WHEN UNIFORM WEAK CONVERGENCE FAILS: EMPIRICAL PROCESSES FOR DEPENDENCE FUNCTIONS VIA EPI- AND HYPOGRAPHS[1]

By Axel Bücher, Johan Segers and Stanislav Volgushev

Ruhr-Universität Bochum and Université catholique de Louvain

For copulas whose partial derivatives are not continuous everywhere on the interior of the unit cube, the empirical copula process does not converge weakly with respect to the supremum distance. This makes it hard to verify asymptotic properties of inference procedures for such copulas. To resolve the issue, a new metric for locally bounded functions is introduced and the corresponding weak convergence theory is developed. Convergence with respect to the new metric is related to epi- and hypoconvergence and is weaker than uniform convergence. Still, for continuous limits, it is equivalent to locally uniform convergence, whereas under mild side conditions, it implies L^p convergence. Even in cases where uniform convergence fails, weak convergence with respect to the new metric is established for empirical copula and tail dependence processes. No additional assumptions are needed for tail dependence functions, and for copulas, the assumptions reduce to existence and continuity of the partial derivatives almost everywhere on the unit cube. The results are applied to obtain asymptotic properties of minimum distance estimators, goodness-of-fit tests and resampling procedures.

1. Introduction. The theory of Hoffman-Jørgensen weak convergence in the space of bounded functions is a great success story in mathematical statistics (van der Vaart and Wellner, 1996; Kosorok, 2008). Measurability assumptions are reduced to a minimum, no smoothness assumptions on the trajectories are needed, it applies in a vast variety of circumstances, and the topology of uniform convergence is fine enough so that, through the continuous mapping theorem and functional delta method, it implies weak convergence of a countless list of interesting statistical functionals.

But precisely because of the strength of uniform convergence, there are circumstances where it does not hold. These arise when the candidate limit

[1]Supported by the collaborative research center "Statistical modeling of nonlinear dynamic processes" (SFB 823) of the German Research Foundation (DFG), by IAP research network grant No. P7/06 of the Belgian government (Belgian Science Policy) and by contract "Projet d'Actions de Recherche Concertées" No. 12/17-045 of the "Communauté française de Belgique".

AMS 2000 subject classifications: Primary 60F05, 62G30; secondary 62G32, 62M09

Keywords and phrases: bootstrap, copula, epigraph, hypograph, stable tail dependence function, minimum distance estimation, weak convergence

process (in the sense of convergence of finite-dimensional distributions) has discontinuous trajectories. For uniform convergence to take place, the locations of the discontinuities must be matched exactly. For the empirical distribution function based on a random sample from a discrete distribution, this holds. However, in other cases, it does not. Think of empirical distributions based on residuals of some sort, that is, observations that are themselves approximations of some latent random variables. Because of the measurement error in the ordinates, jump locations fail to be located exactly, and uniform convergence fails.

The two examples which interest us in this paper concern empirical copula processes and empirical tail dependence function processes. The latent variables arise through the probability integral transform, mapping the observable random variables to a uniform scale by means of the marginal cumulative distribution functions. The latter being unknown, they have to be replaced by estimators, often the marginal empirical distribution functions. The resulting estimator of the copula or tail dependence function is rank-based, and the candidate limit process involves the first-order partial derivatives of the object to be estimated, reflecting the uncertainty in the ordinates. If these are not continuous on a sufficiently large set, weak convergence with respect to the supremum distance fails (Fermanian, Radulović and Wegkamp, 2004; Einmahl, Krajina and Segers, 2012). A bivariate empirical copula process of which the finite-dimensional distributions converge but that does not converge in the uniform topology is worked out in Section 4. More generally, the phenomenon occurs for max-stable dependence structures based on discrete spectral measures, including Marshall–Olkin copulas, for Archimedean copulas with nonsmooth generators, and for mixture models involving such non-smooth components.

Apart from ignoring the issue, dismissing it as unimportant or pathological, there are three solutions to establish weak convergence in such cases:

1. Bypass the empirical process step and prove weak convergence of the statistic of interest in a direct way.
2. Restrict the domain of the empirical process to a subset where the trajectories of the candidate limit process are continuous.
3. Establish weak convergence of the empirical process with respect to a weaker, but still strong enough metric.

If it works, the first alternative is usually the easiest: for instance, rather than writing Spearman's rho as an integral of the empirical copula, it can be represented more simply as a U-statistic. In a more elaborative variation of this theme, asymptotic normality of integrals of the empirical tail dependence function is established in Einmahl, Krajina and Segers (2012) in situations where the underlying empirical process does not converge. However, the disadvantage is that the unifying power of empirical process theory is lost and that each problem has to be dealt with in an *ad hoc* way.

Cutting out the problem points is natural and leads in Genest, Nešlehová and Rémillard (2013) to an intriguing convergence result for empirical copula processes based on count data. However, in order to yield weak convergence of statistics (suprema, integrals), additional information on the behaviour on the left-out regions out is needed, guaranteeing the asymptotic negligibility of the contributions of those regions. The power and simplicity of use of the empirical process result is affected.

For sure, changing the metric and maybe also the underlying function space is the most radical approach. It has the disadvantage of steering away from the familiar framework of uniform convergence of bounded functions. The choice of the new mode of convergence is not obvious. It should be weak enough so that convergence does take place, and still strong enough to enable statistical applications. Ideally, when the limit process has continuous trajectories, it should be equivalent to uniform convergence, so that in standard situations, nothing is lost.

Constructing a new metric was the approach of Skorohod (1956), who introduced a number of metrics on the space of càdlàg functions on $[0, 1]$, the J_1 topology being the most famous one; see also Billingsley (1968) and Whitt (2002). A multivariate extension of the J_1 topology was considered in Neuhaus (1971) and Bickel and Wichura (1971). For our purposes, there are two drawbacks to this extension. First, the topology is restricted to the space of functions that are continuous when points are approached from a certain quadrant and that allow limits when points are approached from the other quadrants. Second, two functions are close in the J_1 topology only if, up to a small perturbation in the coordinates, their supremum distance is small. This implies that their jumps, if any, must match both in location and in magnitude. The latter objection also applies to the Skorohod-type distances introduced in Straf (1972) and Bass and Pyke (1985). Among other things, a sequence of continuous functions cannot converge to a discontinuous function.

A different family of topologies finds its origins in variational analysis and optimization theory, where one seeks to approximate a complicated optimization problem by a simpler one. The metric with which to measure the distance between two functions should be such that the optimum and the set of optimizers of the approximating function is close to the one of the true objective function. In the context of minimization, one identifies a real function f on some nice metric space $\mathbb{T}$ with its *epigraph*, which is the set of all points (x, y) in $\mathbb{T} \times \mathbb{R}$ such that $f(x) \le y$. The epigraph is closed if and only if the function is lower semicontinuous. *Epi-convergence* of functions is then defined as convergence of their epigraphs with respect to an appropriate topology on the family of (closed) subsets of $\mathbb{T} \times \mathbb{R}$. For maximization problems, *hypographs* and *hypo-convergence* are defined in the same way, the inequality sign pointing in the other direction. Most commonly, set convergence is considered with respect to the *Fell hit-and-miss topology*, which, for locally compact, separable spaces, is equivalent to *Painlevé–Kuratowski convergence* of sets (Beer, 1993; Rockafellar and Wets, 1998; Molchanov, 2005). On compact, closed spaces, the latter reduces to convergence in the Hausdorff metric.

It is then tempting to define a mode of convergence by requiring a sequence of functions f_n to both epi- and hypo-converge to a limit function f. However, the limit function f should then be both lower and upper semicontinuous, hence continuous, and in fact, as we will see later, in this case we would be back to (locally) uniform convergence.

Proceeding more carefully, we will require f_n to epi-converge to the lower semicontinuous hull of f and to hypo-converge to the upper semicontinuous hull of f. In topological terms, the epigraphs of f_n should converge to the closure of the epigraph of f, and the hypographs of f_n should converge to the closure of the hypograph of f. Being a combination of both epi- and hypo-convergence, we coin it *hypi*-convergence. It is to be distinguished with epi/hypo-convergence, a different concept arising in connection with saddle points (Attouch and Wets, 1983). Hypi-convergence of functions is similar as but not identical to set convergence of the intersections of the closures of their epi- and hypographs, the latter being considered in Vervaat (1981).

If the domain $\mathbb{T}$ is locally compact and separable, hypi-convergence is metrizable. Since different functions sharing the same lower and upper semicontinuous hulls cannot be distinguished from another, the hypi-metric is in fact only a semimetric. In order to apply weak convergence theoy for metric spaces, we are then led to work with equivalence classes. This construction is akin to the one of L^p metric spaces, whose elements, at least formally, are equivalences classes of functions that coincide almost everywhere.

Broadly speaking, hypi-convergence is intermediate between uniform convergence and L^p convergence. Hypi-convergence implies uniform convergence on compact subsets of the domain at every point of which the limit function is continuous. Hence, for continuous limits, we are back to uniform convergence. But even without continuity, hypi-convergence implies convergence of global extrema. Moreover, for limit functions which are continuous almost everywhere, hypi-convergence implies L^p convergence on compact sets.

In a similar way as one can consider weak epi-convergence of random lower semicontinuous functions (Geyer, 1994; Molchanov, 2005), we develop Hoffman-Jørgensen weak convergence theory with respect to the hypi-(semi)metric. Thanks to an extension of the continuous mapping theorem for semimetric spaces, we are able to leverage the above properties of hypi-convergence to yield weak convergence of finite-dimensional distributions, Kolmogorov–Smirnov type statistics, and procedures related to L^p spaces, notably minimum distance estimators and Cramér–von Mises statistics.

We investigate weak convergence properties of empirical copula processes and empirical tail dependence functions with respect to the hypi-semimetric. Weak hypi-convergence of the empirical copula process is established for copulas whose partial derivatives exist and are continuous everywhere except for an arbitrary Lebesgue null set of the unit cube. This contains the theory of weak convergence in the supremum distance under stronger smoothness conditions as a special case. Additionally, the bootstrap is shown to be consistent. To demonstrate the usefulness, we extend results on minimum distance estimation (Tsukahara, 2005) and on power curves for goodness-of-fit tests under local alternatives (Genest, Quessy and Rémillard, 2007). Similar results are shown for tail dependence functions, extending Bücher and Dette (2011) and Einmahl, Krajina and Segers (2012).

The structure of the paper is as follows. The hypi-topology is introduced in Section 2. Weak convergence in hypi-space is the topic of Section 3. We give tools how to check weak hypi-convergence, including a variant on Slutsky's lemma, and how to derive statistically meaningful results from it through the continuous mapping theorem. The new framework is applied for empirical copula processes in Section 4 and for empirical tail dependence function processes in Section 5. In order to preserve the flow of the text, a number of auxiliary results and all proofs are deferred to a sequence of appendices.

2. Hypi-convergence of locally bounded functions.

We introduce a mode of convergence for real-valued, locally bounded functions on a locally compact, separable metric space (subsection 2.1). For continuous limits, the metric is equivalent to locally uniform convergence, but for discontinuous limits, it is strictly weaker, while still implying L^p convergence (subsection 2.2). The proofs for the results in this section are given in Appendix B.1.

2.1. *The hypi-semimetric.*

Let $(\mathbb{T}, d)$ be a locally compact, separable metric space. The space $\mathbb{T} \times \mathbb{R}$ is a locally compact, separable metric space, too, when equipped for instance with the metric $d_{\mathbb{T}\times\mathbb{R}}((x_1, y_1), (x_2, y_2)) = \max\{d(x_1, x_2), |y_1 - y_2|\}$.

Let $\ell^\infty_{\mathrm{loc}}(\mathbb{T})$ denote the space of locally bounded functions $f : \mathbb{T} \to \mathbb{R}$, that is, functions that are uniformly bounded on compacta. If $\mathbb{T}$ is itself compact, we will simply write $\ell^\infty(\mathbb{T})$. Functions $f \in \ell^\infty_{\mathrm{loc}}(\mathbb{T})$ will be identified with subsets of $\mathbb{T} \times \mathbb{R}$ by considering their epigraphs and hypographs:

$$\begin{aligned} \operatorname{epi} f &= \{(x, y) \in \mathbb{T} \times \mathbb{R} : f(x) \le y\}, \\ \operatorname{hypo} f &= \{(x, y) \in \mathbb{T} \times \mathbb{R} : y \le f(x)\}. \end{aligned}$$

Except for being locally bounded, functions f in $\ell^\infty_{\mathrm{loc}}(\mathbb{T})$ can be arbitrarily rough. A minimal amount of regularity will come from the lower and upper semicontinuous hulls $f_\wedge \le f \le f_\vee$:

$$f_\wedge(x) = \sup_{\varepsilon>0} \inf\{f(x') : d(x', x) < \varepsilon\}, \tag{2.1}$$

$$f_\vee(x) = \inf_{\varepsilon>0} \sup\{f(x') : d(x', x) < \varepsilon\}, \tag{2.2}$$

functions which are elements of $\ell^\infty_{\mathrm{loc}}(\mathbb{T})$, too. A convenient link between epi- and hypographs on the one hand and lower and upper semicontinuous hulls on the other hand is that

$$\operatorname{cl}(\operatorname{epi} f) = \operatorname{epi} f_\wedge, \qquad \operatorname{cl}(\operatorname{hypo} f) = \operatorname{hypo} f_\vee,$$

where 'cl' denotes topological closure, in this case, in the space $\mathbb{T} \times \mathbb{R}$. In particular, a function f is lower (upper) semicontinuous if and only if its epigraph (hypograph) is closed.

Functions being identified with sets, notions of set convergence can be applied to define convergence of functions. We rely on classical theory exposed in, among others, Matheron (1975), Beer (1993), Rockafellar and Wets (1998) and Molchanov (2005). A standard topology on the space of closed subsets of a topological space is the Fell hit-and-miss topology. If the underlying space is locally compact and separable, as in our case, then the Fell topology is metrizable. Moreover, in that case, convergence of a sequence of closed sets in the Fell topology is equivalent to their Painlevé–Kuratowski convergence. Recall that (not necessarily closed) sets A_n of a topological space converge to a set A in the Painlevé–Kuratowski sense if and only if (i) for every $x \in A$ there exists a sequence x_n with $x_n \in A_n$ such that $x_n \to x$ and (ii) whenever $x_{n_k} \in A_{n_k}$ for some subsequence n_k converges to a limit x, we must have $x \in A$. The limit set A is necessarily closed, and Painlevé–Kuratowski convergence of A_n to A is equivalent to Painlevé–Kuratowski convergence of $\operatorname{cl}(A_n)$ to A. If the underlying space is compact, then Painlevé–Kuratowski convergence of closed sets is equivalent to their convergence in the Hausdorff metric.

Let $\mathcal{F}(\mathbb{T} \times \mathbb{R})$ be the space of closed subsets of $\mathbb{T} \times \mathbb{R}$ and let $d_{\mathcal{F}}$ be a metric inducing the Fell topology, or equivalently, Painlevé–Kuratowski convergence. Examples of metrics $d_{\mathcal{F}}$ for the Fell topology are to be found in Rockafellar and Wets (1998), Molchanov (2005) and Ogura (2007). A versatile notion of convergence of functions in optimization theory is epi-convergence: a sequence of functions $f_n : \mathbb{T} \to \mathbb{R}$ is said to epi-converge to a function f if and only if the Painlevé–Kuratowski limit of $\operatorname{epi} f_n$ (or equivalently, its closure) is equal to $\operatorname{epi} f$, that is, if $d_{\mathcal{F}}(\operatorname{cl}(\operatorname{epi} f_n), \operatorname{cl}(\operatorname{epi} f)) \to 0$ as $n \to \infty$. Necessarily, the limit set $\operatorname{epi} f$ is closed, and therefore, f must be lower semicontinuous. Similarly, hypo-convergence of functions is defined as Painlevé–Kuratowski convergence of their hypographs (or their closures), and the limit function is necessarily upper semicontinuous.

If f_n both epi-converges to $f_\wedge$ and hypo-converges to $f_\vee$, then we say that f_n *hypi-converges* to f. This mode of convergence is the one that we propose in this paper. According to the following result, hypi-convergence is metrizable and can be checked conveniently by pointwise criteria.

PROPOSITION 2.1 (Hypi-convergence). *Let $f_n, f \in \ell^\infty_{\mathrm{loc}}(\mathbb{T})$. The following statements are equivalent:*

(i) f_n epi-converges to $f_\wedge$ and hypo-converges to $f_\vee$.
(ii) The following pointwise criteria hold:

$$\begin{cases} \forall x \in \mathbb{T} : \forall x_n \to x : f_\wedge(x) \le \liminf_{n\to\infty} f_n(x_n), \\ \forall x \in \mathbb{T} : \exists x_n \to x : \limsup_{n\to\infty} f_n(x_n) \le f_\wedge(x). \end{cases} \tag{2.3}$$

and

$$\begin{cases} \forall x \in \mathbb{T} : \forall x_n \to x : \limsup_{n\to\infty} f_n(x_n) \le f_\vee(x), \\ \forall x \in \mathbb{T} : \exists x_n \to x : f_\vee(x) \le \liminf_{n\to\infty} f_n(x_n). \end{cases} \tag{2.4}$$

(iii) The distance $d_{\mathrm{hypi}}(f_n, f)$ converges to 0, where d_{hypi} denotes the hypi-semimetric *defined as*

$$d_{\mathrm{hypi}}(f, g) = \max\{d_{\mathcal{F}}(\operatorname{epi} f_\wedge, \operatorname{epi} g_\wedge), d_{\mathcal{F}}(\operatorname{hypo} f_\vee, \operatorname{hypo} g_\vee)\}.$$

(iv) f_n converges to f in the hypi-topology*, which is defined as the coarsest topology on $\ell^\infty_{\mathrm{loc}}(\mathbb{T})$ for which the map*

$$\ell^\infty_{\mathrm{loc}}(\mathbb{T}) \mapsto \mathcal{F}(\mathbb{T} \times \mathbb{R}) \times \mathcal{F}(\mathbb{T} \times \mathbb{R}) : f \mapsto \big(\operatorname{cl}(\operatorname{epi} f), \operatorname{cl}(\operatorname{hypo} f)\big)$$

is continuous, that is, the hypi-open sets in $\ell^\infty_{\mathrm{loc}}(\mathbb{T})$ are the inverse images of open sets in $\mathcal{F}(\mathbb{T} \times \mathbb{R}) \times \mathcal{F}(\mathbb{T} \times \mathbb{R})$.

