ON SOME RESAMPLING PROCEDURES WITH THE EMPIRICAL BETA COPULA

Anna Kiriliouk;
Johan Segers and Hideatsu Tsukahara

DISCUSSION PAPER | 2019 / 12

UCLouvain

ISBA | LIDAM
Louvain Institute of Data Analysis and Modeling in economics and statistics

On Some Resampling Procedures with the Empirical Beta Copula

Anna Kiriliouk, Johan Segers and Hideatsu Tsukahara

Abstract The empirical beta copula is a simple but effective smoother of the empirical copula. Because it is a genuine copula, from which, moreover, it is particularly easy to sample, it is reasonable to expect that resampling procedures based on the empirical beta copula are expedient and accurate. In this paper, after reviewing the literature on some bootstrap approximations for the empirical copula process, we first show the asymptotic equivalence of several bootstrapped processes related to the empirical copula and empirical beta copula. Then we investigate the finite-sample properties of resampling schemes based on the empirical (beta) copula by Monte Carlo simulation. More specifically, we consider interval estimation for some functionals such as rank correlation coefficients and dependence parameters of several well-known families of copulas, constructing confidence intervals by several methods and comparing their accuracy and efficiency. We also compute the actual size and power of symmetry tests based on several resampling schemes for the empirical copula and empirical beta copula.

1 Introduction

Let $\boldsymbol{X}_i = (X_{i1}, \dots, X_{id})$, $i \in \{1, \dots, n\}$, be independent and identically distributed random vectors, and assume that the cumulative distribution function, F, of $\boldsymbol{X}_i$ is

---

Anna Kiriliouk
Faculté des sciences économiques, sociales et de gestion, Université de Namur, Rue de Bruxelles 61, B-5000 Namur, Belgium, e-mail: anna.kiriliouk@unamur.be

Johan Segers
Institut de Statistique, Biostatistique et Sciences Actuarielles, Université catholique de Louvain, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium, e-mail: johan.segers@uclouvain.be

Hideatsu Tsukahara
Faculty of Economics, Seijo University, 6–1–20 Seijo, Setagaya-ku, Tokyo, 157-8511, Japan, e-mail: tsukahar@seijo.ac.jp

continuous. By Sklar's theorem [20], there exists a unique copula, C, such that

$$F(x_1, \ldots, x_d) = C\big(F_1(x_1), \ldots, F_d(x_d)\big),$$

where F_j is the jth marginal distribution function of F. In fact, in the continuous case, we have $C(\boldsymbol{u}) = F\big(F_1^-(u_1), \ldots, F_d^-(u_d)\big)$ for $\boldsymbol{u} = (u_1, \ldots, u_d) \in [0,1]^d$, where $H^-(u) = \inf\{t \in \mathbb{R}\colon H(t) \geqslant u\}$ is the generalized inverse of a distribution function H. The empirical copula $\mathbb{C}_n$ [3] is defined by

$$\mathbb{C}_n(\boldsymbol{u}) := \mathbb{F}_n\big(\mathbb{F}_{n1}^-(u_1), \ldots, \mathbb{F}_{nd}^-(u_d)\big),$$

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

$$\mathbb{F}_n(\boldsymbol{x}) := \frac{1}{n}\sum_{i=1}^n \mathbb{1}\left\{X_{i1} \leqslant x_1, \ldots, X_{id} \leqslant x_d\right\}, \quad \mathbb{F}_{nj}(x_j) := \frac{1}{n}\sum_{i=1}^n \mathbb{1}\left\{X_{ij} \leqslant x_j\right\}.$$

For $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, d\}$, let $R_{ij,n}$ be the rank of X_{ij} among $X_{1j}, \ldots, X_{nj}$; namely,

$$R_{ij,n} = \sum_{k=1}^n \mathbb{1}\{X_{kj} \leqslant X_{ij}\}. \tag{1}$$

Frequently used is a rank-based version of the empirical copula given by

$$\tilde{\mathbb{C}}_n(\boldsymbol{u}) := \frac{1}{n}\sum_{i=1}^n \prod_{j=1}^d \mathbb{1}\left\{\frac{R_{ij,n}}{n} \leqslant u_j\right\}. \tag{2}$$

In the absence of ties, we have

$$\|\tilde{\mathbb{C}}_n - \mathbb{C}_n\|_\infty := \sup_{\boldsymbol{u}\in[0,1]^d} |\tilde{\mathbb{C}}_n(\boldsymbol{u}) - \mathbb{C}_n(\boldsymbol{u})| \leqslant \frac{d}{n}. \tag{3}$$

Both functions $\mathbb{C}_n$ and $\tilde{\mathbb{C}}_n$ are piecewise constant and cannot be genuine copulas. When the sample size is small, they suffer from the presence of ties when used in resampling.

The empirical beta copula [18] is a simple but effective way of correcting and smoothing the empirical copula. Its definition will be given in Section 3. Even though its asymptotic distribution is the same as the one of the usual empirical copula, its accuracy in small samples is usually better, among others because it is itself always a genuine copula. Moreover, drawing random samples from the empirical beta copula is quite straightforward.

Because of these properties, it is reasonable to expect that simple and accurate resampling schemes for the empirical copula process can be constructed based on the empirical beta copula. For tail copulas, a simulation study in [12] showed that the bootstrap based on the empirical beta copula worked significantly better than the direct multiplier bootstrap from [1]. The purpose of this paper is to investigate further both the finite-sample and asymptotic behavior of this resampling method, but then for general copulas.

The paper is structured as follows. In Section 2, we review and discuss the literature on resampling methods for the empirical copula process. The asymptotic properties of two resampling procedures based on the empirical beta copula are investigated in Section 3. In Section 4, extensive simulation studies are conducted to demonstrate the effectiveness of resampling procedures based on the empirical beta copula for constructing confidence intervals for several copula functionals and for testing shape constraints on the copula. We conclude the paper with some discussion and open questions in Section 5. All proofs are grouped together in the Appendix.

2 Review on bootstrapping empirical copula processes

In this section, we give a short review on bootstrapping empirical copula processes, incorporating some newer improvements. We limit ourselves to i.i.d. sequences and note that extensions to stationary time series have been considered in [2], among others.

First we recall a basic result on the weak convergence of the empirical copula process. Let $\ell^\infty([0,1]^d)$ be the Banach space of real-valued, bounded functions on $[0,1]^d$, equipped with the supremum norm $\|\cdot\|_\infty$. The arrow $\rightsquigarrow$ denotes weak convergence in the sense used in [23]. The following condition is the only one needed for our convergence results.

Condition 2.1 For each $j \in \{1,\ldots,d\}$, the copula C has a continuous first-order partial derivative $\dot{C}_j(\boldsymbol{u}) = \partial C(\boldsymbol{u})/\partial u_j$ on the set $\{\boldsymbol{u} \in [0,1]^d : 0 < u_j < 1\}$. □

The following theorem is proved in [17]. Let $\mathbb{U}^C$ denote a C-pinned Brownian sheet, i.e., a centered Gaussian process on $[0,1]^d$ with continuous trajectories and covariance function

$$\operatorname{Cov}\big\{\mathbb{U}^C(\boldsymbol{u}), \mathbb{U}^C(\boldsymbol{v})\big\} = C(\boldsymbol{u}\wedge\boldsymbol{v}) - C(\boldsymbol{u})\,C(\boldsymbol{v}), \qquad \boldsymbol{u},\boldsymbol{v} \in [0,1]^d. \tag{4}$$

Theorem 2.2 *Suppose Condition* 2.1 *holds. Then we have*

$$\mathbb{G}_n := \sqrt{n}(\mathbb{C}_n - C) \rightsquigarrow \mathbb{G}^C, \qquad n \to \infty$$

in $\ell^\infty([0,1]^d)$*, where*

$$\mathbb{G}^C(\boldsymbol{u}) := \mathbb{U}^C(\boldsymbol{u}) - \sum_{j=1}^d \dot{C}_j(\boldsymbol{u})\,\mathbb{U}^C(\mathbf{1}, u_j, \mathbf{1})$$

with u_j *appearing at the* j*-th coordinate.* □

Next we introduce notation for the convergence of conditional laws in probability given the data as defined in [14]; see also [23, Section 2.9]. Let

$$\mathrm{BL}_1 := \{h \colon \ell^\infty([0,1]^d) \to \mathbb{R} \mid \|h\|_\infty \leqslant 1 \text{ and } |h(x) - h(y)| \leqslant \|x - y\|_\infty \\ \text{for all } x, y \in \ell^\infty([0,1]^d)\}. \tag{5}$$

If $\hat{X}_n$ is a sequence of bootstrapped processes in $(\ell^\infty([0,1]^d), \|\cdot\|_\infty)$ with random weights W, then the notation

$$\hat{X}_n \underset{W}{\overset{\mathrm{P}}{\rightsquigarrow}} X, \qquad n \to \infty \tag{6}$$

means that

$$\left.\begin{aligned} \sup_{h \in \mathrm{BL}_1} |\mathrm{E}_W[h(\hat{X}_n)] - \mathrm{E}[h(X)]| &\longrightarrow 0 \quad \text{in outer probability,} \\ \mathrm{E}_W[h(\hat{X}_n)^*] - \mathrm{E}_W[h(\hat{X}_n)_*] &\xrightarrow{\mathrm{P}} 0 \quad \text{for all } h \in \mathrm{BL}_1 \,. \end{aligned}\right\} \tag{7}$$

Here the notation E_W indicates conditional expectation over the weights W given the data $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$, and $h(\hat{X}_n)^*$ and $h(\hat{X}_n)_*$ denote the minimal measurable majorant and maximal measurable minorant, respectively, with respect to the joint data $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n, W$.

