Bernoulli **23**(4B), 2017, 3346–3384
DOI: 10.3150/16-BEJ849

Weak convergence of empirical copula processes indexed by functions

DRAGAN RADULOVIĆ[1], MARTEN WEGKAMP[2] and YUE ZHAO[3]

[1]*Department of Mathematics, Florida Atlantic University, 777 Glades Road, Boca Raton, FL 33431, USA. E-mail: radulovi@fau.edu*
[2]*Department of Mathematics and Department of Statistical Science, Cornell University, 432 Malott Hall, Ithaca, NY 14853, USA. E-mail: marten.wegkamp@cornell.edu*
[3]*Department of Statistical Science, Cornell University, 310 Malott Hall, Ithaca, NY 14853, USA. E-mail: yz453@cornell.edu*

Weak convergence of the empirical copula process indexed by a class of functions is established. Two scenarios are considered in which either some smoothness of these functions or smoothness of the underlying copula function is required.

A novel integration by parts formula for multivariate, right-continuous functions of bounded variation, which is perhaps of independent interest, is proved. It is a key ingredient in proving weak convergence of a general empirical process indexed by functions of bounded variation.

Keywords: Donsker classes; empirical copula process; integration by parts; multivariate functions of bounded variation; weak convergence

1. Introduction

Let F be a distribution function in $\mathbb{R}^d$ with continuous marginals F_j, $j \in \{1, \dots, d\}$ and copula function C; we will make this a blanket assumption throughout the paper. Given an i.i.d. sample $\mathbf{X}_1, \dots, \mathbf{X}_n$ with $\mathbf{X}_i = (X_{i1}, \dots, X_{id})$ distributed according to F, we can construct the empirical distribution function

$$\mathbb{F}_n(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{\mathbf{X}_i \le \mathbf{x}\}, \qquad \mathbf{x} \in \mathbb{R}^d,$$

with marginals $\mathbb{F}_{nj}$, $j \in \{1, \dots, d\}$. The ordinary empirical copula function is defined by

$$\mathbb{C}_n(\mathbf{u}) = \mathbb{F}_n\big(\mathbb{F}_{n1}^{-}(u_1), \dots, \mathbb{F}_{nd}^{-}(u_d)\big), \qquad \mathbf{u} = (u_1, \dots, u_d) \in [0,1]^d.$$

Here, for a distribution function H, its generalized inverse function H^{-1} is defined as

$$H^{-1}(p) = \begin{cases} \inf\{t \in \mathbb{R} : H(t) \ge p\}, & 0 < p \le 1, \\ \sup\{t \in \mathbb{R} : H(t) = 0\}, & p = 0. \end{cases}$$

Then, the ordinary empirical copula process is given by

$$\sqrt{n}(\mathbb{C}_n - C)(\mathbf{u}), \qquad \mathbf{u} \in [0,1]^d. \tag{1}$$

1350-7265

The asymptotic behavior of the ordinary empirical copula process is well studied, see for instance [13,15,23,28]. In this paper, we always assume that the space $\ell^\infty(\mathcal{T})$ for some index set $\mathcal{T}$ is equipped with the supremum norm. [27] obtained weak convergence of the process (1) in $\ell^\infty([0,1]^d)$ under the weak condition that the first-order partial derivatives of the copula function C exist and are continuous on subsets of the unit hypercube; we state this condition precisely in Assumption P. He slightly relaxed the condition used in [13] that required existence and continuity of the first-order partial derivatives of C on the *entire* hypercube. Surely, the condition in [27] is mild, as Theorem 4 in [13] showed that the empirical copula process no longer converges in $\ell^\infty([0,1]^d)$ if the continuity of any of the d first-order partial derivatives fails at a point $\mathbf{u} \in (0,1)^d$. [6] used a weaker semi-metric on $\ell^\infty([0,1]^d)$ and obtain *hypi*-convergence of the empirical copula process, under the assumption stated in their condition 4.3 that the set of points in $[0,1]^d$ where the first-order partial derivatives of the copula function C exist and are continuous has Lebesgue measure one. They showed that hypi-convergence still implies weak convergence of certain Kolmogorov–Smirnov and Cramér–von Mises test statistics. Hypi-convergence is not studied in this paper.

While it can be verified that $\mathbb{C}_n$ is left-continuous with right-hand limits, its cousin

$$\bar{\mathbb{C}}_n(\mathbf{u}) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}\{\mathbb{F}_{n1}(X_{i1}) \le u_1, \dots, \mathbb{F}_{nd}(X_{id}) \le u_d\}, \qquad \mathbf{u} = (u_1,\dots,u_d) \in [0,1]^d$$

is *càdlàg* (right-continuous with left-hand limits) and as such a more standard object in probability theory and, in particular, Lebesgue–Stieltjes integration. We will refer to the process

$$\sqrt{n}(\bar{\mathbb{C}}_n - C)(\mathbf{u}), \qquad \mathbf{u} \in [0,1]^d \tag{2}$$

as the càdlàg version of the ordinary empirical copula process given in (1), or simply the empirical copula process. Given an i.i.d. sample, the difference between the ordinary empirical copula process and its càdlàg version is small; specifically, we will show in Appendix C.1 that, almost surely,

$$\sup_{\mathbf{u}\in[0,1]^d} \left|\sqrt{n}(\mathbb{C}_n - C)(\mathbf{u}) - \sqrt{n}(\bar{\mathbb{C}}_n - C)(\mathbf{u})\right| \le \frac{d}{\sqrt{n}}, \tag{3}$$

whose bivariate version is pointed out in the proof of Theorem 6 in [13]. Hence, the empirical copula process given in (2) converges weakly in $\ell^\infty([0,1]^d)$ under the same weak condition as in [27].

This paper addresses the following question: *Can we generalize the empirical copula process to a process indexed by functions on the unit hypercube, rather than points in the unit hypercube?* Specifically, based on the empirical copula process $\bar{\mathbb{C}}_n$ and functions $g : [0,1]^d \to \mathbb{R}$ belonging to a class $\mathcal{G}$, we consider the generalization

$$\begin{aligned}\bar{\mathbb{Z}}_n(g) &= \sqrt{n}\int_{(0,1]^d} g(\mathbf{u})\,\mathrm{d}(\bar{\mathbb{C}}_n - C)(\mathbf{u}) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \left\{g\big(\mathbb{F}_{n1}(X_{i1}),\dots,\mathbb{F}_{nd}(X_{id})\big) - \mathbb{E}\big[g\big(F_1(X_{i1}),\dots,F_d(X_{id})\big)\big]\right\}.\end{aligned} \tag{4}$$

(In the above, the particular domain of integration $(0,1]^d$ is chosen for later convenience. Because the boundary $[0,1]^d \setminus (0,1]^d$ has measure zero with respect to both $\mathrm{d}\bar{\mathbb{C}}_n$ and $\mathrm{d}C$, it is easily seen that $\bar{\mathbb{Z}}_n(g)$ remains unchanged if the domain of integration $[0,1]^d$ is used instead.) From here on, we use the convention that integral without domain of integration explicitly specified is understood to be over $(0,1]^d$. The generalization (4) is of particular interest because $\bar{\mathbb{Z}}_n(g)$ is a multivariate rank order statistic that is common in the statistics literature. See [24,25] and [23] for early references. Among other considerations, this justifies taking $\bar{\mathbb{C}}_n$ as our starting point. Clearly, (4) reduces to (2) for $g(\mathbf{v}) = 1\{\mathbf{v} \leq \mathbf{u}\}$, and Theorem 6 in [13] states that, for each g that is suitably regular, the statistic (4) has a normal limiting distribution. This leads to the question "*Can we characterize the class $\mathcal{G}$ of functions $g : [0,1]^d \to \mathbb{R}$ for which the process $\bar{\mathbb{Z}}_n$ in* (4) *converges weakly in $\ell^\infty(\mathcal{G})$?*"

To answer this question, we consider two complementary cases, one that requires some smoothness of the underlying copula function C and one that requires smoothness of the indexing functions $g \in \mathcal{G}$. [32] showed that if the functions g are sufficiently smooth, then existence of first-order partial derivatives of C is no longer required for the weak convergence of the process $\bar{\mathbb{Z}}_n$ in $\ell^\infty(\mathcal{G})$. This remarkable fact was established in Corollary 5.4 in [32]. Theorem 7 in our paper corrects a minor mistake in their proof – uniform equicontinuity in lieu of mere continuity of the partial derivatives of $g \in \mathcal{G}$ is required, and demonstrates the weak convergence in a different way under weaker assumptions on $\mathcal{G}$ that require no explicit entropy conditions on $\mathcal{G}$. We stress that many well-known copulas are not differentiable, for example, the Fréchet–Hoeffding copulas, the Marshall–Olkin copula, the Cuadras–Augé copula, the Raftery copula, among many others, see the monograph [20]. Moreover, many of the common goodness-of-fit tests for copulas rely on the weak convergence of the (standard) empirical copula process and thus do not apply in non-differentiable settings.

The scenario where C is sufficiently smooth, while functions in $\mathcal{G}$ are not necessarily differentiable has not been addressed in the literature. In case the underlying copula satisfies Assumption P, we show that under mild conditions on $\mathcal{G}$ the process $\bar{\mathbb{Z}}_n$ converges weakly in $\ell^\infty(\mathcal{G})$. We found a surprisingly simple proof for this fact based on the very general result, Theorem 1 below, which is of interest in its own.

The paper is organized as follows. Section 2 presents a general weak convergence result of empirical processes, indexed by functions of bounded variation, including empirical processes based on stationary sequences satisfying alpha-mixing (or strong-mixing) conditions. We stress that alpha-mixing is the least restrictive form of available mixing assumptions in the literature. In this case, the few results in the literature that treat empirical processes indexed by functions $g \in \mathcal{G}$ all require stringent entropy conditions on $\mathcal{G}$ and on the rate of decay for the mixing coefficients of $\mathbf{X}_i$, see, for instance, [2]. The main culprit is that alpha-mixing does not allow for sharp exponential inequalities for partial sums. The only known cases for which sharp conditions do exist are under more restrictive, beta-mixing dependence. The latter allows for decoupling and yields exponential inequalities not unlike the i.i.d. case [3,10]. Our theory does not stop there and allows for short memory casual linear sequences [11]. [9] proved weak convergence of the standard empirical processes based on stationary sequences that are not necessarily mixing. [8] treated more general processes indexed by classes $\mathcal{G}$ of functions under cumbersome entropy conditions on $\mathcal{G}$. The advantage of the method presented in this paper is that no explicit entropy condition on the set $\mathcal{G}$ is imposed, while only weak convergence of the standard empirical process is required.

Section 3 presents the main results for empirical copula processes indexed by functions. Smoothness of either the copula function C or the indexing functions $g \in \mathcal{G}$ is required.

The proofs of some of the results in Section 3 are deferred to Section 4.

Finally, Appendix A contains a novel integration by parts formula for multivariate, right-continuous functions of bounded variation, which is perhaps of independent interest, Appendix B contains some technical results, and Appendix C provides some bounds on the distance between the ordinary empirical copula process and its càdlàg version.

1.1. Notations

We list in this subsection the notations necessary to address the multivariate extension of the concept of bounded variation and the integration by parts formula in this paper. We mostly follow the notations introduced in Section 3 of [21]. For $\mathbf{x} \in \mathbb{R}^d$, we denote its jth component as x_j, that is, $\mathbf{x} = (x_1, \dots, x_d)$. We let $\mathbf{0} \in \mathbb{R}^d$ be the vector with all components equal to zero, and $\mathbf{1} \in \mathbb{R}^d$ be the vector with all components equal to one. For $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$, we write $\mathbf{a} < \mathbf{b}$ or $\mathbf{a} \le \mathbf{b}$ if these inequalities hold for all d components. For $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$ with $\mathbf{a} \le \mathbf{b}$, the hypercube $[\mathbf{a}, \mathbf{b}]$ is the set $\{\mathbf{x} \in \mathbb{R}^d : \mathbf{a} \le \mathbf{x} \le \mathbf{b}\}$. Thus $[\mathbf{0}, \mathbf{1}] = [0, 1]^d$ is the closed unit hypercube, and in this paper we will work exclusively over this domain unless specified otherwise. Similarly, $(\mathbf{a}, \mathbf{b}] = \{\mathbf{x} \in \mathbb{R}^d : \mathbf{a} < \mathbf{x} \le \mathbf{b}\}$.

For $I, J \subset \{1, \dots, d\}$, we write $|I|$ for the cardinality of I, and $I - J$ for the complement of J with respect to I. A unary minus denotes the complement with respect to $\{1, \dots, d\}$, so that $-I = \{1, \dots, d\} - I$. In expressions involving both the unary minus and other set operations, the unary minus has the highest precedence; for instance, $-I - J = (\{1, \dots, d\} - I) - J$.

For $I \subset \{1, \dots, d\}$, the expression $\mathbf{x}_I$ denotes an $|I|$-tuple of real numbers representing the components x_j for $j \in I$. The domain of $\mathbf{x}_I$ is (typically) the hypercube $[\mathbf{0}_I, \mathbf{1}_I]$. Suppose that $I, J \subset \{1, \dots, d\}$ with $I \cap J = \varnothing$, and $\mathbf{x}, \mathbf{z} \in [\mathbf{0}, \mathbf{1}]$. Then we define the concatenation symbol ':' such that the vector $\mathbf{x}_I : \mathbf{z}_J$ represents the point $\mathbf{y} \in [\mathbf{0}_{I \cup J}, \mathbf{1}_{I \cup J}]$ with $y_j = x_j$ for $j \in I$, and $y_j = z_j$ for $j \in J$. The vector $\mathbf{x}_I : \mathbf{z}_J$ is well defined for $\mathbf{x}_I \in [\mathbf{0}_I, \mathbf{1}_I]$ and $\mathbf{z}_J \in [\mathbf{0}_J, \mathbf{1}_J]$ when $I \cap J = \varnothing$, even if $\mathbf{x}_{-I}$ or $\mathbf{z}_{-J}$ is left unspecified. We also use the concatenation symbol to glue together more than two sets of components. For instance $\mathbf{x}_I : \mathbf{y}_J : \mathbf{z}_K \in [0, 1]^d$ is well defined for $\mathbf{x}_I \in [\mathbf{0}_I, \mathbf{1}_I]$, $\mathbf{y}_J \in [\mathbf{0}_J, \mathbf{1}_J]$ and $\mathbf{z}_K \in [\mathbf{0}_K, \mathbf{1}_K]$ when I, J, K are mutually disjoint sets whose union is $\{1, \dots, d\}$. The main purpose of the concatenation symbol is to construct the argument to a function by taking components from multiple sources.

For a function $f : [0, 1]^d \to \mathbb{R}$, a set $I \subset \{1, \dots, d\}$ and a constant vector $\mathbf{c}_{-I} \in [\mathbf{0}_{-I}, \mathbf{1}_{-I}]$, we can define a function $g : [\mathbf{0}_I, \mathbf{1}_I] \to \mathbb{R}$ as a lower-dimensional projection of f onto $[\mathbf{0}_I, \mathbf{1}_I]$ via $g(\mathbf{x}_I) = f(\mathbf{x}_I : \mathbf{c}_{-I})$. We write $f(\cdot; \mathbf{c}_{-I})$ to denote the function on $[\mathbf{0}_I, \mathbf{1}_I]$ defined in this way with the argument on the left of the semicolon and the constant vector $\mathbf{c}_{-I}$ on the right, so that $f(\mathbf{x}_I; \mathbf{c}_{-I}) = f(\mathbf{x}_I : \mathbf{c}_{-I})$.

2. A general result

The main theorem in this section states that, if $\mathbb{G}_n$ is a stochastic process that converges weakly in $\ell^\infty([0, 1]^d)$ to a continuous Gaussian limit $\mathbb{G}$, then for a large class $\mathcal{F}$ of functions on $[0, 1]^d$, the weak convergence of the stochastic process $\int f\,\mathrm{d}\mathbb{G}_n$, with $f \in \mathcal{F}$, in $\ell^\infty(\mathcal{F})$ follows from the weak convergence of the stochastic process $\mathbb{G}_n$. The proof relies on Proposition 3 that gives a very general integration by parts formula for $\int f\,\mathrm{d}\mathbb{G}_n$. The main idea is to change the integration over $\mathbb{G}_n$ by integration over f. For this reason, we consider functions f for which we can uniquely define signed Borel measures on $[0,1]^d$. The classical Lebesgue–Stieltjes integration theory on $\mathbb{R}$ is based on functions f that are of bounded variation. To consider its multivariate extension, naturally we will need to consider multivariate extensions of the concept of bounded variation.

First, we briefly recall the definition of total variation in the sense of Vitali, and refer to [21] for a lucid presentation. Following [21], a *ladder* $\mathcal{Y}$ of the interval $[0,1]$ is a set containing 0 and finitely many, possibly zero, values in $(0,1)$. Each element $y \in \mathcal{Y}$ has a unique successor y^+, defined as the smallest element in $(y,1) \cap \mathcal{Y}$; if the intersection is empty, we set $y^+ = 1$. A multivariate ladder $\mathcal{Y} = \prod_{j=1}^d \mathcal{Y}_j$ of $[0,1]^d$ is based on d one-dimensional ladders $\mathcal{Y}_j$, not necessarily of the same cardinality, with $\mathcal{Y}_j$ being a ladder of $[\mathbf{0}_{\{j\}}, \mathbf{1}_{\{j\}}]$. Then, the successor $\mathbf{y}^+ = (y_1^+, \dots, y_d^+)$ of $\mathbf{y} = (y_1, \dots, y_d) \in \mathcal{Y}$ is defined by taking each coordinate y_j^+ to be the successor of y_j on the jth coordinate. Because for each coordinate j, the successor y_j^+ of y_j is unique, the successor $\mathbf{y}^+$ of $\mathbf{y}$ is unique as well. Next, we let the d-fold alternating sum of f over the hypercube $(\mathbf{a}, \mathbf{b}]$ be

$$\Delta(f; \mathbf{a}, \mathbf{b}) = \sum_{I \subset \{1,\dots,d\}} (-1)^{|I|} f(\mathbf{a}_I : \mathbf{b}_{-I}). \tag{5}$$

(In the terminology of [1] (see their equation (14)), $\Delta(f; \mathbf{a}, \mathbf{b})$ is the d-dimensional quasi-volume of the hypercube $(\mathbf{a}, \mathbf{b}]$ assigned by the function f. A formal version of this statement will appear later in (8) in terms of the measure $\mathrm{d}f$ defined from the function f.) The variation of a function f on $[0,1]^d$ over a multivariate ladder $\mathcal{Y}$ is

$$V_{\mathcal{Y}}(f) = \sum_{\mathbf{y} \in \mathcal{Y}} \bigl|\Delta\bigl(f; \mathbf{y}, \mathbf{y}^+\bigr)\bigr|.$$

Finally, the total variation of the function f on $[0,1]^d$ in the sense of Vitali, or simply the Vitali variation of f, is

$$V(f) := \sup_{\mathcal{Y}} V_{\mathcal{Y}}(f).$$

Here the supremum is taken over all multivariate ladders $\mathcal{Y} = \prod_{j=1}^d \mathcal{Y}_j$ of $[0,1]^d$.

We will also need to consider total variation in the sense of Krause [17,18] and Hardy [16]. Formally, the variation of a function f on $[0,1]^d$ in the sense of Hardy–Krause (assumed to be anchored at $\mathbf{1}$ unless stated otherwise), or simply the Hardy–Krause variation of f, is

$$V_{\mathrm{HK}}(f) = \sum_{I \subset \{1,\dots,d\} : I \neq \varnothing} V\bigl(f(\cdot; \mathbf{1}_{-I})\bigr). \tag{6}$$

Here $V(f(\cdot; \mathbf{1}_{-I}))$ is the Vitali variation of the function $f(\cdot; \mathbf{1}_{-I})$ on $[\mathbf{0}_I, \mathbf{1}_I]$. (We recall from Section 1.1 that the function $f(\cdot; \mathbf{1}_{-I}) : [\mathbf{0}_I, \mathbf{1}_I] \to \mathbb{R}$ is the lower-dimensional projection of f

onto $[\mathbf{0}_I, \mathbf{1}_I]$ obtained by setting $f(\mathbf{x}_I; \mathbf{1}_{-I}) = f(\mathbf{x}_I : \mathbf{1}_{-I})$.) If $V_{\mathrm{HK}}(f) < \infty$, then we say that the function f is of bounded Hardy–Krause variation.