Note that in (2.3) and (2.4), we can replace f_n by $f_{n,\wedge}$ and $f_{n,\vee}$, respectively. The equivalence of (i) and (ii) follows from well-known pointwise criteria for epi- and hypo-convergence (Molchanov, 2005, Chapter 5, Proposition 3.2(ii)). Statements (iii) and (iv) are just reformulations of what it means to have both epi- and hypo-convergence in (i).

Intuitively, two functions are close in the hypi-semimetric if both their epigraphs and their hypographs are close. Two functions on $\mathbb{T} = [0, 1]$ whose epigraphs are close but whose hypographs are far away are depicted in the upper part of Figure 1: for instance, the point $(0.5, 1)$ belongs to the hypograph of the dotted-line function but is far away from any point in the hypograph of the solid-line function. As a consequence, these two functions are not close in the hypi-semimetric. For comparison, two functions that are close in the hypi-semimetric are depicted in the lower part of Figure 1.

By Proposition 2.1(ii), if f_n hypi-converges to f and if f is continuous at x, then $f_n(x_n) \to f(x)$ whenever $x_n \to x$. The pointwise criteria can be rephrased as follows: for every $x \in \mathbb{T}$, the sequence f_n epi-converges at x to $f_\wedge(x)$ and hypo-converges at x to $f_\vee(x)$, in the sense of equations (A.1) and (A.2) in Appendix A.1.2.

Moreover, it follows that locally uniform convergence of locally bounded functions implies their hypi-convergence. The converse is not true if the hypi-limit is not continuous.

Hypi-convergence of sequences f_n and g_n to f and g respectively does in general not imply hypi-convergence of the sequence of sums, $f_n + g_n$, to the sum of the limits, $f + g$. For instance, let x_n converge to x in $\mathbb{T}$ with $x_n \neq x$ and set $f_n = \mathbb{1}_{\{x_n\}}$ and $g_n = -\mathbb{1}_{\{x\}}$. Still, a sufficient side condition is that at least one of the limit functions is continuous; see Lemma A.4 for a more general result.

2.2. *Leveraging hypi-convergence.* As mentioned already, uniform convergence implies hypi-convergence but not conversely. Nevertheless, at subsets of the domain where the limit function is continuous, the converse does hold. In this sense, working in hypi-space does not necessarily yield weaker results than in the uniform topology.

PROPOSITION 2.2. *Let $K \subset \mathbb{T}$ be compact and let $f \in \ell^\infty_{\mathrm{loc}}(\mathbb{T})$ be continuous in every $x \in K$. If f_n hypi-converges to f in $\ell^\infty_{\mathrm{loc}}(\mathbb{T})$, then $\sup_{x \in K} |f_n(x) - f(x)| \to 0$ as $n \to \infty$.*

Being a combination of epi- and hypo-convergence, hypi-convergence preserves convergence of extrema. Later, we will make use of this property when investigating Kolmogorov–Smirnov type test statistics.

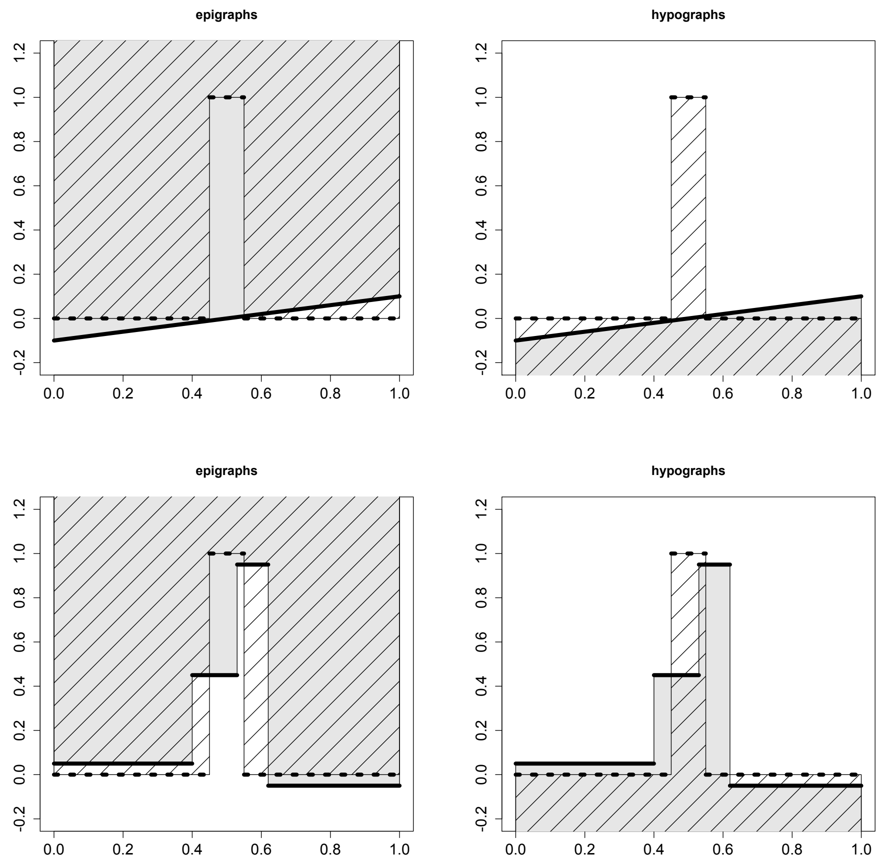

Fig 1. *Top: two functions whose epigraphs are close but whose hypographs are far away. Bottom: two functions of which both the epi- and hypographs are close. The light gray areas represent epigraphs (left column) and hypographs (right column) of the functions depicted by solid lines, whereas shaded areas represent epigraphs (left column) and hypographs (right column) of the functions depicted by dotted lines.*

Proposition 2.3. *Let $G \subset \mathbb{T}$ be an open subset with compact closure. If f_n hypi-converges to f in $\ell^\infty(\mathbb{T})$ and if f is continuous on the boundary of G, then $\inf f_n(G) \to \inf f(G)$ and $\sup f_n(G) \to \sup f(G)$ as $n \to \infty$. If $G = \mathbb{T}$ is compact, then the same conclusions hold true without imposing continuity of f.*

Hypi-convergence implies L^p-convergence for finite p, provided that the limit function is not too rough. Applications later on in the paper concern minimum distance estimators and Cramér–von Mises statistics. Recall that upper and lower semicontinuous functions are necessarily Borel measurable.

PROPOSITION 2.4. *Let μ be a finite Borel measure supported on a compact subset of $\mathbb{T}$. If f_n hypi-converges to f in $\ell^\infty(\mathbb{T})$ and if f is continuous μ-almost everywhere, then, for every $p \in [1, \infty)$, we have $\int |f_{n,\vee} - f_{n,\wedge}|^p \,\mathrm{d}\mu \to 0$ and $\int |f_n^* - f^*|^p \,\mathrm{d}\mu \to 0$ as $n \to \infty$, where f_n^* and f^* represent arbitrary Borel measurable functions on $\mathbb{T}$ such that $f_{n,\wedge} \le f_n^* \le f_{n,\vee}$ and $f_\wedge \le f^* \le f_\vee$.*

3. Weak hypi-convergence of stochastic processes. When applying Hoffman-Jørgensen weak convergence theory, it is customary to work in a metric space. However, d_{hypi} is a semimetric and not a metric: if functions $f, g \in \ell^\infty_{\mathrm{loc}}(\mathbb{T})$ share the same lower and upper semicontinuous hulls, then $d_{\mathrm{hypi}}(f, g) = 0$ even if f and g are different functions.

To obtain a metric space, we consider equivalences classes of functions at hypi-distance zero. For $f \in \ell^\infty_{\mathrm{loc}}(\mathbb{T})$, let $[f]$ be the set of all $g \in \ell^\infty_{\mathrm{loc}}(\mathbb{T})$ such that $d_{\mathrm{hypi}}(f, g) = 0$. Let $L^\infty_{\mathrm{loc}}(\mathbb{T})$ be the space of all such equivalence classes. Then $L^\infty_{\mathrm{loc}}(\mathbb{T})$ becomes a metric space when equipped with the hypi-metric (abusing notation) $d_{\mathrm{hypi}}([f], [g]) := d_{\mathrm{hypi}}(f, g)$. The map $[\,\cdot\,]$ from $\ell^\infty_{\mathrm{loc}}(\mathbb{T})$ into $L^\infty_{\mathrm{loc}}(\mathbb{T})$ sending f to $[f]$ is continuous and it sends open sets to open sets and closed sets to closed sets.

Let X_n and X be maps from probability spaces Ω_n and Ω respectively into $\ell^\infty_{\mathrm{loc}}(\mathbb{T})$. Assume that X is hypi Borel measurable, that is, measurable with respect to the σ-field generated by the hypi-open sets of $\ell^\infty_{\mathrm{loc}}(\mathbb{T})$. Then the map $[X] = [\,\cdot\,](X)$ into $L^\infty_{\mathrm{loc}}(\mathbb{T})$ is Borel measurable, too. Since $L^\infty_{\mathrm{loc}}(\mathbb{T})$ is a metric space, weak convergence theory as in van der Vaart and Wellner (1996) applies: we say that X_n weakly hypi-converges to X in $\ell^\infty_{\mathrm{loc}}(\mathbb{T})$ if and only if $[X_n] \rightsquigarrow [X]$ in $L^\infty_{\mathrm{loc}}(\mathbb{T})$. Simplifying notation, we often omit brackets and write $X_n \rightsquigarrow X$ in $L^\infty_{\mathrm{loc}}(\mathbb{T})$.

In order to prove weak hypi-convergence, we will always combine an initial result on weak convergence of some stochastic process, usually some empirical process and with respect to the supremum distance, with the (extended) continuous mapping theorem (van der Vaart and Wellner, 1996, Theorems 1.3.6 and 1.11.1). The task then consists of proving hypi-continuity of the relevant mappings into $\ell^\infty_{\mathrm{loc}}(\mathbb{T})$ on sufficiently large subsets of their domains. Two situations of particular importance concern convergence to a function in $\ell^\infty_{\mathrm{loc}}(\mathbb{T})$ that is defined as the upper or lower semicontinuous hull of some other function that is originally defined on a dense subset of $\mathbb{T}$ only (Appendix A.1.2) and convergence of sums (Appendix A.1.1), inducing a variant of Slutsky's lemma (Lemma A.10).

In subsection 2.2, convergence in the hypi-semimetric was seen to imply a number of other convergence relations, depending on features of the limit function. In combination with an extension of the continuous mapping theorem to semimetric spaces (Theorem A.9), these continuity properties allow to deduce weak convergence in more familiar spaces from weak hypi-convergence. In contrast, the functional delta method cannot be readily extended to weak hypi-convergence, since addition is not continuous with respect to the hypi-semimetric.

By Proposition 2.2, the map from $(\ell^\infty_{\text{loc}}(\mathbb{T}), d_{\text{hypi}})$ into $(\ell^\infty(K), \|\cdot\|_\infty)$ sending a function f to its restriction $f|_K$ on a compact subset K of $\mathbb{T}$ is continuous at every function f which is continuous in every point x of K; here $\|\cdot\|_\infty$ denotes the supremum norm. As a consequence, weak hypi-convergence implies weak convergence with respect to the supremum distance insofar the limit process has continuous trajectories. More precisely, we have the following result, proved in Appendix B.2.

Corollary 3.1. *Let X_n and X be maps from probability spaces Ω_n and Ω respectively into $\ell^\infty_{\text{loc}}(\mathbb{T})$ such that X is hypi Borel measurable. If $[X_n] \rightsquigarrow [X]$ in $L^\infty_{\text{loc}}(\mathbb{T})$ and if $K \subset \mathbb{T}$ is a nonempty, compact set such that, with probability one, X is continuous in every $x \in K$, then $X_n|_K \rightsquigarrow X|_K$ in $(\ell^\infty(K), \|\cdot\|_\infty)$.*

Taking K to be finite, we find that weak hypi-convergence implies weak convergence of finite-dimensional distributions at points where the limit process is continuous almost surely.

For finite Borel measures μ on $\mathbb{T}$ with compact support, Proposition 2.4 states the continuity of the embedding from the set of Borel measurable functions of $\ell^\infty_{\text{loc}}(\mathbb{T})$ equipped with the hypi-topology into $L^p(\mu)$, for every $1 \le p < \infty$. A technical nuisance is that in order to view $L^p(\mu)$ as a metric space, we have to consider equivalence classes of functions that are equal μ-almost everywhere; notation $[\,\cdot\,]_\mu$.

Corollary 3.2. *Let X_n and X be maps from probability spaces Ω_n and Ω respectively into $\ell^\infty_{\text{loc}}(\mathbb{T})$ such that X is hypi Borel measurable. Let μ be a finite Borel measure on $\mathbb{T}$ with compact support. If $[X_n] \rightsquigarrow [X]$ in $L^\infty_{\text{loc}}(\mathbb{T})$ and if X is μ-almost everywhere continuous with probability one, then $\int |X_{n,\vee} - X_{n,\wedge}|^p \,\mathrm{d}\mu$ converges to 0 in outer probability and both $[X_{n,\vee}]_\mu$ and $[X_{n,\wedge}]_\mu$ converge weakly in $L^p(\mu)$ to $[X_\vee]_\mu = [X_\wedge]_\mu$, for every $p \in [1, \infty)$.*

4. Empirical copula processes. Let $\boldsymbol{X_i} = (X_{i1}, \ldots, X_{id})$, with $i \in \mathbb{N}$, be a strictly stationary sequence of d-variate random vectors. (No confusion should arise from the use of the symbol 'd' for both the metric on $\mathbb{T}$ and the dimension of the random vectors.) Throughout this section, the joint distribution function F of $\boldsymbol{X_i}$ is assumed to have continuous marginal distributions $F_1, \ldots, F_d$ and its copula is denoted by C. Further, for $j = 1, \ldots, d$, let $U_{ij} = F_j(X_{ij})$ and set $\boldsymbol{U_i} = (U_{i1}, \ldots, U_{id})$. Note that $\boldsymbol{U}_i$ is distributed according to C. Consider the empirical distribution functions

$$F_n(\boldsymbol{x}) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{\boldsymbol{X}_i \le \boldsymbol{x}\}, \quad G_n(\boldsymbol{u}) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{\boldsymbol{U}_i \le \boldsymbol{u}\}$$

for $\boldsymbol{x} \in \mathbb{R}^d$ and $\boldsymbol{u} \in [0,1]^d$. For a distribution function H on the reals, let

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

denote the (left-continuous) generalized inverse function of H.

The object of interest is the empirical copula, defined by

$$C_n(\boldsymbol{u}) = F_n(F_{n1}^-(u_1), \ldots, F_{nd}^-(u_d)), \qquad \boldsymbol{u} \in [0,1]^d,$$

where F_{nj} denotes the jth marginal empirical distribution function. For convenience, we will abbreviate the notation for the empirical copula by $C_n(\boldsymbol{u}) = F_n(\boldsymbol{F}_n^-(\boldsymbol{u}))$, with $\boldsymbol{F}_n^-(\boldsymbol{u}) = (F_{n1}^-(u_1), \ldots, F_{nd}^-(u_d))$.

Often, the empirical copula is defined as the distribution function of the vector of rescaled ranks, and/or it is turned into a genuine copula via linear interpolation. Since these variants differ from the empirical copula by at most a deterministic $O(1/n)$ term, uniformly over $[0,1]^d$, they do not affect the asymptotic distribution of the empirical copula process, defined by

$$\mathbb{C}_n = \sqrt{n}(C_n - C).$$

The asymptotic behavior of $\mathbb{C}_n$, especially its weak convergence in the space $\ell^\infty([0,1]^d)$ equipped with the supremum norm $\|\cdot\|_\infty$, has been investigated by several authors under various conditions (Rüschendorf, 1976; Fermanian, Radulović and Wegkamp, 2004; Segers, 2012; Bücher and Volgushev, 2013).

The main arguments to derive the limit of $\mathbb{C}_n$ are as follows. For the sake of a clear explanation let us assume for the moment that the random vectors $(\boldsymbol{X}_i)_{i \in \mathbb{N}}$ form an i.i.d. sequence, even though the same arguments work for many time series models with short-range dependence. Observing that $C_n = F_n(\boldsymbol{F}_n^-) = G_n(\boldsymbol{G}_n^-)$, we can decompose $\mathbb{C}_n$ into two terms:

$$\mathbb{C}_n = \sqrt{n}\{G_n(\boldsymbol{G}_n^-) - C\} = \alpha_n(\boldsymbol{G}_n^-) + \sqrt{n}\{C(\boldsymbol{G}_n^-) - C\}, \tag{4.1}$$

where $\alpha_n = \sqrt{n}(G_n - C)$ denotes the usual empirical process associated to the sequence $(\boldsymbol{U}_i)_{i\in\mathbb{N}}$.

Deriving the limit of the first term in (4.1) is standard: since $\alpha_n \rightsquigarrow \alpha$ in $\ell^\infty([0,1]^d)$ with respect to the supremum norm, for a C-Brownian bridge α, and since $\sup_{0\le u_j\le 1}|G_{nj}^-(u_j) - u_j| = o_P(1)$, we obtain that $\alpha_n(\boldsymbol{G}_n^-) \rightsquigarrow \alpha$ in $(\ell^\infty([0,1]^d), \|\cdot\|_\infty)$, too.

Regarding the second term in (4.1), the argumentation is harder. Set $\beta_n = (\beta_{n1},\dots,\beta_{nd})$, where $\beta_{nj} = \sqrt{n}(G_{nj}^- - \mathrm{id}_{[0,1]})$ denotes the quantile process of the jth coordinate and where id_A is the identity map on a set A. It follows from the functional delta method applied to the inverse mapping $H \mapsto H^-$ that $\|\beta_{nj} + \alpha_{nj}\|_\infty = o_P(1)$, where $\alpha_{nj}(u_j) = \alpha_n(1,\dots,1,u_j,1,\dots,1)$, with $u_j \in [0,1]$ at the jth position. Therefore, $\beta_{nj} \rightsquigarrow -\alpha_j$ in $(\ell^\infty([0,1]), \|\cdot\|_\infty)$, where, similarly, $\alpha_j(u_j)$ is defined as $\alpha(1,\dots,1,u_j,1,\dots,1)$. Now,

$$\sqrt{n}\{C(\boldsymbol{G}_n^-) - C\} = \sqrt{n}\{C(\mathrm{id}_{[0,1]^d} + \beta_n/\sqrt{n}) - C\}, \tag{4.2}$$

which can be handled under suitable differentiability conditions on C. To conclude upon weak convergence with respect to the supremum distance, the weakest assumption so far has been stated in Segers (2012).