In the sequel, the random weights W can signify different things: a multinomial random vector when drawing from the data with replacement, independent and identically distributed multipliers in the multiplier bootstrap, or vectors of order statistics from the uniform distribution when resampling from the empirical beta copula. In (6), the symbol W will then be changed accordingly.

2.1 Straightforward bootstrap

Let $(W_{n1}, \ldots, W_{nn})$ be a multinomial random vector with probabilities $(1/n, \ldots, 1/n)$, independent of the sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$. Set

$$\mathbb{C}_n^*(\boldsymbol{u}) = \mathbb{F}_n^*\big(\mathbb{F}_{n1}^{*-}(u_1), \ldots, \mathbb{F}_{nd}^{*-}(u_d)\big),$$

where

$$\begin{aligned} \mathbb{F}_n^*(\boldsymbol{x}) &:= \frac{1}{n} \sum_{i=1}^n W_{ni} \prod_{j=1}^d \mathbb{1}\{X_{ij} \leqslant x_j\}, \\ \mathbb{F}_{nj}^*(x_j) &:= \frac{1}{n} \sum_{i=1}^n W_{ni} \, \mathbb{1}\{X_{ij} \leqslant x_j\}, \quad j \in \{1, \ldots, d\}. \end{aligned}$$

We can also define the bootstrapped version of the rank-based empirical copula

$$\tilde{\mathbb{C}}_n^*(\boldsymbol{u}) = \frac{1}{n} \sum_{i=1}^n W_{ni} \prod_{j=1}^d \mathbb{1}\left\{ \frac{R_{ij,n}^*}{n} \leqslant u_j \right\}, \tag{8}$$

where

$$R^*_{ij,n} = \sum_{k=1}^{n} W_{nk} \mathbb{1}\left\{X_{kj} \leqslant X_{ij}\right\}. \tag{9}$$

Since a bootstrap sample will have ties with a (large) positive probability, the bound (3) is no longer valid for $\mathbb{C}^*_n$ and $\tilde{\mathbb{C}}^*_n$. But we can prove the following.

Proposition 2.3 *We have*

$$\|\mathbb{C}^*_n - \tilde{\mathbb{C}}^*_n\|_\infty = O_p\left(n^{-1}\log n\right), \qquad n \to \infty. \tag{10}$$

The proof of Proposition 2.3 is given in the Appendix. Convergence in probability of the conditional laws

$$\sqrt{n}(\mathbb{C}^*_n - \mathbb{C}_n) \underset{W}{\overset{\mathrm{P}}{\rightsquigarrow}} \mathbb{G}_C, \quad n \to \infty$$

in the space $\ell^\infty([0,1]^d)$ was shown in [5] under the condition that all partial derivatives $\dot{C}_j$ exist and are continuous on $[0,1]^d$ and in [2] under the weaker Condition 2.1. Because of (3) and Proposition 2.3, it also holds that

$$\tilde{\alpha}_n := \sqrt{n}(\tilde{\mathbb{C}}^*_n - \tilde{\mathbb{C}}_n) \underset{W}{\overset{\mathrm{P}}{\rightsquigarrow}} \mathbb{G}_C, \quad n \to \infty. \tag{11}$$

2.2 Multiplier bootstrap with estimated partial derivatives

The multiplier bootstrap for the empirical copula proposed by [16] has proved useful for many problems. In [1] it was found to have a better finite-sample performance than other resampling methods for the empirical copula process. We present a modified version given by [1] that we employ for the simulation studies in Section 4.

Let $\xi_1, \ldots, \xi_n$ be independent and identically distributed non-negative random variables, independent of the data, with $\mathrm{E}(\xi_i) = \mu$, $\mathrm{Var}(\xi_i) = \tau^2 > 0$ and $\|\xi_i\|_{2,1} := \int_0^\infty \sqrt{\mathrm{P}(|\xi_i| > x)}\,\mathrm{d}x < \infty$. Put $\overline{\xi}_n := n^{-1}\sum_{i=1}^n \xi_i$, and set

$$\mathbb{C}^\circ_n(\boldsymbol{u}) := \frac{1}{n}\sum_{i=1}^{n} \frac{\xi_i}{\overline{\xi}_n} \prod_{j=1}^{d} \mathbb{1}\left\{X_{ij} \leqslant \mathbb{F}^-_{nj}(u_j)\right\},$$

$$\tilde{\mathbb{C}}^\circ_n(\boldsymbol{u}) := \frac{1}{n}\sum_{i=1}^{n} \frac{\xi_i}{\overline{\xi}_n} \prod_{j=1}^{d} \mathbb{1}\left\{\mathbb{F}_{nj}(X_{ij}) \leqslant u_j\right\}.$$

Define $\beta^\circ_n := \sqrt{n}(\mu/\tau)(\mathbb{C}^\circ_n - \mathbb{C}_n)$ and $\tilde{\beta}^\circ_n := \sqrt{n}(\mu/\tau)(\tilde{\mathbb{C}}^\circ_n - \tilde{\mathbb{C}}_n)$. Using Theorem 2.6 in [14] and the a.s. convergence $\|\mathbb{F}^-_{nj} - I\|_\infty \to 0$, where I is the identity function on $[0,1]$, we can show that

$$\beta_n^\circ \underset{\xi}{\overset{\mathrm{P}}{\rightsquigarrow}} \mathbb{U}^C \qquad \text{and} \qquad \tilde{\beta}_n^\circ \underset{\xi}{\overset{\mathrm{P}}{\rightsquigarrow}} \mathbb{U}^C, \qquad n \to \infty.$$

Hence if $\hat{\dot{C}}_j(\boldsymbol{u})$ is the estimate for $\dot{C}_j(\boldsymbol{u})$, applying finite differencing to the empirical copula at a spacing proportional to $n^{-1/2}$, then the processes

$$\begin{cases} \alpha_n^{\mathrm{pdm}\circ}(\boldsymbol{u}) := \beta_n^\circ(\boldsymbol{u}) - \sum_{j=1}^d \hat{\dot{C}}_j(\boldsymbol{u})\, \beta_n^\circ(\mathbf{1}, u_j, \mathbf{1}) \\ \tilde{\alpha}_n^{\mathrm{pdm}\circ}(\boldsymbol{u}) := \tilde{\beta}_n^\circ(\boldsymbol{u}) - \sum_{j=1}^d \hat{\dot{C}}_j(\boldsymbol{u})\, \tilde{\beta}_n^\circ(\mathbf{1}, u_j, \mathbf{1}) \end{cases}$$

give *conditional* approximations of $\mathbb{G}^C$. Namely, we have

$$\alpha_n^{\mathrm{pdm}\circ} \underset{\xi}{\overset{\mathrm{P}}{\rightsquigarrow}} \mathbb{G}_C \quad \text{and} \quad \tilde{\alpha}_n^{\mathrm{pdm}\circ} \underset{\xi}{\overset{\mathrm{P}}{\rightsquigarrow}} \mathbb{G}_C, \qquad n \to \infty.$$

3 Resampling with the empirical beta copula

The *empirical beta copula* [18] is defined as

$$\mathbb{C}_n^\beta(\boldsymbol{u}) = \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^d F_{n,R_{ij,n}}(u_j), \qquad \boldsymbol{u} \in [0,1]^d,$$

where $R_{ij,n}$ denote the ranks as in (1) and where, for $u \in [0,1]$ and $r \in \{1,\ldots,n\}$,

$$F_{n,r}(u) = \sum_{s=r}^n \binom{n}{s} u^s (1-u)^{n-s} \tag{12}$$

is the cumulative distribution function of the beta distribution $\mathcal{B}(r, n+1-r)$. In this section, we examine the asymptotic properties of two resampling procedures based on the empirical beta copula.