From (6), it is clear that a function f on $[0,1]^d$ is of bounded Hardy–Krause variation if and only if for each $I \subset \{1,\dots,d\}$ with $I \neq \varnothing$, the Vitali variation $V(f(\cdot;\mathbf{1}_{-I}))$ of the function $f(\cdot;\mathbf{1}_{-I})$ on $[\mathbf{0}_I, \mathbf{1}_I]$ is bounded. For each $I \subset \{1,\dots,d\}$ with $I \neq \varnothing$, we let the mixed partial derivative of $f(\mathbf{x}_I; \mathbf{1}_{-I})$ taken once with respect to x_j for every $j \in I$ be denoted as $\partial^I f(\mathbf{x}_I; \mathbf{1}_{-I})$. Proposition 14 in [21] states that if $\partial^I f(\mathbf{x}_I; \mathbf{1}_{-I})$ is continuous on $[\mathbf{0}_I, \mathbf{1}_I]$, then

$$V\big(f(\cdot;\mathbf{1}_{-I})\big) = \int_{[\mathbf{0}_I,\mathbf{1}_I]} \big|\partial^I f(\mathbf{x}_I;\mathbf{1}_{-I})\big| \,\mathrm{d}\mathbf{x}_I.$$

Hence a sufficient, but by no means necessary, condition for a function f on $[0,1]^d$ to be of bounded Hardy–Krause variation is that for each $I \subset \{1,\dots,d\}$ with $I \neq \varnothing$, the mixed partial derivative $\partial^I f(\mathbf{x}_I; \mathbf{1}_{-I})$ is continuous on $[\mathbf{0}_I, \mathbf{1}_I]$. A more general characterization of functions of bounded Hardy–Krause variation is provided by a section of part (b) of Theorem 3 in [1] which states that if $\mathrm{d}f$ is a finite signed Borel measure on $[0,1]^d$, then there exists a unique right-continuous (see Assumption F for our definition of right-continuity) function f on $[0,1]^d$ of bounded Hardy–Krause variation for which

$$\mathrm{d}f\big([\mathbf{0},\mathbf{x}]\big) = f(\mathbf{x}), \qquad \mathbf{x} \in [0,1]^d. \tag{7}$$

We will mostly consider functions satisfying the following assumption:

Assumption F. *$f : [0,1]^d \to \mathbb{R}$ is right-continuous (to be precise, following [1], we say a function is right-continuous if it is coordinatewise right-continuous in each coordinate, at every point) and is of bounded Hardy–Krause variation, that is, $V_{\mathrm{HK}}(f) < \infty$.*

By part (a) of Theorem 3 in [1], which is a converse of part (b) of the same theorem we just mentioned above, if a function f satisfies Assumption F, then there exists a unique, finite signed Borel measure $\mathrm{d}f$ on $[0,1]^d$ for which (7) holds. From (7), it is easy to see that, for $\mathbf{a},\mathbf{b} \in [0,1]^d$ with $\mathbf{a} \le \mathbf{b}$, the measure $\mathrm{d}f$ assigns weight

$$\mathrm{d}f\big((\mathbf{a},\mathbf{b}]\big) = \int_{(\mathbf{a},\mathbf{b}]} \mathrm{d}f(\mathbf{x}) = \sum_{I \subset \{1,\dots,d\}} (-1)^{|I|} f(\mathbf{a}_I : \mathbf{b}_{-I}) \tag{8}$$

to the hypercube $(\mathbf{a},\mathbf{b}]$. (We note the similarity between (5) and (8), but also point out that $\mathrm{d}f$ is a Borel measure that is defined on all Borel sets, not just hypercubes.) In fact, we can conclude from [1] a more general result that we will also use later: if a function f satisfies Assumption F, then for arbitrary $I \subset \{1,\dots,d\}$ with $I \neq \varnothing$ and arbitrary $\mathbf{c} \in [0,1]^d$, to the lower-dimensional projection $f(\cdot;\mathbf{c}_{-I})$ on $[\mathbf{0}_I, \mathbf{1}_I]$ there corresponds a unique, finite signed Borel measure $\mathrm{d}f(\cdot;\mathbf{c}_{-I})$ on $[\mathbf{0}_I, \mathbf{1}_I]$ such that

$$\mathrm{d}f\big([\mathbf{0}_I,\mathbf{x}_I];\mathbf{c}_{-I}\big) = f(\mathbf{x}_I;\mathbf{c}_{-I}), \qquad \mathbf{x}_I \in [\mathbf{0}_I,\mathbf{1}_I]. \tag{9}$$

The validity of this claim is verified in Appendix B. Hence, for $\mathbf{a}, \mathbf{b} \in [0,1]^d$ with $\mathbf{a}_I \le \mathbf{b}_I$, the measure $\mathrm{d}f(\cdot; \mathbf{c}_{-I})$ assigns weight

$$\mathrm{d}f\big((\mathbf{a}_I, \mathbf{b}_I]; \mathbf{c}_{-I}\big) = \int_{(\mathbf{a}_I, \mathbf{b}_I]} \mathrm{d}f(\mathbf{x}_I; \mathbf{c}_{-I}) = \sum_{I' \subset I} (-1)^{|I'|} f(\mathbf{a}_{I'} : \mathbf{b}_{I-I'} : \mathbf{c}_{-I}) \tag{10}$$

to the hypercube $(\mathbf{a}_I, \mathbf{b}_I]$. To reiterate, we will identify a function and its lower-dimensional projections uniquely with the measures satisfying (7) and (9), and hence (8) and (10), respectively.

The main result of this section is presented next.

Theorem 1. *Let $\mathbb{G}_n$ be a stochastic process on $[0,1]^d$ such that its sample paths satisfy Assumption F almost surely, and that $\mathbb{G}_n$ converges weakly in $\ell^\infty([0,1]^d)$ to a continuous Gaussian limit $\mathbb{G}$. Assume in addition that $\mathbb{G}_n(\mathbf{u}) = 0$ almost surely, if $u_j = 0$ for some $j \in \{1, \dots, d\}$, and that $\mathbb{G}_n(\mathbf{1}) = 0$. Let $\mathcal{F}$ be a class of functions f on $[0,1]^d$ satisfying Assumption F, and additionally $\sup_{f \in \mathcal{F}} V_{\mathrm{HK}}(f) < \infty$. Then the empirical process $\int f \, \mathrm{d}\mathbb{G}_n$, indexed by $f \in \mathcal{F}$, converges weakly in $\ell^\infty(\mathcal{F})$ to a Gaussian limit.*

Remark 1. The proof of Theorem 1 reveals (through the limiting process $\widetilde{\mathbb{G}}(f)$, $f \in \mathcal{F}$ introduced in the proof) that the limiting process of $\int f \, \mathrm{d}\mathbb{G}_n$ can be characterized as

$$\widetilde{\mathbb{G}}(f) = \sum_{I \subset \{1,\dots,d\}: I \neq \varnothing} (-1)^{|I|} \int_{(\mathbf{0}_I, \mathbf{1}_I]} \mathbb{G}(\mathbf{x}_I; \mathbf{1}_{-I}) \, \mathrm{d}f(\mathbf{x}_I; \mathbf{1}_{-I})$$

for $f \in \mathcal{F}$, based on the limit $\mathbb{G}$ of $\mathbb{G}_n$.

Remark 2. We have stated earlier that if $\mathrm{d}f$ is a finite signed Borel measure on $[0,1]^d$, then there exists a unique right-continuous function f on $[0,1]^d$ of bounded Hardy–Krause variation for which (7) holds. Hence, a natural class $\mathcal{F}$ to consider is the set of such functions f arising from all signed Borel measures $\mathrm{d}f$ on $[0,1]^d$ whose total variations are uniformly bounded. By Example 2.10.4 in [31], such a collection $\mathcal{F}$ is universally Donsker, and we also have $\sup_{f \in \mathcal{F}} V_{\mathrm{HK}}(f) < \infty$ by equation (10) in Theorem 3 and Lemma 2 in [1].

Remark 3. If $\mathcal{F}$ is the particular class of functions as discussed in Remark 2 and so $\mathcal{F}$ is universally Donsker, and if in addition $\mathbb{G}_n$ is the standard empirical process from an i.i.d. sample, then standard empirical process theory straightforwardly yields the conclusion of Theorem 1. However, we stress that Theorem 1 is much more general because it requires weak convergence only of the process $\mathbb{G}_n$ (and the condition on $\mathcal{F}$). As particular instances of the applications of Theorem 1, in Corollary 2 we will apply the theorem to empirical processes based on alpha-mixing sequences, and in Section 3.1 we will apply the theorem to empirical copula processes, in the latter case by replacing the process $\mathbb{G}_n$ with $\sqrt{n}(\bar{\mathbb{C}}_n - C)$. We also note that the mild boundary conditions imposed in the second sentence of Theorem 1 are trivially satisfied in these cases.

Proof of Theorem 1. See Section 2.1. □

Now we apply Theorem 1 to empirical processes based on alpha-mixing sequences. Such a result is new, as weak convergence of empirical processes for dependent variables indexed by functions is sparse in the literature and typically requires rather restrictive beta-mixing conditions. For a stationary sequence of random variables in $[0,1]^d$, [22] proved weak convergence of the standard empirical process $\sqrt{n}(\mathbb{F}_n - F)(\mathbf{x})$ in $\ell^\infty([0,1]^d)$ under alpha-mixing conditions only.

Corollary 2. *Let $\mathbf{X}_i$, $i \in \mathbb{Z}$, be a stationary sequence of random variables in $[0,1]^d$ with continuous distribution function F and with alpha-mixing coefficients*

$$\alpha_k := \sup\bigl\{\bigl|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)\bigr|, A \in \sigma(\mathbf{X}_j, j \le i), B \in \sigma(\mathbf{X}_{k+j}, j \ge i), i \in \mathbb{Z}\bigr\}$$

satisfying

$$\alpha_k = O\bigl(k^{-a}\bigr) \qquad \text{for some } a > 1 \text{ and } k \to \infty.$$

Let $\mathbb{F}_n$ be the empirical distribution function based on $\mathbf{X}_i, \dots, \mathbf{X}_n$ and let $\mathbb{G}_n = \sqrt{n}(\mathbb{F}_n - F)$ be the standard empirical process in $\ell^\infty([0,1]^d)$. Let $\mathcal{F}$ be a class of functions f on $[0,1]^d$ satisfying Assumption F, and additionally $\sup_{f \in \mathcal{F}} V_{\mathrm{HK}}(f) < \infty$. Then the empirical process $\int f \,\mathrm{d}\mathbb{G}_n$, indexed by $f \in \mathcal{F}$, converges weakly in $\ell^\infty(\mathcal{F})$ to a Gaussian limit.

Proof. Theorem 7.3 in [22] establishes the weak convergence of the process $\mathbb{G}_n$ in $\ell^\infty([0,1]^d)$ to a continuous Gaussian limit. The corollary follows immediately from Theorem 1. □

2.1. Proof of Theorem 1

The proof of Theorem 1 relies on the following integration by parts formula for Lebesgue–Stieltjes integration (Proposition 3) and a technical result regarding functions of bounded Hardy–Krause variation (Lemma 4). For Lebesgue–Stieltjes integration, integration by parts in arbitrary dimensions, in contrast to the well-known one-dimensional formula, has been neglected in the literature.

Proposition 3. *Let g be a function on $[0,1]^d$ that satisfies Assumption F. Assume in addition that $g(\mathbf{u}) = 0$ if $u_j = 0$ for some $j \in \{1, \dots, d\}$, and that $g(\mathbf{1}) = 0$. For any other function f on $[0,1]^d$ satisfying Assumption F, we have*

$$\int_{(0,1]^d} f \,\mathrm{d}g = \sum_{I \subset \{1,\dots,d\}: I \neq \varnothing} (-1)^{|I|} \int_{(\mathbf{0}_I, \mathbf{1}_I]} g(\mathbf{x}_I-; \mathbf{1}_{-I}) \,\mathrm{d}f(\mathbf{x}_I; \mathbf{1}_{-I}). \tag{11}$$

Remark. In (11), the left-continuous function $g(\cdot-; \mathbf{1}_{-I})$ on $(\mathbf{0}_I, \mathbf{1}_I]$ is defined as

$$g(\mathbf{x}_I-; \mathbf{1}_{-I}) = \lim_{\mathbf{y}_I < \mathbf{x}_I, \mathbf{y}_I \uparrow \mathbf{x}_I} g(\mathbf{y}_I; \mathbf{1}_{-I}),$$

that is, it denotes the left-hand limit of the function $g(\cdot; \mathbf{1}_{-I})$ on all coordinates $I \subset \{1, \dots, d\}$ jointly. The same convention will be used in Theorem 15. The existence of this left-hand limit is explained in the remarks following Theorem 15.

Proof of Proposition 3. The result follows from the general formula (40) (which is equivalent to (39)) in Theorem 15 in Appendix A, by setting $\mathbf{a} = \mathbf{c} = \mathbf{0}$ and $\mathbf{b} = \mathbf{d} = \mathbf{1}$. To see this, note first that the summands in the second term on the right-hand side of (40) are identically zero by the assumption (imposed in the second sentence) of Proposition 3. Next, the summands in the first term on the right-hand side of (40) are identically zero when $I_2 \neq \varnothing$ because in this case each term $g(\mathbf{x}_{I_1}-; \mathbf{0}_{I_2} : \mathbf{1}_{I_3})$ as the integrand equals zero, again by the assumption (this time imposed in the first half of the second sentence) of Proposition 3. □

For any function f on $[0,1]^d$ that satisfies Assumption F, and any $I \subset \{1,\dots,d\}$ with $I \neq \varnothing$, we let $\mathrm{d}f(\cdot; \mathbf{1}_{-I})$ be the unique, finite signed Borel measure on $[\mathbf{0}_I, \mathbf{1}_I]$ satisfying (9) with $\mathbf{c}_{-I} = \mathbf{1}_{-I}$. We let $\mathrm{d}f(\cdot; \mathbf{1}_{-I}) = \mathrm{d}f^+(\cdot; \mathbf{1}_{-I}) - \mathrm{d}f^-(\cdot; \mathbf{1}_{-I})$ be the Jordan decomposition of the measure $\mathrm{d}f(\cdot; \mathbf{1}_{-I})$. Then, following the terminology of [1], we let the measure $|\mathrm{d}f|(\cdot; \mathbf{1}_{-I}) = \mathrm{d}f^+(\cdot; \mathbf{1}_{-I}) + \mathrm{d}f^-(\cdot; \mathbf{1}_{-I})$ be the variation measure on $[\mathbf{0}_I, \mathbf{1}_I]$, and let $|\mathrm{d}f|([\mathbf{0}_I, \mathbf{1}_I]; \mathbf{1}_{-I})$ be the total variation, both corresponding to the measure $\mathrm{d}f(\cdot; \mathbf{1}_{-I})$.

Lemma 4. *Assume that a function f on $[0,1]^d$ satisfies Assumption F. Then we have*

$$2^d \cdot V_{\mathrm{HK}}(f) \geq \max_{I \subset \{1,\dots,d\}: I \neq \varnothing} \int_{(\mathbf{0}_I, \mathbf{1}_I]} |\mathrm{d}f|(\mathbf{x}_I; \mathbf{1}_{-I}).$$

Proof. We fix arbitrary $I \subset \{1,\dots,d\}$ with $I \neq \varnothing$. We let $V_{\mathrm{HK}}(f(\cdot; \mathbf{1}_{-I}))$ be the variation of the function $f(\cdot; \mathbf{1}_{-I})$ on $[\mathbf{0}_I, \mathbf{1}_I]$ in the sense of Hardy–Krause (anchored at $\mathbf{1}_I$), which can be expressed analogously to (6) as

$$V_{\mathrm{HK}}\big(f(\cdot; \mathbf{1}_{-I})\big) = \sum_{I' \subset I: I' \neq \varnothing} V\big(f(\cdot; \mathbf{1}_{I-I'} : \mathbf{1}_{-I})\big). \tag{12}$$

Note that $f(\cdot; \mathbf{1}_{I-I'} : \mathbf{1}_{-I}) = f(\cdot; \mathbf{1}_{-I'})$, so obviously the non-negative summands on the right hand side of (6) include all the non-negative summands on the right-hand side of (12); hence we have

$$V_{\mathrm{HK}}\big(f(\cdot; \mathbf{1}_{-I})\big) \leq V_{\mathrm{HK}}(f). \tag{13}$$

Similarly, we let $V_{\mathrm{HK0}}(f(\cdot; \mathbf{1}_{-I}))$ be the variation of the function $f(\cdot; \mathbf{1}_{-I})$ on $[\mathbf{0}_I, \mathbf{1}_I]$ in the sense of Hardy–Krause but now anchored at $\mathbf{0}_I$, which is defined analogously to (12) but with $\mathbf{1}_{I-I'}$ replaced by $\mathbf{0}_{I-I'}$ on the right-hand side as

$$V_{\mathrm{HK0}}\big(f(\cdot; \mathbf{1}_{-I})\big) = \sum_{I' \subset I: I' \neq \varnothing} V\big(f(\cdot; \mathbf{0}_{I-I'} : \mathbf{1}_{-I})\big).$$

We then have

$$\begin{aligned}\int_{(\mathbf{0}_I, \mathbf{1}_I]} |\mathrm{d}f|(\mathbf{x}_I; \mathbf{1}_{-I}) &= |\mathrm{d}f|\big((\mathbf{0}_I, \mathbf{1}_I]; \mathbf{1}_{-I}\big) \leq |\mathrm{d}f|\big([\mathbf{0}_I, \mathbf{1}_I] \setminus \{\mathbf{0}_I\}; \mathbf{1}_{-I}\big) \\ &= |\mathrm{d}f|\big([\mathbf{0}_I, \mathbf{1}_I]; \mathbf{1}_{-I}\big) - |\mathrm{d}f|\big(\{\mathbf{0}_I\}; \mathbf{1}_{-I}\big)\end{aligned}$$

$$= V_{\mathrm{HK0}}\big(f(\cdot;\mathbf{1}_{-I})\big) \le \big(2^{|I|}-1\big)V_{\mathrm{HK}}\big(f(\cdot;\mathbf{1}_{-I})\big)$$
$$\le \big(2^{d}-1\big)V_{\mathrm{HK}}(f).$$

Here the transition to the third line follows by equation (10) in Theorem 3 in [1] and the fact that $|\mathrm{d}f|(\{\mathbf{0}_I\};\mathbf{1}_{-I}) = |f(\mathbf{0}_I:\mathbf{1}_{-I})|$ (the latter in turn follows from the proof of part (a) of Theorem 3 in [1]), the second inequality follows by Lemma 2 in [1], and the last inequality follows by (13). The lemma then follows. □

Proof of Theorem 1. We assume that $\sup_{f\in\mathcal{F}} V_{\mathrm{HK}}(f) \le T < \infty$. By Lemma 4, we then have

$$2^d \cdot T \ge \sup_{f\in\mathcal{F}}\Big[\max_{I\subset\{1,\dots,d\}:I\neq\varnothing}\int_{(\mathbf{0}_I,\mathbf{1}_I]}|\mathrm{d}f|(\mathbf{x}_I;\mathbf{1}_{-I})\Big]. \tag{14}$$

We let

$$\mathcal{X} = \big\{X:[0,1]^d\to\mathbb{R}, \|X\|_\infty<\infty, X \text{ is right-continuous}, \\ X(\mathbf{u}) = 0 \text{ if } u_j = 0 \text{ for some } j\in\{1,\dots,d\}, X(\mathbf{1}) = 0\big\}$$

be the metric subspace of $\ell^\infty([0,1]^d)$ (equipped with the same supremum norm) that is the collection of right-continuous functions in $\ell^\infty([0,1]^d)$ that in addition satisfy the conditions imposed in the second sentence of Theorem 1. The stochastic process $\mathbb{G}_n$ and its limit $\mathbb{G}$ then take values in $\mathcal{X}$ almost surely.

