

Directional differentiability for supremum-type functionals: Statistical applications

JAVIER CÁRCAMO^{*}, ANTONIO CUEVAS[†] and
LUIS-ALBERTO RODRÍGUEZ[‡]

Departamento de Matemáticas, Universidad Autónoma de Madrid, 28049 Madrid, Spain.
E-mail: ^{*}javier.carcamo@uam.es; [†]antonio.cuevas@uam.es; [‡]luisalberto.rodriguez@uam.es

We show that various functionals related to the supremum of a real function defined on an arbitrary set or a measure space are Hadamard directionally differentiable. We specifically consider the supremum norm, the supremum, the infimum, and the amplitude of a function. The (usually non-linear) derivatives of these maps adopt simple expressions under suitable assumptions on the underlying space. As an application, we improve and extend to the multidimensional case the results in Raghavachari (*Ann. Statist.* **1** (1973) 67–73) regarding the limiting distributions of Kolmogorov–Smirnov type statistics under the alternative hypothesis. Similar results are obtained for analogous statistics associated with copulas. We additionally solve an open problem about the Berk–Jones statistic proposed by Jager and Wellner (In *A Festschrift for Herman Rubin* (2004) 319–331 IMS). Finally, the asymptotic distribution of maximum mean discrepancies over Donsker classes of functions is derived.

Keywords: Berk–Jones statistic; copulas; Delta method; empirical processes; Hadamard directional derivative; Kolmogorov distance; Kolmogorov–Smirnov statistic; Kuiper statistic; maximum mean discrepancy

1. Introduction

The general framework. The supremum or uniform norm has been systematically used in statistics to quantify the deviation between an observed phenomenon and a theoretical model. A well-known case is the goodness-of-fit problem, where the Kolmogorov distance (i.e., the uniform distance between distribution functions) is one of the main tools to carry out the testing procedures. In this context, the prototypical example is the Kolmogorov–Smirnov test in which the supremum norm of the difference between the empirical distribution function of the sample and the reference distribution function is employed. The sup-norm has also been notably considered in the literature of almost all fields of statistics such as robustness, density estimation, regression and classification, among others. The reason for the extensive use of this distance might rely on different factors: it has a clear and simple interpretation; it takes into account the global behaviour of the functions; and, in general, it is easy to compute.

The aim of this work is to discuss the (directional) differentiability of the supremum norm – as well as various related functionals that commonly appear in statistics – viewed as a real functional from the space of bounded functions defined on an arbitrary set or a measure space. We consider

the supremum norm, the supremum, the infimum, and the amplitude of a real function. As an application, we use an extended version of the functional Delta method to derive the asymptotic distribution of many statistics that can be expressed in terms of these maps. In this way, we provide a simple and unified approach and the appropriate framework to deal with such type of statistics.

The problem under study. Throughout this work, \mathfrak{X} is a nonempty set and $\ell^\infty(\mathfrak{X})$ is the real Banach space of bounded functions $f : \mathfrak{X} \rightarrow \mathbb{R}$, equipped with the supremum norm, $\|f\|_\infty := \sup_{x \in \mathfrak{X}} |f(x)|$. If additionally $(\mathfrak{X}, \mathcal{A}, \mu)$ is a measure space, where \mathcal{A} is a σ -algebra and μ a positive measure, we denote by $\ell^\infty(\mathfrak{X}, \mathcal{A}, \mu)$ the set of classes of equivalence of measurable and essentially bounded functions $f : \mathfrak{X} \rightarrow \mathbb{R}$ with the norm $\|f\|_{\ell^\infty(\mu)} := \text{ess sup}_{x \in \mathfrak{X}} |f(x)|$, where

$$\text{ess sup}_{x \in \mathfrak{X}} f := \inf\{C \in \mathbb{R} : \mu(\{x \in \mathfrak{X} : f(x) > C\}) = 0\}.$$

Important examples of this general setting are $\mathfrak{X} = \mathbb{R}^d$ or $\bar{\mathbb{R}}^d$ ($d \geq 1$), with $\bar{\mathbb{R}} \equiv [-\infty, +\infty]$ the extended real line, and $\mathfrak{X} = \mathcal{F}$, a class of real functions. To avoid unnecessary repetitions, unless specifically mentioned, from now on we will only consider the supremum.

For $q \in \ell^\infty(\mathfrak{X})$, the quantity of interest that we want to estimate is $\phi(q)$, where ϕ is any of the following functionals:

$$\begin{aligned} \delta(f) &:= \|f\|_\infty, & \sigma(f) &:= \sup_{\mathfrak{X}} f, & \iota(f) &:= \inf_{\mathfrak{X}} f, & \text{and} \\ \alpha(f) &:= \text{amp}_{\mathfrak{X}} f, & f &\in \ell^\infty(\mathfrak{X}), \end{aligned} \tag{1.1}$$

with $\text{amp}_{\mathfrak{X}} f := \sup_{\mathfrak{X}} f - \inf_{\mathfrak{X}} f$ (the amplitude of the function f).

We will assume that q can be estimated by \mathbb{Q}_n , a random element taking values in $\ell^\infty(\mathfrak{X})$ a.s. satisfying

$$r_n(\mathbb{Q}_n - q) \rightsquigarrow \mathbb{Q} \text{ in } \ell^\infty(\mathfrak{X}), \text{ as } n \rightarrow \infty, \tag{1.2}$$

where r_n is a sequence of real numbers such that $r_n \rightarrow \infty$, \mathbb{Q} is a tight Borel random variable in $\ell^\infty(\mathfrak{X})$, and we use the arrow ‘ \rightsquigarrow ’ to denote the weak convergence of probability measures in the sense of Hoffmann–Jørgensen (see van der Vaart and Wellner [54]). The scaling r_n usually goes to infinity as the square root of n , but its behaviour could be different in some examples. In van der Vaart and Wellner [54] the theory of weak convergence is developed for a net of probability spaces, that is, a family of spaces indexed by a *directed set*. We recall that a directed set A is a non-empty set with a partial order relation ‘ \leq ’ satisfying that for every $a, b \in A$, there is $c \in A$ such that $a \leq c$ and $b \leq c$. The results obtained in this paper could also be stated in terms of nets. Nevertheless, this generalization is not relevant for the applications considered in this work and it will not be considered in what follows.