3.1 Standard bootstrap for the empirical beta copula

Let $(W_{n1},\ldots,W_{nn})$ be a multinomial random vector with success probabilities $(1/n,\ldots,1/n)$, independent of the original sample. Set

$$\mathbb{C}_n^{\beta*}(\boldsymbol{u}) = \frac{1}{n} \sum_{i=1}^n W_{ni} \prod_{j=1}^d F_{n,R^*_{ij,n}}(u_j),$$

where $R^*_{ij,n}$ are the bootstrapped ranks in (9). Let $S_j \sim \mathrm{Bin}(n, u_j)$, for $j = 1,\ldots,d$, be d independent binomial random variables. Let E_S denote expectation with respect to $(S_1, \ldots, S_d)$, conditionally on the sample and the multinomial random vector. It follows that

$$\mathbb{C}_n^{\beta *}(\boldsymbol{u}) = \frac{1}{n} \sum_{i=1}^{n} W_{ni} \prod_{j=1}^{d} \mathrm{E}_S \left[\mathbb{1} \left\{ \frac{S_j}{n} \geqslant \frac{R_{ij,n}^*}{n} \right\} \right] = \mathrm{E}_S \left[\tilde{\mathbb{C}}_n^*(S_1/n, \ldots, S_d/n) \right],$$

where $\tilde{\mathbb{C}}_n^*$ is the bootstrapped rank-based empirical copula in (8). Similarly, the empirical beta copula is

$$\mathbb{C}_n^{\beta}(\boldsymbol{u}) = \frac{1}{n} \sum_{i=1}^{n} \prod_{j=1}^{d} F_{n, R_{ij,n}}(u_j) = \mathrm{E}_S \left[\tilde{\mathbb{C}}_n(S_1/n, \ldots, S_d/n) \right],$$

where $\tilde{\mathbb{C}}_n$ is the rank-based empirical copula in (2). Consider the bootstrapped processes $\tilde{\alpha}_n$ defined in (11) and $\alpha_n^{\beta} := \sqrt{n}(\mathbb{C}_n^{\beta *} - \mathbb{C}_n^{\beta})$. We find

$$\alpha_n^{\beta}(\boldsymbol{u}) = \mathrm{E}_S[\tilde{\alpha}_n(S_1/n, \ldots, S_d/n)]. \tag{13}$$

From the weak convergence of the bootstrapped process $\tilde{\alpha}_n$, we will prove the following proposition. As a consequence, consistency of the bootstrapped process $\tilde{\alpha}_n$ of the (rank-based) empirical copula in (11) entails consistency of the one for the empirical beta copula.

Proposition 3.1 *Under Condition* 2.1*, we have*

$$\sup_{\boldsymbol{u} \in [0,1]^d} |\alpha_n^{\beta}(\boldsymbol{u}) - \tilde{\alpha}_n(\boldsymbol{u})| = o_P(1), \qquad n \to \infty, \tag{14}$$

and thus $\alpha_n^{\beta} \underset{W}{\overset{\mathrm{P}}{\rightsquigarrow}} \mathbb{G}_C$ *as* $n \to \infty$. □

3.2 Bootstrap by drawing samples from the empirical beta copula

The original motivation of [18] was resampling; the uniform random variables generated independently and rearranged in the order specified by the componentwise ranks of the original sample might in some sence be considered as a bootstrap sample. Although this idea turned out to be not entirely correct, it was still how the empirical beta copula was discovered originally. In the same spirit, it is natural to study the bootstrap method based on drawing samples from the empirical beta copula $\mathbb{C}_n^{\beta}$.

It is in fact very simple to generate a random variate $\boldsymbol{V}$ from $\mathbb{C}_n^{\beta}$.

Algorithm 3.2 Given the ranks $R_{ij,n} = r_{ij}$, $j = 1, \ldots, d$, of the original sample:

1. Generate I from the discrete uniform distribution on $\{1, \ldots, n\}$.

2. Generate independently $V_j^\# \sim \mathcal{B}(r_{Ij}, n+1-r_{Ij}),\ j \in \{1, \ldots, d\}$.
3. Set $\boldsymbol{V}^\# = (V_1^\#, \ldots, V_d^\#)$. □

Repeating the above algorithm n times independently, we get a sample of n independent random vectors drawn from $\mathbb{C}_n^\beta$, conditionally on the data $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$. Let this sample be denoted by $\boldsymbol{V}_i^\# = (V_{i1}^\#, \ldots, V_{id}^\#)$, $i = 1, \ldots, n$. We can think of this procedure as a kind of *smoothed bootstrap* (see [4], [19, Section 3.5]) because the empirical beta copula may be thought of as a smoothed version of the empirical copula.

The joint and marginal empirical distribution functions of the bootstrap sample are

$$\mathbb{G}_n^\#(\boldsymbol{u}) = \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^d \mathbb{1}\{V_{ij}^\# \leqslant u_j\}, \quad \mathbb{G}_{nj}^\#(u_j) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{V_{ij}^\# \leqslant u_j\}.$$

The ranks of the bootstrap sample are given by

$$R_{ij,n}^\# = n\, \mathbb{G}_{nj}^\#(V_{ij}^\#) = \sum_{k=1}^n \mathbb{1}\{V_{kj}^\# \leqslant V_{ij}^\#\}. \tag{15}$$

These yield bootstrapped versions of the Deheuvels empirical copula, the rank-based empirical copula and the empirical beta copula:

$$\mathbb{C}_n^\#(\boldsymbol{u}) := \mathbb{G}_n^\#\big(\mathbb{G}_{n1}^{\#-}(u_1), \ldots, \mathbb{G}_{nd}^{\#-}(u_d)\big), \quad \tilde{\mathbb{C}}_n^\#(\boldsymbol{u}) := \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^d \mathbb{1}\{R_{ij,n}^\#/n \leqslant u_j\},$$
$$\mathbb{C}_n^{\beta\#}(\boldsymbol{u}) := \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^d F_{n,R_{ij,n}^\#}(u_j).$$

Proposition 3.3 *Assume Condition* 2.1. *Then as $n \to \infty$, we have conditional weak convergence in probability as defined in* (6) *with respect to the random vectors $\boldsymbol{V}_1^\#, \ldots, \boldsymbol{V}_n^\#$ of the bootstrapped empirical copula processes*

$$\alpha_n^\# := \sqrt{n}(\mathbb{C}_n^\# - \mathbb{C}_n), \quad \tilde{\alpha}_n^\# := \sqrt{n}(\tilde{\mathbb{C}}_n^\# - \tilde{\mathbb{C}}_n), \quad \alpha_n^{\beta\#} := \sqrt{n}(\mathbb{C}_n^{\beta\#} - \mathbb{C}_n^\beta),$$

to the limit process $\mathbb{G}_C$ defined in Theorem 2.2. □

3.3 Approximating sampling distributions of rank statistics by resampling from the empirical beta copula

Statistical inference for C often involves rank statistics. One way to justify this is to appeal to the invariance of C under coordinatewise continuous strictly increasing transformations. Let us hence consider a rank statistic $T(\boldsymbol{R}_1, \ldots, \boldsymbol{R}_n)$, where $\boldsymbol{R}_i := (R_{i1,n}, \ldots, R_{id,n})$ is a vector consisting of the coordinatewise ranks of $\boldsymbol{X}_i$. Below we suggest a way of approximating its distribution by drawing a sample from $\mathbb{C}_n^{\beta}$ and computing "bootstrap replicates". This also avoids problems with ties encountered when drawing with replacement from the original data. Specifically, our procedure goes as follows.

Algorithm 3.4 (Smoothed beta bootstrap) Given $\boldsymbol{R}_1, \ldots, \boldsymbol{R}_n$:

1. Apply Algorithm 3.2 n times independently to obtain a bootstrap sample $\boldsymbol{V}_1^{\#}, \ldots, \boldsymbol{V}_n^{\#}$ drawn from $\mathbb{C}_n^{\beta}$, compute their ranks $\boldsymbol{R}_1^{\#}, \ldots, \boldsymbol{R}_n^{\#}$ as in (15) and put $T^{\#} := T(\boldsymbol{R}_1^{\#}, \ldots, \boldsymbol{R}_n^{\#})$.
2. Repeat Step 1 a moderate to large number, B, of times to get bootstrap replicates $T_1^{\#}, \ldots, T_B^{\#}$.
3. Use $T_1^{\#}, \ldots, T_B^{\#}$ to approximate the sampling distribution of $T(\boldsymbol{R}_1, \ldots, \boldsymbol{R}_n)$. $\square$

The validity of this procedure follows from our claim in the preceding subsection. Because all the related empirical copula processes are asymptotically equivalent, we need to look into the small-sample performance of the methods. In Subsection 4.2, we construct confidence intervals for some copula functionals by popular rank statistics.

4 Simulation Studies

We assess the performance of the bootstrap methods presented in Sections 2 and 3 in a wide range of applications. In all of the experiments below, the number of Monte Carlo runs and the number of bootstrap replications are both set to 1000. The nominal confidence level is always 0.95 and we use Clayton, Gumbel-Hougaard, Frank and Gauss copula families, see e.g. [15]. Most simulations are done in R with the package copula [9], except for Subsection 4.2, where MATLAB code was used.