CONDITION 4.1. *For $j = 1,\dots,d$ the partial derivatives $\dot{C}_j(\boldsymbol{u})$ exist and are continuous on $\{\boldsymbol{u} \in [0,1]^d : u_j \in (0,1)\}$.*

Under Condition 4.1,

$$\sqrt{n}\{C(\boldsymbol{G}_n^-) - C\}(\boldsymbol{u}) \rightsquigarrow -\sum_{j=1}^d \dot{C}_j(\boldsymbol{u})\alpha_j(u_j), \tag{4.3}$$

where $\dot{C}_j(\boldsymbol{u})$ can be defined for instance as 0 if $u_j \in \{0,1\}$. Hence,

$$\mathbb{C}_n(\boldsymbol{u}) \rightsquigarrow \mathbb{C}(\boldsymbol{u}) = \alpha(\boldsymbol{u}) - \sum_{j=1}^d \dot{C}_j(\boldsymbol{u})\alpha_j(u_j) \tag{4.4}$$

in $\ell^\infty([0,1]^d)$ with respect to the supremum distance.

Condition 4.1 ensures that the limit process $\mathbb{C}$ in (4.4) has continuous trajectories. Actually, if $\mathbb{C}_n$ is to converge weakly with respect to the supremum distance, then the weak limit must have continuous trajectories with probability one. The reason is that the mapping

$$\Delta : \ell^\infty([0,1]^d) \to [0,\infty) : f \mapsto \sup_{\boldsymbol{u}\in[0,1]^d} |f_\vee(\boldsymbol{u}) - f_\wedge(\boldsymbol{u})|$$

is continuous with respect to $\|\cdot\|_\infty$ and that $0 \le \Delta\mathbb{C}_n \le d/\sqrt{n} \to 0$ almost surely. The expression for $\mathbb{C}$ in (4.4) then suggests that $\mathbb{C}_n$ does not converge weakly in $(\ell^\infty([0,1]), \|\cdot\|_\infty)$ if Condition 4.1 does not hold.

EXAMPLE 4.2 (Mixture model). For $\lambda \in (0,1)$, consider the bivariate copula given by

$$C(u_1,u_2) = (1-\lambda)\,u_1u_2 + \lambda\,\min(u_1,u_2).$$

For $u_1 \neq u_2$, the partial derivatives are

$$\begin{aligned}\dot{C}_1(u_1,u_2) &= (1-\lambda)\,u_2 + \lambda\,\mathbb{1}(u_1 < u_2),\\ \dot{C}_2(u_1,u_2) &= (1-\lambda)\,u_1 + \lambda\,\mathbb{1}(u_2 < u_1).\end{aligned}$$

On the diagonal $u_1 = u_2$, the partial derivatives do not exist. Still, by the decomposition in (4.1), the finite-dimensional distributions of $\mathbb{C}_n$ can be seen to converge to the ones of the process $\tilde{\mathbb{C}}$ defined as

$$\tilde{\mathbb{C}}(u_1,u_2) = \alpha(u_1,u_2) - \dot{C}_1(u_1,u_2)\,\alpha_1(u_1) - \dot{C}_2(u_1,u_2)\,\alpha_2(u_2),$$

if $u_1 \neq u_2$, whereas, on the diagonal $u_1 = u_2 = u$,

$$\tilde{\mathbb{C}}(u,u) = \alpha(u,u) - (1-\lambda)\,u\,\{\alpha_1(u)+\alpha_2(u)\} - \lambda\,\max\bigl(\alpha_1(u),\alpha_2(u)\bigr),$$

the distribution of which is non-Gaussian.

Now suppose that $\mathbb{C}_n \rightsquigarrow \mathbb{C}$ in $(\ell^\infty([0,1]^d, \|\cdot\|_\infty)$ for some $\mathbb{C}$. Then the finite-dimensional distributions of $\mathbb{C}$ must be equal to the ones of $\tilde{\mathbb{C}}$. Additionally, the trajectories of $\mathbb{C}$ must be continuous almost surely, and thus the law of the random variable $\mathbb{C}(u_1,u_2)$ must depend continuously on the coordinates (u_1,u_2). However, by the above expressions for $\tilde{\mathbb{C}}$, continuity cannot hold at points on the diagonal. This yields a contradiction, and therefore, $\mathbb{C}_n$ cannot converge weakly with respect to the supremum distance.

By considering weak hypi-convergence, we can go far beyond Condition 4.1. Condition 4.3 imposes the regularity needed to deal with the left-hand side of (4.2) in the hypi-semimetric.

CONDITION 4.3. *The set $\mathcal{S}$ of points in $[0,1]^d$ where the partial derivatives of the copula C exist and are continuous has Lebesgue measure* 1.

Since a copula is monotone in each of its arguments, its partial derivatives automatically exist almost everywhere. Condition 4.3 then only concerns continuity of these partial derivatives. In practice, Condition 4.3 poses no restriction at all. Still, there do exist copulas that do not satisfy Condition 4.3. It can be shown that a bivariate example is given by the copula with Lebesgue density

$$c(u,v) = \frac{\mathbb{1}_{A\times A}(u,v)}{\lambda_1(A)} + \frac{\mathbb{1}_{B\times B}(u,v)}{\lambda_1(B)}$$

where λ_1 denotes the one-dimensional Lebesgue measure, $A \subset [0,1]$ is a closed set which is at the same time nowhere dense and satisfies $\lambda_1(A) \in (0,1)$ and where $B = [0,1] \setminus A$.

For broad applicability, we relax the assumption of serial independence and replace it by the following condition, which holds for i.i.d. sequences as well as for stationary sequences under various weak dependence conditions (Rio, 2000; Doukhan, Fermanian and Lang, 2009; Dehling and Durieu, 2011).

Condition 4.4. *The empirical process $\alpha_n = \sqrt{n}(G_n - C)$ converges weakly in $(\ell^\infty([0,1]^d), \|\cdot\|_\infty)$ to some limit process α which has continuous sample paths, almost surely.*

Under Condition 4.3, the term on the right-hand side of (4.3) is defined only on $\mathcal{S}$. We extend it to the whole of $[0,1]^d$ by taking lower semicontinuous hulls as in Appendix A.1.2. Let $|\cdot|$ denote the Euclidean norm and let $\mathcal{C}(A)$ be the set of continuous real-valued functions on a domain A. Recall our convention of omitting the brackets $[\,\cdot\,]$ when working in $L^\infty_{\mathrm{loc}}(\mathbb{T})$.

Theorem 4.5. *Suppose that Condition 4.4 holds and that C satisfies Condition 4.3. Then,*

$$\mathbb{C}_n \rightsquigarrow \mathbb{C} = \alpha + \mathrm{d}C_{(-\alpha_1,\ldots,-\alpha_d)} \tag{4.5}$$

in $(L^\infty([0,1]^d), d_{\mathrm{hypi}})$, where, for $a = (a_1,\ldots,a_d) \in \{\mathcal{C}([0,1])\}^d$,

$$\mathrm{d}C_a(\boldsymbol{u}) = \sup_{\varepsilon>0} \inf\left\{\sum_{j=1}^d \dot{C}_j(\boldsymbol{v})\, a_j(v_j) : \boldsymbol{v} \in \mathcal{S},\ |\boldsymbol{v}-\boldsymbol{u}| < \varepsilon\right\}.$$

By Section 2, Theorem 4.5 has several useful consequences. First, it implies weak convergence with respect to the supremum distance of the restriction of the empirical copula process to compact subsets of the union of $\mathcal{S}$ and the boundary of $[0,1]^d$, see Proposition 2.2. This is akin to the convergence results for multilinear empirical copulas for count data in Genest, Nešlehová and Rémillard (2013). Furthermore, we obtain weak convergence of the empirical copula process in $(L^p([0,1]^d), \|\cdot\|_p)$ for any $1 \le p < \infty$. To the best of our knowledge, this result is new and opens the door to L^p-type inference procedures for a broad class of copulas. Three possible applications are treated in the following subsections.

4.1. *A bootstrap device.* In this section we suppose that $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ are serially independent. We show that the bootstrap based on resampling with replacement (Fermanian, Radulović and Wegkamp, 2004) and the bootstrap based on the multiplier central limit theorem (Bücher and Dette, 2010) provide valid approximations for $\mathbb{C}$ with respect to the hypi-semimetric. Our multiplier bootstrap is different from the approach in Rémillard and Scaillet (2009), which requires estimation of the first-order partial derivatives of C.

Let $M \in \mathbb{N}$ be some large integer and, for each $m \in \{1, \ldots, M\}$, let $\boldsymbol{X}_1^{[m]}, \ldots, \boldsymbol{X}_n^{[m]}$ be drawn with replacement from the sample. The resampling bootstrap empirical copula process is defined as

$$\mathbb{C}_n^{[m]} = \sqrt{n}(C_n^{[m]} - C_n),$$

where $C_n^{[m]}$ denotes the empirical copula calculated from the bootstrap sample $\boldsymbol{X}_1^{[m]}, \ldots, \boldsymbol{X}_n^{[m]}$. Note that we can represent $C_n^{[m]}$ by $F_n^{[m]}(\boldsymbol{F}_n^{[m]-})$, where

$$F_n^{[m]}(\boldsymbol{x}) = \frac{1}{n} \sum_{i=1}^n W_{ni}^{[m]} \mathbb{1}(\boldsymbol{X}_i \leq \boldsymbol{x})$$

and where $W_n^{[m]} = (W_{n1}^{[m]}, \ldots, W_{nn}^{[m]})$ denotes a multinomial random vector with n trials, n possible outcomes, and success probabilities $(1/n, \ldots, 1/n)$, independent of the sample and independent across $m \in \{1, \ldots, M$.

Regarding the multiplier bootstrap, let $\{\xi_i^{[m]} : i \geq 1, \ m = 1, \ldots, M\}$ be i.i.d. random variables, independent of the sample, with both mean and variance equal to one and such that $\int_0^\infty \sqrt{\mathbb{P}(\xi_i > x)}\, dx < \infty$. Let

$$\tilde{F}_n^{[m]}(\boldsymbol{x}) = \frac{1}{n} \sum_{i=1}^n \xi_i^{[m]} \mathbb{1}(\boldsymbol{X}_i \leq \boldsymbol{x}), \qquad \boldsymbol{x} \in \mathbb{R}^d$$

and define

$$\tilde{\mathbb{C}}_n^{[m]} = \sqrt{n}\{\tilde{F}_n^{[m]}(\tilde{F}_n^{[m]-}) - C_n\}$$

as the multiplier bootstrap empirical copula process. The following proposition shows that both $\mathbb{C}_n^{[1]}, \ldots, \mathbb{C}_n^{[M]}$ and $\tilde{\mathbb{C}}_n^{[1]}, \ldots, \tilde{\mathbb{C}}_n^{[M]}$ can be regarded as asymptotically independent copies of $\mathbb{C}_n$.

PROPOSITION 4.6. *Let $\mathbb{C}^{[1]}, \ldots, \mathbb{C}^{[M]}$ denote independent copies of $\mathbb{C}$. Under Condition 4.3, both $(\mathbb{C}_n, \mathbb{C}_n^{[1]}, \ldots, \mathbb{C}_n^{[M]})$ and $(\mathbb{C}_n, \tilde{\mathbb{C}}_n^{[1]}, \ldots, \tilde{\mathbb{C}}_n^{[M]})$ weakly converge to $(\mathbb{C}, \mathbb{C}^{[1]}, \ldots, \mathbb{C}^{[M]})$ in $(L^\infty([0,1]^d), d_{\mathrm{hypi}})^{M+1}$.*

By hypi-continuity of the supremum and infimum functionals, see Proposition 2.3, the bootstrap approximation can for instance be used to construct asymptotic uniform confidence bands for the copula.

4.2. *Power curves of tests for independence.* In the present section, we derive weak hypi-convergence of the empirical copula process for triangular arrays. We apply it to the problem of comparing statistical tests for independence by local power curves. This comparison has been carried out by Genest, Quessy and Rémillard (2007) under strong differentiability assumptions on copula densities. By considering hypi-convergence, we can extend their results to copulas that do not have a density with respect to the Lebesgue measure.

We consider a triangular array of random vectors $\boldsymbol{X}_1^{(n)}, \dots, \boldsymbol{X}_n^{(n)}$ which are row-wise i.i.d. with continuous marginals and copula $C^{(n)}$. We suppose that there exists a copula C satisfying Condition 4.3 such that

$$\Delta_n = \sqrt{n}\{C^{(n)} - C\} \to \Delta \tag{4.6}$$

uniformly, for some continuous function Δ on $[0,1]^d$. Let $C_n^{(n)}$ denote the empirical copula based on $\boldsymbol{X}_1^{(n)}, \dots, \boldsymbol{X}_n^{(n)}$. Let $\boldsymbol{U}_1^{(n)}, \dots, \boldsymbol{U}_n^{(n)}$ denote the sample obtained by the marginal probability integral transform and let $G_n^{(n)}$ and $\alpha_n^{(n)}$ denote its empirical distribution function and empirical process, respectively. Similarly as before, we have the decomposition

$$\begin{aligned}
\mathbb{C}_n^{(n)} &= \sqrt{n}\{C_n^{(n)} - C^{(n)}\} \\
&= \sqrt{n}\{G_n^{(n)}(\boldsymbol{G}_n^{(n)-}) - C^{(n)}(\boldsymbol{G}_n^{(n)-})\} + \sqrt{n}\{C^{(n)}(\boldsymbol{G}_n^{(n)-}) - C^{(n)}\} \\
&= \alpha_n^{(n)}(\boldsymbol{G}_n^{(n)-}) + \sqrt{n}\{C(\boldsymbol{G}_n^{(n)-}) - C\} + \{\Delta_n(\boldsymbol{G}_n^{(n)-}) - \Delta_n\}.
\end{aligned}$$

We will show in Appendix B.3 that $\alpha_n^{(n)} \rightsquigarrow \alpha$ in $(\ell^\infty([0,1]^d), \|\cdot\|_\infty)$, where α is a C-brownian bridge as in Theorem 4.5. Therefore, the first summand weakly converges to α with respect to the supremum norm. The second summand weakly converges in the hypi-topology to $\mathrm{d}C_{(-\alpha_1,\dots,-\alpha_d)}$, while the last one converges to $\Delta - \Delta \equiv 0$, uniformly. This motivates the following result.

PROPOSITION 4.7. *Under Condition 4.4 and if (4.6) is met with C satisfying Condition 4.3, we have $\mathbb{C}_n^{(n)} \rightsquigarrow \mathbb{C}$ in $(L^\infty([0,1]^d, d_{\mathrm{hypi}})$, where $\mathbb{C}$ is the same process as in Theorem 4.5. Additionally, in $(L^\infty([0,1]^d), d_{\mathrm{hypi}})$,*

$$\sqrt{n}(C_n^{(n)} - C) \rightsquigarrow \mathbb{C} + \Delta.$$

To illustrate the latter result, we investigate the local efficiency of tests for independence as considered in Genest, Quessy and Rémillard (2007).

Instead of imposing conditions (i) and (ii) on page 169 in their paper, we only suppose that (4.6) holds with $C = \Pi$, the independence copula, and $\Delta = \delta\Lambda$, where $\Lambda \in \mathcal{C}([0,1]^d)$ and $\delta \geq 0$. For brevity, we only compare the test statistics

$$T_n = n \int_{[0,1]^d} \{C_n^{(n)} - \Pi\}^2 \mathrm{d}\Pi \quad \text{and} \quad S_n = \sqrt{n}\|C_n^{(n)} - \Pi\|_\infty.$$

From weak hypi-convergence of $\sqrt{n}(C_n^{(n)} - C)$ and Propositions 2.2 and 2.4, we obtain that

$$T_n \rightsquigarrow \mathcal{T}_\delta = \int_{[0,1]^d} (\mathbb{C} + \delta\Lambda)^2 \mathrm{d}\Pi, \qquad S_n \rightsquigarrow \mathcal{S}_\delta = \|\mathbb{C} + \delta\Lambda\|_\infty.$$

Hence, the local power curves of the tests to the level $\alpha \in (0,1)$ in direction Λ are given by

$$\delta \mapsto \mathbb{P}\{\mathcal{T}_\delta > q_{\mathcal{T}_0}(1-\alpha)\}, \qquad \delta \mapsto \mathbb{P}\{\mathcal{S}_\delta > q_{\mathcal{S}_0}(1-\alpha)\},$$

where $q_{\mathcal{T}_0}(1-\alpha)$ and $q_{\mathcal{S}_0}(1-\alpha)$ denote the $(1-\alpha)$-quantiles of $\mathcal{T}_0$ and $\mathcal{S}_0$, respectively. These curves can be compared by analytical calculations as in Genest, Quessy and Rémillard (2007) or by simulation.

4.3. *Minimum L^2-distance estimators.* In this section, we illustrate how the weak hypi-convergence of the empirical copula process can be used to derive the asymptotic distribution for minimum L^2-distance estimators of parametric copulas. More precisely, let $\{C_\theta : \theta \in \Theta\}$ be a parametric family of d-variate copulas. Assume that the parameter space Θ is an open subset of $\mathbb{R}^m$. A minimum L^2-distance estimator $\hat{\theta}_n$ of θ is defined through

$$\|C_n - C_{\hat{\theta}_n}\|_{2,\mu}^2 \leq \inf_{\theta \in \Theta} \|C_n - C_\theta\|_{2,\mu}^2 + \delta_n,$$

where $\|\cdot\|_{2,\mu}$ denotes the $L^2(\mu)$-norm with respect to a probability measure μ on $[0,1]^d$ and where $\delta_n = o_p(n^{-1})$ is some nonnegative, possibly random sequence. This is a slight generalization of the estimator of Tsukahara (2005), who considered the case where μ is Lebesgue measure and who assumed that $\mathbb{C}_n$ converges weakly with respect to the supremum distance. In practice, minimum distance estimators can also be defined with respect to a random weighting measure, for instance C_n itself, and with respect to other L^p-norms (Weiß, 2011).