For any $f\in\mathcal{F}$, we define

$$\bar{\mathbb{G}}_n(f) = \int f\,\mathrm{d}\mathbb{G}_n,$$
$$\widetilde{\mathbb{G}}_n(f) = \Gamma(\mathbb{G}_n, f)$$

based on the functional $\Gamma(\cdot,\cdot):\mathcal{X}\times\mathcal{F}\to\mathbb{R}$ as

$$\Gamma(X,f) := \sum_{I\subset\{1,\dots,d\}:I\neq\varnothing}(-1)^{|I|}\int_{(\mathbf{0}_I,\mathbf{1}_I]}X(\mathbf{x}_I;\mathbf{1}_{-I})\,\mathrm{d}f(\mathbf{x}_I;\mathbf{1}_{-I}). \tag{15}$$

The only difference between the right-hand sides of (15) and (11) after inserting $\mathbb{G}_n$ for X and g, respectively appears in the argument to the function $\mathbb{G}_n(\cdot;\mathbf{1}_{-I})$ on the left of the semicolon.

First, for each $f\in\mathcal{F}$, the functional $\Gamma(\cdot,f):\mathcal{X}\to\mathbb{R}$ is linear and Lipschitz as

$$\begin{aligned}\big|\Gamma(X,f)-\Gamma(Y,f)\big| &\le \sum_{I\subset\{1,\dots,d\}:I\neq\varnothing}\int_{(\mathbf{0}_I,\mathbf{1}_I]}|\mathrm{d}f|(\mathbf{x}_I;\mathbf{1}_{-I})\cdot\|X-Y\|_\infty \\ &\le \big(2^{2d}T\big)\|X-Y\|_\infty.\end{aligned} \tag{16}$$

In the above, we have invoked the property of T as in (14).

Hence, for any fixed $f \in \mathcal{F}$, by the continuous mapping theorem (see, for instance, Theorem 1.3.6 in [31]), and the weak convergence $\mathbb{G}_n \rightsquigarrow \mathbb{G}$, we have

$$\widetilde{\mathbb{G}}_n(f) = \Gamma(\mathbb{G}_n, f) \rightsquigarrow \Gamma(\mathbb{G}, f) := \widetilde{\mathbb{G}}(f)$$

as $n \to \infty$. This result is pointwise in f, that is, it provides *fidi*-convergence of $\widetilde{\mathbb{G}}_n$. Linearity of $\Gamma(\cdot, f)$ yields that the limit $\widetilde{\mathbb{G}}(f)$ is normal.

Next, we define the map $\widetilde{\Gamma} : \mathcal{X} \to \ell^\infty(\mathcal{F})$ as $(\widetilde{\Gamma}(X))(f) = \Gamma(X, f)$ for $X \in \mathcal{X}$ and $f \in \mathcal{F}$. Then the map $\widetilde{\Gamma}$ is Lipschitz with Lipschitz constant $2^{2d}T$ because

$$\big\|\widetilde{\Gamma}(X) - \widetilde{\Gamma}(Y)\big\| = \sup_{f \in \mathcal{F}} \big|\Gamma(X, f) - \Gamma(Y, f)\big| \le \big(2^{2d}T\big)\|X - Y\|_\infty. \tag{17}$$

Here the inequality follows by (16), which in fact holds uniformly over $f \in \mathcal{F}$. The continuous mapping theorem then guarantees that the limit $\widetilde{\mathbb{G}} := \widetilde{\Gamma}(\mathbb{G})$ of $\widetilde{\mathbb{G}}_n = \widetilde{\Gamma}(\mathbb{G}_n)$ is tight in $\ell^\infty(\mathcal{F})$.

Finally, we let the bounded Lipschitz distance between $\bar{\mathbb{G}}_n$ and $\widetilde{\mathbb{G}}$ be

$$d_{\mathrm{BL}}(\bar{\mathbb{G}}_n, \widetilde{\mathbb{G}}) = \sup_h \big|\mathbb{E}\big[h(\bar{\mathbb{G}}_n)\big] - \mathbb{E}\big[h(\widetilde{\mathbb{G}})\big]\big|$$

with the supremum taken over all uniformly bounded, Lipschitz functionals $h : \ell^\infty(\mathcal{F}) \to \mathbb{R}$ with $\sup_{x \in \ell^\infty(\mathcal{F})} |h(x)| \le 1$ and $|h(x) - h(y)| \le \|x - y\|$ for all $x, y \in \ell^\infty(\mathcal{F})$. Using the triangle inequality, we have

$$d_{\mathrm{BL}}(\bar{\mathbb{G}}_n, \widetilde{\mathbb{G}}) \le d_{\mathrm{BL}}(\bar{\mathbb{G}}_n, \widetilde{\mathbb{G}}_n) + d_{\mathrm{BL}}(\widetilde{\mathbb{G}}_n, \widetilde{\mathbb{G}}). \tag{18}$$

For the first term on the right-hand side of (18), we first apply Proposition 3 (with g replaced by $\mathbb{G}_n$, whose sample paths satisfy Assumption F almost surely by assumption) to the term $\bar{\mathbb{G}}_n(f)$ for an arbitrary $f \in \mathcal{F}$, and then invoke the property of T as in (14) to obtain that

$$\begin{aligned}
\|\bar{\mathbb{G}}_n - \widetilde{\mathbb{G}}_n\| &= \sup_{f \in \mathcal{F}} \big|\bar{\mathbb{G}}_n(f) - \widetilde{\mathbb{G}}_n(f)\big| \\
&\le \Big[\sup_{\mathbf{x}} \big|\mathbb{G}_n(\mathbf{x}) - \mathbb{G}_n(\mathbf{x}-)\big|\Big] \cdot \sup_{f \in \mathcal{F}} \Big[\sum_{I \subset \{1,\dots,d\}: I \neq \varnothing} \int_{(\mathbf{0}_I, \mathbf{1}_I]} |\mathrm{d}f|(\mathbf{x}_I; \mathbf{1}_{-I})\Big] \\
&\le \big(2^{2d}T\big) \sup_{\mathbf{x}} \big|\mathbb{G}_n(\mathbf{x}) - \mathbb{G}_n(\mathbf{x}-)\big|;
\end{aligned}$$

then, by the Lipschitz property and boundedness of the functionals h, we conclude that

$$d_{\mathrm{BL}}(\bar{\mathbb{G}}_n, \widetilde{\mathbb{G}}_n) \le \mathbb{E}\Big[\min\Big\{2, \big(2^{2d}T\big) \cdot \sup_{\mathbf{x}} \big|\mathbb{G}_n(\mathbf{x}) - \mathbb{G}_n(\mathbf{x}-)\big|\Big\}\Big]. \tag{19}$$

For the second term on the right hand side of (18), as a consequence of the Lipschitz property of the map $\widetilde{\Gamma}$ with Lipschitz constant $2^{2d}T$ as shown previously in (17), we have

$$\begin{aligned}
d_{\mathrm{BL}}(\widetilde{\mathbb{G}}_n, \widetilde{\mathbb{G}}) &= \sup_h \big|\mathbb{E}\big[h(\widetilde{\mathbb{G}}_n)\big] - \mathbb{E}\big[h(\widetilde{\mathbb{G}})\big]\big| = \sup_h \big|\mathbb{E}\big[h \circ \widetilde{\Gamma}(\mathbb{G}_n)\big] - \mathbb{E}\big[h \circ \widetilde{\Gamma}(\mathbb{G})\big]\big| \\
&\le \big(2^{2d}T\big) d_{\mathrm{BL}}(\mathbb{G}_n, \mathbb{G}).
\end{aligned} \tag{20}$$

In (20) and in (21) below, $d_{\mathrm{BL}}(\mathbb{G}_n, \mathbb{G})$ is the bounded Lipschitz distance between $\mathbb{G}_n$ and $\mathbb{G}$ defined as

$$d_{\mathrm{BL}}(\mathbb{G}_n, \mathbb{G}) = \sup_{\tilde{h}} \bigl| \mathbb{E}\bigl[\tilde{h}(\mathbb{G}_n)\bigr] - \mathbb{E}\bigl[\tilde{h}(\mathbb{G})\bigr] \bigr|$$

with the supremum taken over all uniformly bounded, Lipschitz functionals $\tilde{h} : \ell^\infty([0,1]^d) \to \mathbb{R}$ with $\sup_{x \in \ell^\infty([0,1]^d)} |\tilde{h}(x)| \le 1$ and $|\tilde{h}(x) - \tilde{h}(y)| \le \|x - y\|$ for all $x, y \in \ell^\infty([0,1]^d)$. Combining (18), (19) and (20), we obtain

$$d_{\mathrm{BL}}(\bar{\mathbb{G}}_n, \widetilde{\mathbb{G}}) \le \mathbb{E}\Bigl[\min\Bigl\{2, \bigl(2^{2d} T\bigr) \cdot \sup_{\mathbf{x}} \bigl|\mathbb{G}_n(\mathbf{x}) - \mathbb{G}_n(\mathbf{x}-)\bigr|\Bigr\}\Bigr] + \bigl(2^{2d} T\bigr) d_{\mathrm{BL}}(\mathbb{G}_n, \mathbb{G}). \qquad (21)$$

We then conclude that $d_{\mathrm{BL}}(\bar{\mathbb{G}}_n, \widetilde{\mathbb{G}}) \to 0$ as $n \to \infty$. Since the limit $\widetilde{\mathbb{G}} = \widetilde{\Gamma}(\mathbb{G})$ (as defined earlier) is tight, the desired weak convergence of $\bar{\mathbb{G}}_n(f) = \int f \, \mathrm{d}\mathbb{G}_n$, indexed by $f \in \mathcal{F}$, in $\ell^\infty(\mathcal{F})$ follows. □

3. Empirical copula processes indexed by functions

3.1. Smooth copula functions

Our first result requires the following smoothness condition on the copula function C so that the ordinary empirical copula process $\sqrt{n}(\mathbb{C}_n - C)$ given in (1) based on an i.i.d. sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$ converges weakly in $\ell^\infty([0,1]^d)$ to a Gaussian limit:

Assumption P. *For each $k \in \{1, \ldots, d\}$, the kth first-order partial derivative $\dot{C}_k$ of the copula function C exists and is continuous on the set $V_{d,k} = \{\mathbf{u} \in [0,1]^d : u_k \in (0,1)\}$.*

We then consider the class $\mathcal{G}$ of functions g on $[0,1]^d$ satisfying Assumption F and additionally $\sup_{g \in \mathcal{G}} V_{\mathrm{HK}}(g) < \infty$. In words, we require that the Vitali variations of the functions g and their lower-dimensional projections $g(\cdot; \mathbf{1}_{-I})$ on $[\mathbf{0}_I, \mathbf{1}_I]$ for $I \subset \{1, \ldots, d\} \setminus \varnothing$ are uniformly bounded.

Theorem 5. *Assume that the copula function C satisfies Assumption P. Let $\mathcal{G}$ be a class of functions g on $[0,1]^d$ satisfying Assumption F, and additionally $\sup_{g \in \mathcal{G}} V_{\mathrm{HK}}(g) < \infty$. Then the empirical process $\bar{\mathbb{Z}}_n(g)$ defined in (4), indexed by $g \in \mathcal{G}$, converges weakly in $\ell^\infty(\mathcal{G})$ to a Gaussian limit.*

Remark. From the first remark following Theorem 1, we immediately conclude that the limiting process $\bar{\mathbb{Z}}$ of $\bar{\mathbb{Z}}_n$ can be characterized as

$$\bar{\mathbb{Z}}(g) = \sum_{I \subset \{1,\ldots,d\} : I \neq \varnothing} (-1)^{|I|} \int_{(\mathbf{0}_I, \mathbf{1}_I]} \Bigl\{ \alpha(\mathbf{u}_I; \mathbf{1}_{-I}) - \sum_{k \in I} \dot{C}_k(\mathbf{u}_I; \mathbf{1}_{-I}) \alpha_k(u_k) \Bigr\} \mathrm{d}g(\mathbf{u}_I; \mathbf{1}_{-I})$$

for $g \in \mathcal{G}$. Here α is the limiting C-Brownian bridge in $\ell^\infty([0,1]^d)$ of the standard empirical process $\mathbb{U}_n = \sqrt{n}(\mathbb{H}_n - C)$ for $\mathbb{H}_n$ the empirical distribution function based on pseudo-observations $\mathbf{U}_1, \dots, \mathbf{U}_n$ with $\mathbf{U}_i = (F_1(X_{i1}), \dots, F_d(X_{id}))$, and α_k is the kth marginal of α. Note that $\alpha - \sum_{k=1}^d \dot{C}_k \alpha_k$ (with $(\dot{C}_k \alpha_k)(\mathbf{u})$ interpreted as $\dot{C}_k(\mathbf{u})\alpha_k(u_k)$) is the limiting process in $\ell^\infty([0,1]^d)$ of $\sqrt{n}(\bar{\mathbb{C}}_n - C)$.

Proof of Theorem 5. Under Assumption P, the ordinary empirical copula process $\sqrt{n}(\mathbb{C}_n - C)$ converges weakly in $\ell^\infty([0,1]^d)$ to a continuous Gaussian limit; see for instance Proposition 3.1 in [27] or Corollary 2.5 in [7]. By (3), which holds almost surely for an i.i.d. sample, the empirical copula process $\sqrt{n}(\bar{\mathbb{C}}_n - C)$ converges weakly in $\ell^\infty([0,1]^d)$ to the same limit. The conclusion of the theorem then follows immediately from Theorem 1 by replacing the process $\mathbb{G}_n$ with $\sqrt{n}(\bar{\mathbb{C}}_n - C)$. □

The class of functions $\mathcal{G}$ considered in Theorem 5 is an obvious generalization of the class of indicator functions $\mathbb{1}\{\cdot \le \mathbf{x}\}$ of the half-open rectangles $(\mathbf{0}, \mathbf{x}]$, $\mathbf{x} \in [0,1]^d$. Theorem 5 requires no differentiability of $g \in \mathcal{G}$, only right-continuity and uniformly bounded Hardy–Krause variation.

Now we discuss some generalization of Theorem 5. The proof of the theorem requires that the empirical copula process $\sqrt{n}(\bar{\mathbb{C}}_n - C)$ converges weakly in $\ell^\infty([0,1]^d)$ so that we can apply Theorem 1. For an i.i.d. sample, this is shown in increasing generality by, among others, [13] and [27]. In fact, our method straightforwardly generalizes to certain non-i.i.d. sequence $\mathbf{X}_i$, $i \in \mathbb{Z}$, because Theorem 1 requires weak convergence only of the process $\sqrt{n}(\bar{\mathbb{C}}_n - C)$ in $\ell^\infty([0,1]^d)$. [7], in turn, showed that the latter (or more precisely, the ordinary empirical copula process version (1) of it) is implied by the weak convergence of the standard empirical process $\sqrt{n}(\mathbb{H}_n - C)$ for $\mathbb{H}_n$ the empirical distribution function based on pseudo-observations $\mathbf{U}_1, \dots, \mathbf{U}_n$ with $\mathbf{U}_i = (F_1(X_{i1}), \dots, F_d(X_{id}))$, together with the smoothness condition stated in Assumption P.

Corollary 6. *Assume that the copula function C satisfies Assumption* P. *Moreover, assume that $\sqrt{n}(\mathbb{H}_n - C)$ converges weakly in $\ell^\infty([0,1]^d)$ to a Gaussian limit B_C. Let $\mathcal{G}$ be a class of functions g on $[0,1]^d$ satisfying Assumption* F, *and additionally $\sup_{g \in \mathcal{G}} V_{\mathrm{HK}}(g) < \infty$. Then the empirical process $\bar{\mathbb{Z}}_n(g)$ defined in* (4), *indexed by $g \in \mathcal{G}$, converges weakly in $\ell^\infty(\mathcal{G})$ to a Gaussian limit.*

Proof. The weak convergence of $\sqrt{n}(\mathbb{H}_n - C)$ in $\ell^\infty([0,1]^d)$ to a Gaussian limit B_C implies that the limit process B_C is continuous, $B_C(\mathbf{u}) = 0$ if $u_j = 0$ for some j, and $B_C(\mathbf{1}) = 0$. Corollary 2.5 in [7] shows that the weak convergence of $\sqrt{n}(\mathbb{H}_n - C)$ in $\ell^\infty([0,1]^d)$, the aforementioned properties of its Gaussian limit B_C, and Assumption P together imply that the ordinary empirical copula process $\sqrt{n}(\mathbb{C}_n - C)$ converges weakly in $\ell^\infty([0,1]^d)$ to a continuous Gaussian limit. Moreover, Section C.2 shows that (51) holds under the conditions imposed in the second sentence of the corollary. Hence, the empirical copula process $\sqrt{n}(\bar{\mathbb{C}}_n - C)$ converges weakly in $\ell^\infty([0,1]^d)$ to the same limit as the ordinary empirical copula process. The conclusion of the corollary then follows immediately from Theorem 1, again by replacing the process $\mathbb{G}_n$ with $\sqrt{n}(\bar{\mathbb{C}}_n - C)$. □

Remark. In particular, if the empirical distribution function $\mathbb{H}_n$ in Corollary 6 is based on a stationary sequence $\mathbf{U}_i$, $i \in \mathbb{Z}$ satisfying the (alpha-mixing) conditions in Corollary 2, then the conditions imposed in the second sentence of the corollary hold.

3.2. Smooth index functions

We again assume that the sample $\mathbf{X}_1, \dots, \mathbf{X}_n$ is i.i.d. Our next result requires that $\mathcal{G}$ is a C-Donsker class of differentiable functions $g : [0, 1]^d \to \mathbb{R}$. For any $g \in \mathcal{G}$, we write $\dot{g}_k = \partial_k g$ for the partial derivative of g with respect to the kth coordinate, that is, $\dot{g}_k(\mathbf{u}) = \partial_k g(\mathbf{u}) = \partial g(\mathbf{u})/\partial u_k$, $\mathbf{u} = (u_1, \dots, u_d)$. We assume that the classes $\dot{\mathcal{G}}_k$, $k \in \{1, \dots, d\}$ of partial derivatives

$$\dot{\mathcal{G}}_k = \{\dot{g}_k = \partial_k g, g \in \mathcal{G}\}$$

are uniformly equicontinuous. Interestingly, in this case, the existence of first-order partial derivatives of C is no longer required for the weak convergence of $\bar{\mathbb{Z}}_n$.

Theorem 7. *Assume that*:

- $\mathcal{G}$ *is a* C*-Donsker class of uniformly bounded, continuous functions on* $[0, 1]^d$.
- *For each* $k \in \{1, \dots, d\}$, *the* k*th first-order partial derivative* $\dot{g}_k$ *of* $g \in \mathcal{G}$ *exists on the set* $(0, 1)^d$, *and the class* $\dot{\mathcal{G}}_k$ *is uniformly bounded and uniformly equicontinuous on* $(0, 1)^d$.

Then the empirical process $\bar{\mathbb{Z}}_n(g)$ *defined in* (4), *indexed by* $g \in \mathcal{G}$, *converges weakly in* $\ell^\infty(\mathcal{G})$ *to a Gaussian limit*.

Remark. The proof of Theorem 7 reveals (through the process $\widetilde{\mathbb{Z}}_n(g)$, $g \in \mathcal{G}$ introduced in the proof) that the limiting process $\bar{\mathbb{Z}}$ of $\bar{\mathbb{Z}}_n$ can be characterized as

$$\bar{\mathbb{Z}}(g) = \int g(\mathbf{u})\,\mathrm{d}\alpha(\mathbf{u}) + \sum_{k=1}^{d} \int_{(0,1)^d} \dot{g}_k(\mathbf{u})\alpha_k(u_k)\,\mathrm{d}C(\mathbf{u}) \qquad (22)$$

for $g \in \mathcal{G}$. Here the notations are the same as those in the remark following Theorem 5, and the integral $\int g(\mathbf{u})\,\mathrm{d}\alpha(\mathbf{u})$, $g \in \mathcal{G}$ is understood as the weak limit of $\int g(\mathbf{u})\,\mathrm{d}\mathbb{U}_n(\mathbf{u})$ in $\ell^\infty(\mathcal{G})$.