4.1 Covariance of the limiting process

We compare the estimated covariances of the limiting process $\mathbb{G}^C$ based on the standard and smoothed beta bootstrap methods with the partial derivatives multiplier method, which in [1] is shown to perform better than the straightforward bootstrap or the direct multiplier method. We follow the set-up in [1], evaluating the covariance at four points $\{(i/3, j/3)\}$ for $i, j \in \{1, 2\}$ in the unit square. The variables $\xi_1, \ldots, \xi_n$ are such that $\mathbb{P}[\xi_i = 0] = \mathbb{P}[\xi_i = 2] = 1/2$ for $i \in \{1, \ldots, n\}$. For the bivariate Clayton copula with parameter $\theta = 1$, Table 1 shows the mean squared error of the estimated covariance based on the partial derivative multiplier method $\alpha_n^{\mathrm{pdm}\circ}$, the standard beta bootstrap α_n^{β} and the smoothed beta bootstrap $\alpha_n^{\beta\#}$ for $n = 100$ and $n = 200$. Results for $\alpha_n^{\mathrm{pdm}\circ}$ have been copied from Tables 3 and 4 in [1]. Both methods based on the empirical beta copula outperform the multiplier method in all points but $(1/3, 1/3)$ and $(2/3, 2/3)$.

Table 1 Mean squared error ($\times 10^4$) of the covariance estimates for the bivariate Clayton copula with $\theta = 1$.

		$n = 100$				$n = 200$			
		$\left(\frac{1}{3}, \frac{1}{3}\right)$	$\left(\frac{1}{3}, \frac{2}{3}\right)$	$\left(\frac{2}{3}, \frac{1}{3}\right)$	$\left(\frac{2}{3}, \frac{2}{3}\right)$	$\left(\frac{1}{3}, \frac{1}{3}\right)$	$\left(\frac{1}{3}, \frac{2}{3}\right)$	$\left(\frac{2}{3}, \frac{1}{3}\right)$	$\left(\frac{2}{3}, \frac{2}{3}\right)$
$\alpha_n^{\mathrm{pdm}\circ}$	(1/3, 1/3)	0.8887	0.5210	0.5222	0.3716	0.4595	0.2673	0.2798	0.1961
	(1/3, 2/3)		1.0112	0.1799	0.2988		0.5211	0.1069	0.1577
	(2/3, 1/3)			0.9899	0.2818			0.5092	0.1681
	(2/3, 2/3)				0.6250				0.2992
α_n^{β}	(1/3, 1/3)	0.9992	0.3402	0.3473	0.1956	0.6205	0.2427	0.2383	0.1547
	(1/3, 2/3)		0.7887	0.1294	0.1889		0.4933	0.0857	0.1366
	(2/3, 1/3)			0.7644	0.1821			0.4898	0.1376
	(2/3, 2/3)				0.7108				0.4183
$\alpha_n^{\beta\#}$	(1/3, 1/3)	1.2248	0.2929	0.2924	0.1456	0.6761	0.1874	0.1888	0.1128
	(1/3, 2/3)		0.8461	0.0992	0.1691		0.4814	0.0703	0.1071
	(2/3, 1/3)			0.8856	0.1682			0.4956	0.1149
	(2/3, 2/3)				1.1209				0.5913

4.2 Confidence intervals for rank correlation coefficients

We assess the performance of the straightforward bootstrap and the smoothed beta bootstrap (Subsections 2.1 and 3.3) for constructing confidence intervals for two popular rank correlation coefficients for bivariate distributions, Kendall's τ and Spearman's ρ, which are known to depend only on the copula C associated with F.

The population Kendall's τ is defined by

$$\tau(C) := 4 \int_0^1 \int_0^1 C(u_1, u_2)\, \mathrm{d}C(u_1, u_2) - 1.$$

In terms of

$$Q_{k,i} := \operatorname{sign}[(X_{k,1} - X_{i,1})(X_{k,2} - X_{i,2})], \quad K := \sum_{i=1}^{n-1} \sum_{k=i+1}^{n} Q_{k,i},$$

the sample Kendall's τ is given by $\hat{\tau} := 2K/[n(n-1)]$. Its asymptotic variance may be estimated by

$$\hat{\sigma}_\tau^2 := \frac{2}{n(n-1)} \left[\frac{2(n-2)}{n(n-1)^2} \sum_{i=1}^{n} (C_i - \overline{C})^2 + 1 - \hat{\tau}^2 \right],$$

where $C_i := \sum_{k=1,\, k \neq i}^{n} Q_{k,i}$, $i \in \{1, \ldots, n\}$ and $\overline{C} = n^{-1} \sum_{i=1}^{n} C_i = 2K/n$ (see [10]). An asymptotic confidence interval for τ is thus given by $\hat{\tau} \pm z_{\alpha/2} \hat{\sigma}_\tau$, with $z_{\alpha/2}$ the usual standard normal tail quantile.

This interval can be compared to the confidence intervals obtained by our resampling methods. Table 2 shows the coverage probabilities and the average lengths of the estimated confidence intervals based on the asymptotic distribution, the straightforward bootstrap and the smoothed beta bootstrap for the independence copula ($\tau = 0$) and the Clayton copula with $\theta = 2$ ($\tau = 0.5$) and $\theta = -2/3$ ($\tau = -0.5$). The smoothed beta bootstrap gives the most conservative coverage probabilities, but the shortest length among the three.

Table 2 Coverage probabilities and average lengths of confidence intervals for Kendall's τ for the Clayton copula family computed via the normal approximation, the straightforward bootstrap, and the smoothed beta bootstrap

		$\tau = 0$				$\tau = 0.5$				$\tau = -0.5$			
	n	40	60	80	100	40	60	80	100	40	60	80	100
coverage probability	asymp	0.952	0.930	0.941	0.959	0.946	0.931	0.937	0.943	0.933	0.941	0.939	0.926
	boot	0.957	0.937	0.942	0.963	0.949	0.940	0.949	0.949	0.951	0.947	0.938	0.935
	beta	0.964	0.949	0.949	0.966	0.952	0.947	0.954	0.955	0.963	0.935	0.948	0.939
average length	asymp	0.449	0.355	0.304	0.271	0.364	0.287	0.245	0.217	0.378	0.302	0.257	0.227
	boot	0.450	0.357	0.306	0.272	0.366	0.288	0.246	0.218	0.380	0.304	0.258	0.228
	beta	0.433	0.347	0.299	0.268	0.350	0.279	0.240	0.213	0.365	0.294	0.253	0.224

The population Spearman's ρ and the sample Spearman's rho are given by

$$\rho(C) := 12 \int_0^1 \int_0^1 \left[C(u_1, u_2) - u_1 u_2 \right] du_1 du_2,$$

$$\hat{\rho} := \frac{12}{n(n^2-1)} \sum_{i=1}^{n} \left(R_{i1,n} - \frac{n+1}{2} \right) \left(R_{i2,n} - \frac{n+1}{2} \right).$$

The limiting distribution of $\hat{\rho}$ equals that of $12 \iint \mathbb{G}^C(u_1, u_2) \mathrm{d}u_1 \mathrm{d}u_2$, so it is possible in principle to construct confidence intervals based on this asymptotics. However, unlike the case of $\hat{\tau}$, it is cumbersome and involves the partial derivatives of C, which must be estimated, so we omit it from our study here. In Table 3, one can see that the coverage probabilities are more conservative for the smoothed beta bootstrap than for the straightforward bootstrap, but the average lengths of the estimated confidence intervals are very similar for both methods. This could be due to the fact that $\rho(\mathbb{C}_n^{\beta}) = [(n-1)/(n+1)]\hat{\rho}$, as can be directly computed.

Table 3 Coverage probabilities and average lengths of confidence intervals for Spearman's ρ for the Clayton copula family based on the straightforward bootstrap and the smoothed beta bootstrap.