For $\phi \in \{\delta, \sigma, \iota, \alpha\}$ in (1.1), we are interested in analyzing the asymptotic behaviour of the normalized estimator of $\phi(q)$, that is, the statistic given by

$$D_n(\phi) \equiv D_\phi(q, \mathbb{Q}_n, r_n) := r_n(\phi(\mathbb{Q}_n) - \phi(q)). \tag{1.3}$$

Background. By the continuous mapping theorem, when $q = 0$ (the null function), the weak convergence in (1.2) directly implies that $D_n(\phi) \rightsquigarrow \phi(\mathbb{Q})$. (Note that in this case ‘ \rightsquigarrow ’ is the usual convergence in distribution of random variables.) This situation often corresponds to the case in which $D_n(\phi)$ is a normalized discrepancy – usually measured in terms of the sup-norm – for testing the null hypothesis $H_0 : q = 0$. In this setting, the limiting behaviour of $D_n(\phi)$ if $q \neq 0$ provides information regarding the asymptotic power of the underlying testing procedure. The classical result on the asymptotic distribution of the Kolmogorov–Smirnov statistic under the null hypothesis (see, e.g., van der Vaart [53]) is a well-known example. It is also worth mentioning the usefulness of this approach for testing composite null hypotheses such as $H_0 : q \leq 0$. In this case, the limiting behavior of $D_n(\phi)$ when $q \neq 0$ provides information about both asymptotic power (when $q \not\leq 0$) and asymptotic null behavior (when $q \leq 0$ and $q = 0$). In Beare and Moon [6], Seo [46] and Beare and Shi [7], the focus is on asymptotic null behavior.

Finding the asymptotic distribution of $D_n(\phi)$ in (1.3) when q is not identically zero is a more challenging problem. So far, this problem has been tackled generally for the sup-norm and some particular choices of the function q . To the best of our knowledge, the first remarkable result in this direction was obtained by Raghavachari [39]. This author found the asymptotic distribution of the normalized version of the plug-in estimator of $\phi(F - G)$ (for $\phi \in \{\delta, \sigma, \alpha\}$) in the one-sample and two-sample cases when F and G are continuous univariate distribution functions. The results in Raghavachari [39] have also been summarized in DasGupta [13], Chapter 26. Over the years, the ideas in Raghavachari [39] have been used and replicated by several authors to obtain different results in similar settings. A non-exhaustive list of these references is: Álvarez-Esteban et al. [1,2], Freitag, Lange and Munk [23], Schmoyer [43], among others. In Genest and Nešlehová [24], the authors discussed a test of radial symmetry for copulas in which the key element is the estimation of $\|C - \bar{C}\|_\infty$, where C is a bivariate copula and \bar{C} is its survival copula. Dette et al. [16] used the same technique to find the asymptotic distribution of the estimator of $\|m_1(\beta_1) - m_2(\beta_2)\|_\infty$, where $m_1(\beta_1)$ and $m_2(\beta_2)$ are regression functions with parameters β_1 and β_2 , respectively. In Dette, Kokot and Aue [15], Theorem 6.1, a result in the same spirit as Raghavachari [39] is obtained for convergence of suprema of non-centered processes indexed by directed sets (see Remark 3.2).

The proposed methodology. In all the previous references the approach used to compute the limiting distributions is based on the direct probabilistic analysis of the considered statistics. For instance, the proofs in Raghavachari [39] are essentially based on a careful analysis of the behaviour of the empirical process in the set of points around which the supremum in $\|F - G\|_\infty$ is attained. However, we explore here an alternative, more general, approach. It is based on the idea that the statistics in (1.3) have indeed the usual form, suitable to apply the functional Delta method. Therefore, in light of (1.3), a direct and intuitive approach to find the asymptotic distribution of $D_n(\phi)$ could be to analyze the differentiability of the maps in (1.1) and use the functional Delta method. In fact, as it will become evident in this work, looking at the behaviour and analytic properties of the underlying functional is much more enlightening than working directly with the probability distribution of the statistic.

Though there are many possible ways of defining the concept of differentiability for maps between metric or normed spaces, Hadamard differentiability is perhaps the most convenient in this context as it is appropriate for applying the functional Delta method (see van der Vaart [53],

Section 20). However, there are many important examples of maps which are *not* Hadamard differentiable. This is the case of the functionals in (1.1), which are clearly continuous but non-differentiable. Despite not being fully differentiable, we will show that these maps are *Hadamard directionally differentiable*. This weaker notion of differentiability was introduced by Shapiro [47]. Shapiro [48] and Dümbgen [19] (see also Römisch [41]) independently showed that the Delta method still holds for directional differentiable maps. Recently, this idea has been successfully exploited in the econometric literature; see Beare and Moon [6], Kaido [30], Seo [46], and Beare and Shi [7]. Fang and Santos [20] illustrate in depth the applicability of the directional differentiability to a wide variety of problems in econometrics. See additionally Beare and Fang [5] and Sommerfeld and Munk [50].

Structure and main results. In Section 2, we give the necessary definitions, prove that the maps in (1.1) are Hadamard directional differentiable and determine their derivatives under very general assumptions. In particular, this implies that an extended version of the functional Delta method can be applied for these mappings. As far as the authors know, in the statistical community the Hadamard directional differentiability of the infimum under no additional conditions on the underlying space was first obtained by Römisch [41], Proposition 1, after a personal communication of P. Lachout in 2004. Fang and Santos [20], Lemma S.4.9, also obtained an expression for the Hadamard directional derivative of the supremum for continuous functions defined on a compact metric space.

In Section 2, besides reviewing the different notions of differentiability and the Delta method, we also obtain several original results.

- (a) Theorem 2.1 in Section 2.2 follows the spirit of Römisch [41], Proposition 1, though our proof is slightly different and we include the supremum norm (not covered in Römisch [41]). In the rest of Section 2, we rely on this result to obtain simplified expressions for the derivatives of the mappings in (1.1) when the space \mathfrak{X} is endowed with additional structure.
- (b) In Section 2.3, we assume that \mathfrak{X} is a compact metric space. The main novelty here is that the involved functions are *not* required to be continuous (and we continue to deal with the supremum norm). Fang and Santos [20], Lemma S.4.9, is obtained as a particular case.
- (c) In Sections 2.4 and 2.5, we consider the case in which \mathfrak{X} is a totally bounded metric space and a weakly compact subset of a Banach space, respectively. To the best of our knowledge, the corresponding differentiability results are new in the literature.
- (d) In Section 2.6, we analyze in detail the situation in which $\mathfrak{X} = \bar{\mathbb{R}}^d$ and the functions belong to $\mathcal{D}(\bar{\mathbb{R}}^d) \equiv$ the extension of the Skorohod space in $[0, 1]^d$ (introduced in Neuhaus [36]) to the whole $\bar{\mathbb{R}}^d$. The space $\mathcal{D}(\bar{\mathbb{R}}^d)$ is an important subspace of $\ell^\infty(\bar{\mathbb{R}}^d)$ as it includes the paths of many well-known stochastic processes with jumps in their paths such as multivariate empirical processes. Hence, the functions in $\mathcal{D}(\bar{\mathbb{R}}^d)$ are not necessarily continuous and the expressions of the derivatives of the maps are new.