Proof of Theorem 7. See Section 4.2. □

Discussion of the conditions of Theorem 7

(a) Theorem 7 is slightly more general than Corollary 5.4 in [32]. It corrects a slight mistake in their proof. While they require that the partial derivatives $\dot{g}_k$ are continuous, their proof requires that they are in fact uniformly equicontinuous. In particular, at page 247, line 13 they require convergence, uniformly in g, while their proof of this fact (Lemma 4.1 on the same page) only gives pointwise convergence. While this is easily fixed, the other difference with their result, however, is that we do not assume that the uniform entropy integral $J(1, \mathcal{G}, L_2)$ is finite, which requires an altogether different proof.

(b) It is remarkable that Theorem 7 holds without any condition on C, under rather mild regularity on the functions $g \in \mathcal{G}$. This is in contrast with the smoothness assumption on the copula function C (that is, the condition in [27] as stated in Assumption P) required for the weak convergence of the (standard) empirical copula process (indexed by boxes) in (2).

Arguably the best known examples of non-differentiable copulas are the Marshall–Olkin copula $C(u, v) = \min(u^{1-\alpha} v, uv^{1-\beta})$, and the Fréchet–Hoeffding copulas $C(u, v) = \max(u + v - 1, 0)$ and $C(u, v) = \min(u, v)$. Another example is the Cuadras–Augé copula given by

$$C(u, v) = \bigl\{\min(u, v)\bigr\}^{\theta} \{uv\}^{1-\theta}, \qquad 0 \le \theta \le 1.$$

A common technique to construct a copula from a given function $\delta : [0, 1] \to [0, 1]$ yields non-differentiable copulas as well by setting

$$C(u, v) = \min\bigl[u, v, \bigl\{\delta(u) + \delta(v)\bigr\}/2\bigr]$$

or

$$C(u, v) = \begin{cases} u - \inf\limits_{u \le x \le v} \bigl\{x - \delta(x)\bigr\}, & \text{if } u \le v, \\ v - \inf\limits_{v \le x \le u} \bigl\{x - \delta(x)\bigr\}, & \text{if } u > v. \end{cases}$$

(c) A natural class $\mathcal{G}$ of functions to consider is $C_1^s([0, 1]^d)$, as described in detail by [31], pages 154–157. These are all functions on $[0, 1]^d$ that have uniformly bounded partial derivatives up to order $\lfloor s \rfloor$ and whose highest partial derivatives are Hölder of order $s - \lfloor s \rfloor$. Theorem 2.7.1 and Theorem 2.7.2 in [31] show that the class $C_1^s([0, 1]^d)$ is universally Donsker if $s > d/2$. In particular, this means that for $d = 2$, the process $\bar{\mathbb{Z}}_n$ converge weakly in $\ell^\infty(C_1^s([0, 1]^2))$, provided the smoothness index $s > 1$, in which case $C_1^s([0, 1]^d)$ consists of all functions having uniformly bounded first-order partial derivatives that satisfy a uniform Hölder condition of some arbitrary order.

3.3. Bootstrap empirical copula processes

We provide the bootstrap counterpart of Theorems 5 and 7. Let the bootstrap sample $\mathbf{X}_1^*, \dots, \mathbf{X}_n^*$ with $\mathbf{X}_i^* = (X_{i1}^*, \dots, X_{id}^*)$ be obtained by sampling with replacement from the original i.i.d. sample $\mathbf{X}_1, \dots, \mathbf{X}_n$. We write

$$\mathbb{F}_n^*(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}\bigl\{\mathbf{X}_i^* \le \mathbf{x}\bigr\}, \qquad \mathbf{x} \in \mathbb{R}^d,$$

for the empirical distribution function based on the bootstrap sample, with marginals

$$\mathbb{F}^*_{nj}(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{X^*_{ij} \le t\}, \qquad t \in \mathbb{R}, j \in \{1, \dots, d\}.$$

We let the associated bootstrap ordinary empirical copula function be

$$\mathbb{C}^*_n(\mathbf{u}) = \mathbb{F}^*_n\big(\mathbb{F}^{*-}_{n1}(u_1), \dots, \mathbb{F}^{*-}_{nd}(u_d)\big), \qquad \mathbf{u} \in [0,1]^d, \tag{23}$$

and the càdlàg version of (23), that is, the bootstrap empirical copula function, be

$$\bar{\mathbb{C}}^*_n(\mathbf{u}) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{\mathbb{F}^*_{n1}(X^*_{i1}) \le u_1, \dots, \mathbb{F}^*_{nd}(X^*_{id}) \le u_d\}, \qquad \mathbf{u} \in [0,1]^d. \tag{24}$$

We let $\sqrt{n}(\bar{\mathbb{C}}^*_n - \bar{\mathbb{C}}_n)$ be the bootstrap empirical copula process, and

$$\bar{\mathbb{Z}}^*_n(g) = \sqrt{n}\int g \,\mathrm{d}\big(\bar{\mathbb{C}}^*_n - \bar{\mathbb{C}}_n\big), \qquad g \in \mathcal{G} \tag{25}$$

be the bootstrap empirical copula process indexed by functions. We have the following bootstrap version of Theorems 5 and 7.

Theorem 8. *Let the bootstrap sample $\mathbf{X}^*_1, \dots, \mathbf{X}^*_n$ be obtained by sampling with replacement from the original i.i.d. sample $\mathbf{X}_1, \dots, \mathbf{X}_n$. Assume that the following conditions, which are identical to those stated under Theorem 5, hold:*

- *The copula function C satisfies Assumption P.*
- *$\mathcal{G}$ is a class of functions g on $[0,1]^d$ satisfying Assumption F, and additionally $\sup_{g\in\mathcal{G}} V_{\mathrm{HK}}(g) < \infty$.*

Alternatively, assume that the following conditions, which are variants of those stated under Theorem 7, hold:

- *$\mathcal{G}$ is a C-Donsker class of uniformly bounded, continuous functions on $[0,1]^d$.*
- *For each $k \in \{1, \dots, d\}$, the kth first-order partial derivative $\dot{g}_k$ of $g \in \mathcal{G}$ exists on the set $V_{d,k}$ (we recall the definition of $V_{d,k}$ from Assumption P), and the class $\dot{\mathcal{G}}_k$ is uniformly bounded and uniformly equicontinuous on $V_{d,k}$.*

Then, the conditional distribution of $\bar{\mathbb{Z}}^*_n(g)$ defined in (25), indexed by $g \in \mathcal{G}$, converges weakly in $\ell^\infty(\mathcal{G})$ to the same Gaussian limit as $\bar{\mathbb{Z}}_n(g)$ indexed by $g \in \mathcal{G}$, in probability.

Remark 1. More precisely, we prove that

$$\lim_{n\to\infty} \mathbb{E}\Big[\sup_h \big|\mathbb{E}\big[h(\bar{\mathbb{Z}}_n)\big] - \mathbb{E}^*\big[h\big(\bar{\mathbb{Z}}^*_n\big)\big]\big|\Big] = 0. \tag{26}$$

Here $\mathbb{E}^*$ is the conditional expectation with respect to the bootstrap sample, and the supremum in (26) is taken over all uniformly bounded, Lipschitz functionals $h : \ell^\infty(\mathcal{G}) \to \mathbb{R}$ with

$\sup_{x\in\ell^\infty(\mathcal{G})} |h(x)| \le 1$ and $|h(x) - h(y)| \le \|x - y\|$ for all $x, y \in \ell^\infty(\mathcal{G})$. As usual in the empirical process literature, it is tacitly understood that we take outer probability measures whenever measurability issues arise.

Remark 2. Comparing the required conditions of Theorem 7 and its bootstrap counterpart, the only differences are that for the bootstrap version the kth first-order partial derivative $\dot{g}_k$ of $g \in \mathcal{G}$ should exist on $V_{d,k}$ instead of $(0, 1)^d$, and that the class $\dot{\mathcal{G}}_k$ should be uniformly bounded and uniformly equicontinuous on $V_{d,k}$ instead of $(0, 1)^d$.

Proof of Theorem 8. See Section 4.3. □

3.4. Some applications

Semi-parametric MLE. This type of results is useful in the same way the extension of the empirical process indexed by general Donsker classes from the class of indicator functions $\mathbb{1}\{\cdot \le \mathbf{x}\}$ of the half-spaces $(-\infty, \mathbf{x}]$, $\mathbf{x} \in \mathbb{R}^d$, has proved extremely useful. See, for instance, the monograph [31]. In the context of copula estimation, an important example is the following semi-parametric maximum likelihood estimation problem [29]. Suppose that the copula function C is parametrized by a finite dimensional parameter $\theta \in \Theta$ for Θ a subset of $\mathbb{R}^k$, that C has density c_θ and that the marginal distributions F_j have densities f_j. The log-likelihood function in this setting is

$$\log \ell(\theta) = \sum_{i=1}^n \log c_\theta\bigl(F_1(X_{i1}), \dots, F_d(X_{id})\bigr) + \sum_{i=1}^n \sum_{j=1}^d \log f_j(X_{ij})$$

and therefore a common strategy is to replace the unknown marginals F_j by $\mathbb{F}_{nj}$ and maximize

$$\sum_{i=1}^n \log c_\theta\bigl(\mathbb{F}_{n1}(X_{i1}), \dots, \mathbb{F}_{nd}(X_{id})\bigr)$$

over θ. We assume that we can take the derivative of $\log c_\theta$ with respect to θ, and we let the resulting derivative be the score function ϕ_θ. We then define

$$\Psi(\theta) = \int \phi_\theta(\mathbf{u})\, \mathrm{d}C(\mathbf{u})$$

and

$$\Psi_n(\theta) = \int \phi_\theta(\mathbf{u})\, \mathrm{d}\bar{\mathbb{C}}_n(\mathbf{u}).$$

Example 3.9.35 in [31] shows that the solution $\widehat{\theta}_n$ of $\Psi_n(\theta) = 0$ is asymptotically normal, provided the process $\sqrt{n}(\Psi_n - \Psi)(\theta)$ converges in distribution to a Gaussian $\mathbb{Z}$ with continuous sample paths in $\ell^\infty(\Theta)$ and Ψ is sufficiently regular: $\Psi(\theta) = 0$ has a unique solution θ_0, Ψ is a local homeomorphism at θ_0 and is differentiable at θ_0 with derivative $\dot{\Psi}_{\theta_0}$. Consequently, under the conditions of either Theorem 5 or Theorem 7 (with no assumptions on C in the latter case) with $\mathcal{G}$ replaced by the class of functions ϕ_θ indexed by $\theta \in \Theta$, if

$$\lim_{\|\theta'-\theta\|\to 0} \int (\phi_{\theta'} - \phi_\theta)^2 \, \mathrm{d}C = 0$$

and Ψ satisfies the regularity conditions above, then $\widehat{\theta}_n$ is asymptotically normal.

We do would like to point out that, in practice it is common for the score function ϕ_θ to violate the required conditions of both Theorems 5 and 7. For instance, Example 3.2 in [4] discusses the bivariate Gaussian copula model with θ being the correlation coefficient; the score function in this case is unbounded. We refer the readers to the same reference for a solution in some bivariate cases.

Testing of non-smooth copulas. The Kolmogorov–Smirnov test statistic

$$\sqrt{n} \sup_{\mathbf{u}} \bigl|\mathbb{C}_n(\mathbf{u}) - C(\mathbf{u})\bigr|$$

converges provided C meets the mild smoothness Condition 4.3 of [6]. If we wish to test for a non-smooth C, then Theorem 7 poses a solution by considering

$$\sqrt{n} \sup_{g\in\mathcal{G}} \left| \int g \, \mathrm{d}(\bar{\mathbb{C}}_n - C) \right|$$

for a sufficiently rich class $\mathcal{G}$ instead. For instance, we could consider the class $\mathcal{G} = C_1^s([0,1]^d)$, with $s > d/2$, described in bullet (c) in the discussion following Theorem 7. From a computational point of view, we may consider the class $g(\mathbf{x}) = g_{\mathbf{t}}(\mathbf{x}) = \exp(\langle \mathbf{t}, \mathbf{x}\rangle)$, with $\mathbf{t} \in [-1,1]^d$ so that we compare the moment generating functions (which are defined for any copula, as the random vector with distribution function C is bounded). Indeed, if the function $\int e^{\langle \mathbf{t},\mathbf{u}\rangle} \, \mathrm{d}C(\mathbf{u})$ is piecewise differentiable in $\mathbf{t}$, then this would lead to an easily computable test statistic and a consistent test.

4. Proofs of Theorems 7 and 8

4.1. Notation

Throughout Section 4, we assume without loss of generality that the original observations $\mathbf{X}_i = (X_{i1}, \dots, X_{id})$ are replaced by the pseudo-observations $\mathbf{U}_i = (U_{i1}, \dots, U_{id}) = (F_1(X_{i1}), \dots, F_d(X_{id}))$, $i = 1, \dots, n$. With this simplification, the copula function C remains invariant, but now all marginals F_j, $j \in \{1, \dots, d\}$ are uniform distributions on $[0, 1]$, and additionally we have $F = C$. For the ordinary empirical copula process, this common simplification in the copula literature is justified by, for instance, Lemma 8 in [14]. For the purpose of this paper, it remains true that $\bar{\mathbb{Z}}_n(g)$, $g \in \mathcal{G}$ remains invariant under this simplification because $\bar{\mathbb{C}}_n$, the càdlàg version of the ordinary empirical copula function, remains invariant. Having made this blanket assumption (so in particular $F_j(t) = t$, $j \in \{1, \dots, d\}$), we denote by $\mathbb{U}_n$ the standard empirical process $\sqrt{n}(\mathbb{F}_n - F)$ in $\ell^\infty([0,1]^d)$ with marginals $\mathbb{U}_{nj} = \sqrt{n}(\mathbb{F}_{nj} - F_j)$, $j \in \{1, \dots, d\}$.

4.2. Proof of Theorem 7

Recall that $\dot{g}_k$ is the partial derivative of $g \in \mathcal{G}$ with respect to the kth coordinate. We define the empirical process

$$\widetilde{\mathbb{Z}}_n(g) = \int \left[g + \sum_{k=1}^{d} T_k(g) \right] \mathrm{d}\mathbb{U}_n$$

for $g \in \mathcal{G}$, based on the functions $T_k(g) : [0,1]^d \to \mathbb{R}$ for $k \in \{1, \dots, d\}$ and $g \in \mathcal{G}$ defined as

$$T_k(g)(\mathbf{x}) = \int_{(0,1)^d} \dot{g}_k(\mathbf{u}) \mathbb{1}\{x_k \le u_k\} \, \mathrm{d}C(\mathbf{u}). \tag{27}$$

Lemma 9 shows that $\widetilde{\mathbb{Z}}_n$ converges weakly in $\ell^\infty(\mathcal{G})$ to a Gaussian limit, and it suffices to show that $\bar{\mathbb{Z}}_n$ and $\widetilde{\mathbb{Z}}_n$ are asymptotically equivalent, as $n \to \infty$. Using Fubini's theorem, we have that

$$\begin{aligned}
\int T_k(g) \, \mathrm{d}\mathbb{F}_n &= \frac{1}{n} \sum_{i=1}^{n} \int_{(0,1)^d} \dot{g}_k(\mathbf{x}) \mathbb{1}\{X_{ik} \le x_k\} \, \mathrm{d}C(\mathbf{x}) \\
&= \int_{(0,1)^d} \dot{g}_k(\mathbf{x}) \mathbb{F}_{nk}(x_k) \, \mathrm{d}C(\mathbf{x})
\end{aligned}$$

and similarly

$$\int T_k(g) \mathrm{d}C = \int_{(0,1)^d} \dot{g}_k(\mathbf{x}) F_k(x_k) \, \mathrm{d}C(\mathbf{x}),$$

so that

$$\int T_k(g) \, \mathrm{d}\mathbb{U}_n = \int_{(0,1)^d} \dot{g}_k(\mathbf{x}) \mathbb{U}_{nk}(x_k) \, \mathrm{d}C(\mathbf{x}). \tag{28}$$

It is now easily verified that

$$(\bar{\mathbb{Z}}_n - \widetilde{\mathbb{Z}}_n)(g) = I(g) + II(g)$$

for

$$\begin{aligned}
I(g) &= \int \sqrt{n} \big[g\big(\mathbb{F}_{n1}(x_1), \dots, \mathbb{F}_{nd}(x_d)\big) - g(\mathbf{x}) \big] \, \mathrm{d}\mathbb{F}_n(\mathbf{x}) - \sum_{k=1}^{d} \int_{(0,1)^d} \dot{g}_k(\mathbf{x}) \mathbb{U}_{nk}(x_k) \, \mathrm{d}\mathbb{F}_n(\mathbf{x}), \\
II(g) &= \sum_{k=1}^{d} \int_{(0,1)^d} \dot{g}_k(\mathbf{x}) \mathbb{U}_{nk}(x_k) \, \mathrm{d}n^{-1/2} \mathbb{U}_n(\mathbf{x}).
\end{aligned}$$

Hence, if

$$\sup_{g\in\mathcal{G}}\bigl|I(g)+\;II(g)\bigr|=o_p(1),$$

then $\bar{\mathbb{Z}}_n$ converges weakly in $\ell^\infty(\mathcal{G})$ to the same limit as $\widetilde{\mathbb{Z}}_n$. This is verified in Propositions 10 and 11, and the proof of Theorem 7 is complete. □

Lemma 9. *Under the assumptions of Theorem 7, the empirical process $\widetilde{\mathbb{Z}}_n(g)$, indexed by $g\in\mathcal{G}$, converges weakly in $\ell^\infty(\mathcal{G})$ to a Gaussian limit.*

Proof. It suffices to prove that the class of functions $\{g+\sum_{k=1}^d T_k(g): g\in\mathcal{G}\}$ is C-Donsker. In turn it suffices to show that each of $\int g\,\mathrm{d}\mathbb{U}_n$ and $\int T_k(g)\,\mathrm{d}\mathbb{U}_n$, $k\in\{1,\dots,d\}$ converges weakly in $\ell^\infty(\mathcal{G})$. That $\int g\,\mathrm{d}\mathbb{U}_n$ converges weakly in $\ell^\infty(\mathcal{G})$ is obvious by the assumption that $\mathcal{G}$ is C-Donsker. That for each $k\in\{1,\dots,d\}$, $\int T_k(g)\,\mathrm{d}\mathbb{U}_n$ converges weakly in $\ell^\infty(\mathcal{G})$ can be seen through (28), the assumption that the class $\dot{\mathcal{G}}_k$ is uniformly bounded, and the continuous mapping theorem. □

Proposition 10. *Under the assumptions of Theorem 7, we have*

$$\sup_{g\in\mathcal{G}}\bigl|I(g)\bigr|=o_p(1). \tag{29}$$

Proof. We first consider the first term in $I(g)$. We let A_n be the event on which there exists $\mathbf{x}<\mathbf{1}$ (we recall our convention that the inequalities here hold for all d components) such that $\mathbb{F}_n(\mathbf{x})=1$. Because the copula function C is continuous, the event A_n has probability one, and we focus on this event. Then, we can freely change the domain of integration in the first term in $I(g)$ from $(0,1]^d$ to $(0,1)^d$. We let $L_{n,\mathbf{x}}$ be (random) sets defined as $L_{n,\mathbf{x}}=\{k\in\{1,\dots,d\}: x_k\neq\mathbb{F}_{nk}(x_k)\}$. Then we have, by the mean value theorem,

$$\int\sqrt{n}\bigl[g\bigl(\mathbb{F}_{n1}(x_1),\dots,\mathbb{F}_{nd}(x_d)\bigr)-g(\mathbf{x})\bigr]\,\mathrm{d}\mathbb{F}_n(\mathbf{x})=\int_{(0,1)^d}\sum_{k\in L_{n,\mathbf{x}}}\bigl[\dot{g}_k(\widetilde{\mathbf{X}}_{n,\mathbf{x}})\mathbb{U}_{nk}(x_k)\bigr]\,\mathrm{d}\mathbb{F}_n(\mathbf{x}).$$

Here, if $L_{n,\mathbf{x}}\neq\varnothing$, then $\widetilde{\mathbf{X}}_{n,\mathbf{x}}$ is a (random) point in the interior of the line segment between $\mathbf{x}$ and $(\mathbb{F}_{n1}(x_1),\dots,\mathbb{F}_{nd}(x_d))$, and it always holds that $\widetilde{\mathbf{X}}_{n,\mathbf{x}}\in(0,1)^d$. To see this, note that for all $k\in L_{n,\mathbf{x}}$, we have $x_k\neq\mathbb{F}_{nk}(x_k)$, and so the kth component of $\widetilde{\mathbf{X}}_{n,\mathbf{x}}$ lies within the open interval $(0,1)$; next, for all $k\notin L_{n,\mathbf{x}}$, the kth component of $\widetilde{\mathbf{X}}_{n,\mathbf{x}}$ is simply x_k, which again lies within $(0,1)$ because the domain of integration is $(0,1)^d$. Hence, for each $k\in L_{n,\mathbf{x}}$ the required partial derivative $\dot{g}_k(\widetilde{\mathbf{X}}_{n,\mathbf{x}})$ always exists.