		$\rho = 0$				$\rho = 0.5$				$\rho = -0.5$			
	n	40	60	80	100	40	60	80	100	40	60	80	100
coverage probability	boot	0.956	0.943	0.953	0.951	0.959	0.953	0.949	0.952	0.952	0.954	0.960	0.956
	beta	0.965	0.946	0.957	0.956	0.961	0.958	0.960	0.952	0.969	0.957	0.964	0.958
average length	boot	0.634	0.514	0.444	0.397	0.524	0.424	0.367	0.326	0.519	0.418	0.366	0.324
	beta	0.625	0.510	0.442	0.395	0.522	0.424	0.368	0.325	0.519	0.418	0.367	0.324

4.3 Confidence intervals for a copula parameter

Suppose that the copula of F is parametrized by $\theta \in \Theta \subset \mathbb{R}$, so that $F(x_1, x_2) = C_\theta(F_1(x_1), F_2(x_2))$. When the F_j's are unknown, the resulting problem of estimating θ is semiparametric and is studied in [6, 21]. Assume that C_θ is absolutely continuous with density c_θ, which is differentiable with respect to θ. Replacing the unknown F_j's in the score equation by their (rescaled) empirical counterparts, one gets the estimating equation

$$\sum_{k=1}^{n} \frac{\dot{c}_\theta[\mathbb{F}_{n1}(X_{k,1}), \mathbb{F}_{n2}(X_{k,2})]}{c_\theta[\mathbb{F}_{n1}(X_{k,1}), \mathbb{F}_{n2}(X_{k,2})]} = 0, \tag{16}$$

where $\dot{c}_\theta = \partial c_\theta / \partial \theta$. The solution $\widehat{\theta}$ to (16) is called the *pseudo-likelihood estimator*.

We compare confidence intervals for θ when estimated by the pseudo-likelihood estimator $\widehat{\theta}$ based on the asymptotic variance given in [6], the straightforward bootstrap, the smoothed beta bootstrap and the classic parametric bootstrap. Tables 4 and 5 show the estimated coverage probabilities and average interval lengths of the confidence intervals for the Clayton, Gauss, Frank and Gumbel–Hougaard copula families. For the Clayton copula, the smoothed beta bootstrap gives the shortest intervals both for $\theta = 1$ and $\theta = 2$, but only for $\theta = 2$ the coverage probabilities are too liberal, which is somewhat puzzling. For the Frank and Gumbel–Hougaard copulas, the smoothed beta bootstrap gives the most conservative coverage probabilities, but the shortest length among the four. For the Gauss copula, the asymptotic approximation gives significantly smaller coverage probabilities than the nominal value 0.95.

Table 4 Coverage probabilities and average lengths of confidence intervals for the parameter of the Clayton copula with $\theta = 1$ ($\tau = 1/3$) and $\theta = 2$ ($\tau = 1/2$). Intervals computed via the asymptotic normal approximation, the straightforward bootstrap, the smoothed beta bootstrap, and the parametric bootstrap.

		$\theta = 1$				$\theta = 2$			
	n	40	60	80	100	40	60	80	100
coverage probability	asymp	0.954	0.969	0.960	0.965	0.951	0.940	0.940	0.946
	boot	0.953	0.943	0.944	0.943	0.968	0.952	0.953	0.951
	beta	0.953	0.964	0.957	0.952	0.933	0.904	0.908	0.906
	param	0.924	0.923	0.933	0.948	0.957	0.951	0.955	0.953
average length	asymp	2.011	1.632	1.354	1.237	2.764	2.142	1.821	1.615
	boot	1.894	1.449	1.198	1.046	2.991	2.205	1.841	1.626
	beta	1.517	1.225	1.050	0.935	1.957	1.612	1.420	1.296
	param	1.914	1.448	1.222	1.070	2.821	2.150	1.829	1.617

Table 5 Coverage probabilities and average lengths of confidence intervals for the parameter of the Gaussian copula with $\theta = 1/\sqrt{2}$, the Frank copula with $\theta = 5.75$ and the Gumbel–Hougaard copula with $\theta = 2$. All copulas have $\tau \approx 1/2$. Intervals computed via the asymptotic normal approximation, the straightforward bootstrap, the smoothed beta bootstrap, and the parametric bootstrap.

		Gauss				Frank				Gumbel–Hougaard			
	n	40	60	80	100	40	60	80	100	40	60	80	100
coverage probability	asymp	0.881	0.895	0.910	0.928	0.941	0.950	0.948	0.965	0.954	0.940	0.940	0.955
	boot	0.942	0.944	0.947	0.950	0.957	0.956	0.946	0.963	0.965	0.951	0.953	0.965
	beta	0.968	0.962	0.970	0.953	0.965	0.961	0.952	0.965	0.970	0.951	0.952	0.954
	param	0.903	0.921	0.923	0.930	0.938	0.956	0.941	0.962	0.924	0.926	0.932	0.945
average length	asymp	0.303	0.274	0.213	0.193	5.699	4.487	3.821	3.391	1.425	1.082	0.929	0.816
	boot	0.319	0.257	0.219	0.197	6.139	4.677	3.949	3.464	1.572	1.162	0.968	0.855
	beta	0.341	0.269	0.228	0.203	5.367	4.335	3.735	3.329	1.170	0.947	0.826	0.747
	param	0.292	0.242	0.210	0.191	5.729	4.494	3.848	3.389	1.546	1.170	0.983	0.869

4.4 Testing symmetry of a copula

For a bivariate copula C, consider the problem of testing the symmetry hypothesis $H_0 : C(u_1, u_2) = C(u_2, u_1)$ for all $(u_1, u_2) \in [0,1]^2$. We focus on two test statistics proposed in [7],

$$S_n = \int_{[0,1]^2} [\mathbb{C}_n(u_1, u_2) - \mathbb{C}_n(u_2, u_1)]^2 \, \mathrm{d}\mathbb{C}_n(u_1, u_2),$$
$$R_n = \int_{[0,1]^2} [\mathbb{C}_n(u_1, u_2) - \mathbb{C}_n(u_2, u_1)]^2 \, \mathrm{d}u_1 \, \mathrm{d}u_2,$$

and also include a version of R_n based on the empirical beta copula, i.e.,

$$R_n^\beta = \int_{[0,1]^2} [\mathbb{C}_n^\beta(u_1, u_2) - \mathbb{C}_n^\beta(u_2, u_1)]^2 \, \mathrm{d}u_1 \, \mathrm{d}u_2.$$

Similarly as in Proposition 1 in [7], the statistic R_n^β can be computed via

$$R_n^\beta = \frac{2}{n^2} \sum_{i=1}^n \sum_{j=1}^n \{B_n(R_{i1,n}, R_{j1,n}) B_n(R_{i2,n}, R_{j2,n}) - B_n(R_{i1,n}, R_{j2,n}) B_n(R_{i2,n}, R_{j1,n})\}$$

with $B_n(r,s) = \int_0^1 F_{n,r}(u) F_{n,s}(u) \, \mathrm{d}u$ for $r, s \in \{1, \ldots, n\}$ and $F_{n,r}(u)$ as in (12). For fixed n, the matrix B_n can be precomputed and stored, reducing the computation time for the resampling methods. A similar modification of S_n into S_n^β is obviously possible as well, but would be computationally more demanding.

In order to compute p-values, we need to generate bootstrap samples from a distribution fulfilling the restriction specified by H_0. A natural candidate is a 'symmetrized' version of the empirical beta copula

$$\mathbb{C}_n^{\beta,\mathrm{sym}}(u_1, u_2) := \frac{1}{2} \mathbb{C}_n^\beta(u_1, u_2) + \frac{1}{2} \mathbb{C}_n^\beta(u_2, u_1).$$

When resampling, this simply amounts to interchanging the two coordinates at random in step 3 of Algorithm 3.2. We employ the following three resampling schemes for comparison of actual sizes of the tests.

- The symmetrized smoothed beta bootstrap: we resample from $\mathbb{C}_n^{\beta,\mathrm{sym}}$ to get bootstrap replicates of R_n, R_n^β, and S_n;
- The symmetrized version of the straightforward bootstrap for R_n and S_n;
- *exchTest* in the R package copula [9], which implements the multiplier bootstrap for R_n and S_n as described in [7] and in Section 5 of [13]. For R_n, the grid length in *exchTest* is set to $m = 50$.

Tables 6 and 7 show the actual sizes of the symmetry tests for the Clayton and Gauss copulas. On the whole, the smoothed beta bootstrap works better than *exchTest* or equally well both for R_n and S_n, except when dependence is strong ($\tau = 0.75$) and the sample size is small ($n = 50$), although no method produces a satisfying result in this case. The smoothed beta bootstrap with R_n^β produces actual sizes similar to those with R_n. The statistic S_n performs slightly better than R_n on average, especially for strong positive dependence. The straightforward bootstrap performs poorly in all cases, which is as expected [16].

To compare the power of the tests, the Clayton and Gauss copulas are made asymmetric by Khoudraji's device [11], that is, the asymmetric version of a copula C is defined as

$$K_\delta(u_1, u_2) = u_1^\delta C(u_1^{1-\delta}, u_2), \qquad (u_1, u_2) \in [0, 1]^2.$$

Table 8 shows the empirical power of R_n and R_n^β for $\delta \in \{0.25, 0.5, 0.75\}$ for the three resampling methods. We see that the smoothed beta bootstraps with R_n and R_n^β have higher power than *exchTest* for almost all sample sizes and parameter values considered, and among them, the smoothed beta bootstrap with R_n^β has a slightly higher power in almost all cases.