The versatility of the proposed methodology is illustrated in depth in Sections 3–6, where we derive the asymptotic distribution of various statistics with no additional effort. We base the results on the directional differentiability of the functionals and the weak convergence of the underlying stochastic processes. Hence, this unifying approach allows us to reduce a usually difficult

statistical problem to a much simpler analytical question related to the directional differentiability of the corresponding functional. Using these ideas, we obtain the following applications: In Section 3, we extend and give simpler and shorter proofs of the results in Raghavachari [39] both in the one-sample and two-sample cases. The extension is carried out in different directions. First, no assumption on the involved distribution functions is necessary to derive the asymptotic results. In contrast, in Raghavachari [39] the continuity of the distribution functions is required. Secondly, the results are obtained in a multidimensional setting. We note that the proofs are very simple (compared with those in Raghavachari [39]) because they just rely on the analysis of the differentiability of the functionals and the convergence of the associated processes separately. It should be further remarked that those works that have used the results and ideas in Raghavachari [39] were forced to impose the continuity of the involved functions as an assumption in their statements; see for instance Álvarez-Esteban et al. [2], Equation (11), Freitag, Lange and Munk [23], Section 2, or Dette et al. [16], Assumption 7.4. The regularity limitation of working with continuous functions is not mathematically aesthetic and it is in fact unnecessary, as we will show in this paper. Moreover, in Section 4 we will extend these results to copulas. Also, in Section 5, we apply this technique to solve an open question by Jager and Wellner [28] related to the Berk–Jones statistic. Finally, in Section 6 we derive the asymptotic distribution for the plug-in estimators of maximum mean discrepancies with respect to a Donsker class.

The main results of this paper can also be applied to find the asymptotic distribution of the empirical risk over Donsker classes of functions and estimators of kernel distances. These applications are not included in the present paper due to the limited space available and they will be developed in future works.

2. Main results

In this section, we introduce the definitions of directional differentiability of maps between Banach spaces, recall an extended version of the Delta method for these mappings, and discuss the analytic properties of the functionals introduced in Section 1 according to the mathematical structure of \mathfrak{X} .

2.1. Directional differentiability and the Delta method

In many situations it is common to face the problem of estimating a transformation, $\phi(\theta)$, of a (possibly infinite-dimensional) parameter θ . Typically, θ is unknown but can be estimated by means of T_n and ϕ is a map defined in a metric space. If ϕ is smooth enough in a local neighborhood of θ – for instance, differentiable at θ in a precise sense – the asymptotic distribution of (the normalized version of) $\phi(T_n)$ can be determined by expanding ϕ around θ and using an invariance principle for T_n in the underlying metric space. Of course, this is the key idea behind the (*functional*) *Delta method*, one of the most frequently used methodologies in statistics to compute the limiting distribution of an estimator of a quantity of interest (see van der Vaart and Wellner [54], Section 3.9). This technique is specially fruitful when dealing with the popular plug-in estimators, which, by construction, are functions of the empirical distribution function

of the observed sample. In such cases, the powerful theory of weak convergence of empirical processes provides the suitable mathematical machinery to determine the asymptotic behaviour of this kind of estimators (see Giné and Nickl [25]).

We start with the notion of Gâteaux directional differentiability.

Definition 2.1. Let \mathcal{D} and \mathcal{E} be real Banach spaces with norms $\|\cdot\|_{\mathcal{D}}$ and $\|\cdot\|_{\mathcal{E}}$, respectively. A map $\phi : \mathcal{D} \rightarrow \mathcal{E}$ is said to be *Gâteaux directionally differentiable* at $\theta \in \mathcal{D}$ tangentially to a set $\mathcal{D}_0 \subset \mathcal{D}$ if there exists a map $\phi'_\theta : \mathcal{D}_0 \rightarrow \mathcal{E}$ such that

$$\left\| \frac{\phi(\theta + t_n h) - \phi(\theta)}{t_n} - \phi'_\theta(h) \right\|_{\mathcal{E}} \rightarrow 0, \tag{2.1}$$

for all $h \in \mathcal{D}_0$ and all sequences $\{t_n\} \subset \mathbb{R}$ such that $t_n \downarrow 0$.

It is well-known that Gâteaux differentiability is too weak for the Delta method to hold. To solve this problem, the directions along which we approach to $\phi(\theta)$ in (2.1) have to be allowed to change with n . This naturally leads to the concept of Hadamard directional differentiability. We follow Shapiro [47] for the next definition.

Definition 2.2. In the context of the previous definition, we say that $\phi : \mathcal{D} \rightarrow \mathcal{E}$ is *Hadamard directionally differentiable* at $\theta \in \mathcal{D}$ tangentially to a set $\mathcal{D}_0 \subset \mathcal{D}$ if there exists a map $\phi'_\theta : \mathcal{D}_0 \rightarrow \mathcal{E}$ such that

$$\left\| \frac{\phi(\theta + t_n h_n) - \phi(\theta)}{t_n} - \phi'_\theta(h) \right\|_{\mathcal{E}} \rightarrow 0, \tag{2.2}$$

for all $h \in \mathcal{D}_0$ and all sequences $\{t_n\} \subset \mathbb{R}$ and $\{h_n\} \subset \mathcal{D}$ such that $t_n \downarrow 0$ and $\|h_n - h\|_{\mathcal{D}} \rightarrow 0$.

Obviously, the Hadamard directional differentiability condition (2.2) is stronger than the Gâteaux notion (2.1). The only difference between the directional and the usual differentiability is that the derivative ϕ'_θ is no longer required to be linear in Definitions 2.1 and 2.2. Nevertheless, if equation (2.2) is satisfied, then ϕ'_θ is continuous and positive homogeneous of degree 1 (see Shapiro [47], Proposition 3.1).