Next, we can write the second term in $I(g)$ as

$$\int_{(0,1)^d}\sum_{k=1}^d\bigl[\dot{g}_k(\mathbf{x})\mathbb{U}_{nk}(x_k)\bigr]\,\mathrm{d}n^{-1/2}\mathbb{U}_n(\mathbf{x})=\int_{(0,1)^d}\sum_{k\in L_{n,\mathbf{x}}}\bigl[\dot{g}_k(\mathbf{x})\mathbb{U}_{nk}(x_k)\bigr]\,\mathrm{d}\mathbb{F}_n(\mathbf{x}).$$

The equality holds because when $k \notin L_{n,\mathbf{x}}$, we have $x_k = \mathbb{F}_{nk}(x_k)$ and so $\mathbb{U}_{nk}(x_k) = 0$. Therefore, the term $I(g)$ can be written as

$$I(g) = \int_{(0,1)^d} \sum_{k \in L_{n,\mathbf{x}}} \big[\big(\dot{g}_k(\widetilde{\mathbf{X}}_{n,\mathbf{x}}) - \dot{g}_k(\mathbf{x})\big)\mathbb{U}_{nk}(x_k)\big] \, \mathrm{d}\mathbb{F}_n(\mathbf{x}).$$

For each $k \in \{1, \dots, d\}$, because the class $\dot{\mathcal{G}}_k$ is uniformly equicontinuous on $(0,1)^d$, by Lemma 12 with the rectangular subset $\mathcal{S}$ taken to be $(0,1)^d$, there exists a bounded, non-negative, and monotone increasing function ϕ_k with $\lim_{t \downarrow 0} \phi_k(t) = 0$ such that, for arbitrary $\mathbf{x} \in (0,1)^d$ and arbitrary $k \in L_{n,\mathbf{x}}$, we have

$$\sup_{g \in \mathcal{G}} \big|\dot{g}_k(\widetilde{\mathbf{X}}_{n,\mathbf{x}}) - \dot{g}_k(\mathbf{x})\big| \le \phi_k\big(\|\widetilde{\mathbf{X}}_{n,\mathbf{x}} - \mathbf{x}\|\big) \le \phi_k\big(\big\|\mathbb{F}_n(\mathbf{x}) - \mathbf{x}\big\|\big) \le \phi_k\big(\big\|n^{-1/2}\mathbb{U}_n\big\|_\infty\big),$$

whence

$$\begin{aligned} \sup_{g \in \mathcal{G}} \big|I(g)\big| &\le \sup_{g \in \mathcal{G}} \sup_{\mathbf{x} \in (0,1)^d} \sum_{k \in L_{n,\mathbf{x}}} \big|\big(\dot{g}_k(\widetilde{\mathbf{X}}_{n,\mathbf{x}}) - \dot{g}_k(\mathbf{x})\big)\mathbb{U}_{nk}(x_k)\big| \int_{(0,1)^d} \mathrm{d}\mathbb{F}_n(\mathbf{x}) \\ &\le \sum_{k=1}^{d} \|\mathbb{U}_{nk}\|_\infty \phi_k\big(\big\|n^{-1/2}\mathbb{U}_n\big\|_\infty\big). \end{aligned}$$

The standard empirical process $\mathbb{U}_{nk}$ converges weakly in $\ell^\infty([0,1])$ and hence $\|\mathbb{U}_{nk}\|_\infty = O_p(1)$. By the Glivenko–Cantelli theorem for the standard empirical process $\mathbb{U}_n$ in $\ell^\infty([0,1]^d)$, we have $\|n^{-1/2}\mathbb{U}_n\|_\infty = o_p(1)$, and hence $\phi_k(\|n^{-1/2}\mathbb{U}_n\|_\infty) = o_p(1)$. Hence (29) is verified. □

Proposition 11. *Under the assumptions of Theorem 7, we have*

$$\sup_{g \in \mathcal{G}} \big|II(g)\big| = o_p(1). \tag{30}$$

Proof. It suffices to show that

$$\sup_{g \in \mathcal{G}} \big|II_k(g)\big| = o_p(1)$$

for each $k \in \{1, \dots, d\}$, for

$$II_k(g) = \int_{(0,1)^d} \dot{g}_k(\mathbf{x})\mathbb{U}_{nk}(x_k) \, \mathrm{d}n^{-1/2}\mathbb{U}_n(\mathbf{x}).$$

We define the classes of functions

$$\begin{aligned} \mathcal{D}_n(M) &= \big\{D : D \text{ is a distribution function on } [0,1] \text{ with } \sqrt{n}\|D - I\|_\infty \le M\big\}, \\ \mathcal{H}_{k,n}(M) &= \big\{h : (0,1)^d \to \mathbb{R}, h(\mathbf{x}) = \sqrt{n}(D - I)(x_k) f_k(\mathbf{x}), \\ &\qquad f_k \in \dot{\mathcal{G}}_k, D \in \mathcal{D}_n(M)\big\}. \end{aligned} \tag{31}$$

Fix an arbitrary (small) $\varepsilon \in (0,1)$. There exists $M = M(\varepsilon) < \infty$ such that

$$\liminf_{n\to\infty} \mathbb{P}\big\{\|\mathbb{U}_{nk}\|_\infty \le M\big\} \ge 1-\varepsilon.$$

On the event $\{\|\mathbb{U}_{nk}\|_\infty \le M\}$, we have

$$\sup_{g\in\mathcal{G}} \big|II_k(g)\big| \le \sup_{h\in\mathcal{H}_{k,n}(M)} \bigg|\int_{(0,1)^d} h \,\mathrm{d}n^{-1/2}\mathbb{U}_n\bigg|, \tag{32}$$

and to prove the proposition, it suffices to verify that the term on the right hand side of (32) converges to zero, in probability, as $n \to \infty$. Since the domain of integration in (32) is $(0,1)^d$ and the copula function C is continuous, without loss of generality in our remaining proof we focus on measures on $(0,1)^d$ and supremum $\|\cdot\|_\infty$ taken over $(0,1)^d$. By a straightforward modification of Theorem 2.4.3 in [31], the right-hand side of (32) converges to zero (in probability as $n\to\infty$), if:

1. the class $\mathcal{H}_{k,n}(M)$ has an integrable envelope and
2. for all $\xi > 0$,

$$\log N\big(\xi, \mathcal{H}_{k,n}(M), L_1(\mathbb{F}_n)\big) = o_p(n)$$

   holds. Here $N(\xi, \mathcal{H}_{k,n}(M), L_1(\mathbb{F}_n))$ is the ξ-covering number of $\mathcal{H}_{k,n}(M)$ in $L_1(\mathbb{F}_n)$, that is, the number of closed balls of radius ξ in $L_1(\mathbb{F}_n)$ needed to cover $\mathcal{H}_{k,n}(M)$.

Since $\dot{\mathcal{G}}_k$ is uniformly bounded, $\sup_{f_k\in\dot{\mathcal{G}}_k} \|f_k\|_\infty \le M_k$ for some $M_k < \infty$, and we find

$$\sup_{h\in\mathcal{H}_{k,n}(M)} \|h\|_\infty \le M\cdot M_k,$$

so the envelope condition is fulfilled. We now verify that the metric entropy condition holds. We fix arbitrary $h, h' \in \mathcal{H}_{k,n}(M)$, and write

$$h(\mathbf{x}) = \sqrt{n}(D-I)(x_k) f_k(\mathbf{x}), \qquad h'(\mathbf{x}) = \sqrt{n}\big(D'-I\big)(x_k) f'_k(\mathbf{x})$$

for $f_k, f'_k \in \dot{\mathcal{G}}_k$ and $D, D' \in \mathcal{D}_n(M)$. We can easily deduce that, for any probability measure Q,

$$\begin{aligned}\int_{(0,1)^d} \big|h-h'\big|\,\mathrm{d}Q &\le \sqrt{n}M_k \int_{(0,1)^d} \big|\big(D-D'\big)(x_k)\big|\,\mathrm{d}Q(\mathbf{x}) + M\int_{(0,1)^d} \big|f_k - f'_k\big|\,\mathrm{d}Q \\ &= \sqrt{n}M_k \int_{(0,1)} \big|\big(D-D'\big)(t)\big|\,\mathrm{d}Q_k(t) + M\int_{(0,1)^d} \big|f_k - f'_k\big|\,\mathrm{d}Q.\end{aligned}$$

Here Q_k is the kth marginal of Q. Hence, we conclude that, for any probability measure Q and $\xi > 0$,

$$\begin{aligned}&\log N\big(\xi, \mathcal{H}_{k,n}(M), L_1(Q)\big) \\ &\quad\le \log N\big(\xi/(2\sqrt{n}M_k), \mathcal{D}_n(M), L_1(Q_k)\big) + \log N\big(\xi/(2M), \dot{\mathcal{G}}_k, L_1(Q)\big) \qquad (33) \\ &\quad\le \log N\big(\xi/(2\sqrt{n}M_k), \mathcal{D}_n(M), L_1(Q_k)\big) + \log N_\infty\big(\xi/(2M), \dot{\mathcal{G}}_k\big).\end{aligned}$$

Here $N_\infty(\varepsilon, \dot{\mathcal{G}}_k)$ is the ε-covering number of $\dot{\mathcal{G}}_k$ in $L_\infty((0,1)^d)$. By Lemma 13 and Lemma 14 in Section 4.4, we have, from (33), that

$$\begin{aligned}\log N\bigl(\xi, \mathcal{H}_{k,n}(M), L_1(\mathbb{F}_n)\bigr) &\le \sup_Q \log N\bigl(\xi, \mathcal{H}_{k,n}(M), L_1(Q)\bigr)\\ &\le K_1\sqrt{n} + K_2 = O(\sqrt{n}) = o(n)\end{aligned}$$

with the supremum taken over all probability measures Q, for some finite constants K_1, $K_2 = K_2(\xi)$ independent of n. This completes the proof of (30). □

4.3. Proof of Theorem 8

We first consider the bootstrap counterpart of Theorem 5. We decompose the bootstrap empirical copula process $\sqrt{n}(\bar{\mathbb{C}}_n^* - \bar{\mathbb{C}}_n)$ as

$$\sqrt{n}\bigl(\bar{\mathbb{C}}_n^* - \bar{\mathbb{C}}_n\bigr) = \sqrt{n}\bigl(\bar{\mathbb{C}}_n^* - \mathbb{C}_n^*\bigr) + \sqrt{n}\bigl(\mathbb{C}_n^* - \mathbb{C}_n\bigr) + \sqrt{n}(\mathbb{C}_n - \bar{\mathbb{C}}_n). \tag{34}$$

We tackle the three terms on the right-hand side of (34) in sequence. Section C.3 shows that the first term satisfies (52). By Corollary 2.11 in [7], which applies under our bootstrap scheme and Assumption P, the second term, which is the bootstrap ordinary empirical copula process, converges weakly in $\ell^\infty([0,1]^d)$ to the same limit as the ordinary empirical copula process. Finally, we easily obtain from (3) that the third term satisfies

$$\sup_{\mathbf{u}\in[0,1]^d} \bigl|\sqrt{n}(\mathbb{C}_n - \bar{\mathbb{C}}_n)(\mathbf{u})\bigr| \le \frac{d}{\sqrt{n}}$$

almost surely. Therefore, we conclude that the bootstrap empirical copula process $\sqrt{n}(\bar{\mathbb{C}}_n^* - \bar{\mathbb{C}}_n)$ converges weakly in $\ell^\infty([0,1]^d)$ to the same limit as the ordinary empirical copula process. With this fact, the proof of the bootstrap counterpart of Theorem 5 is similar to the proof of Theorem 5, and is obtained by replacing the process $\mathbb{G}_n$ in Theorem 1 with $\sqrt{n}(\bar{\mathbb{C}}_n^* - \bar{\mathbb{C}}_n)$.

From here on we concentrate on the bootstrap counterpart of Theorem 7. We let $\mathbb{U}_n^* = \sqrt{n}(\mathbb{F}_n^* - \mathbb{F}_n)$ be the bootstrap counterpart of $\mathbb{U}_n = \sqrt{n}(\mathbb{F}_n - F)$ with marginals $\mathbb{U}_{nj}^*$, $j \in \{1,\dots,d\}$, and recall that $F = C$ and the marginal distributions F_j are uniform distributions on $[0,1]$.

We write the bootstrap empirical copula process indexed by functions in (25) as

$$\bar{\mathbb{Z}}_n^*(g) = \sqrt{n}\biggl(\int g\bigl(\mathbb{F}_{n1}^*(x_1),\dots,\mathbb{F}_{nd}^*(x_d)\bigr)\,\mathrm{d}\mathbb{F}_n^*(\mathbf{x}) - \int g\bigl(\mathbb{F}_{n1}(x_1),\dots,\mathbb{F}_{nd}(x_d)\bigr)\,\mathrm{d}\mathbb{F}_n(\mathbf{x})\biggr),$$

and define the empirical process

$$\widetilde{\mathbb{Z}}_n^*(g) = \int\Biggl[g + \sum_{k=1}^d T_k(g)\Biggr]\mathrm{d}\mathbb{U}_n^* = \int g\,\mathrm{d}\mathbb{U}_n^* + \sum_{k=1}^d \int_{(0,1)^d} \dot{g}_k(\mathbf{x})\mathbb{U}_{nk}^*(x_k)\,\mathrm{d}C(\mathbf{x})$$

for $g \in \mathcal{G}$. Here the function $T_k(g)$ is as defined in (27). The process $\widetilde{\mathbb{Z}}_n^*(g)$, $g \in \mathcal{G}$ has a tight Gaussian limit in $\ell^\infty(\mathcal{G})$, by the bootstrap CLT (see, for instance, Theorem 3.6.1 in [31]) and the proof of Lemma 9. Hence it suffices to show

$$\sup_{g \in \mathcal{G}} \left| \bar{\mathbb{Z}}_n^*(g) - \widetilde{\mathbb{Z}}_n^*(g) \right| = o_{p^*}(1). \tag{35}$$

For this, we first observe that, after rearranging terms and invoking the same trick as in the proof of Proposition 10 to change some domains of integration from $(0,1]^d$ to $(0,1)^d$, we have

$$\begin{aligned}
n^{-1/2}\bar{\mathbb{Z}}_n^*(g) &= \int g\big(\mathbb{F}_{n1}^*(x_1), \ldots, \mathbb{F}_{nd}^*(x_d)\big)\, \mathrm{d}\big(\mathbb{F}_n^* - \mathbb{F}_n\big)(\mathbf{x}) \\
&\quad + \int \big\{g\big(\mathbb{F}_{n1}^*(x_1), \ldots, \mathbb{F}_{nd}^*(x_d)\big) - g\big(\mathbb{F}_{n1}(x_1), \ldots, \mathbb{F}_{nd}(x_d)\big)\big\}\, \mathrm{d}\mathbb{F}_n(\mathbf{x}) \\
&= \int g(\mathbf{x})\, \mathrm{d}\big(\mathbb{F}_n^* - \mathbb{F}_n\big)(\mathbf{x}) \\
&\quad + \int_{(0,1)^d} \big\{g\big(\mathbb{F}_{n1}^*(x_1), \ldots, \mathbb{F}_{nd}^*(x_d)\big) - g\big(\mathbb{F}_{n1}(x_1), \ldots, \mathbb{F}_{nd}(x_d)\big)\big\}\, \mathrm{d}C(\mathbf{x}) \\
&\quad + \int_{(0,1)^d} \big\{g\big(\mathbb{F}_{n1}^*(x_1), \ldots, \mathbb{F}_{nd}^*(x_d)\big) - g\big(\mathbb{F}_{n1}(x_1), \ldots, \mathbb{F}_{nd}(x_d)\big)\big\}\, \mathrm{d}(\mathbb{F}_n - C)(\mathbf{x}) \\
&\quad + \int_{(0,1)^d} \big\{g\big(\mathbb{F}_{n1}^*(x_1), \ldots, \mathbb{F}_{nd}^*(x_d)\big) - g(\mathbf{x})\big\}\, \mathrm{d}\big(\mathbb{F}_n^* - \mathbb{F}_n\big)(\mathbf{x})
\end{aligned}$$

almost surely, so that

$$\big(\bar{\mathbb{Z}}_n^* - \widetilde{\mathbb{Z}}_n^*\big)(g) = I^*(g) + II^*(g) + III^*(g) \tag{36}$$

with

$$\begin{aligned}
I^*(g) &= \int_{(0,1)^d} \sqrt{n}\big[g\big(\mathbb{F}_{n1}^*(x_1), \ldots, \mathbb{F}_{nd}^*(x_d)\big) - g\big(\mathbb{F}_{n1}(x_1), \ldots, \mathbb{F}_{nd}(x_d)\big)\big]\, \mathrm{d}C(\mathbf{x}) \\
&\quad - \sum_{k=1}^{d} \int_{(0,1)^d} \dot{g}_k(\mathbf{x}) \mathbb{U}_{nk}^*(x_k)\, \mathrm{d}C(\mathbf{x}), \\
II^*(g) &= \int_{(0,1)^d} \big\{g\big(\mathbb{F}_{n1}^*(x_1), \ldots, \mathbb{F}_{nd}^*(x_d)\big) - g\big(\mathbb{F}_{n1}(x_1), \ldots, \mathbb{F}_{nd}(x_d)\big)\big\}\, \mathrm{d}\mathbb{U}_n(\mathbf{x}), \\
III^*(g) &= \int_{(0,1)^d} \big\{g\big(\mathbb{F}_{n1}^*(x_1), \ldots, \mathbb{F}_{nd}^*(x_d)\big) - g(\mathbf{x})\big\}\, \mathrm{d}\mathbb{U}_n^*(\mathbf{x}).
\end{aligned}$$

Hence it in turn suffices to show that $\sup_{g \in \mathcal{G}} |I^*(g)|$, $\sup_{g \in \mathcal{G}} |II^*(g)|$, $\sup_{g \in \mathcal{G}} |III^*(g)|$ are all $o_{p^*}(1)$.