By not insisting on weak convergence with respect to the supremum distance, we can allow for instance for Marshall–Olkin copulas (Embrechts, Lindskog and McNeil, 2003), more general extreme-value copulas whose spectral measures have atoms (Gudendorf and Segers, 2012), Archimedean copulas with nonsmooth generators (McNeil and Nešlehová, 2009), and mixtures thereof. Since such copula families do not possess densities, the pseudo-likelihood estimator of Genest, Ghoudi and Rivest (1995) is not available.

We apply Theorems A.14 and A.16 on minimum L^2-distance estimators given in Appendix A.3. Define the quantities ∇C_{θ_0} and J_{θ_0} by replacing 'G' by 'C' in Conditions A.13 and A.15.

PROPOSITION 4.8. *Suppose that the copula C satisfies Condition 4.3, that the parametric model satisfies Condition A.11, and that the data-generating process is such that Condition 4.4 holds. Suppose moreover that $\mu([0,1]^d \setminus \mathcal{S}) = 0$, with $\mathcal{S}$ as in Condition 4.3.*

(i) If the model is correctly specified, i.e., if $C = C_{\theta_0}$ for some $\theta_0 \in \Theta$ and if Condition A.13 holds, then

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \rightsquigarrow \left\{ \int \nabla C_{\theta_0} \nabla C_{\theta_0}^T \, d\mu \right\}^{-1} \int \nabla C_{\theta_0} \mathbb{C} \, d\mu.$$

(ii) For incorrectly specified models, i.e., for $C \notin \{C_\theta : \theta \in \Theta\}$ and $\theta_0 = \arg\min_{\theta \in \Theta} \|C_\theta - C\|_{2,\mu}$, if Condition A.15 holds, then

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \rightsquigarrow \left\{ \int \left(\nabla C_{\theta_0} \nabla C_{\theta_0}^T + (C_{\theta_0} - C) J_{\theta_0} \right) d\mu \right\}^{-1} \int \nabla C_{\theta_0} \mathbb{C} \, d\mu.$$

By Condition 4.3, the assumption that the complement of $\mathcal{S}$ is a μ-null set is automatically verified if μ is absolutely continuous with respect to Lebesgue measure.

5. Stable tail dependence functions. In this section, let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$, $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{id})$ be i.i.d. d-variate random vectors with distribution function F and continuous marginal distribution functions $F_1, \ldots, F_d$. We assume that the following limit, called the stable tail dependence function of F,

$$L(\boldsymbol{x}) = \lim_{t \downarrow 0} t^{-1} \mathbb{P}\left\{1 - F_1(X_{11}) \le t x_1 \text{ or } \ldots \text{ or } 1 - F_d(X_{1d}) \le t x_d\right\}, \tag{5.1}$$

exists as a function $L : [0, \infty)^d \to [0, \infty)$.

For $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, d\}$, let R_i^j denote the rank of X_{ij} among $X_{1j}, \ldots, X_{nj}$. Replacing all distribution functions in (5.1) by their empirical counterparts and replacing t by k/n where $k = k_n$ is a positive sequence such that $k_n \to \infty$ and $k_n = o(n)$, we obtain the following nonparametric estimator for L_n, called the empirical (stable) tail dependence function:

$$\hat{L}_n(\boldsymbol{x}) = \frac{1}{k} \sum_{i=1}^{n} \mathbb{1} \left\{ R_i^1 > n + \frac{1}{2} - kx_1 \text{ or } \dots \text{ or } R_i^d > n + \frac{1}{2} - kx_d/n \right\}$$

(Huang, 1992; Drees and Huang, 1998). The inclusion of the term $1/2$ inside the indicators serves to improve the finite sample behavior of the estimator.

In Einmahl, Krajina and Segers (2012), a functional central limit theorem for $\sqrt{k}(\hat{L}_n - L)$ is given in the topology of uniform convergence on compact subsets of $[0,\infty)^d$. The result requires L to have continuous first-order partial derivatives on sufficiently large subsets of $[0,\infty)^d$, similar to Condition 4.1 for copulas. By switching to weak hypi-convergence, we are able to get rid of smoothness conditions altogether.

Similarly as in Section 4, let $\mathcal{S}$ denote the set of all points $\boldsymbol{x} \in [0,\infty)^d$ where L is differentiable. The function L being convex, Theorem 25.5 in Rockafellar (1970) implies that the complement of $\mathcal{S}$ is a Lebesgue null set and that the gradient $(\dot{L}_1, \dots, \dot{L}_d)$ of L is continuous on $\mathcal{S}$. Proceeding as in Appendix A.1.2, we may define, for any $(a_1, \dots, a_d) \in \{C([0,\infty))\}^d$, a function on $[0,\infty)^d$ by

$$\mathrm{d}L_{(a_1,\dots,a_d)}(\boldsymbol{x}) = \sup_{\varepsilon>0} \inf \left\{ \sum_{j=1}^{d} \dot{L}_j(\boldsymbol{y})\, a_j(y_j) : \boldsymbol{y} \in \mathcal{S}, |\boldsymbol{x} - \boldsymbol{y}| < \varepsilon \right\}. \tag{5.2}$$

As in Einmahl, Krajina and Segers (2012), let Λ be the Borel measure on $[0,\infty)^d$ such that $\Lambda(A(\boldsymbol{x})) = L(x)$ where $A(\boldsymbol{x}) = \bigcup_{j=1}^d \{y \in [0,\infty)^d : y_j \leq x_j\}$ for $\boldsymbol{x} \in [0,\infty)^d$. Let $\mathbb{W}$ be a mean-zero Gaussian process on $[0,\infty)^d$ with continuous trajectories and with covariance function $\mathbb{E}[\mathbb{W}(\boldsymbol{x})\,\mathbb{W}(\boldsymbol{y})] = \Lambda(A(\boldsymbol{x}) \cap A(\boldsymbol{y}))$. Let $\Delta_{d-1} = \{\boldsymbol{x} \in [0,1]^d : x_1 + \cdots + x_d = 1\}$ be the unit simplex in $\mathbb{R}^d$. For $f \in \ell^\infty_{\mathrm{loc}}([0,\infty)^d)$ and $j = 1, \dots, d$, define $f_j^0 \in \ell^\infty_{\mathrm{loc}}([0,\infty))$ through $f_j^0(x_j) = f(0, \dots, 0, x_j, 0, \dots, 0)$.

THEOREM 5.1. *Let $\boldsymbol{X}_i$, $i \in \mathbb{N}$, be i.i.d. d-dimensional random vectors with common distribution function F with continuous margins $F_1, \dots, F_d$ and stable tail dependence function L. Suppose that the following conditions hold:*

(C1) For some $\alpha > 0$ we have, uniformly in $\boldsymbol{x} \in \Delta_{d-1}$,

$$t^{-1}\mathbb{P}\{1 - F_1(X_{11}) \leq tx_1 \text{ or } \dots \text{ or } 1 - F_d(X_{1d}) \leq tx_d\} = L(\boldsymbol{x}) + O(t^\alpha), \qquad t \downarrow 0.$$

(C2) We have $k = o(n^{2\alpha/(1+2\alpha)})$ *and* $k \to \infty$ *as* $n \to \infty$.

Then, in $(L^\infty_{\mathrm{loc}}([0,\infty)^d), d_{\mathrm{hypi}})$,

$$\sqrt{k}\,\big(\hat{L}_n - L\big) \rightsquigarrow \mathbb{W} + \mathrm{d}L_{(-\mathbb{W}^0_1,\ldots,-\mathbb{W}^0_d)}, \qquad n \to \infty.$$

Conditions (C1) and (C2) also appear in Theorem 4.6 in Einmahl, Krajina and Segers (2012) and are needed to ensure that the estimator is asymptotically unbiased. The difference with their theorem is that we do not need their Condition (C3) on the partial derivatives of L. Therefore, Theorem 5.1 also covers piecewise linear stable tail dependence functions arising from max-linear models (Wang and Stoev, 2011).

Weak hypi-convergence of $\sqrt{k}(\hat{L}_n - L)$ can be exploited to validate statistical procedures for tail dependence functions in the same way as was done with weak hypi-convergence of empirical copula processes in Section 4. In contrast to copulas, no smoothness conditions on L are needed at all. Applications include the bootstrap (Peng and Qi, 2008) and minimum L^2-distance estimation (Bücher and Dette, 2011). In particular, Proposition 4.8 remains valid upon replacing C by L and $\sqrt{n}$ by $\sqrt{k}$. Weak hypi-convergence implying weak L^2-convergence, Theorem 5.1 also provides an alternative way to prove the asymptotic normality of the M-estimator in Einmahl, Krajina and Segers (2012).

6. Conclusion and outlook. We have introduced the hypi-topology on the space of locally bounded functions defined on a suitable domain. This topology being induced by a semimetric, we considered Hoffman-Jørgensen weak convergence of equivalence classes of functions. Many interesting functionals were shown to be hypi-continuous, which allowed to transfer weak hypi-convergence to weak convergence in more familiar metric spaces. Under mild assumptions, we proved weak hypi-convergence of empirical copula and empirical tail dependence function processes.

An alternative view on the hypi-topology can be gained by identifying $f \in \ell^\infty_{\mathrm{loc}}(\mathbb{T})$ with its completed graph $\Gamma(f) = \mathrm{epi}(f_\wedge) \cap \mathrm{hypo}(f_\vee)$ (Vervaat, 1981). We suspect that for certain domains, hypi-convergence is equivalent to set convergence of completed graphs. For $\mathbb{T} = [0,1]$ the latter can be seen to be equivalent to Skorohod M_2-convergence (Molchanov, 2005, p. 377), whence hypi-convergence can be regarded as a coordinate-free extension of Skorohod M_2-convergence to nonsmooth functions on rather general domains.

In the applications of this paper, we proved weak hypi-convergence by an application of the continuous mapping theorem to a process converging with respect to the supremum distance. An intrinsic alternative would be to investigate weak convergence of vectors of suprema and infima over compact sets (Molchanov, 2005, Proposition 3.15, p. 344).

Weak hypi-convergence might be useful in other statistical applications, as for instance for empirical processes of residuals in time series and regression models where the distribution function of the errors is not continuously differentiable. An extension of the functional delta method would be yet another worthwhile topic for future research.

APPENDIX A: AUXILIARY RESULTS

A.1. Verifying hypi-convergence.

A.1.1. *Pointwise convergence and convergence of sums.* Let $(\mathbb{T}, d)$ be a metric space. For $f : \mathbb{T} \mapsto \mathbb{R}$, define extended real-valued functions $f_\wedge$ and $f_\vee$ as in (2.1) and (2.2), respectively. Since we do not require f to be locally bounded, $f_\wedge$ and $f_\vee$ can attain $-\infty$ and $+\infty$, respectively.

For $f_n : \mathbb{T} \mapsto \mathbb{R}$, we say that f_n epi-converges to $\alpha \in \mathbb{R}$ at $x \in \mathbb{T}$ if the following two conditions are met:

$$\text{(A.1)} \qquad \begin{cases} \text{(i)} & \forall\, x_n \to x : \quad \liminf_{n\to\infty} f_n(x_n) \ge \alpha, \\ \text{(ii)} & \exists\, x_n \to x : \quad \limsup_{n\to\infty} f_n(x_n) \le \alpha. \end{cases}$$

Similarly, f_n hypo-converges to α at x if

$$\text{(A.2)} \qquad \begin{cases} \text{(i)} & \forall\, x_n \to x : \quad \limsup_{n\to\infty} f_n(x_n) \le \alpha, \\ \text{(ii)} & \exists\, x_n \to x : \quad \liminf_{n\to\infty} f_n(x_n) \ge \alpha. \end{cases}$$

If additionally $f : \mathbb{T} \mapsto \mathbb{R}$, then f_n is said to epi- or hypo-converge to f at x if $\alpha = f(x)$ in the preceding conditions. According to Proposition 2.1, f_n hypi-converges to f in $\ell^\infty_{\mathrm{loc}}(\mathbb{T})$ if and only if f_n epi-converges to $f_\wedge$ and hypo-converges to $f_\vee$ at every $x \in \mathbb{T}$. For $x \in \mathbb{T}$ and $\varepsilon > 0$, let $B(x, \varepsilon) = \{y \in \mathbb{T} : d(x, y) < \varepsilon\}$.

LEMMA A.1 (Continuous convergence of hulls). *Let $g_n : \mathbb{T} \mapsto \mathbb{R}$, $x \in \mathbb{T}$ and $\alpha \in \mathbb{R}$. Then*

$$g_n(x_n) \to \alpha \qquad \forall\, x_n \to x$$

implies

$$g_{n,\wedge}(x_n) \to \alpha \quad \textit{and} \quad g_{n,\vee}(x_n) \to \alpha \qquad \forall\, x_n \to x.$$

Proof. We only consider the lower hulls. By definition, we have

$$g_{n,\wedge}(x_n) = \sup_{\varepsilon>0} \inf g_n(B(x_n,\varepsilon)) \geq \inf g_n(B(x_n,1/n)).$$

This allows to choose a $y_n \in B(x_n, 1/n)$ such that

$$g_{n,\wedge}(x_n) \geq \inf g_n(B(x_n,1/n)) \geq g_n(y_n) - \frac{1}{n} \to \alpha,$$

since $d(y_n, x) \leq 1/n + d(x_n, x) \to 0$. Hence,

$$\alpha \leq \liminf_{n\to\infty} g_{n,\wedge}(x_n) \leq \limsup_{n\to\infty} g_{n,\wedge}(x_n) \leq \limsup_{n\to\infty} g_n(x_n) = \alpha$$

which implies the assertion. □

Lemma A.2 (On sums of hulls and hulls of sums). *For $f, g : \mathbb{T} \to \mathbb{R}$ such that $g_\wedge$ and $g_\vee$ are both finite, we have*

$$\begin{aligned} f_\wedge + g_\wedge \leq (f+g)_\wedge \leq f_\wedge + g_\vee, \\ f_\vee + g_\wedge \leq (f+g)_\vee \leq f_\vee + g_\vee. \end{aligned}$$

In particular, if g is continuous in $x \in \mathbb{T}$, then $(f+g)_\wedge(x) = f_\wedge(x) + g(x)$ and $(f+g)_\vee(x) = f_\vee(x) + g(x)$.

Proof. Fix $x \in \mathbb{T}$. For $\varepsilon > 0$, we have

$$\inf_{y\in B(x,\varepsilon)} \{f(y) + g(y)\} \geq \inf_{y\in B(x,\varepsilon)} f(y) + \inf_{y\in B(x,\varepsilon)} g(y).$$

Taking limits as $\varepsilon \downarrow 0$, we find

$$(f+g)_\wedge(x) \geq f_\wedge(x) + g_\wedge(x).$$

By the same inequality, we have

$$\begin{aligned} f_\wedge(x) &= ((f+g) + (-g))_\wedge(x) \\ &\geq (f+g)_\wedge(x) + (-g)_\wedge(x) = (f+g)_\wedge(x) - g_\vee(x), \end{aligned}$$

whence

$$f_\wedge(x) + g_\vee(x) \geq (f+g)_\wedge(x).$$

The inequalities for $(f+g)_\vee$ follow similarly. □

LEMMA A.3 (Epi- and hypo-convergence of hulls of sums). *Let $f_n, g_n : \mathbb{T} \mapsto \mathbb{R}$ and let $x \in \mathbb{T}$ be such that $g_n(x_n) \to \beta \in \mathbb{R}$ for all sequences $x_n \to x$. If $f_{n,\wedge}$ epi-converges to α at x, then $(f_n + g_n)_\wedge$ epi-converges to $\alpha + \beta$ at x. Similarly for upper semicontinuous hulls and hypo-convergence.*

PROOF. It is sufficient to consider epi-convergence, (A.1). To be shown:

(i) $\forall\, x_n \to x :\ \liminf_{n\to\infty}(f_n + g_n)_\wedge(x_n) \geq \alpha + \beta$,
(ii) $\exists\, x_n \to x :\ \limsup_{n\to\infty}(f_n + g_n)_\wedge(x_n) \leq \alpha + \beta$.

By Lemma A.1, $g_{n,\wedge}(x_n)$ and $g_{n,\vee}(x_n)$ both converge to β whenever $x_n \to x$.

First consider (i). Let $x_n \to x$. By Lemma A.2,

$$\begin{aligned}\liminf_{n\to\infty}(f_n + g_n)_\wedge(x_n) &\geq \liminf_{n\to\infty}\{f_{n,\wedge}(x_n) + g_{n,\wedge}(x_n)\}\\ &\geq \liminf_{n\to\infty} f_{n,\wedge}(x_n) + \liminf_{n\to\infty} g_{n,\wedge}(x_n) \geq \alpha + \beta.\end{aligned}$$

Next consider (ii). The assumptions allow to choose a sequence $x_n \to x$ such that

$$\limsup_{n\to\infty} f_{n,\wedge}(x_n) \leq \alpha.$$

Again by Lemma A.2,

$$\begin{aligned}\limsup_{n\to\infty}(f_n + g_n)_\wedge(x_n) &\leq \limsup_{n\to\infty}\{f_{n,\wedge}(x_n) + g_{n,\vee}(x_n)\}\\ &\leq \limsup_{n\to\infty} f_{n,\wedge}(x_n) + \limsup_{n\to\infty} g_{n,\vee}(x_n) \leq \alpha + \beta,\end{aligned}$$

as required. □

LEMMA A.4 (Hypi-convergence of sums). *Let $\mathbb{T}$ be locally compact and separable. If f_n and g_n hypi-converge to f and g in $\ell^\infty_{\mathrm{loc}}(\mathbb{T})$ respectively and if at every point $x \in \mathbb{T}$, at least one of the two functions f or g is continuous, then $f_n + g_n$ hypi-converges to $f + g$.*

PROOF. By Proposition 2.1(ii), we have to verify that for every $x \in \mathbb{T}$, the sequence $f_n + g_n$ epi-converges at x to $(f+g)_\wedge(x)$ and hypo-converges at x to $(f+g)_\vee(x)$ in the sense of (A.1) and (A.2). Fix $x \in \mathbb{T}$. Without loss of generality, assume that g is continuous at x. Then $(f+g)_\wedge(x) = f_\wedge(x) + g(x)$ and $(f+g)_\vee = f_\vee(x) + g(x)$ by Lemma A.2. By Proposition 2.1(ii), f_n epi-converges to $f_\wedge(x)$ at x and hypo-converges to $f_\vee(x)$ at x. Moreover, $g_n(x_n) \to g(x)$ whenever $x_n \to x$. Apply Lemma A.3 to conclude that $f_n + g_n$ epi-converges to $f_\wedge(x) + g(x)$ and hypo-converges to $f_\vee(x) + g(x)$ at x, as required. □

A.1.2. *Upper and lower semicontinuous extensions.* In this section, some useful elementary properties of upper and lower semicontinuous extensions of a function defined on a dense subset of a metric space are recorded. Let $(\mathbb{T}, d)$ be a metric space, let $A \subset \mathbb{T}$ be dense, and let $f : A \mapsto \mathbb{R}$. Extend the domain of f from A to the whole of $\mathbb{T}$ by

$$f_\wedge(x) = \sup_{\varepsilon>0} \inf f(B(x, \varepsilon) \cap A), \tag{A.3}$$

$$f_\vee(x) = \inf_{\varepsilon>0} \sup f(B(x, \varepsilon) \cap A), \tag{A.4}$$

for $x \in \mathbb{T}$; as before, $B(x, \varepsilon) = \{y \in \mathbb{T} : d(x, y) < \varepsilon\}$ is the open ball centered at x of radius ε. These definitions also make sense if $A = \mathbb{T}$, in which case they are the same as (2.1) and (2.2), with the difference that $f_\wedge(x)$ can be $-\infty$ and $f_\vee(x)$ can be $+\infty$. Formally, we should indicate the sets A and $\mathbb{T}$ in the notation in order to distinguish the extensions from the hulls defined in (2.1) and (2.2), but for simplicity, we suppress the sets from the notation.