Remark 2.1. If ϕ is as in Definitions 2.1, and additionally ϕ is locally Lipschitz, that is, there exists a constant $C > 0$ such that $\|\phi(f) - \phi(g)\|_{\mathcal{E}} \leq C \|f - g\|_{\mathcal{D}}$, for all $f, g \in \mathcal{D}$ in a neighborhood of each point of \mathcal{D} , then Hadamard directional differentiability is equivalent to the Gâteaux one (see Shapiro [47], Proposition 3.5). This condition is satisfied by $\delta, \sigma, \iota, \alpha : \ell^\infty(\mathfrak{X}) \rightarrow \mathbb{R}$ defined in (1.1). Hence, to check that the maps considered in Section 1 are Hadamard directionally differentiable at $f \in \ell^\infty(\mathfrak{X})$ we only need to show Gâteaux directional differentiability.

The important fact about Hadamard directional differentiability is that it is the crucial condition to ensure the validity of the following *extended (functional) Delta method*.

Proposition 2.1. Let \mathcal{D} and \mathcal{E} be Banach spaces and $\phi : \mathcal{D}_\phi \subset \mathcal{D} \rightarrow \mathcal{E}$, where \mathcal{D}_ϕ is the domain of ϕ . Assume that ϕ is Hadamard directionally differentiable at $\theta \in \mathcal{D}_\phi$ tangentially to a set

$\mathcal{D}_0 \subset \mathcal{D}$. For some sample spaces Ω_n , let $T_n : \Omega_n \rightarrow \mathcal{D}_\phi$ be maps such that $r_n(T_n - \theta) \rightsquigarrow T$, for some sequence of numbers $r_n \rightarrow \infty$ and a random element T that takes values in \mathcal{D}_0 . Then, $r_n(\phi(T_n) - \phi(\theta)) \rightsquigarrow \phi'_\theta(T)$. If additionally ϕ'_θ can be continuously extended to \mathcal{D} , then we have that $r_n(\phi(T_n) - \phi(\theta)) = \phi'_\theta(r_n(T_n - \theta)) + o_P(1)$.

Remark 2.2. The detailed proof of Proposition 2.1 can be found in Shapiro [48], Theorem 2.1 (see also Römisch [41], Theorem 1, or Fang and Santos [20], Theorem 2.1), but it is essentially the same one as for the traditional Delta method; see van der Vaart [53], Theorem 20.8. The key idea is to apply the *extended continuous mapping theorem* (van der Vaart and Wellner [54], Theorem 1.11.1) to the sequence of functionals defined by $\phi_n(h) := r_n(\phi(\theta + r_n^{-1}h) - \phi(\theta))$, $n \in \mathbb{N}$.

In the present context, let us assume that $\theta_n \rightarrow \theta$ and $r_n(T_n - \theta_n) \rightsquigarrow T$, and we want to determine conditions so that $r_n(\phi(T_n) - \phi(\theta_n)) \rightsquigarrow \phi'_\theta(T)$. As it is pointed out in van der Vaart and Wellner [54], page 375, a stronger form of differentiability is needed to obtain such a “uniform” version of the Delta method.

Definition 2.3. In the context of Definition 2.1, we say that $\phi : \mathcal{D} \rightarrow \mathcal{E}$ is *uniformly Hadamard differentiable* at $\theta \in \mathcal{D}$ tangentially to a set $\mathcal{D}_0 \subset \mathcal{D}$ if there exists a map $\phi'_\theta : \mathcal{D}_0 \rightarrow \mathcal{E}$ such that

$$\left\| \frac{\phi(\theta_n + t_n h_n) - \phi(\theta_n)}{t_n} - \phi'_\theta(h) \right\|_{\mathcal{E}} \rightarrow 0,$$

for all $h \in \mathcal{D}_0$ and all sequences $\{t_n\} \subset \mathbb{R}$, $\{\theta_n\}, \{h_n\} \subset \mathcal{D}$ such that $t_n \downarrow 0$, $\|\theta_n - \theta\|_{\mathcal{D}} \rightarrow 0$, and $\|h_n - h\|_{\mathcal{D}} \rightarrow 0$.

If ϕ is uniformly Hadamard differentiable at θ , $\theta_n \rightarrow \theta$ and $r_n(T_n - \theta_n) \rightsquigarrow T$, we still have that $r_n(\phi(T_n) - \phi(\theta_n)) \rightsquigarrow \phi'_\theta(T)$; see van der Vaart and Wellner [54], Theorem 3.9.5.

2.2. A general result on Hadamard directional differentiability

In the next theorem, we show that the maps introduced in Section 1 are directionally differentiable at every function of $\ell^\infty(\mathfrak{X})$, where \mathfrak{X} is an arbitrary space. In the sequel $\text{sgn}(\cdot)$ denotes the sign function.

Theorem 2.1. *The maps δ, σ, ι and α in (1.1) are Hadamard directionally differentiable at every $f \in \ell^\infty(\mathfrak{X}) \setminus \{0\}$. For $g \in \ell^\infty(\mathfrak{X})$, their derivatives are respectively given by*

$$\begin{aligned} \delta'_f(g) &= \lim_{\epsilon \downarrow 0} \sup_{A_\epsilon(|f|)} (g \cdot \text{sgn}(f)), & \sigma'_f(g) &= \lim_{\epsilon \downarrow 0} \sup_{A_\epsilon(f)} g, \\ \iota'_f(g) &= \lim_{\epsilon \downarrow 0} \inf_{B_\epsilon(f)} g, & \alpha'_f(g) &= \lim_{\epsilon \downarrow 0} \left(\sup_{A_\epsilon(f)} g - \inf_{B_\epsilon(f)} g \right), \end{aligned} \tag{2.3}$$

where, for $\epsilon > 0$ and $h \in \ell^\infty(\mathfrak{X})$, $A_\epsilon(h)$ and $B_\epsilon(h)$ are the superlevel and sublevel sets of h defined by

$$A_\epsilon(h) := \left\{ x \in \mathfrak{X} : h(x) \geq \sup_{\mathfrak{X}} h - \epsilon \right\} \quad \text{and} \quad B_\epsilon(h) := \left\{ x \in \mathfrak{X} : h(x) \leq \inf_{\mathfrak{X}} h + \epsilon \right\}.$$

Moreover, if $(\mathfrak{X}, \mathcal{A}, \mu)$ is a measure space, the result still holds if we substitute the suprema (respectively, infima) by essential suprema (respectively, infima) with respect to μ .