For each $k \in \{1, \ldots, d\}$, because the class $\dot{\mathcal{G}}_k$ is uniformly equicontinuous on $V_{d,k}$, by Lemma 12 with the rectangular subset $\mathcal{S}$ taken to be $V_{d,k}$, there exists a bounded, non-negative, and monotone increasing functions ϕ_k with $\lim_{t\downarrow 0}\phi_k(t)=0$ such that $\sup_{g\in\mathcal{G}}|\dot{g}_k(\mathbf{x})-\dot{g}_k(\mathbf{y})|\le \phi_k(\|\mathbf{x}-\mathbf{y}\|)$ for all $\mathbf{x},\mathbf{y}\in V_{d,k}$. Now, we first consider the first term I^* on the right in (36). We let $L_{n,\mathbf{x}}$ be (random) sets defined as $L_{n,\mathbf{x}}=\{k\in\{1,\dots,d\}:\mathbb{F}_{nk}(x_k)\ne\mathbb{F}^*_{nk}(x_k)\}$. Then we have, by the mean value theorem,

$$\begin{aligned}&\int_{(0,1)^d}\sqrt{n}\big[g\big(\mathbb{F}^*_{n1}(x_1),\dots,\mathbb{F}^*_{nd}(x_d)\big)-g\big(\mathbb{F}_{n1}(x_1),\dots,\mathbb{F}_{nd}(x_d)\big)\big]\,\mathrm{d}C(\mathbf{x})\\&\quad=\int_{(0,1)^d}\sum_{k\in L_{n,\mathbf{x}}}\big[\dot{g}_k(\widetilde{\mathbf{X}}_{n,\mathbf{x}})\mathbb{U}^*_{nk}(x_k)\big]\,\mathrm{d}C(\mathbf{x}).\end{aligned}$$

Here, if $L_{n,\mathbf{x}}\ne\varnothing$, then $\widetilde{\mathbf{X}}_{n,\mathbf{x}}$ is a (random) point in the interior of the line segment between $(\mathbb{F}_{n1}(x_1),\dots,\mathbb{F}_{nd}(x_d))$ and $(\mathbb{F}^*_{n1}(x_1),\dots,\mathbb{F}^*_{nd}(x_d))$, and it always holds that $\widetilde{\mathbf{X}}_{n,\mathbf{x}}\in V_{d,k}$ for each $k\in L_{n,\mathbf{x}}$. To see this, note that for all $k\in L_{n,\mathbf{x}}$, we have $\mathbb{F}_{nk}(x_k)\ne\mathbb{F}^*_{nk}(x_k)$, and so the kth component of $\widetilde{\mathbf{X}}_{n,\mathbf{x}}$ lies within the open interval $(0,1)$. Hence, for each $k\in L_{n,\mathbf{x}}$ the required partial derivative $\dot{g}_k(\widetilde{\mathbf{X}}_{n,\mathbf{x}})$ always exists. Therefore, we can write $I^*(g)$ as

$$I^*(g)=\int_{(0,1)^d}\sum_{k\in L_{n,\mathbf{x}}}\big[\big(\dot{g}_k(\widetilde{\mathbf{X}}_{n,\mathbf{x}})-\dot{g}_k(\mathbf{x})\big)\mathbb{U}^*_{nk}(x_k)\big]\,\mathrm{d}C(\mathbf{x}).$$

Next, we reason as in Proposition 10 and use the property of the functions ϕ_k's and the fact that both $\mathbb{U}_{nk}$ and $\mathbb{U}^*_{nk}$ converge weakly. More precisely,

$$\sup_{g\in\mathcal{G}}\big|I^*(g)\big|\le\sum_{k=1}^d\big\|\mathbb{U}^*_{nk}\big\|_\infty\phi_k\big(\big\|\mathbb{F}^*_n-\mathbb{F}_n\big\|_\infty+\|\mathbb{F}_n-I\|_\infty\big)=o_{p^*}(1),$$

which follows because $\|\mathbb{U}^*_{nk}\|_\infty=O_{p^*}(1)$ and $\phi_k(\|\mathbb{F}^*_n-\mathbb{F}_n\|_\infty+\|\mathbb{F}_n-I\|_\infty)=o_{p^*}(1)$.

For the second term II^* on the right in (36), we write

$$\begin{aligned}\sup_{g\in\mathcal{G}}\big|II^*(g)\big|&\le\sum_{k=1}^d\sup_{g\in\mathcal{G}}\bigg|\int_{(0,1)^d}\dot{g}_k(\mathbf{x})\mathbb{U}^*_{nk}(x_k)\,\mathrm{d}n^{-1/2}\mathbb{U}_n(\mathbf{x})\bigg|\\&\quad+2\sum_{k=1}^d\big\|\mathbb{U}^*_{nk}\big\|_\infty\phi_k\big(\big\|\mathbb{F}^*_n-\mathbb{F}_n\big\|_\infty+\|\mathbb{F}_n-I\|_\infty\big)\\&=\sum_{k=1}^d\sup_{g\in\mathcal{G}}\bigg|\int_{(0,1)^d}\dot{g}_k(\mathbf{x})\mathbb{U}^*_{nk}(x_k)\,\mathrm{d}n^{-1/2}\mathbb{U}_n(\mathbf{x})\bigg|+o_{p^*}(1).\end{aligned}$$

We fix arbitrary $k\in\{1,\dots,d\}$. For each fixed $\varepsilon>0$, there exists a $M=M(\varepsilon)<\infty$ such that the event

$$\big\{\big\|\mathbb{U}^*_{nk}\big\|_\infty\le M/2\big\}\cap\big\{\|\mathbb{U}_{nk}\|_\infty\le M/2\big\}\tag{37}$$

holds with (bootstrap) probability at least $1-\varepsilon$ as $n \to \infty$. By the decomposition

$$\dot{g}_k(\mathbf{x})\mathbb{U}^*_{nk}(x_k) = \sqrt{n}\big(\mathbb{F}^*_{nk}(x_k) - x_k\big)\dot{g}_k(\mathbf{x}) - \sqrt{n}\big(\mathbb{F}_{nk}(x_k) - x_k\big)\dot{g}_k(\mathbf{x}),$$

we conclude that on the event in (37), we have

$$\dot{g}_k\mathbb{U}^*_{nk} = h - h'$$

for $h, h' \in \mathcal{H}_{k,n}(M)$, with the class $\mathcal{H}_{k,n}(M)$ as defined in (31). Hence, on the event in (37),

$$\sup_{g\in\mathcal{G}}\left|\int_{(0,1)^d} \dot{g}_k(\mathbf{x})\mathbb{U}^*_{nk}(x_k)\,\mathrm{d}n^{-1/2}\mathbb{U}_n(\mathbf{x})\right| \le 2\cdot \sup_{h\in\mathcal{H}_{k,n}(M)}\left|\int_{(0,1)^d} h\,\mathrm{d}n^{-1/2}\mathbb{U}_n\right|.$$

The right-hand side of the above inequality is $o_p(1)$ by the same reasoning as in the proof of Proposition 11. Hence, $\sup_{g\in\mathcal{G}}|II^*(g)| = o_{p^*}(1)$.

For the third term III^* on the right in (36), we argue as for the term II^* above, now using the weak convergence of $\mathbb{U}^*_n$ in lieu of $\mathbb{U}_n$. In particular, for each fixed $\varepsilon > 0$, choose $M < \infty$ for which

$$\bigcap_{k=1}^{d}\big\{\big\|\sqrt{n}\big(\mathbb{F}^*_{nk} - I\big)\big\|_\infty \le M\big\}$$

holds with (bootstrap) probability at least $1-\varepsilon$ as $n \to \infty$. On this event, $\sqrt{n}(\mathbb{F}^*_{nk} - I)\dot{g}_k$ belongs to $\mathcal{H}_{k,n}(M)$, $k \in \{1,\dots,d\}$, and

$$\sup_{g\in\mathcal{G}}\big|III^*(g)\big| \le \sum_{k=1}^{d}\sup_{h\in\mathcal{H}_{k,n}(M)}\left|\int_{(0,1)^d} h\,\mathrm{d}n^{-1/2}\mathbb{U}^*_n\right| + 2\sum_{k=1}^{d}\big\|\sqrt{n}\big(\mathbb{F}^*_{nk} - I\big)\big\|_\infty\cdot\phi_k\big(\big\|\mathbb{F}^*_n - I\big\|_\infty\big).$$

The second term on the right-hand side of the above inequality is $o_{p^*}(1)$. Furthermore, the proof of Theorem 2.4.3 in [31] or the proof of the uniform Glivenko–Cantelli theorem (Theorem 2.8.1 in [31]) shows that

$$\sup_{h\in\mathcal{H}_{k,n}(M)}\left|\int_{(0,1)^d} h\,\mathrm{d}n^{-1/2}\mathbb{U}^*_n\right| = o_{p^*}(1)$$

because all functions $h \in \mathcal{H}_{k,n}(M)$ are uniformly bounded and the required entropy condition is met with ease. Indeed, as we have shown in the proof of Proposition 11, for all $\xi > 0$, and with the supremum taken over all probability measures Q on $(0,1)^d$,

$$\sup_Q \log N\big(\xi, \mathcal{H}_{k,n}(M), L_1(Q)\big) = O(\sqrt{n}).$$

Hence, $\sup_{g\in\mathcal{G}}|III^*(g)| = o_{p^*}(1)$.

By (36) and the bounds on the terms I^*, II^* and III^*, we conclude that (35) holds, and the theorem follows from the weak convergence of $\widetilde{\mathbb{Z}}^*_n$. □

4.4. Technical results

This subsection contains technical lemmata needed for the proofs of Propositions 10 and 11, which are further required for the proofs of Theorems 7 and 8.

Lemma 12. *Let $\mathcal{S}$ be a rectangular subset of $[0,1]^d$. Let $\mathcal{F}$ be a class of functions $f:\mathcal{S}\to\mathbb{R}$ that are uniformly equicontinuous on $\mathcal{S}$. Then there exists a monotone increasing function $\phi_{\mathcal{F}}:[0,\infty)\to[0,\infty)$ such that*

$$\lim_{x\downarrow 0}\phi_{\mathcal{F}}(x)=0 \tag{38}$$

and

$$\sup_{f\in\mathcal{F}}\bigl|f(\mathbf{x})-f(\mathbf{y})\bigr|\le\phi_{\mathcal{F}}\bigl(\|\mathbf{x}-\mathbf{y}\|\bigr)\qquad\text{for all } \mathbf{x},\mathbf{y}\in\mathcal{S}.$$

In addition, $\phi_{\mathcal{F}}$ is finite valued.

Proof. Let

$$\phi_{\mathcal{F}}(a)=\sup_{\mathbf{x},\mathbf{y}\in\mathcal{S}:\|\mathbf{x}-\mathbf{y}\|\le a}\ \sup_{f\in\mathcal{F}}\bigl|f(\mathbf{x})-f(\mathbf{y})\bigr|.$$

Clearly, $\phi_{\mathcal{F}}$ is monotone increasing. In addition, (38) must hold, for otherwise $\mathcal{F}$ is not uniformly equicontinuous. Next, for any $\mathbf{x}',\mathbf{y}'\in\mathcal{S}$, we let $\delta=\|\mathbf{x}'-\mathbf{y}'\|$, and observe that

$$\sup_{f\in\mathcal{F}}\bigl|f(\mathbf{x}')-f(\mathbf{y}')\bigr|\le\sup_{\mathbf{x},\mathbf{y}\in\mathcal{S}:\|\mathbf{x}-\mathbf{y}\|\le\delta}\ \sup_{f\in\mathcal{F}}\bigl|f(\mathbf{x})-f(\mathbf{y})\bigr|=\phi_{\mathcal{F}}(\delta)=\phi_{\mathcal{F}}\bigl(\|\mathbf{x}'-\mathbf{y}'\|\bigr).$$

It remains to show that $\phi_{\mathcal{F}}$ is finite valued. Clearly, $\phi_{\mathcal{F}}(0)=0$. Now we fix arbitrary $a>0$ and show that $\phi_{\mathcal{F}}(a)<\infty$. By (38), we can choose $\delta\in(0,a]$ small enough so that $\phi_{\mathcal{F}}(\delta)<\infty$. Because the class $\mathcal{F}$ is defined on the bounded set $\mathcal{S}$, there exists some finite absolute constant C ($C=\sqrt{d}$, the Euclidean distance between $\mathbf{0}$ and $\mathbf{1}$, suffices) such that, for each pair of $\mathbf{x},\mathbf{y}\in\mathcal{S}$, we can construct a δ-chain

$$\{\mathbf{x}=\mathbf{x}_{\mathbf{x},\mathbf{y},0},\mathbf{x}_{\mathbf{x},\mathbf{y},1},\dots,\mathbf{x}_{\mathbf{x},\mathbf{y},k_{\mathbf{x},\mathbf{y}}}=\mathbf{y}\}$$

within $\mathcal{S}$ such that $k_{\mathbf{x},\mathbf{y}}\le\lceil C/\delta\rceil$ and

$$\|\mathbf{x}_{\mathbf{x},\mathbf{y},i}-\mathbf{x}_{\mathbf{x},\mathbf{y},i-1}\|\le\delta$$

for each $i=1,\dots,k_{\mathbf{x},\mathbf{y}}$. Then, by the construction of $\phi_{\mathcal{F}}$, we have

$$\begin{aligned}\phi_{\mathcal{F}}(a)&=\sup_{\mathbf{x},\mathbf{y}\in\mathcal{S}:\|\mathbf{x}-\mathbf{y}\|\le a}\ \sup_{f\in\mathcal{F}}\bigl|f(\mathbf{x})-f(\mathbf{y})\bigr|\\&\le\sup_{\mathbf{x},\mathbf{y}\in\mathcal{S}:\|\mathbf{x}-\mathbf{y}\|\le a}\ \sup_{f\in\mathcal{F}}\sum_{i=1}^{k_{\mathbf{x},\mathbf{y}}}\bigl|f(\mathbf{x}_{\mathbf{x},\mathbf{y},i})-f(\mathbf{x}_{\mathbf{x},\mathbf{y},i-1})\bigr|\end{aligned}$$

$$\leq \sup_{\mathbf{x},\mathbf{y}\in\mathcal{S}:\|\mathbf{x}-\mathbf{y}\|\leq a} \sum_{i=1}^{k_{\mathbf{x},\mathbf{y}}} \phi_{\mathcal{F}}(\delta)$$
$$\leq \lceil C/\delta \rceil \cdot \phi_{\mathcal{F}}(\delta) < \infty. \qquad \square$$

Lemma 13. *Let $\mathcal{S}$ be a rectangular subset of $[0,1]^d$. Let $\mathcal{F}$ be a class of functions $f:\mathcal{S}\to\mathbb{R}$ that are uniformly bounded and uniformly equicontinuous on $\mathcal{S}$. Then the class $\mathcal{F}$ is totally bounded in $L_\infty(\mathcal{S})$.*

Proof. Since by Lemma 12 there exists a bounded, non-negative and monotone increasing function ϕ such that $\lim_{t\downarrow 0}\phi(t)=0$ and $\sup_{f\in\mathcal{F}}|f(\mathbf{x})-f(\mathbf{y})|\leq\phi(\|\mathbf{x}-\mathbf{y}\|)$ for all $\mathbf{x},\mathbf{y}\in\mathcal{S}$, and since the domain $\mathcal{S}$ is bounded, we can construct, for each $\varepsilon>0$, a regular δ-grid of $\mathcal{S}$ with $\delta=\delta(\varepsilon)$ strictly positive and $\phi(\sqrt{d}\delta)\leq\varepsilon/2$. Since $\mathcal{F}$ is uniformly bounded, it is easy to see that using this finite grid with δ chosen above, there are finitely many functions $g_1,\dots,g_M$ such that $\sup_{f\in\mathcal{F}}\min_{1\leq k\leq M}\|f-g_k\|_\infty<\varepsilon$ (here the supremum in $\|\cdot\|_\infty$ is taken over $\mathcal{S}$). (Using the line of reasoning as in the proof of Lemma 2.3 in [30], we can actually provide a cleverer upper bound on M.) $\square$

Lemma 14. *The class $\mathcal{F}$ of monotone functions $f:\mathbb{R}\to[0,1]$ satisfies*

$$\log N\big(\varepsilon,\mathcal{F},L_r(Q)\big)\leq K\frac{1}{\varepsilon}$$

for all $\varepsilon>0$, all probability measures Q and all $r\geq 1$, for a constant K that depends on r only.

Proof. See Theorem 2.7.5 in [31]. $\square$

Appendix A: Integration by parts

In this section, we present a general yet simple integration by parts formula for Lebesgue–Stieltjes integration in arbitrary dimensions in formula (39) in Theorem 15. For Lebesgue–Stieltjes integration, the integration by parts formula in one dimension is well-known; see for instance, Theorem 14.1 in Chapter 3 of [26]. However, its multivariate extension does not appear to be adequately addressed. To the best of our knowledge, the only general integration by parts formula for Lebesgue–Stieltjes integration in arbitrary dimensions is Proposition A.1 in [12], page 149, which is essentially (42) in our derivation of Theorem 15. We improve upon this formula on two fronts:

- The integrands in (42) are expressed in terms of the measure dg defined from the function g through formula (9) (with f replaced by g), instead of the function g itself. Furthermore, if we expand the integrands in (42) directly in terms of the function g, significant cancellation will occur; this can be clearly seen through the derivation of our Theorem 15, or through comparing the number of terms in (42) after the expansion and in our formula (39). Both of these reasons make Proposition A.1 in [12] unnecessarily complicated. In contrast, our formula (39) is much simplified and essentially optimal.

- The condition under which the integration by parts formula holds is now expressed succinctly as a result of the connection, recently established in [1], between functions of bounded Hardy–Krause variation and signed measures.

Recently, Theorem A.6 and Corollary A.7 in [4] derive a simple bivariate integration by parts formula for Lebesgue–Stieltjes integration; however, their proof is quite specific to the case $d = 2$ and does not appear to be easily generalizable to higher dimensions. We will demonstrate how we recover Corollary A.7 in [4] from our general formula. We refer the readers to Section 1.1 for notations, which in turn follow [21]. To obtain a more general formula, we consider functions on $[\mathbf{a}, \mathbf{b}] \subset \mathbb{R}^d$ for $\mathbf{a} \le \mathbf{b}$ instead of the unit hypercube $[0, 1]^d$ on which we focused for the remainder of the paper.

Theorem 15 (Integration by parts). *We let* $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$ *with* $\mathbf{a} \le \mathbf{b}$, *and functions* $f, g : [\mathbf{a}, \mathbf{b}] \to \mathbb{R}$. *We assume that* f, g *both satisfy Assumption* F *with* $[0, 1]^d$ *replaced by* $[\mathbf{a}, \mathbf{b}]$. *We let the domain of integration be* $(\mathbf{c}, \mathbf{d}] \subset \mathbb{R}^d$ *with* $\mathbf{a} \le \mathbf{c} \le \mathbf{d} \le \mathbf{b}$. *Then we have the integration by parts formula*:

$$\int_{(\mathbf{c},\mathbf{d}]} f\, \mathrm{d}g = \sum_{\substack{I_1, I_2, I_3 \subset \{1,\dots,d\}: \\ I_1 + I_2 + I_3 = \{1,\dots,d\}}} (-1)^{|I_1|+|I_2|} \int_{(\mathbf{c}_{I_1}, \mathbf{d}_{I_1}]} g(\mathbf{x}_{I_1}-; \mathbf{c}_{I_2} : \mathbf{d}_{I_3})\, \mathrm{d}f(\mathbf{x}_{I_1}; \mathbf{c}_{I_2} : \mathbf{d}_{I_3}). \tag{39}$$

In (39) *the '+' symbol within* $I_1 + I_2 + I_3$ *denotes the disjoint union, so the summation is taken over all partitions of the set* $\{1, \dots, d\}$ *into the sets* I_1, I_2, I_3.

Remarks.