Table 6 Actual sizes of symmetry tests based on R_n and S_n for the Clayton copula ($\theta \in \{-1/3, 2/3, 2, 6\}$), with p-values computed by the multiplier bootstrap (exchTest), the straightforward bootstrap (boot), the smoothed beta bootstrap (beta) and of the test based on R_n^β, with p-values computed by the smoothed beta bootstrap (beta2). The nominal size is $\alpha = 0.05$

	S_n					R_n					
	n	50	100	200	400		n	50	100	200	400
$\tau = -1/5$	exchTest	0.055	0.033	0.039	0.040	$\tau = -1/5$	exchTest	0.044	0.035	0.039	0.051
	boot	0.021	0.024	0.031	0.034		boot	0.009	0.019	0.027	0.044
	beta	0.057	0.038	0.039	0.041		beta	0.046	0.035	0.042	0.059
							beta2	0.050	0.037	0.041	0.059
$\tau = 0.25$	exchTest	0.039	0.029	0.036	0.036	$\tau = 0.25$	exchTest	0.030	0.022	0.040	0.031
	boot	0.009	0.015	0.026	0.034		boot	0.001	0.011	0.020	0.030
	beta	0.039	0.039	0.044	0.046		beta	0.042	0.032	0.041	0.043
							beta2	0.033	0.033	0.044	0.045
$\tau = 0.5$	exchTest	0.033	0.020	0.026	0.019	$\tau = 0.5$	exchTest	0.015	0.014	0.030	0.031
	boot	0.001	0.008	0.015	0.017		boot	0.001	0.005	0.017	0.025
	beta	0.030	0.029	0.039	0.028		beta	0.020	0.022	0.040	0.047
							beta2	0.019	0.023	0.045	0.046
$\tau = 0.75$	exchTest	0.025	0.026	0.018	0.014	$\tau = 0.75$	exchTest	0.000	0.007	0.007	0.012
	boot	0.000	0.002	0.001	0.008		boot	0.000	0.000	0.001	0.004
	beta	0.006	0.017	0.026	0.029		beta	0.000	0.006	0.017	0.029
							beta2	0.000	0.007	0.021	0.036

Table 7 Actual sizes of symmetry tests based on R_n and S_n for the Gauss copula, with p-values computed by the multiplier bootstrap (exchTest), the straightforward bootstrap (boot), the smoothed beta bootstrap (beta) and of the test based on R_n^{β}, with p-values computed by the smoothed beta bootstrap (beta2). The nominal size is $\alpha = 0.05$

	S_n						R_n				
	n	50	100	200	400		n	50	100	200	400
$\tau = -0.5$	exchTest	0.047	0.032	0.038	0.039	$\tau = -0.5$	exchTest	0.026	0.037	0.037	0.041
	boot	0.022	0.023	0.032	0.040		boot	0.007	0.014	0.027	0.033
	beta	0.044	0.030	0.035	0.043		beta	0.020	0.030	0.036	0.040
							beta2	0.022	0.034	0.041	0.042
$\tau = 0.25$	exchTest	0.028	0.035	0.031	0.040	$\tau = 0.25$	exchTest	0.029	0.025	0.030	0.041
	boot	0.007	0.015	0.023	0.038		boot	0.008	0.015	0.025	0.037
	beta	0.033	0.038	0.039	0.045		beta	0.040	0.030	0.033	0.048
							beta2	0.037	0.031	0.037	0.048
$\tau = 0.5$	exchTest	0.034	0.018	0.025	0.029	$\tau = 0.5$	exchTest	0.011	0.014	0.019	0.034
	boot	0.003	0.005	0.015	0.026		boot	0.001	0.005	0.006	0.028
	beta	0.032	0.033	0.041	0.044		beta	0.016	0.024	0.031	0.042
							beta2	0.018	0.023	0.033	0.047
$\tau = 0.75$	exchTest	0.018	0.017	0.011	0.008	$\tau = 0.75$	exchTest	0.001	0.001	0.005	0.009
	boot	0.000	0.000	0.002	0.006		boot	0.000	0.000	0.000	0.003
	beta	0.006	0.015	0.021	0.029		beta	0.001	0.001	0.010	0.028
							beta2	0.001	0.001	0.011	0.027

5 Concluding Remarks

We studied the performance of resampling procedures based on the empirical beta copula. All the related empirical copula processes are proved to have an asymptotically equivalent behavior. Comparative analysis through Monte Carlo experiments shows that, on the whole, the smoothed beta bootstrap works fairly well and gives us a potent alternative to the existing resampling schemes, although its effectiveness somewhat varies from one copula to another.

Higher-order asymptotics for the various nonparametric copula estimators might help us understand what happens with various resampling procedures [8, 19], although calculating such expansions seems to be a formidable task.

Appendix: Mathematical Proofs

Proof of Proposition 2.3

Let $\mathcal{N}_n = \{i = 1, \dots, n : W_{ni} \geqslant 1\}$ be the set of indices that are sampled at least once. Then $\mathbb{F}^*_{nj}$ is a discrete distribution function with atoms $\{X^*_{ij} : i \in \mathcal{N}_n\}$ and probabilities $n^{-1} W_{ni}$.

Since $n^{-1} R^*_{ij,n} = \mathbb{F}^*_{nj}(X_{ij})$, we have

$$\begin{aligned}
\left|\mathbb{C}^*_n(\boldsymbol{u}) - \tilde{\mathbb{C}}^*_n(\boldsymbol{u})\right| &\leqslant \frac{1}{n} \sum_{i \in \mathcal{N}_n} W_{ni} \left| \prod_{j=1}^{d} \mathbb{1}\{X_{ij} \leqslant \mathbb{F}^{*-}_{nj}(u_j)\} - \prod_{j=1}^{d} \mathbb{1}\{\mathbb{F}^*_{nj}(X_{ij}) \leqslant u_j\} \right| \\
&\leqslant \frac{1}{n} \sum_{i \in \mathcal{N}_n} W_{ni} \sum_{j=1}^{d} \left| \mathbb{1}\{X_{ij} \leqslant \mathbb{F}^{*-}_{nj}(u_j)\} - \mathbb{1}\{\mathbb{F}^*_{nj}(X_{ij}) \leqslant u_j\} \right| \\
&= \frac{1}{n} \sum_{i \in \mathcal{N}_n} W_{ni} \sum_{j=1}^{d} \left| \mathbb{1}\{X_{ij} = \mathbb{F}^{*-}_{nj}(u_j)\} - \mathbb{1}\{\mathbb{F}^*_{nj}(X_{ij}) = u_j\} \right|,
\end{aligned}$$

where in the last equality, we used the fact that $x < G^-(u)$ if and only if $G(x) < u$ for any (right-continuous) distribution function G, any real x, and any $u \in [0, 1]$. For each $j \in \{1, \dots, d\}$, $\mathbb{F}^*_{nj}(X_{ij}) = u_j$ implies $X_{ij} = \mathbb{F}^{*-}_{nj}(u_j)$ since $\mathbb{F}^*_{nj}$ jumps at X_{ij}, $i \in \mathcal{N}_n$, and there is at most a single $i \in \mathcal{N}_n$ such that $X_{ij} = \mathbb{F}^{*-}_{nj}(u_j)$. Thus we have

$$\left|\mathbb{C}^*_n(\boldsymbol{u}) - \tilde{\mathbb{C}}^*_n(\boldsymbol{u})\right| \leqslant \frac{d}{n} \max_{i=1,\dots,n} W_{ni}.$$

By coupling the multinomial random vector $(W_{n1}, \dots, W_{nn})$ to a vector of independent Poisson(1) random variables $(W'_{n1}, \dots, W'_{nn})$ as in [23, pp. 346–348], it can be shown that $\max_{i=1,\dots,n} W_{ni} = O_p(\log n)$ as $n \to \infty$. Equation (10) follows.