Proof. We first start with σ as the conclusion for the rest of the maps can be derived from this case. Let us fix $f \in \ell^\infty(\mathfrak{X}) \setminus \{0\}$. For $n \in \mathbb{N}$ and each sequence of real numbers $\{s_n\}$ such that $s_n \uparrow \infty$, we consider $\sigma_n(f) : \ell^\infty(\mathfrak{X}) \rightarrow \mathbb{R}$ defined by

$$\sigma_n(f, g) := \sup_{\mathfrak{X}}(s_n f + g) - s_n \sup_{\mathfrak{X}} f, \quad g \in \ell^\infty(\mathfrak{X}). \tag{2.4}$$

From Remarks 2.1 and 2.2, it suffices to show that $\sigma_n(f, g) \rightarrow \sigma'_f(g)$, as $n \rightarrow \infty$, with $\sigma'_f(g)$ defined in (2.3). For $\epsilon > 0$ and $x \notin A_\epsilon(f)$, we have that

$$s_n f(x) + g(x) - s_n \sup_{\mathfrak{X}} f \leq \sup_{\mathfrak{X}} g - s_n \epsilon.$$

Hence, for all $\epsilon > 0$, we obtain that

$$\begin{aligned} \limsup_{n \rightarrow \infty} \sigma_n(f, g) &= \limsup_{n \rightarrow \infty} \left(\sup_{A_\epsilon(f)} (s_n f + g) - s_n \sup_{\mathfrak{X}} f \right) \\ &\leq \sup_{A_\epsilon(f)} g. \end{aligned} \tag{2.5}$$

Conversely, let us define

$$h(\epsilon) := \sup_{A_\epsilon(f)} g, \quad \epsilon > 0. \tag{2.6}$$

Observe that h is non-decreasing and thus the limit as ϵ decreases to 0 exists and, by definition, coincides with $\sigma'_f(g)$. For each $m \in \mathbb{N}$, there exists $x_m \in A_{1/m}(f)$ satisfying

$$g(x_m) \geq h(1/m) - 1/m \quad \text{and} \quad f(x_m) \geq \sup_{\mathfrak{X}} f - 1/m. \tag{2.7}$$

From (2.7), for each s_n , we have that

$$\begin{aligned} h(1/m) &\leq g(x_m) + 1/m \\ &= s_n f(x_m) + g(x_m) - s_n f(x_m) + 1/m \\ &\leq \sigma_n(f, g) + (s_n + 1)/m. \end{aligned} \tag{2.8}$$

Now (2.8) implies that, for all $n \in \mathbb{N}$,

$$\limsup_{\epsilon \downarrow 0} \sup_{A_\epsilon(f)} g = \lim_{m \rightarrow \infty} h(1/m) \leq \sigma_n(f, g). \tag{2.9}$$

The proof corresponding to σ follows from (2.5) and (2.9).

Now, we consider the map δ in (1.1). Assume that $f \in \ell^\infty(\mathfrak{X})$ with $\|f\|_\infty > 0$. For $g \in \ell^\infty(\mathfrak{X})$, we have to show that $\delta_n(f, g) \rightarrow \delta'_f(g)$, as $n \rightarrow \infty$, where $\delta_n(f, g) := \|s_n f + g\|_\infty - s_n \|f\|_\infty$ and $s_n \uparrow \infty$. First, for $\epsilon < \|f\|_\infty/2$ and $s_n > 2\|g\|_\infty/\|f\|_\infty$, it is readily checked that $s_n|f| + \text{sgn}(f) \cdot g \geq 0$ globally on $A_\epsilon(|f|)$. We hence conclude that

$$\lim_{n \rightarrow \infty} \delta_n(f, g) = \lim_{n \rightarrow \infty} \sigma_n(|f|, g \cdot \text{sgn}(f)) = \sigma'_{|f|}(g \cdot \text{sgn}(f)) = \delta'_f(g).$$

The proof for ι and α follows from the duality between supremum and infimum. Finally, the case in which \mathfrak{X} is a measure space can be treated in a similar way so it is therefore omitted. \square

As pointed out in the Introduction, Römisch [41], Proposition 1, provides the same result as Theorem 2.1 for the infimum. Obviously, the derivatives of the supremum and amplitude of a function can be derived from the infimum by duality. The additional contribution of Theorem 2.1 is the differentiability of the supremum norm operator, δ . Also, the proof we have included here is slightly different to the one in Römisch [41]. The expressions in (2.3) will be used throughout Sections 2.3–2.6 to obtain simplified expressions of the derivatives.

Theorem 2.1 ensures that the functionals in (1.1) are Hadamard directionally differentiable. Nevertheless, in general these maps are *not* uniformly Hadamard differentiable (see Definition 2.3) as the following example shows.

Example 2.1. Let \mathfrak{X} be the interval $[0, 1]$ in \mathbb{R} and we consider the function $f \equiv 1$. For $x \in [0, 1]$ and $n \in \mathbb{N}$, let $f_n(x) := 1 + x/n$, $g(x) := 1 - x$, and $s_n = n$. We have that $f_n \rightarrow f$ in $\ell^\infty(\mathfrak{X})$ and it is easy to check that $\sigma_n(f_n, g) = 0$, where σ_n is given in (2.4). However, $\sigma'_f(g) = \sup_{[0,1]} g = 1$. We conclude that σ is not uniformly Hadamard differentiable, and therefore neither are the rest of the maps in (1.1).

Following the same ideas as in the proof of Theorem 2.1, the following partial result can be proved.

Corollary 2.1. *Let δ , σ , ι and α be as in (1.1). For each $f, g \in \ell^\infty(\mathfrak{X})$ and all sequences $\{t_n\} \subset \mathbb{R}$, $\{f_n\}, \{g_n\} \subset \ell^\infty(\mathfrak{X})$ such that $t_n \downarrow 0$, $f_n \rightarrow f$ and $g_n \rightarrow g$ in $\ell^\infty(\mathfrak{X})$, we have that*

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{\delta(f_n + t_n g_n) - \delta(f_n)}{t_n} &\leq \delta'_f(g), & \limsup_{n \rightarrow \infty} \frac{\sigma(f_n + t_n g_n) - \sigma(f_n)}{t_n} &\leq \sigma'_f(g), \\ \liminf_{n \rightarrow \infty} \frac{\iota(f_n + t_n g_n) - \iota(f_n)}{t_n} &\geq \iota'_f(g), & \limsup_{n \rightarrow \infty} \frac{\alpha(f_n + t_n g_n) - \alpha(f_n)}{t_n} &\leq \alpha'_f(g), \end{aligned} \tag{2.10}$$

where $\delta'_f, \sigma'_f, \iota'_f$ and α'_f are given in (2.3).