Clearly, $f_\wedge(x) \le f(x) \le f_\vee(x)$ for every $x \in A$. For any open set $U \subset \mathbb{T}$, we have

$$\inf f_\wedge(U) = \inf f(U \cap A),$$
$$\sup f_\vee(U) = \sup f(U \cap A).$$

The functions $f_\wedge$ and $f_\vee$ from $\mathbb{T}$ into $[-\infty, +\infty]$ are lower and upper semicontinuous, respectively. If every x in A admits a neighbourhood on which f is bounded, then $f_\wedge$ and $f_\vee$ are real-valued.

If f is continuous at $x \in A$, then $f_\wedge(x) = f_\vee(x) = f(x)$, and $f_\wedge$ and $f_\vee$, seen as functions on $\mathbb{T}$, are continuous at x, too. If f is continuous on the whole of A, then its domain does not really matter insofar as the extension is concerned.

Lemma A.5. *Let $E \subset A \subset \mathbb{T}$ be such that E is dense in $\mathbb{T}$. Let $f : A \mapsto \mathbb{R}$ and consider the restriction $f_E : E \mapsto \mathbb{R}$ of f to E and the extensions $(f_E)_\wedge$ and $(f_E)_\vee$ of f_E to $\mathbb{T}$. If f is continuous, then $(f_E)_\wedge = f_\wedge$ and $(f_E)_\vee = f_\vee$.*

Proof. It is sufficient to show that

$$\inf f(U \cap E) = \inf f(U \cap A)$$

for every open set $U \subset \mathbb{T}$: indeed, if the above equality holds, then for $x \in \mathbb{T}$,

$$(f_E)_\wedge(x) = \sup_{\varepsilon>0} \inf f(B(x, \varepsilon) \cap E) = \sup_{\varepsilon>0} \inf f(B(x, \varepsilon) \cap A) = f_\wedge(x).$$

So let $U \subset \mathbb{T}$ be open. Since $E \subset A$, clearly $\inf f(U \cap E) \geq \inf f(U \cap A)$. On the other hand, if $x \in U \cap A$, then since E is dense and U is open we can find a sequence $x_n \in U \cap E$ such that $\lim_{n\to\infty} x_n = x$. By continuity, $\lim_{n\to\infty} f(x_n) = f(x)$. As a consequence, $f(x) \geq \inf f(U \cap E)$. Since $x \in U \cap A$ was arbitrary, we have $\inf f(U \cap A) \geq \inf f(U \cap E)$.

The proof that $(f_E)_\vee = f_\vee$ is completely analogous. □

PROPOSITION A.6. *Let $A \subset \mathbb{T}$ be dense. Let $f : A \to \mathbb{R}$ be continuous and let $f_\wedge$ be its lower semi-continuous extension from A to $\mathbb{T}$. Assume that $f_\wedge$ is real-valued. If the functions $f_n : \mathbb{T} \to \mathbb{R}$ converge pointwise on A to f and if $\liminf_n f_n(x_n) \geq f_\wedge(x)$ whenever $x_n \in \mathbb{T}$ converges to $x \in \mathbb{T}$, then f_n epi-converges to $f_\wedge$. Similarly for hypo-convergence to $f_\vee$.*

PROOF. In view of Molchanov (2005, Proposition 3.2(ii), page 337), we have to show that for $x \in \mathbb{T}$ we can find $x_n \to x$ such that $\limsup_n f_n(x_n) \leq h_\wedge(x)$. If $x \in A$ this is true by assumption of pointwise convergence: just take $x_n = x$ and note that $f_\wedge(x) = f(x)$ by continuity of f. So take $x \in \mathbb{T} \setminus A$.

Let $(y_m)_m$ be a sequence in A converging to x such that $f_\wedge(x) = \lim_m f(y_m)$. For every m, we have $\lim_n f_n(y_m) = f(y_m)$ by pointwise convergence. Find integers $1 = n_0 < n_1 < n_2 < \dots$ as follows: for $m \geq 1$, let n_m be the smallest integer larger than n_{m-1} such that $|f_n(y_m) - f(y_m)| \leq 1/m$ for all $n \geq n_m$. Now define a sequence x_n as follows: set $x_n = y_m$ if $n_m \leq n < n_{m-1}$. Then $x_n \to x$ and

$$\begin{aligned} |f_n(x_n) - f_\wedge(x)| &\leq |f_n(y_m) - f(y_m)| + |f(y_m) - f_\wedge(x)| \\ &\leq 1/m + |f(y_m) - f_\wedge(x)| \end{aligned}$$

whenever $n_m \leq n < n_{m-1}$. It follows that $\lim_n f_n(x_n) = f_\wedge(x)$. □

COROLLARY A.7. *Let $A \subset \mathbb{T}$ be dense. Let $f : A \to \mathbb{R}$ be continuous and suppose that its lower and upper semi-continuous extensions from A to $\mathbb{T}$ $f_\wedge$ and $f_\vee$ are real-valued. Let $f^* : \mathbb{T} \mapsto \mathbb{R}$ be such that $f_\wedge \leq f^* \leq f_\vee$. Then $f_\wedge = (f^*)_\wedge$ and $f_\vee = (f^*)_\vee$ on $\mathbb{T}$. Moreover, if the functions $f_n : \mathbb{T} \to \mathbb{R}$ are locally bounded and verify*

$$\forall x \in \mathbb{T} : \forall x_n \to x : f_\wedge(x) \leq \liminf_{n\to\infty} f_n(x_n) \leq \limsup_{n\to\infty} f_n(x_n) \leq f_\vee(x),$$

then f_n hypi-converges to f^*.

PROOF. Since $f_\wedge$ and $f_\vee$ are real-valued and lower and upper semi-continuous respectively, the function f^* is locally bounded. By continuity, the functions f, $f_\wedge$, $f_\vee$ and f^* coincide on A and $f_\wedge = (f^*)_\wedge$ and $f_\vee = (f^*)_\vee$ on $\mathbb{T}$. Thanks to Proposition A.6, f_n epi-converges to $f_\wedge$ and hypo-converges to $f_\vee$. It follows that f_n hypi-converges to f^*. □

A.2. Weak convergence and semimetric spaces. The continuous mapping theorem is one of the workhorses in weak convergence theory (van der Vaart and Wellner, 1996, Theorem 1.3.6). We formulate and prove a version that is adapted to semimetric spaces (Theorem A.9). In particular, the map under consideration is not required to be defined on equivalence classes of points at distance zero but rather on the original semimetric space itself. This allows us to extend Slutsky's lemma to hypi-convergence (Lemma A.10).

Let $(\mathbb{D}, d)$ be a semimetric space. For $x \in \mathbb{D}$, put $[x] = \mathrm{cl}\{x\}$, the set of $y \in \mathbb{D}$ such that $d(x, y) = 0$. Since $d(x', y') = d(x, y)$ whenever $x' \in [x]$ and $y' \in [y]$, we can, abusing notation, define a metric $d([x], [y]) := d(x, y)$ on the quotient space $[\mathbb{D}] = \{[x] : x \in \mathbb{D}\}$. Let $[\,\cdot\,]$ denote the map $\mathbb{D} \mapsto [\mathbb{D}] : x \mapsto [x]$. Obviously, $[\,\cdot\,]$ is continuous. The image of an open (closed) subset of $\mathbb{D}$ under $[\,\cdot\,]$ is open (closed) in $[\mathbb{D}]$.

Let $\mathcal{B}(\mathbb{D})$ and $\mathcal{B}([\mathbb{D}])$ be the Borel σ-fields on $(\mathbb{D}, d)$ and $([\mathbb{D}], d)$, respectively, that is, the smallest σ-fields containing the open sets. There is a one-to-one correspondence between both σ-fields: for $B \in \mathcal{B}(\mathbb{D})$, the set $[B] = \{[x] : x \in B\}$ is a Borel set in $\mathbb{D}$, and conversely, every Borel set B of $\mathbb{D}$ can be written as $\bigcup_{[x]\in[B]}[x]$; in particular $x \in B$ if and only if $[x] \in [B]$. A Borel law L on $(\mathbb{D}, \mathcal{B}(\mathbb{D}))$ induces a Borel law $L \circ [\,\cdot\,]^{-1}$ on $([\mathbb{D}], \mathcal{B}([\mathbb{D}]))$ and vice versa.

One of the merits of Hoffman-Jørgensen weak convergence is that measurability requirements are relaxed. In the context of semimetric spaces, measurability issues require, perhaps, some extra care.

Lemma A.8 (Measurability). *Let $(\mathbb{D}, d)$ and $(\mathbb{E}, e)$ be semimetric spaces. Let $g : \mathbb{D} \mapsto \mathbb{E}$ be arbitrary. Then the set D_g of $x \in \mathbb{D}$ such that g is not continuous in x is Borel measurable. More generally, $g^{-1}(B) \setminus D_g$ is a Borel set in $\mathbb{D}$ for every Borel set B in $\mathbb{E}$.*

Proof. As in Theorem 1.3.6 in van der Vaart and Wellner (1996), write $D_g = \bigcup_{m\geq 1} \bigcap_{k \geq 1} G_{m,k}$ with

$$G_{m,k} = \{x \in \mathbb{D} : \text{there exist } y, z \in \mathbb{D} \text{ such that } d(x, y) < 1/k,\ d(x, z) < 1/k \text{ and } e(g(y), g(z)) > 1/m\}.$$

Every set $G_{m,k}$ is open. Therefore, D_g is a Borel set.

Let $\mathcal{A}$ be the collection of subsets A of $\mathbb{E}$ such that $g^{-1}(A) \setminus D_g$ is a Borel set in $\mathbb{D}$. Since D_g and $\mathbb{D} \setminus D_g$ are Borel sets in $\mathbb{D}$, the collection $\mathcal{A}$ is a σ-field on $\mathbb{E}$. Hence it is sufficient to show that $\mathcal{A}$ contains the open sets.

The restriction $g|_{\mathbb{D}\setminus D_g}$ of g to $\mathbb{D}\setminus D_g$ is continuous. Hence, for $G \subset \mathbb{E}$ open, the set $g^{-1}(G) \setminus D_g = (g|_{\mathbb{D}\setminus D_g})^{-1}(G)$ is open in $\mathbb{D} \setminus D_g$ with respect to the relative topology. But then $g^{-1}(G) \setminus D_g$ can be written as the intersection of an open set in $\mathbb{D}$ and $\mathbb{D} \setminus D_g$, which is a Borel set. $\square$

In our version of the continuous mapping theorem, the map g is defined on $\mathbb{D}$ and not on $[\mathbb{D}]$, that is, even if $d(x, y) = 0$, it may occur that $g(x) \neq g(y)$. Therefore, we cannot directly apply Theorem 1.3.6 in van der Vaart and Wellner (1996). Nevertheless, the proof is inspired from the proof of that theorem.

THEOREM A.9 (Continuous mapping). *Let $(\mathbb{D}, d)$ be a semimetric space and let $(\mathbb{E}, e)$ be a metric space. Let $g : \mathbb{D} \mapsto \mathbb{E}$ be arbitrary and let D_g be the set of $x \in \mathbb{D}$ such that g is not continuous in x. Let $(\Omega_\alpha, \mathcal{A}_\alpha, P_\alpha)$, $\alpha \in A$, be a net of probability spaces and let $X_\alpha : \Omega_\alpha \mapsto \mathbb{D}$ be arbitrary maps; let $(\Omega, \mathcal{A}, P)$ be a probability space and let $X : \Omega \mapsto \mathbb{D}$ be Borel measurable. If $[X_\alpha] \rightsquigarrow [X]$ in $[\mathbb{D}]$ and if $X(\Omega) \subset \mathbb{D} \setminus D_g$, then $g(X) : \Omega \mapsto \mathbb{E}$ is Borel measurable and $g(X_\alpha) \rightsquigarrow g(X)$ in $\mathbb{E}$.*

PROOF. Let L be the law of X on $\mathbb{D}$, that is, $L(B) = P(X \in B)$ for $B \in \mathcal{B}(\mathbb{D})$. Then $L \circ [\,\cdot\,]^{-1}$ is the law of $[X] = [\,\cdot\,](X)$ on $[\mathbb{D}]$. Clearly, $L(D_g) = 0$.

Since $X(\Omega) \subset \mathbb{D} \setminus D_g$, we have, for every Borel set B of $\mathbb{E}$,

$$\{g(X) \in B\} = \{X \in g^{-1}(B)\} = \{X \in g^{-1}(B) \setminus D_g\}.$$

The right-hand side is a measurable subset of Ω by Lemma A.8. Hence $g(X) : \Omega \mapsto \mathbb{E}$ is Borel measurable. In addition, we see that the law of $g(X)$ on $\mathbb{E}$ is given by $L(g^{-1}(\,\cdot\,) \setminus D_g)$.

For ease of notation, we write $\mathbb{P}^*$ rather than $\mathbb{P}^*_\alpha$, stars denoting outer probabilities. According to the Portmanteau theorem (van der Vaart and Wellner, 1996, Theorem 1.3.4(iii)), a necessary and sufficient condition for weak convergence is that $\limsup_\alpha \mathbb{P}^*(g(X_\alpha) \in F) \le \mathbb{P}(g(X) \in F)$ for every closed set $F \subset \mathbb{E}$.

Let $F \subset \mathbb{E}$ be closed. We have

$$\begin{aligned} \mathbb{P}^*\big(g(X_\alpha) \in F\big) &= \mathbb{P}^*\big(X_\alpha \in g^{-1}(F)\big) \\ &\le \mathbb{P}^*\big(X_\alpha \in \mathrm{cl}(g^{-1}(F))\big) = \mathbb{P}^*\big([X_\alpha] \in [\mathrm{cl}(g^{-1}(F))]\big). \end{aligned}$$

The last equality follows from the fact that for every Borel set B of $\mathbb{D}$, we have $x \in B$ iff $[x] \in [B]$. Since $\mathrm{cl}(g^{-1}(F))$ is closed in $\mathbb{D}$, the set $[\mathrm{cl}(g^{-1}(F))]$ is closed in $[\mathbb{D}]$. By weak convergence $[X_\alpha] \rightsquigarrow L \circ [\,\cdot\,]^{-1}$, we find

$$\limsup_\alpha \mathbb{P}^*\big(g(X_\alpha) \in F\big) \le L \circ [\,\cdot\,]^{-1}\big([\mathrm{cl}(g^{-1}(F))]\big) \overset{(1)}{=} L\big(\mathrm{cl}(g^{-1}(F))\big)$$

$$\overset{(2)}{\le} L(g^{-1}(F) \cup D_g) \overset{(3)}{=} L(g^{-1}(F) \setminus D_g) \overset{(4)}{=} \mathbb{P}\big(g(X) \in F\big).$$

Explanations: (1) since $[\,\cdot\,]^{-1}([B]) = B$ for every Borel set B of $\mathbb{D}$; (2) since $\mathrm{cl}(g^{-1}(F)) \subset g^{-1}(F) \cup D_g$, which follows from the fact that if g is continuous in x and if $g(x) \in \mathbb{E} \setminus F$ (open), then there is a neighborhood U of x in $\mathbb{D}$ such that $g(U) \subset \mathbb{E} \setminus F$ and thus x does not belong to $\mathrm{cl}(g^{-1}(F))$; (3) since $L(D_g) = 0$; (4) since the law of $g(X)$ is $L(g^{-1}(\,\cdot\,) \setminus D_g)$. □

LEMMA A.10 (Slutsky). *Let $(\mathbb{T}, d)$ be a locally compact, separable metric space. Let $X_n, Y_n : \Omega_n \mapsto \ell^\infty_{\mathrm{loc}}(\mathbb{T})$ be arbitrary maps and let $X : \Omega \mapsto \ell^\infty_{\mathrm{loc}}(\mathbb{T})$ be Borel measurable with respect to the hypi-semimetric. If $[X_n] \rightsquigarrow [X]$ and $[Y_n] \rightsquigarrow [0]$ in $L^\infty_{\mathrm{loc}}(\mathbb{T})$, then $[X_n + Y_n] \rightsquigarrow [X]$ in $L^\infty_{\mathrm{loc}}(\mathbb{T})$.*

PROOF. Let $\mathbb{D}$ be the product space $\ell^\infty_{\mathrm{loc}}(\mathbb{T}) \times \ell^\infty_{\mathrm{loc}}(\mathbb{T})$ equipped with the box semimetric $d_{\mathbb{D}}((f_1, f_2), (f_1', f_2')) = \max\{d_{\mathrm{hypi}}(f_1, f_1'), d_{\mathrm{hypi}}(f_2, f_2')\}$. Let $\mathbb{E}$ be the metric space $L^\infty_{\mathrm{loc}}(\mathbb{T})$ equipped with the hypi-metric on equivalence classes. Consider the map $g : \mathbb{D} \mapsto \mathbb{E}$ given by $g(f_1, f_2) = [f_1 + f_2]$.