Proof of Proposition 3.1

Fix $\varepsilon > 0$. We know from (11) that $\tilde{\alpha}_n$ converges weakly in $\ell^\infty([0, 1]^d)$ to a Gaussian process with continuous trajectories. Write $\boldsymbol{S} = (S_1, \dots, S_d)$ and for a point $\boldsymbol{x} \in \mathbb{R}^d$, put $|\boldsymbol{x}|_\infty = \max(|x_1|, \dots, |x_d|)$. Further, put $\|f\|_\infty = \sup\{|f(\boldsymbol{u})| : \boldsymbol{u} \in [0, 1]^d\}$ for $f : [0, 1]^d \to \mathbb{R}$. Then

$$\begin{aligned}
|\alpha^\beta_n(\boldsymbol{u}) - \tilde{\alpha}_n(\boldsymbol{u})| &\leqslant \mathrm{E}_S\left[|\tilde{\alpha}_n(S_1/n, \dots, S_d/n) - \tilde{\alpha}_n(\boldsymbol{u})|\right] \\
&\leqslant 2\|\tilde{\alpha}_n\|_\infty \, \mathrm{P}_S[|\boldsymbol{S}/n - \boldsymbol{u}|_\infty > \varepsilon] + \sup_{\substack{\boldsymbol{v},\boldsymbol{w} \in [0,1]^d \\ |\boldsymbol{v}-\boldsymbol{w}|_\infty \leqslant \varepsilon}} |\tilde{\alpha}_n(\boldsymbol{v}) - \tilde{\alpha}_n(\boldsymbol{w})|.
\end{aligned}$$

Let $Y_n(\varepsilon)$ denote the supremum on the right-hand side. By Tchebysheff's inequality, the probability in the first term on the right-hand side is bounded by a constant multiple of $n^{-1/2}\varepsilon^{-1}$ and thus tends to zero uniformly in $\boldsymbol{u} \in [0, 1]^d$. Since $\|\tilde{\alpha}_n\|_\infty =$

$O_p(1)$, we get $\|\alpha_n^\beta - \tilde{\alpha}_n\|_\infty = o_p(1) + Y_n(\varepsilon)$ as $n \to \infty$. By weak convergence of $\tilde{\alpha}_n$ in $\ell^\infty([0,1])$ to a process with continuous trajectories, we can find, for any $\eta > 0$, a sufficiently small $\varepsilon > 0$ such that $\limsup_{n\to\infty} \mathrm{P}[Y_n(\varepsilon) > \eta] \leqslant \eta$. Equation (14) follows. □

Proof of Proposition 3.3

Step 1. — Recall the C-pinned Brownian sheet $\mathbb{U}^C$ defined prior to Theorem 2.2. We will show that

$$\gamma_n^\# := \sqrt{n}(\mathbb{G}_n^\# - \mathbb{C}_n^\beta) \underset{V}{\overset{\mathrm{P}}{\rightsquigarrow}} \mathbb{U}^C, \qquad n \to \infty. \tag{17}$$

Because of (7), there are two claims to show: convergence in the bounded Lipschitz metric (Step 1.1) and asymptotic measurability (Step 1.2).

Step 1.1 — Let $\mathcal{P}$ denote the set of all Borel probability measures on $[0,1]^d$. For $P \in \mathcal{P}$, let $\mathbb{U}^P$ denote a tight, P-Brownian bridge on $[0,1]^d$. Specifically, $\mathbb{U}^P$ is a centered Gaussian process with covariance function $\mathrm{E}[\mathbb{U}^P(\boldsymbol{u})\mathbb{U}^P(\boldsymbol{v})] = F(\boldsymbol{u} \wedge \boldsymbol{v}) - F(\boldsymbol{u})F(\boldsymbol{v})$, where F is the cumulative distribution function associated to P, and whose trajectories are uniformly continuous a.s. with respect to the standard deviation semimetric [23, Example 1.5.10]

$$d_P(\boldsymbol{u},\boldsymbol{v}) = \bigl(\mathrm{E}[\{\mathbb{U}^P(\boldsymbol{u}) - \mathbb{U}^P(\boldsymbol{v})\}^2]\bigr)^{1/2} = \bigl(F(\boldsymbol{u}) - 2F(\boldsymbol{u}\wedge\boldsymbol{v}) + F(\boldsymbol{v})\bigr)^{1/2}. \tag{18}$$

Further, for $P \in \mathcal{P}$, let $\mathbb{U}_{n,P}$ denote the empirical process based on independent random sampling from P. We view $\mathbb{U}_{n,P}$ as a random element of $\ell^\infty([0,1]^d)$ via the empirical and true cumulative distribution functions. Let $P_n^\beta \in \mathcal{P}$ be the (random) probability measure associated to the empirical beta copula $\mathbb{C}_n^\beta$. Recall BL_1 in (5).

We need to show that, as $n \to \infty$.

$$\sup_{h\in\mathrm{BL}_1} \Bigl| \mathrm{E}^*_{P_n^\beta}[h(\mathbb{U}_{n,P_n^\beta})] - \mathrm{E}[h(\mathbb{U}^C)] \Bigr| \longrightarrow 0 \qquad \text{in outer probability.}$$

By the triangle inequality, it is sufficient to show the pair of convergences

$$\sup_{h\in\mathrm{BL}_1} \Bigl| \mathrm{E}^*_{P_n^\beta}[h(\mathbb{U}_{n,P_n^\beta})] - \mathrm{E}[h(\mathbb{U}^{P_n^\beta})] \Bigr| \longrightarrow 0, \tag{19}$$

$$\sup_{h\in\mathrm{BL}_1} \Bigl| \mathrm{E}[h(\mathbb{U}^{P_n^\beta})] - \mathrm{E}[h(\mathbb{U}^C)] \Bigr| \longrightarrow 0, \tag{20}$$

as $n \to \infty$, in outer probability. We do this in Steps 1.1.1 and 1.1.2, respectively.

Step 1.1.1. — Identify a point $\boldsymbol{u} \in [0,1]^d$ with the indicator function $\mathbb{1}_{(-\infty,\boldsymbol{u}]}$ on $\mathbb{R}^d$. The resulting class $\mathcal{F} = \{\mathbb{1}_{(-\infty,\boldsymbol{u}]} : \boldsymbol{u} \in [0,1]^d\}$ being bounded (by 1) and VC [23, Example 2.6.1], it satisfies the uniform entropy condition (2.5.1) in [23]; see Theorem 2.6.7 in the same book. By Theorem 2.8.3 therein, we obtain the uniform Donsker property

$$\sup_{P\in\mathcal{P}} \sup_{h\in \mathrm{BL}_1} \left| \mathrm{E}^*_P[h(\mathbb{U}_{n,P})] - \mathrm{E}[h(\mathbb{U}^P)] \right| \to 0, \qquad n\to\infty. \tag{21}$$

The supremum over h in (19) is bounded by the double supremum over P and h in (21). The convergence in (19) is thus proved.

Step 1.1.2. — We need to show that, almost surely, $\mathbb{U}^{P_n^\beta} \rightsquigarrow \mathbb{U}^C$ as $n\to\infty$. All processes involved are tight, centered Gaussian processes, with covariance function determined in (4) via $\mathbb{C}_n^P$ or C. The strong consistency of the empirical copula together with [18, Proposition 2.8] yields

$$\|\mathbb{C}_n^\beta - C\|_\infty \to 0, \qquad n\to\infty, \text{ a.s.} \tag{22}$$

and this property implies (20). First, Eq. (22) implies the a.s. convergence of the covariance function of $\mathbb{U}^{P_n^\beta}$ to the one of $\mathbb{U}^C$ and thus the a.s. convergence of the finite-dimensional distributions. Second, the asymptotic tightness a.s. follows from the uniform continuity of the trajectories with respect to their respective intrinsic standard deviation semimetrics (18) and the uniform convergence a.s. of these standard deviation semimetrics, again by (22).

Step 1.2 — The asymptotic measurability of $\gamma_n^\#$ follows from the *unconditional* (i.e., jointly in $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n,\boldsymbol{V}_1^\#,\ldots,\boldsymbol{V}_n^\#$) weak convergence $\gamma_n^\# \rightsquigarrow \mathbb{U}^C$ as $n\to\infty$. This claims falls apart into convergence of the finite-dimensional distributions and asymptotic tightness. The former can be shown via the Lindeberg central limit theorem for triangular arrays conditionally on $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n$ in a way similar to the proof of Theorem 23.4 in [22]. The latter follows as in Theorems 2.5.2 and 2.8.3 in [23, p. 128 and 171] conditionally on $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n$ using the fact that the class of indicator functions of cells in $\mathbb{R}^d$ is VC [23, Example 2.5.4].

Step 2. — Consider the map Φ that sends a cumulative distribution function H on $[0,1]^d$ whose marginals do not assign mass at zero to the function $\boldsymbol{u}\mapsto H(H_1^-(u_1),\ldots,H_d(u_d))$. We have $\mathbb{C}_n^\# = \Phi(\mathbb{G}_n^\#)$ and $\mathbb{C}_n^\beta = \Phi(\mathbb{C}_n^\beta)$. By [2, Theorem 2.4], the map Φ is Hadamard differentiable at the true copula C tangentially to a certain set $\mathbb{D}_0$ at which the distribution of $\mathbb{U}^C$ is concentrated. By (17), the form of the Hadamard derivative Φ'_C of Φ at C together with the functional delta method for the bootstrap [23, Theorem 3.9.11] yield conditional weak convergence in probability of $\alpha_n^\# = \sqrt{n}\{\Phi(\mathbb{G}_n^\#) - \Phi(\mathbb{C}_n^\beta)\}$ to $\mathbb{G}_C = \Phi'_C(\mathbb{U}^C)$.