In general, the reverse inequalities in (2.10) fail to hold because it is not possible to control the term $(\phi(f_n) - \phi(f))/t_n$ ($\phi \in \{\delta, \sigma, \iota, \alpha\}$), for all sequences $\{t_n\} \subset \mathbb{R}$ and $\{f_n\} \subset \ell^\infty(\mathfrak{X})$ such that $t_n \downarrow 0$ and $f_n \rightarrow f$.

2.3. Compact metric spaces

In some occasions the limit in ϵ of the derivatives in (2.3) can be removed. For example, if \mathfrak{X} is a compact metric space, the derivatives can be characterized by means of convergent sequences in \mathfrak{X} as the following corollary shows.

Corollary 2.2. *In the context of Theorem 2.1, let us further assume that (\mathfrak{X}, d) is a compact metric space. The derivatives in (2.3) can be expressed as*

$$\begin{aligned} \delta'_f(g) &= \sup_{A_0(|f|)} (g \cdot \text{sgn}(f))_{|f|}^\blacktriangle, & \sigma'_f(g) &= \sup_{A_0(f)} g_f^\blacktriangle, \\ \iota'_f(g) &= \inf_{B_0(f)} g_f^\blacktriangledown, & \alpha'_f(g) &= \sup_{A_0(f)} g_f^\blacktriangle - \inf_{B_0(f)} g_f^\blacktriangledown, \end{aligned} \tag{2.11}$$

where for $h, l \in \ell^\infty(\mathfrak{X})$,

$$A_0(h) := \left\{ x \in \mathfrak{X} : \text{there exists } \{x_n\} \subset \mathfrak{X} \text{ with } x_n \rightarrow x \text{ and } h(x_n) \rightarrow \sup_{\mathfrak{X}} h \right\}, \tag{2.12}$$

$$B_0(h) := \left\{ x \in \mathfrak{X} : \text{there exists } \{x_n\} \subset \mathfrak{X} \text{ with } x_n \rightarrow x \text{ and } h(x_n) \rightarrow \inf_{\mathfrak{X}} h \right\},$$

$$h_l^\blacktriangle(x) := \sup \left\{ \limsup_{n \rightarrow \infty} h(x_n) : x_n \rightarrow x \text{ and } l(x_n) \rightarrow \sup_{\mathfrak{X}} l \right\}, \quad x \in A_0(l), \tag{2.13}$$

$$h_l^\blacktriangledown(x) := \inf \left\{ \liminf_{n \rightarrow \infty} h(x_n) : x_n \rightarrow x \text{ and } l(x_n) \rightarrow \inf_{\mathfrak{X}} l \right\}, \quad x \in B_0(l).$$

Proof. We only give a detailed proof for σ because the rest of the cases are analogous. We consider the sequence $\{x_m\}$ satisfying (2.7) obtained in Theorem 2.1. As (\mathfrak{X}, d) is compact, we can extract a convergent subsequence $x_{m_k} \rightarrow x$ in \mathfrak{X} , as $k \rightarrow \infty$. From (2.7), we have that $x \in A_0(f)$ and, recalling (2.6), from Theorem 2.1, we obtain that

$$\sigma'_f(g) = \lim_{k \rightarrow \infty} h(1/m_k) \leq \limsup_{k \rightarrow \infty} g(x_{m_k}) \leq g_f^\blacktriangle(x) \leq \sup_{A_0(f)} g_f^\blacktriangle. \tag{2.14}$$

In the other direction, let $x \in A_0(f)$ and $\{x_n\} \subset \mathfrak{X}$ such that $x_n \rightarrow x$ and $f(x_n) \rightarrow \sup_{\mathfrak{X}} f$. For each $\epsilon > 0$, we have that $x_n \in A_\epsilon(f)$, for n large enough. We therefore conclude that

$$\limsup_{n \rightarrow \infty} g(x_n) \leq \sup_{A_\epsilon(f)} g, \quad \text{for all } \epsilon > 0. \tag{2.15}$$

The conclusion follows from (2.14), (2.15) and Theorem 2.1. □

Remark 2.3. From the proof of Corollary 2.2 we see that the result is still valid for sequentially compact topological spaces. As this extension is not important for the applications in this work we will omit this framework in the following.

In the following, if (\mathfrak{X}, d) is a metric space we denote by $\mathcal{C}(\mathfrak{X}, d)$ the subset of continuous functions in $\ell^\infty(\mathfrak{X})$. We observe that if $g \in \mathcal{C}(\mathfrak{X}, d)$, then $g_f^\blacktriangle(x) = g(x)$ ($x \in A_0(f)$) and $g_f^\blacktriangledown(x) = g(x)$ ($x \in B_0(f)$), where g_f^\blacktriangle and g_f^\blacktriangledown are defined as in (2.13). If we further assume that $f \in \mathcal{C}(\mathfrak{X}, d)$, we have that $A_0(|f|) = M^+(|f|)$, $A_0(f) = M^+(f)$ and $B_0(f) = M^-(f)$, where for $h \in \ell^\infty(\mathfrak{X})$,

$$M^+(h) := \left\{ x \in \mathfrak{X} : h(x) = \sup_{\mathfrak{X}} h \right\} \quad \text{and} \quad M^-(h) := \left\{ x \in \mathfrak{X} : h(x) = \inf_{\mathfrak{X}} h \right\}. \quad (2.16)$$

This observation yields the following corollary.

Corollary 2.3. *Let (\mathfrak{X}, d) be a compact metric space and let δ, σ, ι and α be the maps defined in (1.1). The maps σ, ι and α are Hadamard directionally differentiable at any $f \in \ell^\infty(\mathfrak{X})$ tangentially to the set $\mathcal{C}(\mathfrak{X}, d)$ with derivatives, for $g \in \mathcal{C}(\mathfrak{X}, d)$,*

$$\sigma'_f(g) = \sup_{A_0(f)} g, \quad \iota'_f(g) = \inf_{B_0(f)} g \quad \text{and} \quad \alpha'_f(g) = \sup_{A_0(f)} g - \inf_{B_0(f)} g. \quad (2.17)$$

If additionally $f \in \mathcal{C}(\mathfrak{X}, d) \setminus \{0\}$, we have that

$$\begin{aligned} \delta'_f(g) &= \sup_{M^+(|f|)} (g \cdot \text{sgn}(f)), & \sigma'_f(g) &= \sup_{M^+(f)} g, \\ \iota'_f(g) &= \inf_{M^-(f)} g, & \alpha'_f(g) &= \sup_{M^+(f)} g - \inf_{M^-(f)} g, \end{aligned} \quad (2.18)$$

where $M^+(\cdot)$ and $M^-(\cdot)$ are defined in (2.16).