By Lemma A.4 and by continuity of the map $\ell^\infty_{\mathrm{loc}}(\mathbb{T}) \mapsto L^\infty_{\mathrm{loc}}(\mathbb{T}) : f \mapsto [f]$ with respect to d_{hypi}, the map g is continuous at pairs (f_1, f_2) such that f_2 is continuous, in particular if $f_2 = 0$. In view of the Continuous Mapping Theorem A.9, it is sufficient to show that $[(X_n, Y_n)] \rightsquigarrow [(X, 0)]$ in $[\mathbb{D}]$, where $[(f_1, f_2)]$ is the equivalence class of pairs (f_1', f_2') in $\mathbb{D}$ at distance zero from (f_1, f_2), and where $[\mathbb{D}]$ is the quotient space of all such equivalence classes equipped with the natural metric induced by $d_{\mathbb{D}}$.

The distance (in $[\mathbb{D}]$) between $[(X_n, Y_n)]$ and $[(X_n, 0)]$ is equal to the distance (in $\mathbb{D}$) between (X_n, Y_n) and $(X_n, 0)$, which is equal to $d_{\mathrm{hypi}}(Y_n, 0) = d_{\mathrm{hypi}}([Y_n], [0])$. Since $[Y_n] \rightsquigarrow [0]$, the latter distance converges to zero in outer probability; see Lemma 1.10.2(iii) in van der Vaart and Wellner (1996). By item (i) of the same lemma, the desired conclusion $[(X_n, Y_n)] \rightsquigarrow [(X, 0)]$ is a consequence from $[(X_n, 0)] \rightsquigarrow [(X, 0)]$, to be proven. But the latter convergence follows from the Continuous Mapping Theorem A.9 and continuity of the map $\ell^\infty_{\mathrm{loc}}(\mathbb{T}) \mapsto \mathbb{D} : f \mapsto [(f, 0)]$. □

A.3. Minimum L^2-distance estimators. In this section, we present a general treatment of minimum L^2-distance estimators that encompasses both the copula and the tail dependence function setting. As we will see, all the results depend only on the geometry of L^2-spaces. While the findings here are very much in the spirit of previous work in this direction, see, e.g., Tsukahara (2005), Beran (1984) and the references cited therein, we were not able to find a concise treatment that was suitable for our needs. For that reason, this section contains a rigorous, self-contained statement of some basic results about minimum L^2-distance estimation.

Let $\Theta \subset \mathbb{R}^m$ be open and let $\{G_\theta : \theta \in \Theta\}$ be a parametric family of square-integrable functions defined on some compact $K \subset \mathbb{R}^d$ with respect to some probability measure μ on K. Let $\|\cdot\|_{2,\mu}$ denote the L^2-norm. Consider a non-parametric estimator of G, say $\hat{G}_n$, assumed to be square-integrable too. Consider a minimum-distance estimator $\hat{\theta}_n$ of θ which satisfies

$$\|\hat{G}_n - G_{\hat{\theta}_n}\|_{2,\mu}^2 \le \inf_{\theta\in\Theta} \|\hat{G}_n - G_\theta\|_{2,\mu}^2 + \delta_n \tag{A.5}$$

for a nonnegative, possibly random sequence $\delta_n = o_p(1)$.

In what follows, $G \in L^2(K,\mu)$ denotes the true unknown function corresponding to the data-generating process. Our first assumption is needed to ensure weak consistency of $\hat{\theta}_n$. The assumption of uniform boundedness is obviously fulfilled for copulas and stable tail dependence functions.

CONDITION A.11. *The map $\theta \mapsto \|G_\theta - G\|_{2,\mu}$ has a unique minimum θ_0. Additionally, this minimum is well-separated, that is, for every open set $O \subset \Theta$ with $\theta_0 \in O$, we have $\inf_{\theta\notin O} \|G_\theta - G\|_{2,\mu} > \|G_{\theta_0} - G\|_{2,\mu}$. Moreover $\sup_{\theta\in\Theta} \|G_\theta\|_{2,\mu} < \infty$.*

Next, consider a generic condition on the weak convergence of $\hat{G}_n$. As before, weak convergence in $L^2(K,\mu)$ has to be considered in the metric spaces of equivalence classes of functions that are equal almost everywhere, but for convenience, we suppress this from the notation. For copulas and stable tail dependence functions, it is a consequence from weak hypi-convergence to a limit process which is μ-almost everywhere continuous.

CONDITION A.12. *There exists a positive sequence $\alpha_n \to \infty$ such that $\delta_n = o_p(\alpha_n^{-2})$ and, in $L^2(K,\mu)$,*

$$\mathbb{G}_n := \alpha_n(\hat{G}_n - G) \rightsquigarrow \mathbb{G}, \qquad n \to \infty.$$

Our first result deals with weak convergence of the estimator $\hat{\theta}_n$ under the assumption that G belongs to $\{G_\theta : \theta \in \Theta\}$.

CONDITION A.13. *There exists a* $\theta_0 \in \Theta$ *such that* $G = G_{\theta_0}$. *Additionally, for some* $\mathbb{R}^m$*-valued function* ∇G_{θ_0} *whose entries lie in* $L^2(K, \mu)$,

$$G_\theta - G_{\theta_0} = \nabla G_{\theta_0}^T(\theta - \theta_0) + o(|\theta - \theta_0|), \qquad \theta \to \theta_0,$$

in $L^2(K, \mu)$. *Moreover, the matrix* $\int \nabla G_{\theta_0} \nabla G_{\theta_0}^T \, d\mu$ *is non-singular.*

Condition A.13 can be viewed as an assumption on differentiability in quadratic mean. It typically follows from pointwise differentiability. Assume for instance that, for μ-almost all $\boldsymbol{x} \in K$, the function $\theta \mapsto G_\theta(\boldsymbol{x})$ is continuously differentiable in a neighborhood of θ_0 with derivative $dG(\theta, \boldsymbol{x})$. Then the L^2 expansion in Condition A.13 follows from majorized convergence if all entries of the vector $dG(\theta, \boldsymbol{x})$ are uniformly bounded in $\boldsymbol{x}$ and θ, or from van der Vaart (1998, Proposition 2.29) if the L^2-norms of the entries of $dG(\theta, \cdot)$ are finite and continuous in θ.

THEOREM A.14. *Under Conditions A.11, A.12 and A.13 we have*

$$\alpha_n(\hat{\theta}_n - \theta_0) \rightsquigarrow \left\{ \int \nabla G_{\theta_0} \nabla G_{\theta_0}^T \, d\mu \right\}^{-1} \int \nabla G_{\theta_0} \mathbb{G}_n \, d\mu.$$

PROOF. Let the functions $\mathbb{M}_n$ and $\mathbb{M}$ from Θ into $\mathbb{R}$ be defined by

$$\text{(A.6)} \qquad \mathbb{M}_n(\theta) = -\|G_\theta - \hat{G}_n\|_{2,\mu}^2, \qquad \mathbb{M}(\theta) = -\|G_\theta - G\|_{2,\mu}^2.$$

The proof will proceed by an application of Theorem 3.2.16 in van der Vaart and Wellner (1996). To this end, observe that

$$\mathbb{M}_n(\theta) - \mathbb{M}(\theta) = 2 \int (G_\theta - G)(\hat{G}_n - G) \, d\mu - \int (\hat{G}_n - G)^2 \, d\mu.$$

In particular, this implies that

$$\sup_{\theta \in \Theta} |\mathbb{M}_n(\theta) - \mathbb{M}(\theta)| \le 2\|G - \hat{G}_n\|_{2,\mu} \sup_{\theta \in \Theta} \|G_\theta - G\|_{2,\mu} + \|G - \hat{G}_n\|_{2,\mu}^2$$

which converges to 0 in (outer) probability by Condition A.12. Together with Condition A.11, an application of Corollary 3.2.3 in van der Vaart and Wellner (1996) yields the consistency of $\hat{\theta}_n$ for θ_0. Next, by Condition A.13,

$$\begin{aligned} \mathbb{M}(\theta) - \mathbb{M}(\theta_0) &= -\int (G_\theta - G_{\theta_0})^2 d\mu \\ &= (\theta - \theta_0)^T \int \nabla G_{\theta_0} \nabla G_{\theta_0}^T \, d\mu \, (\theta - \theta_0) + o(|\theta - \theta_0|^2). \end{aligned}$$

Finally, consider an arbitrary random sequence $\theta_n = \theta_0 + o_P^*(1)$ with $\theta_n \in \Theta$ for all $n \in \mathbb{N}$. We then obtain from Conditions A.12 and A.13 (note that the first equality holds without Condition A.13)

$$\begin{aligned}\text{(A.7)}\quad &\alpha_n\{\mathbb{M}_n(\theta_n) - \mathbb{M}(\theta_n)\} - \alpha_n\{\mathbb{M}_n(\theta_0) - \mathbb{M}(\theta_0)\}\\ &\quad = 2\int (G_{\theta_n} - G_{\theta_0})\,\mathbb{G}_n\,d\mu = 2\int \nabla G_{\theta_0}^T\,\mathbb{G}_n\,d\mu\,(\theta_n - \theta_0) + o_P^*(|\theta_n - \theta_0|).\end{aligned}$$

All assumptions of Theorem 3.2.16 in van der Vaart and Wellner (1996) are verified with $r_n = \alpha_n$, $Z_n = 2\int \nabla G_{\theta_0}\,\mathbb{G}_n\,d\mu$ and $V = -2\int \nabla G_{\theta_0}\,\nabla G_{\theta_0}^T\,d\mu$ (mind the footnote on page 300). The assertion follows. □

Next, consider misspecified models, that is, models where G does not belong to $\{G_\theta : \theta \in \Theta\}$. In this case, it is still possible to derive a weak convergence result for the estimator $\hat\theta_n$. The main difference is that G_{θ_0} is now the best approximation of G by functions from $\{G_\theta : \theta \in \Theta\}$. Additionally, we need a stronger smoothness assumption on the map $\theta \mapsto G_\theta$.

CONDITION A.15. *Assume that there exists an $\mathbb{R}^m$-valued function ∇G_{θ_0} and an $\mathbb{R}^{m\times m}$-valued function J_{θ_0} with all entries being elements of $L^2(K,\mu)$ such that*

$$G_{\theta_0+h} - G_{\theta_0} = \nabla G_{\theta_0}^T h + \frac{1}{2} h^T J_{\theta_0} h + o(|h|^2), \qquad h \to 0,$$

in $L^2(K,\mu)$. Additionally, assume non-singularity of the matrix

$$\int \bigl(\nabla G_{\theta_0}\,\nabla G_{\theta_0}^T + (G_{\theta_0} - G)\,J_{\theta_0}\bigr)\,d\mu.$$

Again, this assumption can be derived from conditions on pointwise differentiability of $\theta \mapsto G_\theta(\boldsymbol{x})$ in a similar way as explained after Condition A.13.

THEOREM A.16. *Under Conditions A.11, A.12 and A.15 we have*

$$\alpha_n(\hat\theta_n - \theta_0) \rightsquigarrow \left\{\int \bigl(\nabla G_{\theta_0}\,\nabla G_{\theta_0}^T + (G_{\theta_0} - G)\,J_{\theta_0}\bigr)\,d\mu\right\}^{-1} \int \nabla G_{\theta_0}\,\mathbb{G}_n\,d\mu.$$

PROOF. Recall the definitions of $\mathbb{M}_n$ and $\mathbb{M}$ in (A.6). Consistency of $\hat\theta_n$ for θ_0 follows by exactly the same arguments as in the proof of Theorem A.14. Moreover, under Condition A.15, we have

$$\begin{aligned}(G_{\theta_0+h} - G)^2 = (G_{\theta_0} - G)^2 &+ 2(G_{\theta_0} - G)\,\nabla G_{\theta_0}^T\,h\\ &+ \frac{1}{2}\,2h^T\bigl\{\nabla G_{\theta_0}\,\nabla G_{\theta_0}^T + (G_{\theta_0} - G)\,J_{\theta_0}\bigr\}\,h + o(|h|^2).\end{aligned}$$

Since the above representation holds in $L^2(K,\mu)$, we obtain a corresponding two-term Taylor expansion for $\mathbb{M}(\theta)-\mathbb{M}(\theta_0)$. Next, observe that the expression for $\alpha_n\{\mathbb{M}_n(\theta_n)-\mathbb{M}(\theta_n)\}-\alpha_n\{\mathbb{M}_n(\theta_0)-\mathbb{M}(\theta_0)\}$ in (A.7) continues to hold under the present assumptions. We thus have verified all the assumptions of Theorem 3.2.16 in van der Vaart and Wellner (1996) with $r_n=\alpha_n$, $Z_n=2\int \nabla G_{\theta_0}\,\mathbb{G}_n\,d\mu$ and $V=-2\int\{\nabla G_{\theta_0}\,\nabla G_{\theta_0}^T+(G_{\theta_0}-G)\,J_{\theta_0}\}d\mu$. The assertion follows. $\square$

APPENDIX B: PROOFS

B.1. Proofs for Section 2.

PROOF OF PROPOSITION 2.2. Suppose that $\sup_{x\in K}|f_n(x)-f(x)|$ does not converge to 0. Then there exists $\varepsilon>0$, a subsequence n_k, and points $x_{n_k}\in K$ such that

$$|f_{n_k}(x_{n_k})-f(x_{n_k})|\geq\varepsilon,\qquad k\geq 1. \tag{B.1}$$

Passing to a further subsequence if necessary, we may assume that x_{n_k} converges to some point $x\in K$ as $k\to\infty$. Since f is continuous in x, the pointwise criteria in Proposition 2.1(ii) together with hypi-convergence of f_{n_k} to f imply that $f_{n_k}(x_{n_k})\to f(x)$ as $k\to\infty$. Since also $f(x_{n_k})\to f(x)$, we arrive at a contradiction with (B.1). $\square$

PROOF OF PROPOSITION 2.3. It is sufficient to consider convergence of infima. By definition, $f_{n,\wedge}$ epi-converges to $f_\wedge$. Let $\bar G$ denote the topological closure of G in $\mathbb{T}$. Since $\inf f_\wedge(\bar G)=\inf f_\wedge(G)$ by continuity of f on the boundary of G, we have $\inf f_{n,\wedge}(G)\to\inf f_\wedge(G)$ (Molchanov, 2005, Chapter 5, Proposition 3.2(iii)). If $G=\mathbb{T}$, the same holds true without imposing continuity of f. Finally, since G is open, we have $\inf f_{n,\wedge}(G)=\inf f_n(G)$ and similarly for f. $\square$

LEMMA B.1. *If f_n hypi-converges to f, then for every compact $K\subset\mathbb{T}$,*

$$\limsup_{n\to\infty}\sup_{x\in K}\max\{|f_{n,\wedge}(x)|,|f_{n,\vee}(x)|\}\leq\max\left\{\sup_{x\in K}f_\vee(x),-\inf_{x\in K}f_\wedge(x)\right\}.$$

PROOF. Since $f_{n,\wedge}\leq f_{n,\vee}$, we have

$$\sup_{x\in K}\max\{|f_{n,\wedge}(x)|,|f_{n,\vee}(x)|\}\leq\max\left\{\sup_{x\in K}f_{n,\vee}(x),-\inf_{x\in K}f_{n,\wedge}(x)\right\}$$

and thus

$$\limsup_{n\to\infty} \sup_{x\in K} \max\{|f_{n,\wedge}(x)|, |f_{n,\vee}(x)|\}$$
$$\leq \max\left\{\limsup_{n\to\infty} \sup_{x\in K} f_{n,\vee}(x), -\liminf_{n\to\infty} \inf_{x\in K} f_{n,\wedge}(x)\right\}.$$

The statement follows from epi-convergence of $f_{n,\wedge}$ to $f_\wedge$, hypo-convergence of $f_{n,\vee}$ to $f_\vee$, and Proposition 3.2(iii), Chapter 5, in Molchanov (2005). □

Proof of Proposition 2.4. By the assumptions and by the pointwise criteria in Proposition 2.1(ii), it follows that $f_{n,\wedge}(x)$ and $f_{n,\vee}(x)$ converge to $f(x) = f_\wedge(x) = f_\vee(x)$ at μ-almost every $x \in \mathbb{T}$. Let $K \subset \mathbb{T}$ be compact such that $\mu(\mathbb{T}\setminus K) = 0$. By Lemma B.1, the functions $|f_{n,\wedge} - f_\wedge|^p$ and $|f_{n,\vee} - f_\vee|^p$ are uniformly bounded on K. Apply the bounded convergence theorem to see that both integrals $\int |f_{n,\wedge} - f_\wedge|^p \,\mathrm{d}\mu$ and $\int |f_{n,\vee} - f_\vee|^p \,\mathrm{d}\mu$ converge to zero. Minkowski's inequality yields $\int |f_{n,\vee} - f_{n,\wedge}|^p \,\mathrm{d}\mu \to 0$. Similarly, we can replace $f_\wedge$ ($f_{n,\wedge}$) and $f_\vee$ ($f_{n,\vee}$) by any measurable function f^* (f_n^*) such that $f_\wedge \leq f^* \leq f_\vee$ ($f_{n,\wedge} \leq f_n^* \leq f_{n,\vee}$), in particular by f (f_n) itself, should it be Borel measurable. □

B.2. Proofs for Section 3.

Proof of Corollary 3.1. By Proposition 2.2, the restriction map

$$\cdot|_K : \ell^\infty_{\mathrm{loc}}(\mathbb{T}) \mapsto \ell^\infty(K) : f \mapsto f|_K$$

is continuous in functions $f \in \ell^\infty_{\mathrm{loc}}(\mathbb{T})$ that are continuous on every $x \in K$ with respect to the hypi-semimetric on the domain $\ell^\infty_{\mathrm{loc}}(\mathbb{T})$ and the supremum distance on the image space $\ell^\infty(K)$. Since with probability one, X is continuous in every $x \in K$, we can, changing versions if necessary, assume that the range of X is actually disjoint with the set of discontinuity points of g. Apply the Continuous Mapping Theorem A.9 to conclude. □

Proof of Corollary 3.2. Consider the maps $[(\,\cdot\,)_\wedge]_\mu$ and $[(\,\cdot\,)_\vee]_\mu$ from $\ell^\infty_{\mathrm{loc}}(\mathbb{T})$ into $L^p(\mu)$ sending f to $[f_\wedge]_\mu$ and $[f_\vee]_\mu$, respectively. If f is continuous μ-almost everywhere, then $f_\wedge = f = f_\vee$ μ-almost everywhere, implying $[f_\wedge]_\mu = [f_\vee]_\mu$; moreover, by Proposition 2.4, the maps $[(\,\cdot\,)_\wedge]_\mu$ and $[(\,\cdot\,)_\vee]_\mu$ are continuous at such f.