- In (39), the left-continuous function $g(\cdot -; \mathbf{c}_{I_2} : \mathbf{d}_{I_3})$ on $(\mathbf{c}_{I_1}, \mathbf{d}_{I_1}]$ is defined as

$$g(\mathbf{x}_{I_1}-; \mathbf{c}_{I_2} : \mathbf{d}_{I_3}) = \lim_{\mathbf{y}_{I_1} < \mathbf{x}_{I_1}, \mathbf{y}_{I_1} \uparrow \mathbf{x}_{I_1}} g(\mathbf{y}_{I_1}; \mathbf{c}_{I_2} : \mathbf{d}_{I_3}),$$

  that is, it denotes the left-hand limit of the function $g(\cdot; \mathbf{c}_{I_2} : \mathbf{d}_{I_3})$ on all coordinates $I_1 \subset \{1, \dots, d\}$ jointly. The alert reader may notice that we have implicitly assumed the existence of this limit. In fact, this is guaranteed by Assumption F imposed on g, which implies the identification of the function g with a signed Borel measure $\mathrm{d}g$ to which g gives rise through formula (7) (with f replaced by g). Then, the existence of the left-hand limit of the function $g(\cdot; \mathbf{c}_{I_2} : \mathbf{d}_{I_3})$ follows from the continuity of the measure $\mathrm{d}g$.
- If either f or g is continuous, then we can replace $\mathbf{x}_{I_1}-$ in (39) by $\mathbf{x}_{I_1}$. This is obvious if g is continuous. If instead f is continuous, then this follows because g can be decomposed as a sum of two *completely monotone* functions (see Appendix B), which implies by Lemma 1 in [19] that the set of discontinuities of g is concentrated on a countable number of coordinate hyperplanes, and consequently this set has measure zero with respect to $\mathrm{d}f$.
- In Theorem 15, we used the convention that if $I_1 = \varnothing$, then

$$\int_{(\mathbf{c}_{I_1}, \mathbf{d}_{I_1}]} g(\mathbf{x}_{I_1}-; \mathbf{c}_{I_2} : \mathbf{d}_{I_3})\, \mathrm{d}f(\mathbf{x}_{I_1}; \mathbf{c}_{I_2} : \mathbf{d}_{I_3}) = g(\mathbf{c}_{I_2} : \mathbf{d}_{I_3}) f(\mathbf{c}_{I_2} : \mathbf{d}_{I_3}) = (fg)(\mathbf{c}_{I_2} : \mathbf{d}_{I_3}),$$

that is, when there is no integration variable, the integration operation simply disappears.[1] This convention allows for a more condensed expression, and we will use the same convention below. Alternatively, to avoid invoking this convention, (39) can be equivalently expressed as

$$\begin{aligned}\int_{(\mathbf{c},\mathbf{d}]} f\,\mathrm{d}g &= \sum_{\substack{I_1,I_2,I_3\subset\{1,\dots,d\}:\\ I_1\neq\varnothing, I_1+I_2+I_3=\{1,\dots,d\}}} (-1)^{|I_1|+|I_2|}\int_{(\mathbf{c}_{I_1},\mathbf{d}_{I_1}]} g(\mathbf{x}_{I_1}-;\mathbf{c}_{I_2}:\mathbf{d}_{I_3})\,\mathrm{d}f(\mathbf{x}_{I_1};\mathbf{c}_{I_2}:\mathbf{d}_{I_3}) \\ &\quad + \sum_{I\subset\{1,\dots,d\}} (-1)^{|I|}(fg)(\mathbf{c}_I:\mathbf{d}_{-I}).\end{aligned} \tag{40}$$

Note that the summation in the last line of (40) is just $\mathrm{d}(fg)((\mathbf{c},\mathbf{d}])$, the weight of the hypercube $(\mathbf{c},\mathbf{d}]$ assigned by the measure $\mathrm{d}(fg)$ defined from the product (fg). We remark here that the function (fg) is right-continuous and is of bounded Hardy–Krause variation (by the closure property of bounded Hardy–Krause variation under multiplication; see, for instance, Proposition 11 in [21]), so the measure $\mathrm{d}(fg)$ is well defined.

- In one dimension, the formula reduces to

$$\begin{aligned}\int_{(c,d]} f(x)\,\mathrm{d}g(x) &= -\int_{(c,d]} g(x-)\,\mathrm{d}f(x) + \big[(fg)(d)-(fg)(c)\big] \\ &= -\int_{(c,d]} g(x)\,\mathrm{d}f(x) + \big[(fg)(d)-(fg)(c)\big] \\ &\quad + \int_{(c,d]} \big\{g(x)-g(x-)\big\}\,\mathrm{d}f(x).\end{aligned}$$

  Here d denotes the right end point of domain of integration, rather than dimension. The second term on the right is the generalized length of the interval $(c,d]$, based on the right-continuous function (fg), and the third term equals $\sum_x\{g(x)-g(x-)\}\{f(x)-f(x-)\}$ with the summation taken over all (countable) common points x of discontinuity of f and g. This term vanishes if either f or g is continuous.
- Note that no further simplification of (39) is possible, as clearly all terms in the sum, now with integrands expressed directly in terms of the function g (instead of the measure to which it corresponds), are distinct. In contrast, Proposition A.1 in [12] still contains duplicates that are hidden in the measures that appear as integrands. In d dimensions, there are a total of 3^d terms in the sum of (39), because each of the d coordinates belongs to exactly one of the sets I_1, I_2, I_3. If $d=2$ and if either f or g is continuous, these $3^2=9$ terms

[1] In [21], the constant $f(c)$ on the right-hand side of equation (13) is alternatively written as $\Delta_{u=\varnothing}(f;x,c)$, which appears in the first line of the equation array in the proof that follows and which resembles a measure. We follow suit and pretend that the constant $f(\mathbf{c}_{I_2}:\mathbf{d}_{I_3})$ in fact gives rise to a measure.

exactly correspond to Corollary A.7 in [4], as we demonstrate now. In this case, (40) (which is equivalent to (39) with our convention) claims

$$\int_{(\mathbf{c},\mathbf{d}]} f\,\mathrm{d}g = \sum_{\substack{I_1,I_2,I_3\subset\{1,2\}:\\ I_1\neq\varnothing, I_1+I_2+I_3=\{1,2\}}} (-1)^{|I_1|+|I_2|}\int_{(\mathbf{c}_{I_1},\mathbf{d}_{I_1}]} g(\mathbf{x}_{I_1};\mathbf{c}_{I_2}:\mathbf{d}_{I_3})\,\mathrm{d}f(\mathbf{x}_{I_1};\mathbf{c}_{I_2}:\mathbf{d}_{I_3}) + \sum_{I\subset\{1,2\}}(-1)^{|I|}(fg)(\mathbf{c}_I:\mathbf{d}_{-I}). \tag{41}$$

We work out the terms on the right-hand side of (41) one by one. We start from the second line on the right-hand side of (41). For $I=\varnothing$, we have

$$(-1)^0(fg)(\mathbf{c}_\varnothing:\mathbf{d}_{\{1,2\}}) = (fg)(d_1,d_2).$$

For $I=\{1\}$, we have

$$(-1)^1(fg)(\mathbf{c}_{\{1\}}:\mathbf{d}_{\{2\}}) = -(fg)(c_1,d_2).$$

For $I=\{2\}$, we have

$$(-1)^1(fg)(\mathbf{c}_{\{2\}}:\mathbf{d}_{\{1\}}) = -(fg)(d_1,c_2).$$

For $I=\{1,2\}$, we have

$$(-1)^2(fg)(\mathbf{c}_{\{1,2\}}:\mathbf{d}_\varnothing) = (fg)(c_1,c_2).$$

We now switch to the first line on the right-hand side of (41). For $I_1=\{1\}$, $I_2=\varnothing$, and so $I_3=\{2\}$, we have

$$(-1)^{1+0}\int_{(\mathbf{c}_{\{1\}},\mathbf{d}_{\{1\}}]} g(\mathbf{x}_{\{1\}};\mathbf{d}_{\{2\}})\,\mathrm{d}f(\mathbf{x}_{\{1\}};\mathbf{d}_{\{2\}}) =: -\int_{(c_1,d_1]} g(x_1,d_2)f(\mathrm{d}x_1,d_2).$$

For $I_1=\{1\}$, $I_2=\{2\}$, and so $I_3=\varnothing$, we have

$$(-1)^{1+1}\int_{(\mathbf{c}_{\{1\}},\mathbf{d}_{\{1\}}]} g(\mathbf{x}_{\{1\}};\mathbf{c}_{\{2\}})\,\mathrm{d}f(\mathbf{x}_{\{1\}};\mathbf{c}_{\{2\}}) =: \int_{(c_1,d_1]} g(x_1,c_2)f(\mathrm{d}x_1,c_2).$$

For $I_1=\{2\}$, $I_2=\varnothing$, and so $I_3=\{1\}$, we have

$$(-1)^{1+0}\int_{(\mathbf{c}_{\{2\}},\mathbf{d}_{\{2\}}]} g(\mathbf{x}_{\{2\}};\mathbf{d}_{\{1\}})\,\mathrm{d}f(\mathbf{x}_{\{2\}};\mathbf{d}_{\{1\}}) =: -\int_{(c_2,d_2]} g(d_1,x_2)f(d_1,\mathrm{d}x_2).$$

For $I_1=\{2\}$, $I_2=\{1\}$, and so $I_3=\varnothing$, we have

$$(-1)^{1+1}\int_{(\mathbf{c}_{\{2\}},\mathbf{d}_{\{2\}}]} g(\mathbf{x}_{\{2\}};\mathbf{c}_{\{1\}})\,\mathrm{d}f(\mathbf{x}_{\{2\}};\mathbf{c}_{\{1\}}) =: \int_{(c_2,d_2]} g(c_1,x_2)f(c_1,\mathrm{d}x_2).$$

Finally, for $I_1 = \{1,2\}$ and so $I_2 = I_3 = \varnothing$, we have

$$(-1)^{2+0} \int_{(\mathbf{c}_{\{1,2\}}, \mathbf{d}_{\{1,2\}}]} g(\mathbf{x}_{\{1,2\}})\, \mathrm{d}f(\mathbf{x}_{\{1,2\}}) = \int_{(\mathbf{c},\mathbf{d}]} g(x_1, x_2)\, \mathrm{d}f(x_1, x_2) = \int_{(\mathbf{c},\mathbf{d}]} g\, \mathrm{d}f.$$

Thus in the end, after collecting all terms and replacing the dummy integration variables x_1 and x_2 by u and v respectively, we have

$$\begin{aligned}\int_{(\mathbf{c},\mathbf{d}]} f\, \mathrm{d}g = {}& \int_{(\mathbf{c},\mathbf{d}]} g\, \mathrm{d}f + (fg)(d_1, d_2) - (fg)(c_1, d_2) - (fg)(d_1, c_2) + (fg)(c_1, c_2) \\ & - \int_{(c_1,d_1]} g(u, d_2) f(\mathrm{d}u, d_2) + \int_{(c_1,d_1]} g(u, c_2) f(\mathrm{d}u, c_2) \\ & - \int_{(c_2,d_2]} g(d_1, v) f(d_1, \mathrm{d}v) + \int_{(c_2,d_2]} g(c_1, v) f(c_1, \mathrm{d}v).\end{aligned}$$

The terms in the above equation are arranged as and correspond exactly to the terms in Corollary A.7 in [4] (note that their $\Delta(fg, c_1, c_2, d_1, d_2)$ term is exactly the sum of the second to the fifth terms on the right hand side of the above equation).

We remark here that, when $d = 2$ and neither f nor g is continuous, with just a little more algebra, we can also recover Theorem A.6 in [4] from our equation (40). The extra terms appearing in Theorem A.6 in [4] as compared to their Corollary A.7 will come from the replacement of the integrand $g(\mathbf{x}_{I_1}; \mathbf{c}_{I_2} : \mathbf{d}_{I_3})$ by $g(\mathbf{x}_{I_1}-; \mathbf{c}_{I_2} : \mathbf{d}_{I_3})$ in our equation (41).

- If g corresponds to sample paths of the empirical copula process (so $[\mathbf{a}, \mathbf{b}] = [\mathbf{c}, \mathbf{d}] = [0, 1]^d$), then (39) can be further simplified to formula (11) in Proposition 3, because the assumptions in the second sentence of Proposition 3 are fulfilled.

Proof of Theorem 15. We recall how the functions f and g, and the lower-dimensional projections of the former, are identified with measures satisfying (7) and (9). Using equation (13) (or more precisely, using the left-hand side of the first line of the equation array in the proof that follows) in [21] (see also Lemma A.2 in [12]), we have

$$f(\mathbf{x}) = \sum_{I \subset \{1,\dots,d\}} (-1)^{|I|}\, \mathrm{d}f\big((\mathbf{x}_I, \mathbf{d}_I]; \mathbf{d}_{-I}\big).$$

Plugging this into $\int_{(\mathbf{c},\mathbf{b}]} f\, \mathrm{d}g$, we have

$$\begin{aligned}\int_{(\mathbf{c},\mathbf{d}]} f\, \mathrm{d}g &= \sum_{I \subset \{1,\dots,d\}} (-1)^{|I|} \Big[\int_{(\mathbf{c},\mathbf{d}]} \mathrm{d}f\big((\mathbf{x}_I, \mathbf{d}_I]; \mathbf{d}_{-I}\big)\, \mathrm{d}g(\mathbf{x})\Big] \\ &= \sum_{I \subset \{1,\dots,d\}} (-1)^{|I|} \Big[\int_{(\mathbf{c},\mathbf{d}]} \int_{(\mathbf{c}_I, \mathbf{d}_I]} \mathbb{1}\{\mathbf{c}_I < \mathbf{x}_I < \mathbf{y}_I \le \mathbf{d}_I\}\, \mathrm{d}f(\mathbf{y}_I; \mathbf{d}_{-I})\, \mathrm{d}g(\mathbf{x})\Big].\end{aligned}$$

Applying Fubini's theorem, we obtain

$$\begin{aligned}\int_{(\mathbf{c},\mathbf{d}]} f\,\mathrm{d}g &= \sum_{I\in\{1,\dots,d\}} (-1)^{|I|}\left[\int_{(\mathbf{c}_I,\mathbf{d}_I]}\int_{(\mathbf{c},\mathbf{d}]} \mathbb{1}\{\mathbf{c}_I < \mathbf{x}_I < \mathbf{y}_I \le \mathbf{d}_I\}\,\mathrm{d}g(\mathbf{x})\,\mathrm{d}f(\mathbf{y}_I;\mathbf{d}_{-I})\right] \\ &= \sum_{I\subset\{1,\dots,d\}} (-1)^{|I|}\left[\int_{(\mathbf{c}_I,\mathbf{d}_I]} \mathrm{d}g\big((\mathbf{c}_I,\mathbf{y}_I)\times(\mathbf{c}_{-I},\mathbf{d}_{-I}]\big)\,\mathrm{d}f(\mathbf{y}_I;\mathbf{d}_{-I})\right].\end{aligned} \tag{42}$$

Up to this point, we have essentially derived a variant of Proposition A.1 in [12]. However, (42) can be significantly simplified. Continuing from (42), we have

$$\begin{aligned}&\int_{(\mathbf{c},\mathbf{d}]} f\,\mathrm{d}g \\ &\quad= \sum_{I\subset\{1,\dots,d\}} (-1)^{|I|}\left[\int_{(\mathbf{c}_I,\mathbf{d}_I]} \lim_{\mathbf{x}_I<\mathbf{y}_I,\mathbf{x}_I\uparrow\mathbf{y}_I} \mathrm{d}g\big((\mathbf{c}_I,\mathbf{x}_I]\times(\mathbf{c}_{-I},\mathbf{d}_{-I}]\big)\,\mathrm{d}f(\mathbf{y}_I;\mathbf{d}_{-I})\right] \\ &\quad= \sum_{I\subset\{1,\dots,d\}} (-1)^{|I|}\left[\int_{(\mathbf{c}_I,\mathbf{d}_I]} \lim_{\mathbf{x}_I<\mathbf{y}_I,\mathbf{x}_I\uparrow\mathbf{y}_I} \sum_{I_1\subset I}\sum_{I_2\subset -I}(-1)^{|I_1|+|I_2|}\right. \\ &\qquad\left.\times\, g(\mathbf{x}_{I-I_1}:\mathbf{c}_{I_1};\mathbf{c}_{I_2}:\mathbf{d}_{-I-I_2})\,\mathrm{d}f(\mathbf{y}_I;\mathbf{d}_{-I})\right] \\ &\quad= \sum_{I\subset\{1,\dots,d\}}\sum_{I_1\subset I}\sum_{I_2\subset -I}(-1)^{|I|+|I_1|+|I_2|} \\ &\qquad\times\left[\int_{(\mathbf{c}_I,\mathbf{d}_I]} g(\mathbf{y}_{I-I_1}-:\mathbf{c}_{I_1};\mathbf{c}_{I_2}:\mathbf{d}_{-I-I_2})\,\mathrm{d}f(\mathbf{y}_I;\mathbf{d}_{-I})\right].\end{aligned} \tag{43}$$

The term in the square bracket in the last line of (43) can be further simplified as

$$\sum_{I_3\subset I_1}(-1)^{|I_3|}\int_{(\mathbf{c}_{I-I_1},\mathbf{d}_{I-I_1}]} g(\mathbf{y}_{I-I_1}-;\mathbf{c}_{I_1+I_2}:\mathbf{d}_{-I-I_2})\,\mathrm{d}f(\mathbf{y}_{I-I_1};\mathbf{c}_{I_3}:\mathbf{d}_{-I+I_1-I_3}). \tag{44}$$

This claim is easily verified for step functions g (in $\mathbf{y}_{I-I_1}$) and the general case follows by approximating g by a sum of step functions. After replacing the term in the square bracket in the last line of (43) by (44), we obtain

$$\begin{aligned}\int_{(\mathbf{c},\mathbf{d}]} f\,\mathrm{d}g = &\sum_{I\subset\{1,\dots,d\}}\sum_{I_1\subset I}\sum_{I_2\subset -I}\sum_{I_3\subset I_1}(-1)^{|I|+|I_1|+|I_2|+|I_3|} \\ &\times\left[\int_{(\mathbf{c}_{I-I_1},\mathbf{d}_{I-I_1}]} g(\mathbf{x}_{I-I_1}-;\mathbf{c}_{I_1+I_2}:\mathbf{d}_{-I-I_2})\,\mathrm{d}f(\mathbf{x}_{I-I_1};\mathbf{c}_{I_3}:\mathbf{d}_{-I+I_1-I_3})\right].\end{aligned} \tag{45}$$

We now simplify the summation in (45). We let $u_1 = I - I_1, u_2 = I_3, u_3 = -I - I_2, u_4 = I_2, u_5 = I_1 - I_3$, so $u_1 + u_2 + u_3 + u_4 + u_5 = \{1,\dots,d\}$. Then, the summation over

$I \subset \{1,\dots,d\}$, $I_1 \subset I$, $I_2 \subset -I$, $I_3 \subset I_1$ becomes a summation over u_1, u_2, u_3, u_4, u_5 with $u_1 + u_2 + u_3 + u_4 + u_5 = \{1,\dots,d\}$, and furthermore $(-1)^{|I|+|I_1|+|I_2|+|I_3|}$ becomes $(-1)^{|u_1|+3|u_2|+|u_4|+2|u_5|} = (-1)^{|u_1|+|u_2|+|u_4|}$. Consequently, (45) becomes

$$\int_{(\mathbf{c},\mathbf{d}]} f \,\mathrm{d}g = \sum_{\substack{u_1,u_2,u_3,u_4,u_5 \subset \{1,\dots,d\}:\\ u_1+u_2+u_3+u_4+u_5=\{1,\dots,d\}}} (-1)^{|u_1|+|u_2|+|u_4|} \times \left[\int_{(\mathbf{c}_{u_1},\mathbf{d}_{u_1}]} g(\mathbf{x}_{u_1}-; \mathbf{c}_{u_2+u_4+u_5} : \mathbf{d}_{u_3}) \,\mathrm{d}f(\mathbf{x}_{u_1}; \mathbf{c}_{u_2} : \mathbf{d}_{u_3+u_4+u_5}) \right]. \tag{46}$$

Note that in the square bracket in (46), u_4 and u_5 always appear together as the union $u_4 + u_5$, and so the term in the square bracket is uniquely determined by u_1, u_2, u_3 (in which case $u_4 + u_5$ is also determined). Now we evaluate, for given u_1, u_2, u_3, the coefficient

$$\sum_{\substack{u_4,u_5 \subset \{1,\dots,d\}:\\ u_4+u_5=\{1,\dots,d\}-u_1-u_2-u_3}} (-1)^{|u_1|+|u_2|+|u_4|} = (-1)^{|u_1|+|u_2|} \sum_{\substack{u_4,u_5 \subset \{1,\dots,d\}:\\ u_4+u_5=\{1,\dots,d\}-u_1-u_2-u_3}} (-1)^{|u_4|}.$$

One moment's thought reveals that the summation on the right-hand side of the above equality is zero, unless $u_4 + u_5 = \varnothing$, in which case it is one.[2] Finally, applying this result to (46) yields (39). □

Appendix B: Measures defined from lower dimensional projections of functions of bounded Hardy–Krause variation

We assume that a function f satisfies Assumption F. We let $I \subset \{1,\dots,d\}$ with $I \neq \varnothing$ and $\mathbf{c} \in [0,1]^d$ be arbitrary, and let $f(\cdot; \mathbf{c}_{-I})$ be the lower dimensional projection of the function f onto $[\mathbf{0}_I, \mathbf{1}_I]$. We show that, corresponding to the function $f(\cdot; \mathbf{c}_{-I})$ there exists a unique, finite signed Borel measure $\mathrm{d}f(\cdot; \mathbf{c}_{-I})$ on $[\mathbf{0}_I, \mathbf{1}_I]$ satisfying (9). By part (a) of Theorem 3 in [1], because $f(\cdot; \mathbf{c}_{-I})$ is obviously right-continuous, it suffices to show that $f(\cdot; \mathbf{c}_{-I})$ is of bounded Hardy–Krause variation on $[\mathbf{0}_I, \mathbf{1}_I]$. By definition of the Hardy–Krause variation, in turn it suffices to show that for each $I' \subset I$ with $I' \neq \varnothing$, the Vitali variation of the function $f(\cdot; \mathbf{1}_{I-I'} : \mathbf{c}_{-I})$ on $[\mathbf{0}_{I'}, \mathbf{1}_{I'}]$ is bounded. Without loss of generality we show this for an arbitrary $I' \subset I$ with $I' \neq \varnothing$.