The expression of the derivative σ'_f in (2.18) for continuous functions defined on a compact metric space has been previously obtained in Fang and Santos [20], Lemma S.4.9. Observe that equalities in (2.17) are valid even when the function f is not continuous (as in the more general Corollary 2.2). Note also that $M^+(|f|)$ (respectively, $M^+(f)$ and $M^-(f)$) in (2.16) is the set of extremal points corresponding to the sup-norm (respectively, the supremum and infimum) of f .

Another interesting question is to find conditions under which the derivatives of the maps are linear, i.e., the cases in which the mappings are fully Hadamard differentiable. This kind of results can be traced back to Banach [4] (see also Leonard and Taylor [31,32], and the references therein). In these works the supremum norm differentiability was investigated from the point of view of functional analysis within the space $\mathcal{C}(\mathfrak{X}, d)$, with (\mathfrak{X}, d) a compact metric space. The following result, a direct consequence of Corollary 2.3, provides general outcomes in a different context. We denote by $\text{Card}(A)$ the cardinality of the set A .

Corollary 2.4. *Assume that (\mathfrak{X}, d) is a compact metric space and let $f \in \ell^\infty(\mathfrak{X}) \setminus \{0\}$. Let $A_0(\cdot)$ and $B_0(\cdot)$ be the sets in (2.12). For the maps defined in (1.1) we have that:*

- (a) *The map δ is (fully) Hadamard differentiable at f tangentially to the set $\mathcal{C}(\mathfrak{X}, d)$ if and only if $\text{Card}(A_0(|f|)) = 1$ and $\{\limsup_{n \rightarrow \infty} \text{sgn}(f(x_n)) : x_n \rightarrow x \text{ and } |f(x_n)| \rightarrow \|f\|_\infty\} = \{c\}$. In such a case, $\delta'_f(g) = cg(x^*)$, where $A_0(|f|) = \{x^*\}$.*

- (b) The map σ is (fully) Hadamard differentiable at f tangentially to the set $\mathcal{C}(\mathfrak{X}, d)$ if and only if $\text{Card}(A_0(f)) = 1$. In such a case, $\sigma'_f(g) = g(x^+)$, where $A_0(f) = \{x^+\}$.
- (c) The map ι is (fully) Hadamard differentiable at f tangentially to the set $\mathcal{C}(\mathfrak{X}, d)$ if and only if $\text{Card}(B_0(f)) = 1$. In such a case, $\iota'_f(g) = g(x^-)$, where $B_0(f) = \{x^-\}$.
- (d) The map α is (fully) Hadamard differentiable at f tangentially to the set $\mathcal{C}(\mathfrak{X}, d)$ if and only if $\text{Card}(A_0(f)) = \text{Card}(B_0(f)) = 1$. In such a case, $\alpha'_f(g) = g(x^+) - g(x^-)$, where $A_0(f) = \{x^+\}$ and $B_0(f) = \{x^-\}$.

Note that when $f \in \mathcal{C}(\mathfrak{X}, d)$, we have that $A_0(|f|) = M^+(|f|)$ in (2.16) and the condition $\text{Card}(A_0(|f|)) = 1$ means that f is a *peaking function*, that is, there exists $x^* \in \mathfrak{X}$ such that $|f(x^*)| = \|f\|_\infty$ and $|f(x^*)| > |f(x)|$, for all $x \in \mathfrak{X}$ with $x \neq x^*$.

From a statistical point of view, identifying the cases in which the maps are Hadamard differentiable has two important consequences when the limit in (1.2) is Gaussian: firstly, as the linear derivatives are (essentially) the evaluation at an appropriate point, by the extended Delta method (see Proposition 2.1), the asymptotic distribution of the statistic in (1.3) is normal; secondly, the standard bootstrap for (1.3) is consistent if and only if the underlying map ϕ is fully Hadamard differentiable (see Fang and Santos [20], Theorem 3.1).

2.4. Totally bounded metric spaces

If \mathbb{Q} is a tight Borel measurable map into $\ell^\infty(\mathfrak{X})$ as in (1.2), then there is a pseudo-metric on \mathfrak{X} such that the sample paths of \mathbb{Q} are uniformly continuous and \mathfrak{X} is totally bounded (see van der Vaart and Wellner [54], Lemma 1.5.9). For statistical applications it is therefore important to determine conditions under which the derivatives in (2.3) have similar expressions as those in Corollary 2.3 when the underlying space is totally bounded.

We recall that if (\mathfrak{X}, d) is a totally bounded metric space, $(\bar{\mathfrak{X}}, d)$ is a compact metric space, where $\bar{\mathfrak{X}}$ is the completion of \mathfrak{X} with respect to d . Further, the space $\mathcal{C}_u(\mathfrak{X}, d)$ of bounded and uniformly continuous functions $f : \mathfrak{X} \rightarrow \mathbb{R}$ is isometric to $\mathcal{C}(\bar{\mathfrak{X}}, d)$. Each $f \in \mathcal{C}_u(\mathfrak{X}, d)$ has a unique extension to a function $\bar{f} \in \mathcal{C}(\bar{\mathfrak{X}}, d)$. For $x \in \bar{\mathfrak{X}} \setminus \mathfrak{X}$, this extension is defined by $\bar{f}(x) = \lim_{n \rightarrow \infty} f(x_n)$, with $\{x_n\} \subset \mathfrak{X}$ such that $x_n \rightarrow x$ (in fact, Cauchy-continuity is enough to check that \bar{f} is well-defined, but uniform continuity suffices for our purposes).

