{T}(A^o)$ is close to $\widetilde{\mathbb{T}}_n$ in this situation ?

Computational cost: Step 1 requires n computation of the $\mathbb{Z}_n$ process. Since the function F is pre-computed, Step 2 requires $2^d \left(\frac{n^{1/d}(n^{1/d}-1)}{2}\right)^d \leqslant n^2$ summations. For Step 3, we need to check if L

rectangles overlap, which requires at most $L^2/2$ such verifications. Each of which could be performed with $2d$ operations. Thus, for Step 3, we need $10dL^2n^L/(L!3^{2dL})$ simple computer operations at most. For a typical (larger) case $n = 800$, $d = 2$, and $L = 4$, the number of computations needed for Step 2 and Step 3 is bounded by 10^6. Since an ATV test typically requires 10^3 bootstrap resamplings, the total number of summations needed is of the order 10^9. Typical desktop computer (using C++ or Fortran code) evaluates ATV in 2-5 seconds. We would like to caution that Step 1, although negligible if coded in C++ or Fortran, tends to be very slow if performed using more elaborate programming languages like R or Mathematica.

Convergency. In Step 3, we choose K so that we could claim that $\mathbf{A}^o$ is very close to the optimal value $\mathbf{A}^*$ (such that $\widetilde{\mathbb{T}}_n = \mathcal{T}(\mathbf{A}^*)$) with very high probability. To be more precise, we define the distance between two rectangles by $d(B_i, B_k) = \max(\max_{j\leqslant d}(|a_{i,j} - a_{k,j}|), \max_{j\leqslant d}(|b_{i,j} - b_{k,j}|))$ and a neighborhood of $\mathbf{A}^*$ by

$$\mathcal{N}(\mathbf{A}^*) = \{U_{i=1}^L B_i \ : \ d(B_i, A_i^*) \leqslant \frac{1}{n^{1/d}} \text{ for every } i\}.$$

Then, the cardinality of $\mathcal{N}(\mathbf{A}^*)$ is 3^{2dL} (for each corner of a rectangle in $\mathcal{N}(\mathcal{A})$, we have 3 choices $b_{i,j} = x_{i,j} + \frac{\delta}{n^{1/d}}$ where $\delta \in \{-1, 0, 1\}$). Simple computation and the choice of K above now yields that $P(\mathbf{X}^o \in \mathcal{N}(\mathbf{X}^*)) > 1 - e^{-9.5} > 0.9999$.

JDF : This results justifies the choice of K. Dragan, you should provide details to recover this inequality, in my opinion. At least a reference, or a sketch of the proof...

In other words, with a very high probability, we can claim that the approximated collection of rectangles (i.e. $\mathbf{A}^o$) consists of rectangles that are at most one grid unit (i.e. $\frac{1}{n^{1/d}}$) away from the optimal collection of rectangles $\mathcal{A}^*$. Since there are L such rectangles it is easy to show that

$$|\check{\mathbb{T}}_n - \widetilde{\mathbb{T}}_n| = O_P(n^{-1/(2d)}L_n).$$

The improvement. It is possible to enhance the proposed algorithm by including an additional Step 4, which would concentrate on local search. Implementation of more sophisticated algorithms like Accelerated Random Search algorithm (Appel et al., 2004) would allow us to quickly search entire $\mathcal{N}(\mathbf{A}^o)$ which very likely contains the actual optimum $\mathbf{A}^*$. We experimented with this approach, and although it produced slightly larger values for statistic $\widetilde{\mathbb{T}}_n$, the overall performance did not significantly changed. This is most likely due to the above estimate $O_P(L_n n^{-1/(2d)})$. We suspect that this additional Step 4 would be of more value for dimensions larger d than 2. For a good review of optimization schemes relevant to this scenario we refer to the paper by Hvattuma and Gloverb (2009), where the authors describe eight optimization schemes and contrasts their performance on numerous test functions in higher dimensions (up to dimension 64).

References

[1] M.J. Appel, R. Labarre and D. Radulović (2004). On Accelerated Random Search, *SIAM J. Optim.* **14(3)**, 708 − 731.