Changing versions if necessary, we can assume that all trajectories of X are continuous μ-almost everywhere. It follows that $[X_\wedge]_\mu = [X_\vee]_\mu$. By the Continuous Mapping Theorem A.9, both $[X_{n,\wedge}]_\mu$ and $[X_{n,\vee}]_\mu$ converge weakly to $[X_\wedge]_\mu = [X_\vee]_\mu$ in $L^p(\mu)$.

Finally, the map $\ell^\infty_{\mathrm{loc}}(\mathbb{T}) \mapsto \mathbb{R}$ sending f to $\int |f_\vee - f_\wedge|^p \, \mathrm{d}\mu$ is continuous in functions f that are continuous μ-almost everywhere, in which case the integral is zero; this is a consequence of Proposition 2.4. Again, we can apply the Continuous Mapping Theorem A.9 to conclude that $\int |X_{n,\wedge} - X_{n,\vee}|^p \, \mathrm{d}\mu$ converges weakly to 0. The limit being constant, the convergence also takes place in outer probability (van der Vaart and Wellner, 1996, Lemma 1.10.2(iii)). □

B.3. Proofs for Section 4.

Proof of Theorem 4.5. For the sake of a clear exposition, we split the proof into two propositions. First, Proposition B.3 shows that Condition 4.3 implies a certain abstract hypi-differentiability property stated in Condition B.2. Then, Proposition B.4 shows that the latter condition suffices to obtain the conclusion of Theorem 4.5. □

Condition B.2 (Hypi-differentiability of C). *Define the set*

$$\mathcal{W}(t) := \left\{ a \in \{\ell^\infty([0,1])\}^d : \boldsymbol{u} + ta(\boldsymbol{u}) \in [0,1]^d \ \forall \boldsymbol{u} \in [0,1]^d \right\},$$

where $a(\boldsymbol{u}) = (a_1(u_1), \ldots, a_d(u_d))$. Whenever $t_n \searrow 0$, $t_n \neq 0$, and $a_n = (a_{n1}, \ldots, a_{nd}) \in \{\ell^\infty([0,1])\}^d$ converges uniformly to $a \in \mathcal{W} := \{\mathcal{C}([0,1])\}^d$ (i.e., $\|a_{nj} - a_j\|_\infty \to 0$ for all $j = 1, \ldots, d$) such that $a_n \in \mathcal{W}(t_n)$ for all $n \in \mathbb{N}$, the functions

$$[0,1]^d \mapsto \mathbb{R} : u \mapsto t_n^{-1}\{C(\boldsymbol{u} + t_n a_n(\boldsymbol{u})) - C(\boldsymbol{u})\}.$$

converge in $(\ell^\infty([0,1]^d), d_{\mathrm{hypi}})$ to some limit $\mathrm{d}C_a$.

Proposition B.3. *A copula C satisfying Condition 4.3 also satisfies the hypi-differentiability Condition B.2 with derivative*

$$\mathrm{d}C_a(\boldsymbol{u}) = \sup_{\varepsilon > 0} \inf \left\{ \sum_{j=1}^d \dot{C}_j(\boldsymbol{v})\, a_j(v_j) : \boldsymbol{v} \in \mathcal{S}, \, |\boldsymbol{v} - \boldsymbol{u}| < \varepsilon \right\}, \qquad \boldsymbol{u} \in [0,1]^d.$$

Conversely, it is an open problem whether there exists a copula that satisfies Condition B.2 but violates Condition 4.3. According to the next proposition, Condition B.2 can replace Condition 4.3 in Theorem 4.5.

Proposition B.4. *Suppose that Condition 4.4 holds and that C satisfies Condition B.2. Then,*

$$\mathbb{C}_n \rightsquigarrow \mathbb{C} = \alpha + \mathrm{d}C_{(-\alpha_1, \ldots, -\alpha_d)}$$

in $(L^\infty([0,1]^d), d_{\mathrm{hypi}})$.

PROOF OF PROPOSITION B.3. Let $t_n \searrow 0$ and let $a_n \in \mathcal{W}(t_n)$ converge uniformly to $a \in \mathcal{W}$. As in Condition B.2, we use the notation $a_n(\boldsymbol{u}) = (a_{n1}(u_1), \dots, a_{nd}(u_d))$. We have to prove epi- and hypo-convergence of

$$\boldsymbol{u} \mapsto F_n(\boldsymbol{u}) := t_n^{-1}\{C(\boldsymbol{u} + t_n a_n(\boldsymbol{u})) - C(\boldsymbol{u})\}$$

to $F_\wedge$ and $F_\vee$, respectively, where $F = \mathrm{d}C_a$. By Corollary A.7 it suffices to show that

(i) $\forall \boldsymbol{u} \in [0,1]^d : \forall \boldsymbol{u}_n \to \boldsymbol{u} : \liminf_{n\to\infty} F_n(\boldsymbol{u}_n) \geq F_\wedge(\boldsymbol{u})$,
(ii) $\forall \boldsymbol{u} \in [0,1]^d : \forall \boldsymbol{u}_n \to \boldsymbol{u} : \limsup_{n\to\infty} F_n(\boldsymbol{u}_n) \leq F_\vee(\boldsymbol{u})$.

We begin with the proof of (i) and fix a point $\boldsymbol{u} \in [0,1]^d$ and a sequence $\boldsymbol{u}_n \to \boldsymbol{u}$. Choose $\varepsilon > 0$ and let $|\cdot|_1$ denote the L_1-norm on $\mathbb{R}^d$. Due to Lemma B.5 we may choose

$$\boldsymbol{u}_n^\star \in \{\boldsymbol{v} \in [0,1]^d : |\boldsymbol{u}_n - \boldsymbol{v}|_1 \leq \varepsilon t_n/2\}$$

and

$$\boldsymbol{u}_n^\circ \in \{\boldsymbol{v} \in [0,1]^d : |\boldsymbol{u}_n + t_n a_n(\boldsymbol{u}_n) - \boldsymbol{v}|_1 \leq \varepsilon t_n/2\}$$

such that, for the path

$$\gamma_n(s) = (1-s)\boldsymbol{u}_n^\star + s\boldsymbol{u}_n^\circ, \quad s \in [0,1],$$

the set $\{s \in [0,1] : \gamma_n(s) \notin \mathcal{S}\}$ has Lebesgue-measure zero. Define

$$f_n(s) = t_n^{-1} C(\gamma_n(s)), \qquad s \in [0,1],$$

and note that

$$\begin{aligned} &|\{f_n(1) - f_n(0)\} - F_n(\boldsymbol{u}_n)| \\ &\quad = t_n^{-1}\,|C(\boldsymbol{u}_n^\circ) - C(\boldsymbol{u}_n + t_n a_n(\boldsymbol{u}_n)) - C(\boldsymbol{u}_n^\star) + C(\boldsymbol{u}_n)| \\ &\qquad \leq t_n^{-1}\,\{|\boldsymbol{u}_n + t_n a_n(\boldsymbol{u}_n) - \boldsymbol{u}_n^\circ|_1 + |\boldsymbol{u}_n^\star - \boldsymbol{u}_n|_1\} \leq \varepsilon \end{aligned}$$

by Lipschitz-continuity of C. Lipschitz-continuity of C also implies absolute continuity of the function f_n, which allows us to choose $\boldsymbol{v}_n \in \gamma_n([0,1]) \cap \mathcal{S}$

such that

$$
\begin{aligned}
\varepsilon + F_n(u_n) &\ge f_n(1) - f_n(0) = \int_0^1 f_n'(s)\,ds \\
&= \sum_{j=1}^d t_n^{-1}(u_{nj}^\circ - u_{nj}^\star) \int_0^1 \dot{C}_j(\gamma_n(s))\,ds \\
&= \sum_{j=1}^d \left[a_{nj}(u_{nj}) + t_n^{-1}\left\{u_{nj}^\circ - u_{nj}^\star - t_n a_{nj}(u_{nj})\right\}\right] \int_0^1 \dot{C}_j(\gamma_n(s))\,ds \\
&\ge \inf_{s:\gamma_n(s)\in\mathcal{S}} \sum_{j=1}^d a_{nj}(u_{nj})\dot{C}_j(\gamma_n(s)) - \varepsilon \\
&\ge \sum_{j=1}^d a_{nj}(v_{nj})\dot{C}_j(\boldsymbol{v}_n) + \sum_{j=1}^d \{a_{nj}(u_{nj}) - a_{nj}(v_{nj})\}\dot{C}_j(\boldsymbol{v}_n) - 2\varepsilon \\
&\ge \sum_{j=1}^d a_{nj}(v_{nj})\dot{C}_j(\boldsymbol{v}_n) - 3\varepsilon = F(\boldsymbol{v}_n) - 3\varepsilon = F_\wedge(\boldsymbol{v}_n) - 3\varepsilon
\end{aligned}
$$

for sufficiently large n, where we have used the bounds $0 \le \dot{C}_j \le 1$, uniform convergence of a_{nj} to a_j, uniform continuity of a_j and the fact that F is continuous in $\boldsymbol{v}_n$. Hence, by lower semi-continuity of $F_\wedge$,

$$\liminf_{n\to\infty} F_n(\boldsymbol{u}_n) \ge F_\wedge(\boldsymbol{u}) - 4\varepsilon$$

and as $\varepsilon > 0$ was arbitrary the assertion in (i) follows.

The proof of (ii) is analogous. In the main inequality chain, all signs can be reversed if the infimum is replaced by a supremum and upon noting that on $\mathcal{S}$, the functions F, $F_\wedge$ and $F_\vee$ are equal. □

LEMMA B.5. *Let $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^d$ be two distinct points and denote by $H_{\boldsymbol{u}}$ and $H_{\boldsymbol{v}}$ the hyperplanes being orthogonal to $\boldsymbol{u} - \boldsymbol{v}$ and passing through $\boldsymbol{u}$ and $\boldsymbol{v}$, respectively. For $\delta > 0$, set $H_{\boldsymbol{u}}^\delta = H_{\boldsymbol{u}} \cap B_1(\boldsymbol{u}, \delta)$ and $H_{\boldsymbol{v}}^\delta = H_{\boldsymbol{v}} \cap B_1(\boldsymbol{v}, \delta)$, where $B_1(\boldsymbol{u}, \delta)$ denotes the unit ball of radius δ centered at $\boldsymbol{u}$ with respect to the $\|\cdot\|_1$-norm. Finally, let $\mathcal{Z}$ denote the cylinder with top area equal to $H_{\boldsymbol{u}}^\delta$ and bottom area equal to $H_{\boldsymbol{v}}^\delta$, i.e.,*

$$\mathcal{Z} = \{\boldsymbol{y} + s(\boldsymbol{v} - \boldsymbol{u}) : \boldsymbol{y} \in H_{\boldsymbol{u}}^\delta,\ s \in [0,1]\}.$$

Let $\mathcal{D}$ be a Lebesgue-null set in $\mathbb{R}^d$ and define, for any $\boldsymbol{y} \in H_{\boldsymbol{u}}^\delta$,

$$\mathcal{Z}_{\boldsymbol{y}}^{\mathcal{D}} = \{s \in \mathbb{R} : \boldsymbol{y} + s(\boldsymbol{v} - \boldsymbol{u}) \in \mathcal{Z} \cap \mathcal{D}\}.$$

Then $\mathcal{Z}_{\boldsymbol{y}}^{\mathcal{D}}$ is a one-dimensional Lebesgue-null set for almost all $\boldsymbol{y} \in H_{\boldsymbol{u}}^\delta$.

PROOF OF LEMMA B.5. By affine transformation, it suffices to prove the result for the standard cylinder, i.e., for $\boldsymbol{u} = (0, \dots, 0)$ and $\boldsymbol{v} = (0, \dots, 0, 1)$, such that

$$\mathcal{Z} = \{\boldsymbol{x} \in \mathbb{R}^{d-1} : |\boldsymbol{x}|_1 \le r\} \times [0,1]$$

for some $r > 0$. By a special case of Fubini's theorem known as Cavalieri's principle,

$$\lambda_d(\mathcal{Z} \cap \mathcal{D}) = \int_{\mathbb{R}^{d-1}} \lambda_1(\{s \in \mathbb{R} : (\boldsymbol{x}, s) \in \mathcal{Z} \cap \mathcal{D}\}) \, d\boldsymbol{x} = \int_{\mathbb{R}^{d-1}} \lambda_1(\mathcal{Z}^{\mathcal{D}}_{(\boldsymbol{x},0)}) \, d\boldsymbol{x}$$

where λ_k denotes k-dimensional Lebesgue measure. Since the expression on the left-hand side is equal to 0, the assertion follows. $\square$

PROOF OF PROPOSITION B.4. Recall the definition of $\beta_n = (\beta_{n1}, \dots, \beta_{nj})$, with $\beta_{nj} = \sqrt{n}(G^{-}_{nj} - \mathrm{id}_{[0,1]})$. It follows from Condition 4.4 and the functional delta method for the inverse mapping, also known as Vervaat's Lemma, that

$$(\alpha_n, \beta_n) = (\alpha_n, \beta_{n1}, \dots, \beta_{nd}) \rightsquigarrow (\alpha, -\alpha_1, \dots, -\alpha_d)$$

in $\ell^\infty([0,1]^d) \times \{\ell^\infty([0,1])\}^d$, with respect to the supremum distance in each coordinate. Note that we can write $\mathbb{C}_n = g_n(\alpha_n, \beta_n)$, where $g_n : \ell^\infty([0,1]^d) \times \mathcal{W}(1/\sqrt{n}) \to (L^\infty([0,1]^d), d_{\mathrm{hypi}})$ is defined as

$$g_n(a, b) = a(\mathrm{id}_{[0,1]^d} + b/\sqrt{n}) + \sqrt{n}\{C(\mathrm{id}_{[0,1]^d} + b/\sqrt{n}) - C\}. \tag{B.2}$$

Exploiting Condition B.2 and Lemma A.4 (recall that α is continuous almost surely), the assertion follows from the extended continuous mapping Theorem, see Theorem 1.11.1 in van der Vaart and Wellner (1996). $\square$

PROOF OF PROPOSITION 4.6. We restrict ourselves to the case $M = 1$, the proof in the general case being just a notationally more involved adaptation. Moreover, we only consider the resampling bootstrap as the assertion about the multiplier bootstrap can be proved analogously.

Passing, without loss of generality, to uniform $(0,1)$ margins, we have $\mathbb{C}^{[1]}_n = \sqrt{n}\{G^{[1]}_n(\boldsymbol{G}^{[1]-}_n) - G_n(\boldsymbol{G}^{-}_n)\}$, where, for any $\boldsymbol{u} \in [0,1]^d$,

$$G^{[1]}_n(\boldsymbol{u}) = \frac{1}{n} \sum_{i=1}^{n} W^{[1]}_{ni} \, \mathbb{1}(\boldsymbol{U}_i \le \boldsymbol{u}).$$

Let $\alpha_n^{[1]}$ denote the corresponding bootstrap empirical process of the (unobservable) sample $\boldsymbol{U}_1, \ldots, \boldsymbol{U}_n$, formally defined, for any $\boldsymbol{u} \in [0,1]^d$, as

$$\alpha_n^{[1]}(\boldsymbol{u}) = \sqrt{n}\{G_n^{[1]}(\boldsymbol{u}) - G_n(\boldsymbol{u})\} = \frac{1}{\sqrt{n}} \sum_{i=1}^n (W_{ni}^{[1]} - 1)\mathbb{1}(\boldsymbol{U}_i \le \boldsymbol{u})$$
$$= \frac{1}{\sqrt{n}} \sum_{i=1}^n (W_{ni}^{[1]} - 1)\{\mathbb{1}(\boldsymbol{U}_i \le \boldsymbol{u}) - C(\boldsymbol{u})\}.$$

Assembling results from van der Vaart and Wellner (1996), we have

$$(\alpha_n, \alpha_n^{[1]}) \rightsquigarrow (\alpha, \alpha^{[1]}) \quad \text{in} \;\; (\ell^\infty([0,1]^d), \|\cdot\|_\infty)^2, \tag{B.3}$$

where $\alpha^{[1]}$ denotes an independent copy of α. This follows from Corollary 2.9.3 in van der Vaart and Wellner (1996) and by Poissonization as in the proof of Theorem 3.6.1 in that reference.

The remaining part the proof consists of transferring the result in (B.3) to weak convergence of $(\mathbb{C}_n, \mathbb{C}_n^{[1]})$ in $(L^\infty([0,1]^d), d_{\text{hypi}})^2$ by similar arguments as in the proof of Theorem 4.5. First, exactly along the lines of the proof of Proposition B.3, we can show that Condition 4.3 implies the following abstract differentiability condition with $\mathrm{d}C_{b-a}$ as defined in Theorem 4.5. Recall the definitions of $\mathcal{W}(t)$ and $\mathcal{W}$ from Condition B.2.

Condition B.6. *Whenever $t_n \searrow 0$, $t_n \neq 0$ and $a_n, b_n \in \mathcal{W}(t_n)$ converge uniformly to $a, b \in \mathcal{W}$, respectively, the functions*

$$[0,1]^d \mapsto \mathbb{R} : u \mapsto t_n^{-1}\{C(\boldsymbol{u} + t_n b_n(\boldsymbol{u})) - C(\boldsymbol{u} + t_n a_n(\boldsymbol{u}))\}.$$

converge in $(\ell^\infty([0,1]^d, d_{\text{hypi}})$ to some limit, say $\mathrm{d}C_{b-a}$, which only depends on the difference $b - a$.