We essentially proceed as in the proof (up to roughly the top of page 160) of Lemma 2 in [1]. We let $f(\mathbf{x}) = f(\mathbf{0}) + f^+(\mathbf{x}) + f^-(\mathbf{x})$ be the Jordan decomposition of the function f on $[0,1]^d$ in the sense defined in Theorem 2 in [1]. Then, both f^+ and f^- are finite-valued (which can

[2] If $u_4 + u_5 \neq \varnothing$ and $|u_4 + u_5|$ is odd, for each $u \subset u_4 + u_5$, the term $(-1)^{|u_4|}$ with $u_4 = u$ cancels with the term $(-1)^{|u_4|}$ with $u_4 = u_4 + u_5 - u$, because exactly one of $|u|$ and $|u_4 + u_5 - u|$ is odd. If $u_4 + u_5 \neq \varnothing$ and $|u_4 + u_5|$ is even, and if $i \in u_4 + u_5$, we can separately consider the case $i \in u_4$ and $i \notin u_4$ to effectively reduce the number of coordinates to consider from even to odd, and apply the previous argument again.

be seen from equations (26) and (27) in the proof of Theorem 2 in [1]), and both f^+ and f^- as well as all their lower dimensional projections are completely monotone in the sense defined at the bottom of page 153 of [1] (alternatively, see Section 3 of [19]).

By the closure property of bounded Vitali variation under summation, it now suffices to show that the Vitali variations $V^{\pm}(f^{\pm}(\cdot;\mathbf{1}_{I-I'}:\mathbf{c}_{-I}))$ of the functions $f^{\pm}(\cdot;\mathbf{1}_{I-I'}:\mathbf{c}_{-I})$ on $[\mathbf{0}_{I'},\mathbf{1}_{I'}]$ are bounded. Without loss of generality, we show $V(f^+(\cdot;\mathbf{1}_{I-I'}:\mathbf{c}_{-I}))<\infty$. By definition,

$$\begin{aligned} V\bigl(f^+(\cdot;\mathbf{1}_{I-I'}:\mathbf{c}_{-I})\bigr) &= \sup_{\mathcal{Y}_{I'}} V_{\mathcal{Y}_{I'}}\bigl(f^+(\cdot;\mathbf{1}_{I-I'}:\mathbf{c}_{-I})\bigr) \\ &= \sup_{\mathcal{Y}_{I'}} \sum_{\mathbf{y}_{I'}\in\mathcal{Y}_{I'}} \biggl|\sum_{I''\subset I'}(-1)^{|I''|} f^+\bigl(\mathbf{y}_{I''}:\mathbf{y}^+_{I'-I''}:\mathbf{1}_{I-I'}:\mathbf{c}_{-I}\bigr)\biggr|. \end{aligned}$$

Here the supremum is taken over all multivariate ladders $\mathcal{Y}_{I'}=\prod_{j\in I'}\mathcal{Y}_j$ of $[\mathbf{0}_{I'},\mathbf{1}_{I'}]$ for $\mathcal{Y}_j$ a ladder of $[\mathbf{0}_{\{j\}},\mathbf{1}_{\{j\}}]$, and in the first line $V_{\mathcal{Y}_{I'}}(f^+(\cdot;\mathbf{1}_{I-I'}:\mathbf{c}_{-I}))$ is the variation of the function $f^+(\cdot;\mathbf{1}_{I-I'}:\mathbf{c}_{-I})$ over the multivariate ladder $\mathcal{Y}_{I'}$. Next, in the second line the quantity inside the absolute value function is the $|I'|$-fold alternating sum of the function $f^+(\cdot;\mathbf{1}_{I-I'}:\mathbf{c}_{-I})$ over the hypercube $(\mathbf{y}_{I'},\mathbf{y}^+_{I'}]$ for $\mathbf{y}^+_{I'}$ the successor of $\mathbf{y}_{I'}\in\mathcal{Y}_{I'}$; because f^+ (as a function on $[0,1]^d$) is completely monotone, this quantity is always non-negative. Consequently, we have

$$V\bigl(f^+(\cdot;\mathbf{1}_{I-I'}:\mathbf{c}_{-I})\bigr) = \sup_{\mathcal{Y}_{I'}} \sum_{\mathbf{y}_{I'}\in\mathcal{Y}_{I'}} \sum_{I''\subset I'}(-1)^{|I''|} f^+\bigl(\mathbf{y}_{I''}:\mathbf{y}^+_{I'-I''}:\mathbf{1}_{I-I'}:\mathbf{c}_{-I}\bigr).$$

The double summation on the right-hand side of the above equation forms a telescoping sum, and after simplification always reduces to the right hand side of (47), no matter what ladder $\mathcal{Y}_{I'}$ is chosen. Therefore, we obtain

$$V\bigl(f^+(\cdot;\mathbf{1}_{I-I'}:\mathbf{c}_{-I})\bigr) = \sum_{I''\subset I'}(-1)^{|I''|} f^+(\mathbf{0}_{I''}:\mathbf{1}_{I-I''}:\mathbf{c}_{-I}). \tag{47}$$

Because each term in the summation on the right-hand side of (47) is finite, we conclude that $V(f^+(\cdot;\mathbf{1}_{I-I'}:\mathbf{c}_{-I}))$ is finite as well.

Appendix C: Closeness of the ordinary empirical copula process and its càdlàg version

In this section we show that, under appropriate conditions, the ordinary empirical copula process and its càdlàg version are uniformly sufficiently close on the unit hypercube, and hence the weak convergence of one in $\ell^\infty([0,1]^d)$ implies the weak convergence of the other in $\ell^\infty([0,1]^d)$ to the same limit. In both Sections C.1 and C.2, we consider the process $\sqrt{n}(\mathbb{C}_n-C)$ given in (1) and its càdlàg version $\sqrt{n}(\bar{\mathbb{C}}_n-C)$ given in (2); specifically, in Section C.1, we prove (3) for an i.i.d. sample, and in Section C.2, we consider the non-i.i.d. case concerning Corollary 6. Finally, in Section C.3, we consider the bootstrap empirical copula process from an i.i.d. sample.

C.1. The i.i.d. case

We fix an arbitrary $\mathbf{u} \in [0, 1]^d$. Adapting the proof of Lemma 7.2 in [5] up to the bottom of page 517 for our case, we have

$$\left|\mathbb{C}_n(\mathbf{u}) - \bar{\mathbb{C}}_n(\mathbf{u})\right| \le \sum_{j=1}^{d} \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\left\{X_{ij} = \mathbb{F}_{nj}^{-}(u_j)\right\}. \tag{48}$$

Up to this point, no i.i.d. assumption has been imposed. Now we concentrate on the event that, for each $j \in \{1, \dots, d\}$, the i.i.d. sample of the jth coordinate $X_{1j}, \dots, X_{nj}$ do not coincide. This event is independent of the choice of $\mathbf{u}$, and has probability one by the i.i.d. and continuity assumptions. Because the coordinate random variables $X_{1j}, \dots, X_{nj}$ do not coincide, regardless of the value of $\mathbb{F}_{nj}^{-}(u_j)$, we have

$$\sum_{i=1}^{n} \mathbb{1}\left\{X_{ij} = \mathbb{F}_{nj}^{-}(u_j)\right\} \le 1,$$

and hence we conclude from (48) that

$$\left|\mathbb{C}_n(\mathbf{u}) - \bar{\mathbb{C}}_n(\mathbf{u})\right| \le \frac{d}{n}. \tag{49}$$

Because our choice of $\mathbf{u} \in [0, 1]^d$ is arbitrary, we conclude that Inequality (49) still holds after taking the supremum of its left-hand side over $\mathbf{u} \in [0, 1]^d$. That Inequality (3) holds with probability one then follows.

C.2. The non-i.i.d. case concerning Corollary 6

As in Corollary 6, we let $\mathbb{H}_n$ be the empirical distribution function based on pseudo-observations $\mathbf{U}_1, \dots, \mathbf{U}_n$ with $\mathbf{U}_i = (F_1(X_{i1}), \dots, F_d(X_{id}))$. We assume that the standard empirical process $\mathbb{U}_n = \sqrt{n}(\mathbb{H}_n - C)$ converges weakly in $\ell^\infty([0, 1]^d)$ to a Gaussian limit, which is the condition imposed in the second sentence of Corollary 6 and which implies that the limit process is continuous. As in Corollary 6, here we allow for a non-i.i.d. sample $\mathbf{X}_1, \dots, \mathbf{X}_n$.

If again the event on which for each $j \in \{1, \dots, d\}$, the coordinate random variables $X_{1j}, \dots, X_{nj}$ do not coincide has probability one, then our proof for the i.i.d. case in Section C.1 still applies, and the almost sure bound established in (49) over $\mathbf{u} \in [0, 1]^d$ (which leads to (3)) is stronger than that in (51), the conclusion of this section. However, without the i.i.d. assumption, the standard empirical process $\mathbb{U}_n$ may converge weakly even if ties occur; for instance, this is the case for the random repetition process, which satisfies a beta-mixing condition, described in Section 4.2 of [5]. To tackle the case concerning Corollary 6, we follow up Inequality (48) (recall that up to this point no i.i.d. assumption has been imposed) by adapting the rest of the proof of Lemma 7.2 in [5] for our case. We then have, for an arbitrary $\mathbf{u} \in [0, 1]^d$,

$$\left|\mathbb{C}_n(\mathbf{u}) - \bar{\mathbb{C}}_n(\mathbf{u})\right| \le \frac{d}{n} + \frac{d}{\sqrt{n}} \omega_n\left(\frac{1}{n}\right). \tag{50}$$

Here in the last line ω_n is the modulus of continuity of $\mathbb{U}_n$. Multiplying (50) by $\sqrt{n}$ and taking the supremum over $\mathbf{u} \in [0,1]^d$, we obtain

$$\sup_{\mathbf{u}\in[0,1]^d} \left|\sqrt{n}(\mathbb{C}_n - C)(\mathbf{u}) - \sqrt{n}(\bar{\mathbb{C}}_n - C)(\mathbf{u})\right| \le \frac{d}{\sqrt{n}} + d \cdot \omega_n\left(\frac{1}{n}\right).$$

By the fact that $\mathbb{U}_n$ converges weakly in $\ell^\infty([0,1]^d)$ to a continuous limit, $\omega_n(1/n) = o_p(1)$. Hence, we conclude that

$$\sup_{\mathbf{u}\in[0,1]^d} \left|\sqrt{n}(\mathbb{C}_n - C)(\mathbf{u}) - \sqrt{n}(\bar{\mathbb{C}}_n - C)(\mathbf{u})\right| = o_p(1). \tag{51}$$

C.3. The bootstrap empirical copula process from an i.i.d. sample

We recall the bootstrap setting described in Section 3.3. In particular, we recall the bootstrap ordinary empirical copula function $\mathbb{C}_n^*$ given in (23) and the bootstrap empirical copula function $\bar{\mathbb{C}}_n^*$ given in (24). Translating (48) to the bootstrap case and taking the supremum over $\mathbf{u} \in [0,1]^d$, we obtain

$$\begin{aligned}\sup_{\mathbf{u}\in[0,1]^d} \left|\mathbb{C}_n^*(\mathbf{u}) - \bar{\mathbb{C}}_n^*(\mathbf{u})\right| &\le \sup_{\mathbf{u}\in[0,1]^d} \sum_{j=1}^d \frac{1}{n} \sum_{i=1}^n \mathbb{1}\left\{X_{ij}^* = \mathbb{F}_{nj}^{*-}(u_j)\right\} \\ &\le \sum_{j=1}^d \frac{1}{n} \sup_{t\in[0,1]} \sum_{i=1}^n \mathbb{1}\left\{X_{ij}^* = \mathbb{F}_{nj}^{*-}(t)\right\}.\end{aligned}$$

As t varies over the interval $[0,1]$, the generalized inverse function $\mathbb{F}_{nj}^{*-}(t)$ can only take one of the at most n distinct values of the bootstrap sample $X_{1j}^*, \dots, X_{nj}^*$. For a particular $t \in [0,1]$, the summation $\sum_{i=1}^n \mathbb{1}\{X_{ij}^* = \mathbb{F}_{nj}^{*-}(t)\}$ counts the number of times the value $\mathbb{F}_{nj}^{*-}(t)$ is picked from the original i.i.d. sample of the jth coordinate $X_{1j}, \dots, X_{nj}$. Now we again concentrate on the event with probability one that, for each $j \in \{1, \dots, d\}$, the i.i.d. sample $X_{1j}, \dots, X_{nj}$ do not coincide. Then, the random variable $W_{nj} = \sup_{t\in[0,1]} \sum_{i=1}^n \mathbb{1}\{X_{ij}^* = \mathbb{F}_{nj}^{*-}(t)\}$ follows the distribution of the maximum among all coordinates of a n-dimensional multinomial random vector with parameters n and $(1/n, \dots, 1/n)$. Summing W_{nj} over $j \in \{1, \dots, d\}$, we conclude that

$$\sup_{\mathbf{u}\in[0,1]^d} \left|\sqrt{n}\left(\mathbb{C}_n^* - \bar{\mathbb{C}}_n^*\right)(\mathbf{u})\right| \le \sum_{j=1}^d \frac{1}{n} W_{nj} = o_p(1). \tag{52}$$

Acknowledgement

The research of Wegkamp is supported in part by the NSF DMS-13-10119 Grant. The authors sincerely thank the referees and the Associate Editor for their many constructive remarks.

References

[1] Aistleitner, C. and Dick, J. (2015). Functions of bounded variation, signed measures, and a general Koksma–Hlawka inequality. *Acta Arith*. **167** 143–171. MR3312093

[2] Andrews, D.W.K. and Pollard, D.P. (1994). An introduction to functional central limit theorems for dependent stochastic processes. *Int*. *Stat*. *Rev*. **62** 119–132.

[3] Arcones, M.A. and Yu, B. (1994). Central limit theorems for empirical and U-processes of stationary mixing sequences. *J*. *Theoret*. *Probab*. **7** 47–71. MR1256391

[4] Berghaus, B., Bücher, A. and Volgushev, S. (2016). Weak convergence of the empirical copula process with respect to weighted metrics. *Bernoulli*. To appear. Available at arXiv:1411.5888.

[5] Bücher, A. and Segers, J. (2014). Extreme value copula estimation based on block maxima of a multivariate stationary time series. *Extremes* **17** 495–528. MR3252823

[6] Bücher, A., Segers, J. and Volgushev, S. (2014). When uniform weak convergence fails: Empirical processes for dependence functions and residuals via epi- and hypographs. *Ann*. *Statist*. **42** 1598–1634. MR3262462

[7] Bücher, A. and Volgushev, S. (2013). Empirical and sequential empirical copula processes under serial dependence. *J*. *Multivariate Anal*. **119** 61–70. MR3061415

[8] Dehling, H., Durieu, O. and Tusche, M. (2014). Approximating class approach for empirical processes of dependent sequences indexed by functions. *Bernoulli* **20** 1372–1403. MR3217447

[9] Dehling, H., Durieu, O. and Volny, D. (2009). New techniques for empirical processes of dependent data. *Stochastic Process*. *Appl*. **119** 3699–3718. MR2568292

[10] Doukhan, P., Massart, P. and Rio, E. (1995). Invariance principles for absolutely regular empirical processes. *Ann*. *Inst*. *Henri Poincaré B*, *Probab*. *Stat*. **30** 63–82.

[11] Doukhan, P. and Surgailis, D. (1998). Functional central limit theorem for the empirical process of short memory linear processes. *C*. *R*. *Acad*. *Sci*. *Paris Sér*. *I Math*. **326** 87–92. MR1649521

[12] Fermanian, J.-D. (1998). Contributions à l'Analyse Nonparamétrique des Fonctions de Hasard sur Données Multivariées et Censurées. Ph.D. thesis, Univ. Paris VI.

[13] Fermanian, J.-D., Radulović, D. and Wegkamp, M. (2004). Weak convergence of empirical copula processes. *Bernoulli* **10** 847–860. MR2093613

[14] Fermanian, J.-D., Radulović, D. and Wegkamp, M. (2015). Asymptotic total variation tests for copulas. *Bernoulli* **21** 1911–1945. MR3352066

[15] Gaenssler, P. and Stute, W. (1987). *Seminar on Empirical Processes*. *DMV Seminar* **9**. Basel: Birkhäuser. MR0902803

[16] Hardy, G.H. (1905). On double Fourier series, and especially those which represent the double zeta-function with real and incommensurable parameters. *Q*. *J*. *Math*. **37** 53–79.

[17] Krause, M. (1903). Über Fouriersche Reihen mit zwei veränderlichen Grössen. *Leipziger Ber*. **55** 164–197.

[18] Krause, M. (1903). Über mittelwertsätze im gebiete der doppelsummen and doppelintegrale. *Leipziger Ber*. **55** 239–263.

[19] Leonov, A.S. (1998). Remarks on the total variation of functions of several variables and on a multidimensional analogue of Helly's choice principle. *Mathematical Notes* **63** 61–71.

[20] Nelsen, R.B. (1999). *An Introduction to Copulas*. *Lecture Notes in Statistics* **139**. New York: Springer. MR1653203

[21] Owen, A.B. (2005). Multidimensional variation for quasi-Monte Carlo. In *Contemporary Multivariate Analysis and Design of Experiments*. *Ser*. *Biostat*. **2** 49–74. Hackensack, NJ: World Sci. Publ. MR2271076

[22] Rio, E. (2000). *Théorie asymptotique des processus aléatoires faiblement dépendants*. *Mathématiques* & *Applications* (*Berlin*) [*Mathematics* & *Applications*] **31**. Berlin: Springer. MR2117923

[23] Rüschendorf, L. (1976). Asymptotic distributions of multivariate rank order statistics. *Ann. Statist.* **4** 912–923. MR0420794
[24] Ruymgaart, F.H. (1974). Asymptotic normality of nonparametric tests for independence. *Ann. Statist.* **2** 892–910. MR0386140
[25] Ruymgaart, F.H., Shorack, G.R. and van Zwet, W.R. (1972). Asymptotic normality of nonparametric tests for independence. *Ann. Math. Stat.* **43** 1122–1135. MR0339397
[26] Saks, S. (1937). *Theory of the Integral*. New York: Hafner.
[27] Segers, J. (2012). Asymptotics of empirical copula processes under non-restrictive smoothness assumptions. *Bernoulli* **18** 764–782. MR2948900
[28] Stute, W. (1984). The oscillation behavior of empirical processes: The multivariate case. *Ann. Probab.* **12** 361–379. MR0735843
[29] Tsukahara, H. (2005). Semiparametric estimation in copula models. *Canad. J. Statist.* **33** 357–375. MR2193980
[30] van de Geer, S.A. (2000). *Empirical Processes in M-Estimation*. Cambridge: Cambridge Univ. Press.
[31] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes*. New York: Springer. MR1385671
[32] van der Vaart, A.W. and Wellner, J.A. (2007). Empirical processes indexed by estimated functions. In *Asymptotics: Particles, Processes and Inverse Problems. Institute of Mathematical Statistics Lecture Notes – Monograph Series* 234–252. Beachwood, OH: IMS.

Received June 2015 and revised March 2016