[2] D. Berg (2009). Copula goodness-of-fit testing: An overview and power comparison. *European J. Finance* **15**, $675-701$
[3] P.J. Bickel and D.A. Freedman (1981). Some Asymptotic Theory for the Bootstrap. *Ann. Statist.* **9**(6), $1196-1217$.
[4] P.J. Bickel and D.A. Freedman (1983). Bootstrapping Regression Models with Many Parameters. In *A Festschrift for Erich L. Lehmann*, Editors P. J. Bickel, K. Doksum, J. L. Hodges. Wadsworth Statistics/Probability Series.
[5] A. Bücher and H. Dette (2010). A note on bootstrap approximations for the empirical copula process. *Statist. Probab. Lett.* **80**, $1925-1932$.
[6] S. Csörgó and D.M. Mason (1989). Bootstrapping Empirical Functions. *Ann. Statist.* **17**(4), $1447-1471$.
[7] P. Deheuvels (1979). La fonction de dépendance empirique et ses propriétés. Un test non paramétrique d'indépendance. *Acad. Roy. Belg. Bull. Cl. Sci. (5)* **65**, $274-292$.
[8] J. Dobrić and F. Schmid (2005). Testing goodness-of-.t for parametric families of copulas - Application to financial data. *Comm. Statist.: Simulation and Computation* **34**(4), $1053-1068$.
[9] J.-D. Fermanian (2005). Goodness-of-fit tests for copulas. *J. Multivariate Anal.* **95**(1), $119-152$.
[10] J.-D. Fermanian, D. Radulović and M.H. Wegkamp (2004). Weak convergence of empirical copula processes. *Bernoulli* **10**, $847-860$.
[11] J.-D. Fermanian (2012). An overview of the goodness-of-fit test problem for copulas. In *Copulae in Mathematical and Quantitative Finance*, P. Jaworski, F. Durant and W. Hrdle (ed.), $68-89$, Springer.
[12] J.-D. Fermanian and M.H. Wegkamp (2012). Time-dependent copulas. *J. Multivariate Anal.* **110**, $19-29$.
[13] C. Genest, K. Ghoudi and L.-P. Rivest (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika* **82**, $543-552$.
[14] C. Genest and B. Rémillard (2008). Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. *Ann. Henri Poincaré (Probabilités et Statistiques)* **44**(6), $1096-1127$.
[15] C. Genest, B. Rémillard and D. Beaudoin (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance Math. Econom.* **44**, $199-213$.
[16] T.H. Hildebrandt (1963). *Introduction to the Theory of Integration.* Academic Press, New York, London.
[17] W. Hoeffding (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* **19**, $293-325$.
[18] L.M. Hvattuma and F. Gloverb (2009). Finding local optima of high-dimensional functions using direct search methods. *European Journal of Operational Research*, **95**(1), $31-45$.
[19] E. Liebscher (2008). Construction of asymmetric multivariate copulas. *Journal of Multivariate Analysis* **99**, $2234-2250$.
[20] D. Mason (1981). Bounds for Weighted Empirical Distribution Functions. *Ann. Probab.* **9**, $881-884$.
[21] R.B. Nelsen (2006). *An Introduction to Copulas.* Springer.
[22] D. Radulović (1998). Can we bootstrap even if CLT fails? *J. Theoret. Probab.* **11**, $813-830$.
[23] D. Radulović (2012). Direct Bootstrapping Technique and its Application to a Novel Goodness of Fit Test. *J. Multivariate Anal.* **107**, $181-199$.
[24] D. Radulović (2013). High Dimensional CLT and its Applications. In *High Dimensional Probability VI, Progress in Probability*, **66**, $357-373$.

[25] B. Rémillard and O. Scaillet (2009). Testing for equality between two copulas *J. Multivariate Anal.* **100**, 377 − 386.

[26] O. Scaillet (2007). Kernel-based goodness-of-fit tests with fixed design smoothing parameters. *J. Multivariate Anal.* **98**, 533 − 543.

[27] J. Segers (2012). Asymptotic of empirical copula processes under nonrestrictive smoothness assumptions. *Bernoulli* **18**, 764 − 782.

[28] J.H. Shih and T.A. Louis (1995). Inferences on the association parameter in copula models for bivariate survival data. *Biometrics* **51**, 1384 − 1399.

[29] G.R. Shorack and J.A. Wellner (2009). *Empirical Processes with Applications to Statistics*, 2nd Edition, SIAM.

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

[31] A.W. van der Vaart (1998). *Asymptotic Statistics.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge.

[32] A.W. van der Vaart and J. Wellner (2000). *Weak convergence and empirical processes.* Springer Series in Statistics. Springer.