Now, from (B.3) and the continuous mapping Theorem, we obtain

$$(\alpha_n, \gamma_n) = \sqrt{n}\left\{(G_n, G_n^{[1]}) - (C, C)\right\} \rightsquigarrow (\alpha, \alpha + \alpha^{[1]}) = (\alpha, \gamma)$$

with respect to the supremum norm. In analogy to the definition of β_n in Section 4, set $\delta_n = (\delta_{n1}, \ldots, \delta_{nd})$ with the empirical quantile process $\delta_{nj} = \sqrt{n}\{G_{nj}^{[1]-} - \mathrm{id}_{[0,1]}\}$. Note that, by the functional delta for the inverse mapping,

$$(\alpha_n, \beta_{n1}, \ldots, \beta_{nd}, \gamma_n, \delta_{n1}, \ldots, \delta_{nd}) \rightsquigarrow (\alpha, -\alpha_1, \ldots, -\alpha_d, \gamma, -\gamma_1, \ldots, -\gamma_d)$$

with respect to the supremum distance. We can write

$$\begin{aligned}
\mathbb{C}_n^{[1]} &= \gamma_n(\boldsymbol{G}_n^{[1]-}) - \alpha_n(\boldsymbol{G}_n^{-}) + \sqrt{n}\{C(\boldsymbol{G}_n^{[1]-}) - C(\boldsymbol{G}_n^{-})\} \\
&= \gamma_n(\mathrm{id}_{[0,1]} + \delta_n/\sqrt{n}) - \alpha_n(\mathrm{id}_{[0,1]} + \beta_n/\sqrt{n}) \\
&\qquad + \sqrt{n}\{C(\mathrm{id}_{[0,1]} + \delta_n/\sqrt{n}) - C(\mathrm{id}_{[0,1]} + \beta_n/\sqrt{n})\} \\
&=: h_n(\alpha_n, \beta_n, \gamma_n, \delta_n),
\end{aligned}$$

and hence $(\mathbb{C}_n, \mathbb{C}_n^{[1]}) = (g_n(\alpha_n, \beta_n), h_n(\alpha_n, \beta_n, \gamma_n, \delta_n))$, where g_n is defined in (B.2). By a similar reasoning as in the proof of Theorem 4.5, Condition B.6 together with Lemma A.3 and the extended continuous mapping Theorem finally implies that

$$\begin{aligned}
(\mathbb{C}_n, \mathbb{C}_n^{[1]}) &\rightsquigarrow (\mathbb{C}, \gamma - \alpha + \mathrm{d}C_{(-(\gamma_1-\alpha_1),\dots,-(\gamma_d-\alpha_d))}) \\
&\qquad = (\mathbb{C}, \alpha^{[1]} + \mathrm{d}C_{(-\alpha_1^{[1]},\dots,-\alpha_d^{[1]})}) = (\mathbb{C}, \mathbb{C}^{[1]})
\end{aligned}$$

in $(L^\infty([0,1]^d), d_{\mathrm{hypi}})^2$ as asserted. □

PROOF OF PROPOSITION 4.7. Both assertions follow from the extended continuous mapping Theorem by the same reasoning as in the proof of Theorem 4.5 if we show that $\alpha_n^{(n)} \rightsquigarrow \alpha$ in $(\ell^\infty([0,1]^d), \|\cdot\|_\infty)$. The latter can be deduced from Theorem 2.8.9 in van der Vaart and Wellner (1996):

- the required measurability conditions follow from Example 2.3.4 in the latter reference;
- the uniform entropy condition follows from the fact that lower orthants form a VC-class;
- Condition (2.8.5) in the Theorem is implied by assumption (4.6) on the copula $C^{(n)}$;
- the conditions on the envelope $F \equiv 1$ are trivially satisfied.

□

B.4. Proofs for Section 5.

PROOF OF THEOREM 5.1. As in Einmahl, Krajina and Segers (2012), set

$$\boldsymbol{V}_i = (V_{i1}, \dots, V_{id}) = (1 - F_1(X_{i1}), \dots, 1 - F_d(X_{id})) \quad (i = 1, \dots, n)$$

and $S_n(\boldsymbol{x}) = (S_{n1}(x_1), \dots, S_{nd}(x_d))$, $\boldsymbol{x} \in [0,\infty)^d$, where, for $j = 1, \dots, d$,

$$S_{nj}(x_j) = \frac{n}{k} Q_{nj}\left(\frac{kx_j}{n}\right) \quad (x_j \geq 0), \quad Q_{nj}(v_j) = V_{\lceil nv_j \rceil:n,j} \quad (v_j \in [0,1]),$$

and where $V_{1:n,j} \le \ldots, \le V_{n:n,j}$ are the order statistics of $V_{1j}, \ldots, V_{nj}$ and where $\lceil x \rceil$ is the smallest integer not smaller than x. Moreover, set

$$\tilde{L}_n(\boldsymbol{x}) = \frac{1}{k} \sum_{i=1}^{n} \mathbb{1}\left(V_{i1} < \frac{kx_1}{n} \text{ or } \ldots \text{ or } V_{id} < \frac{kx_d}{n} \right)$$
$$\tilde{\mu}_n(\boldsymbol{x}) = \mathbb{E}[\tilde{L}_n(\boldsymbol{x})] = \frac{n}{k} \mathbb{P}\left(V_{i1} < \frac{kx_1}{n} \text{ or } \ldots \text{ or } V_{id} < \frac{kx_d}{n} \right),$$

and note that

$$\tilde{L}_n\{S_n(\boldsymbol{x})\} = \frac{1}{k} \sum_{i=1}^{n} \mathbb{1}\left(R_i^1 > n + 1 - kx_1 \text{ or } \ldots \text{ or } R_i^d > n + 1 - kx_d/n \right).$$

Let d_∞ denote a metric on $\ell^\infty_{\text{loc}}([0,\infty)^d)$ inducing the topology of uniform convergence on compacta, for instance

$$d_\infty(f,g) = \sum_{T=1}^{\infty} 2^{-T} \min\left\{ 1, \sup_{\boldsymbol{x} \in [0,T]^d} |f(\boldsymbol{x}) - g(\boldsymbol{x})| \right\}$$

for $f, g \in \ell^\infty_{\text{loc}}([0,\infty)^d)$. We re-define α_n and $\beta_n = (\beta_{n1}, \ldots, \beta_{nd})$ to mean

$$\alpha_n(\boldsymbol{x}) = \sqrt{k}\left\{ \tilde{L}_n(\boldsymbol{x}) - \tilde{\mu}_n(\boldsymbol{x}) \right\}, \qquad \beta_{nj}(x_j) = \sqrt{k}\left\{ S_{n1}(x_j) - x_j \right\}$$

for $\boldsymbol{x} = (x_1, \ldots, x_d) \in [0,\infty)^d$. It follows from the results in Einmahl, Krajina and Segers (2012, p. 1786) that

$$(\alpha_n, \beta_n) := (\alpha_n, \beta_{n1}, \ldots, \beta_{nd}) \rightsquigarrow \left(\mathbb{W}(\boldsymbol{x}), -\mathbb{W}_1^0, \ldots, -\mathbb{W}_d^0 \right) \tag{B.4}$$

in $(\ell^\infty_{\text{loc}}([0,\infty)^d), d_\infty) \times (\ell^\infty_{\text{loc}}([0,\infty)), d_\infty)^d$. Note that the limit processes on the right-hand side of (B.4) have continuous trajectories, almost surely.

Now, write

$$\sqrt{k}(\hat{L}_n - L) = A_{n1} + A_{n2} + B_{n1} + B_{n2},$$

where

$$A_{n1} = \sqrt{k}(\hat{L}_n - \tilde{L}_n \circ S_n), \qquad A_{n2} = \sqrt{k}(\tilde{\mu}_n \circ S_n - L \circ S_n),$$
$$B_{n1} = \sqrt{k}(\tilde{L}_n \circ S_n - \tilde{\mu}_n \circ S_n), \qquad B_{n2} = \sqrt{k}(L \circ S_n - L).$$

It has been shown in the proof of Theorem 4.6 in Einmahl, Krajina and Segers (2012) that A_{n1} and A_{n2} are $o_P(1)$ with respect to d_∞. Hence, as a consequence of Slutsky's Lemma A.10, it remains to be shown that $B_n :=$

$B_{n1} + B_{n2}$ weakly converges to $\mathbb{W} + \mathrm{d}L_{(-\mathbb{W}_1^0,\dots,-\mathbb{W}_d^0)}$ in $L^\infty_{\mathrm{loc}}([0,\infty)^d)$. Note that we can write $B_n = h_n(\alpha_n, \beta_n)$, where, using the notation of $\mathcal{V}(t)$ from Lemma B.7 below, $h_n : \ell^\infty_{\mathrm{loc}}([0,\infty)^d) \times \mathcal{V}(1/\sqrt{k}) \to L^\infty_{\mathrm{loc}}([0,\infty)^d)$ is defined as

$$h_n(a,b) = a(\mathrm{id}_{[0,\infty)^d} + b/\sqrt{k}) + \sqrt{k}\{L(\mathrm{id}_{[0,\infty)^d} + b/\sqrt{k}) - L\}.$$

Exploiting Lemma B.7 and Lemma A.4, the assertion follows from the extended continuous mapping Theorem, see Theorem 1.11.1 in van der Vaart and Wellner (1996). □

LEMMA B.7. *Let L be an arbitrary stable tail dependence function. Define the set*

$$\mathcal{V}(t) = \left\{(a_1, ..., a_d) \in \{\ell^\infty_{\mathrm{loc}}([0,\infty))\}^d : \boldsymbol{x} + ta(\boldsymbol{x}) \in [0,\infty)\ \forall \boldsymbol{x} \in [0,\infty)^d\right\},$$

where we set $a(\boldsymbol{x}) = (a_1(x_1), \dots, a_d(x_d))$. For every $a \in \{\mathcal{C}([0,\infty))\}^d$, the map dL_a defined in (5.2) has the following property: whenever $t_n \searrow 0$, $t_n \neq 0$, and $a_n = (a_{n1}, \dots, a_{nd}) \to a$ with respect to the d-fold product metric $d_\infty \times \cdots \times d_\infty$ such that $a_n \in \mathcal{V}(t_n)$ for all $n \in \mathbb{N}$, the functions

$$[0,\infty)^d \mapsto \mathbb{R} : \boldsymbol{x} \mapsto t_n^{-1}\{L(\boldsymbol{x} + t_n a_n(\boldsymbol{x})) - L(\boldsymbol{x})\}$$

converge to a limit, say $\mathrm{d}L_a$, in $(\ell^\infty_{\mathrm{loc}}([0,\infty)^d), d_{\mathrm{hypi}})$.

PROOF OF LEMMA B.7. Recall that stable tail dependence functions are convex. Thus Theorem 25.5 in Rockafellar (1970) implies that the complement of the set $\mathcal{S}$ as defined on page 20 has Lebesgue measure zero. Moreover, L is Lipschitz-continuous with respect to the L_1-norm, whence its partial derivatives take values in $[0,1]$, insofar they exist. The proof now follows analogously to the proof of Proposition B.3. □

REFERENCES

ATTOUCH, H. and WETS, R. J. B. (1983). A convergence theory for saddle functions. *Trans. Amer. Math. Soc.* **280** 1–41. MR712247 (85f:49024)

BASS, R. F. and PYKE, R. (1985). The space $D(A)$ and weak convergence for set-indexed processes. *Ann. Probab.* **13** 860–884. MR799425 (86m:60053)

BEER, G. (1993). *Topologies on closed and closed convex sets. Mathematics and its Applications* **268**. Kluwer Academic Publishers Group, Dordrecht. MR1269778 (95k:49001)

BERAN, R. (1984). Minimum distance procedures. In *Nonparametric methods. Handbook of Statist.* **4** 741–754. North-Holland, Amsterdam. MR831734

BICKEL, P. J. and WICHURA, M. J. (1971). Convergence criteria for multiparameter stochastic processes and some applications. *Ann. Math. Statist.* **42** 1656–1670. MR0383482 (52 ##4363)

Billingsley, P. (1968). *Convergence of probability measures.* John Wiley & Sons Inc., New York. MR0233396 (38 ##1718)

Bücher, A. and Dette, H. (2010). A note on bootstrap approximations for the empirical copula process. *Statist. Probab. Lett.* **80** 1925–1932. MR2734261 (2011i:62057)

Bücher, A. and Dette, H. (2011). Multiplier bootstrap approximations of tail copulas – with applications. *arXiv:1102.0110.* to appear in Bernoulli.

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

Dehling, H. and Durieu, O. (2011). Empirical processes of multidimensional systems with multiple mixing properties. *Stochastic Process. Appl.* **121** 1076–1096. MR2775107 (2012d:60093)

Doukhan, P., Fermanian, J.-D. and Lang, G. (2009). An empirical central limit theorem with applications to copulas under weak dependence. *Stat. Inference Stoch. Process.* **12** 65–87. MR2486117 (2011a:62284)

Drees, H. and Huang, X. (1998). Best attainable rates of convergence for estimators of the stable tail dependence function. *J. Multivariate Anal.* **64** 25–47. MR1619974 (99m:62042)

Einmahl, J. H. J., Krajina, A. and Segers, J. (2012). An M-estimator for tail dependence in arbitrary dimensions. *Ann. Statist.* **40** 1764–1793. MR3015043

Embrechts, P., Lindskog, F. and McNeil, A. J. (2003). Modelling Dependence with Copulas and Applications to Risk Management. In *Handbook of Heavy Tailed Distributions in Finance* (S. Rachev, ed.) 329–384. Elsevier.

Fermanian, J.-D., Radulović, D. and Wegkamp, M. (2004). Weak convergence of empirical copula processes. *Bernoulli* **10** 847–860. MR2093613 (2005i:60037)

Genest, C., Ghoudi, K. and Rivest, L. P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika* **82** 543–552. MR1366280 (96j:62081)

Genest, C., Nešlehová, J. and Rémillard, B. (2013). On the multilinear empirical copula process. *Bernoulli* in press.

Genest, C., Quessy, J.-F. and Rémillard, B. (2007). Asymptotic local efficiency of Cramér-von Mises tests for multivariate independence. *Ann. Statist.* **35** 166–191. MR2332273 (2008m:62107)

Geyer, C. J. (1994). On the asymptotics of constrained M-estimation. *Ann. Statist.* **22** 1993–2010. MR1329179 (96e:62031)

Gudendorf, G. and Segers, J. (2012). Nonparametric estimation of multivariate extreme-value copulas. *J. Statist. Plann. Inference* **142** 3073–3085. MR2956794

Huang, X. (1992). Statistics of bivariate extreme values. PhD thesis, Tinbergen Institute Research Series, Netherlands.

Kosorok, M. R. (2008). *Introduction to empirical processes and semiparametric inference. Springer Series in Statistics.* Springer, New York. MR2724368 (2012b:62005)

Matheron, G. (1975). *Random sets and integral geometry.* John Wiley & Sons, New York-London-Sydney With a foreword by Geoffrey S. Watson, Wiley Series in Probability and Mathematical Statistics. MR0385969 (52 ##6828)

McNeil, A. J. and Nešlehová, J. (2009). Multivariate Archimedean copulas, d-monotone functions and l_1-norm symmetric distributions. *Ann. Statist.* **37** 3059–3097. MR2541455 (2010j:62033)

Molchanov, I. (2005). *Theory of random sets. Probability and its Applications (New York).* Springer-Verlag London Ltd., London. MR2132405 (2006b:60004)

Neuhaus, G. (1971). On weak convergence of stochastic processes with multidimensional time parameter. *Ann. Math. Statist.* **42** 1285–1295. MR0293706 (45 ##2783)

OGURA, Y. (2007). On some metrics compatible with the Fell-Matheron topology. *Internat. J. Approx. Reason.* **46** 65–73. MR2362225 (2008h:68121)

PENG, L. and QI, Y. (2008). Bootstrap approximation of tail dependence function. *J. Multivariate Anal.* **99** 1807–1824. MR2444820 (2009m:62157)

RÉMILLARD, B. and SCAILLET, O. (2009). Testing for equality between two copulas. *J. Multivariate Anal.* **100** 377–386. MR2483426 (2010g:62161)

RIO, E. (2000). *Théorie asymptotique des processus aléatoires faiblement dépendants. Mathématiques & Applications (Berlin) [Mathematics & Applications]* **31**. Springer-Verlag, Berlin. MR2117923 (2005k:60001)

ROCKAFELLAR, R. T. (1970). *Convex analysis. Princeton Mathematical Series, No. 28.* Princeton University Press, Princeton, N.J. MR0274683 (43 ##445)

ROCKAFELLAR, R. T. and WETS, R. J. B. (1998). *Variational analysis. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **317**. Springer-Verlag, Berlin. MR1491362 (98m:49001)

RÜSCHENDORF, L. (1976). Asymptotic distributions of multivariate rank order statistics. *Ann. Statist.* **4** 912–923. MR0420794 (54 ##8806)

SEGERS, J. (2012). Asymptotics of empirical copula processes under non-restrictive smoothness assumptions. *Bernoulli* **18** 764–782. MR2948900

SKOROHOD, A. V. (1956). Limit theorems for stochastic processes. *Teor. Veroyatnost. i Primenen.* **1** 289–319. MR0084897 (18,943c)

STRAF, M. L. (1972). Weak convergence of stochastic processes with several parameters. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory* 187–221. Univ. California Press, Berkeley, Calif. MR0402847 (53 ##6661)

TSUKAHARA, H. (2005). Semiparametric estimation in copula models. *Canad. J. Statist.* **33** 357–375. MR2193980 (2007c:62021)

VAN DER VAART, A. W. (1998). *Asymptotic statistics. Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge University Press, Cambridge. MR1652247 (2000c:62003)

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak convergence and empirical processes. Springer Series in Statistics.* Springer-Verlag, New York. With applications to statistics. MR1385671 (97g:60035)

VERVAAT, W. (1981). Une compactification des espaces fonctionnels C et D; une alternative pour la démonstration de théorèmes limites fonctionnels. *C. R. Acad. Sci. Paris Sér. I Math.* **292** 441–444. MR611412 (82c:60008)

WANG, Y. and STOEV, S. A. (2011). Conditional sampling for spectrally discrete max-stable random fields. *Adv. in Appl. Probab.* **43** 461–483. MR2848386 (2012g:60173)

WEISS, G. (2011). Copula parameter estimation by maximum-likelihood and minimum-distance estimators: a simulation study. *Comput. Statist.* **26** 31–54. MR2773799

WHITT, W. (2002). *Stochastic-process limits. Springer Series in Operations Research.* Springer-Verlag, New York. An introduction to stochastic-process limits and their application to queues. MR1876437 (2003f:60005)

RUHR-UNIVERSITÄT BOCHUM
FAKULTÄT FÜR MATHEMATIK
UNIVERSITÄTSSTRASSE 150
44780 BOCHUM, GERMANY
E-MAIL: stanislav.volgushev@rub.de
axel.buecher@rub.de

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
INSTITUT DE STATISTIQUE
VOIE DU ROMAN PAYS, 20
B-1348 LOUVAIN-LA-NEUVE
E-MAIL: johan.segers@uclouvain.be
axel.buecher@rub.de