Since $|\tilde{\mathbb{C}}_n - \mathbb{C}_n| \leqslant d/n$ and $|\tilde{\mathbb{C}}_n^\# - \mathbb{C}_n^\#| \leqslant d/n$ by (3), we obtain the conditional weak convergence in probability of $\tilde{\alpha}_n^\#$ to $\mathbb{G}_C$.

Finally, since $\alpha_n^{\beta\#}(\boldsymbol{u}) = \mathrm{E}_S[\tilde{\alpha}_n^\#(S_1/n,\ldots,S_d/n)]$ as in (13), we arrive at the conditional weak convergence in probability of $\alpha_n^{\beta\#}$ to $\mathbb{G}_C$ in a way similar to the proof of Proposition 3.1. □

References

1. A. Bücher and H. Dette. A note on bootstrap approximations for the empirical copula process. *Statistics and Probability Letters*, 80:1925–1932, 2010.
2. A. Bücher and S. Volgushev. Empirical and sequential empirical copula processes under serial dependence. *Journal of Multivariate Analysis*, 119:61–70, 2013.
3. P. Deheuvels. La fonction de dépendence empirique et ses propriétés, un test non paramétrique d'indépendance. *Bulletin de la classe des sciences, Académie Royale de Belgique, 5e série*, 65:274–292, 1979.
4. B. Efron. *The Jackknife, the Bootstrap and Other Resampling Plans*. Society for Industrial and Applied Mathematics, Philadelphia, 1982.
5. J.-D. Fermanian, D. Radulović, and M. J. Wegkamp. Weak convergence of empirical copula processes. *Bernoulli*, 10:847–860, 2004.
6. C. Genest, K. Ghoudi, and L.-P. Rivest. A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82:543–552, 1995.
7. C. Genest, J. Nešlehová, and J.-F. Quessy. Tests of symmetry for bivariate copulas. *Annals of the Institute of Statistical Mathematics*, 64:811âĂŞ–834, 2012.
8. P. Hall. *The Bootstrap and Edgeworth Expansion*. Springer-Verlag, New York, 1992.
9. M. Hofert, I. Kojadinovic, M. Maechler, J. Yan, and J. G. Nešlehová. Package 'copula', R package version 0.999-19, 2018.
10. M. Hollander, D. A. Wolfe, and E. Chicken. *Nonparametric Statistical Methods*. John Wiley & Sons, Hoboken, New Jersey, third edition, 2014.
11. A. Khoudraji. *Contributions à l'éude des copules et à la modélisation de valeurs extrêmes bivariés*. PhD thesis, Université Laval, Québec, Canada, 1995.
12. A. Kiriliouk, J. Segers, and L. Tafakori. An estimator of the stable tail dependence function-based on the empirical beta copula. *Extremes*, 21:581–600, 2018.
13. I. Kojadinovic and J. Yan. A non-parametric test of exchangeability for extreme-value and left-tail decreasing bivariate copulas. *Scandinavian Journal of Statistics*, 39:480–496, 2012.
14. M. R. Kosorok. *Introduction to Empirical Processes and Semiparametric Inference*. Springer, New York, 2008.
15. R. B. Nelsen. *An Introduction to Copulas*. Springer-Verlag, New York, second edition, 2006.
16. B. Rémillard and O. Scaillet. Testing for equality between two copulas. *Journal of Multivariate Analysis*, 100:377–386, 2009.
17. J. Segers. Asymptotics of empirical copula processes under nonrestrictive smoothness assumptions. *Bernoulli*, 18:764–782, 2012.
18. J. Segers, M. Sibuya, and H. Tsukahara. The empirical beta copula. *Journal of Multivariate Analysis*, 155:35–51, 2017.
19. J. Shao and D. Tu. *The Jackknife and Bootstrap*. Springer-Verlag, New York, 1995.
20. M. Sklar. Fonctions de répartition á n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8:229–231, 1959.
21. H. Tsukahara. Semiparametric estimation in copula models. *Canadian Journal of Statistics*, 33:357–375, 2005. [Erratum: *Canadian Journal of Statistics*, **39**, 734–735 (2011)].
22. A. W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, Cambridge, 1998.
23. A. W. van der Vaart and J. A. Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer-Verlag, New York, 1996.

Table 8 Actual power of symmetry tests based on R_n with p-values computed by the multiplier bootstrap (exchTest), the straightforward bootstrap (boot), the smoothed beta bootstrap (beta) and of the test based on R_n^β with p-values computed by the smoothed beta bootstrap (beta2), for the Clayton and Gauss copulas, made asymmetric by Khoudraji's device. The nominal size is $\alpha = 0.05$

	Clayton						Gauss				
$\delta = 0.25$	n	50	100	200	400		n	50	100	200	400
$\tau = 0.25$	exchTest	0.025	0.031	0.042	0.049	$\tau = 0.25$	exchTest	0.033	0.028	0.034	0.050
	boot	0.006	0.017	0.029	0.042		boot	0.006	0.010	0.032	0.047
	beta	0.034	0.039	0.056	0.050		beta	0.044	0.042	0.048	0.057
	beta2	0.034	0.041	0.055	0.054		beta2	0.046	0.033	0.048	0.059
$\tau = 0.5$	exchTest	0.047	0.090	0.197	0.449	$\tau = 0.5$	exchTest	0.052	0.078	0.188	0.401
	boot	0.004	0.041	0.145	0.407		boot	0.007	0.028	0.136	0.366
	beta	0.060	0.100	0.216	0.469		beta	0.061	0.088	0.198	0.433
	beta2	0.062	0.103	0.227	0.486		beta2	0.065	0.098	0.212	0.441
$\tau = 0.75$	exchTest	0.199	0.630	0.985	1.000	$\tau = 0.75$	exchTest	0.205	0.637	0.973	1.000
	boot	0.038	0.380	0.949	1.000		boot	0.051	0.338	0.921	1.000
	beta	0.227	0.637	0.981	1.000		beta	0.208	0.614	0.974	1.000
	beta2	0.242	0.667	0.986	1.000		beta2	0.225	0.639	0.986	1.000
$\delta = 0.5$	n	50	100	200	400		n	50	100	200	400
$\tau = 0.25$	exchTest	0.028	0.029	0.050	0.055	$\tau = 0.25$	exchTest	0.044	0.053	0.053	0.083
	boot	0.008	0.011	0.031	0.054		boot	0.009	0.019	0.034	0.068
	beta	0.035	0.033	0.056	0.064		beta	0.051	0.059	0.055	0.090
	beta2	0.038	0.037	0.059	0.064		beta2	0.052	0.062	0.056	0.091
$\tau = 0.5$	exchTest	0.069	0.127	0.269	0.576	$\tau = 0.5$	exchTest	0.100	0.203	0.388	0.730
	boot	0.015	0.068	0.219	0.539		boot	0.027	0.119	0.326	0.695
	beta	0.077	0.140	0.299	0.593		beta	0.105	0.213	0.410	0.741
	beta2	0.075	0.144	0.306	0.602		beta2	0.116	0.225	0.416	0.750
$\tau = 0.75$	exchTest	0.385	0.814	0.997	1.000	$\tau = 0.75$	exchTest	0.478	0.914	1.000	1.000
	boot	0.125	0.644	0.993	1.000		boot	0.198	0.792	1.000	1.000
	beta	0.393	0.824	0.997	1.000		beta	0.475	0.916	1.000	1.000
	beta2	0.425	0.833	0.998	1.000		beta2	0.507	0.923	1.000	1.000
$\delta = 0.75$	n	50	100	200	400		n	50	100	200	400
$\tau = 0.25$	exchTest	0.032	0.039	0.046	0.054	$\tau = 0.25$	exchTest	0.043	0.046	0.056	0.076
	boot	0.008	0.020	0.030	0.047		boot	0.016	0.027	0.036	0.060
	beta	0.036	0.043	0.051	0.060		beta	0.047	0.053	0.061	0.081
	beta2	0.039	0.045	0.054	0.058		beta2	0.053	0.053	0.060	0.081
$\tau = 0.5$	exchTest	0.042	0.089	0.129	0.266	$\tau = 0.5$	exchTest	0.088	0.169	0.317	0.655
	boot	0.012	0.045	0.100	0.251		boot	0.030	0.113	0.271	0.604
	beta	0.053	0.094	0.132	0.279		beta	0.099	0.184	0.342	0.660
	beta2	0.053	0.102	0.140	0.287		beta2	0.102	0.194	0.354	0.670
$\tau = 0.75$	exchTest	0.144	0.372	0.693	0.962	$\tau = 0.75$	exchTest	0.369	0.636	0.947	1.000
	boot	0.051	0.268	0.645	0.958		boot	0.133	0.510	0.918	1.000
	beta	0.156	0.382	0.720	0.965		beta	0.303	0.637	0.950	1.000
	beta2	0.169	0.402	0.727	0.966		beta2	0.334	0.664	0.954	1.000