In this setting, it is straightforward to check that Corollary 2.2 still holds if we substitute the sets $A_0(\cdot)$ and $B_0(\cdot)$ by

$$\begin{aligned} \bar{A}_0(h) &:= \left\{ x \in \bar{\mathfrak{X}} : \text{there exists } \{x_n\} \subset \mathfrak{X} \text{ with } x_n \rightarrow x \text{ and } h(x_n) \rightarrow \sup_{\mathfrak{X}} h \right\}, \\ \bar{B}_0(h) &:= \left\{ x \in \bar{\mathfrak{X}} : \text{there exists } \{x_n\} \subset \mathfrak{X} \text{ with } x_n \rightarrow x \text{ and } h(x_n) \rightarrow \inf_{\mathfrak{X}} h \right\}, \end{aligned} \tag{2.19}$$

for $h \in \ell^\infty(\mathfrak{X})$. In particular, the following corollary, important for statistical applications in which \mathfrak{X} is a class of functions (see Section 6), holds.

Corollary 2.5. *Let (\mathfrak{X}, d) be a totally bounded metric space and let δ, σ, ι and α be the maps defined in (1.1).*

- (a) The maps σ, ι and α are Hadamard directionally differentiable at $f \in \ell^\infty(\mathfrak{X})$ tangentially to the set $\mathcal{C}_u(\mathfrak{X}, d)$ with derivatives, for $g \in \mathcal{C}_u(\mathfrak{X}, d)$,

$$\sigma'_f(g) = \sup_{\bar{A}_0(f)} \bar{g}, \quad \iota'_f(g) = \inf_{\bar{B}_0(f)} \bar{g} \quad \text{and} \quad \alpha'_f(g) = \sup_{\bar{A}_0(f)} \bar{g} - \inf_{\bar{B}_0(f)} \bar{g},$$

where $\bar{A}_0(\cdot)$ and $\bar{B}_0(\cdot)$ are defined in (2.19).

- (b) If additionally $f \in \mathcal{C}_u(\mathfrak{X}, d) \setminus \{0\}$, we have that

$$\begin{aligned} \delta'_f(g) &= \sup_{\bar{M}^+(|f|)} (\bar{g} \cdot \text{sgn}(\bar{f})), & \sigma'_f(g) &= \sup_{\bar{M}^+(f)} \bar{g}, & \iota'_f(g) &= \inf_{\bar{M}^-(f)} \bar{g}, \\ \alpha'_f(g) &= \sup_{\bar{M}^+(f)} \bar{g} - \inf_{\bar{M}^-(f)} \bar{g}, \end{aligned}$$

where for $h \in \mathcal{C}_u(\mathfrak{X}, d)$,

$$\bar{M}^+(h) := \left\{ x \in \mathfrak{X} : \bar{h}(x) = \sup_{\mathfrak{X}} h \right\} \quad \text{and} \quad \bar{M}^-(h) := \left\{ x \in \mathfrak{X} : \bar{h}(x) = \inf_{\mathfrak{X}} h \right\}. \quad (2.20)$$

Remark 2.4. Corollary 2.4 still holds if (\mathfrak{X}, d) is a totally bounded metric space and we replace $\mathcal{C}(\mathfrak{X}, d)$, $A_0(\cdot)$ and $B_0(\cdot)$ with $\mathcal{C}_u(\mathfrak{X}, d)$, $\bar{A}_0(\cdot)$ and $\bar{B}_0(\cdot)$ (defined in (2.19)), respectively.

2.5. Weakly compact sets

The compactness assumption on \mathfrak{X} in Corollaries 2.2 and 2.3 could be too demanding in some infinite-dimensional settings. A simple inspection of the proof of Corollary 2.2 shows that a similar result can be stated when \mathfrak{X} is a weakly compact subset of a Banach space by using Eberlein-Šmulian theorem (see Conway [12], page 163). In such a case, Corollary 2.2 still holds by substituting the sets $A_0(h)$ and $B_0(h)$ in (2.12) and the quantities $h_f^\blacktriangle(x)$ and $h_f^\blacktriangledown(x)$ in (2.13) respectively, by

$$\begin{aligned} A_0^w(h) &:= \left\{ x \in \mathfrak{X} : \text{there exists } \{x_n\} \subset \mathfrak{X} \text{ with } x_n \rightharpoonup x \text{ and } h(x_n) \rightarrow \sup_{\mathfrak{X}} h \right\}, \\ B_0^w(h) &:= \left\{ x \in \mathfrak{X} : \text{there exists } \{x_n\} \subset \mathfrak{X} \text{ with } x_n \rightharpoonup x \text{ and } h(x_n) \rightarrow \inf_{\mathfrak{X}} h \right\}, \end{aligned} \quad (2.21)$$

and

$$\begin{aligned} h_f^{\blacktriangle, w}(x) &:= \sup \left\{ \limsup_{n \rightarrow \infty} h(x_n) : x_n \rightharpoonup x \text{ and } f(x_n) \rightarrow \sup_{\mathfrak{X}} f \right\}, \quad x \in A_0^w(f), \\ h_f^{\blacktriangledown, w}(x) &:= \inf \left\{ \liminf_{n \rightarrow \infty} h(x_n) : x_n \rightharpoonup x \text{ and } f(x_n) \rightarrow \inf_{\mathfrak{X}} f \right\}, \quad x \in B_0^w(f), \end{aligned} \quad (2.22)$$

where $x_n \rightharpoonup x$ stands for the weak convergence in the corresponding space. We recall that if $\{x_n\} \subset \mathcal{B}$ with \mathcal{B} a Banach space, $x_n \rightharpoonup x$ means that $\varphi(x_n) \rightarrow \varphi(x)$ for all $\varphi \in \mathcal{B}^*$, the topological dual space of \mathcal{B} formed by linear and continuous functionals from \mathcal{B} to \mathbb{R} . If $\mathcal{B} = \mathcal{H}$ is

a Hilbert space with inner product $\langle \cdot, \cdot \rangle$, the weak convergence amounts to $\langle x_n, y \rangle \rightarrow \langle x, y \rangle$, for all $y \in \mathcal{H}$.

In this context, we have analogous results as Corollaries 2.3 and 2.5 by changing the set of tangency points.

Definition 2.4. If \mathfrak{X} is a subset of a vector space, a function $g : \mathfrak{X} \rightarrow \mathbb{R}$ is said to be *prelinear* on \mathfrak{X} if $\sum_{i=1}^r \lambda_i g(x_i) = 0$ whenever $\sum_{i=1}^r \lambda_i x_i = 0$, for $r < \infty$, $\lambda_i \in \mathbb{R}$ and $x_i \in \mathfrak{X}$ ($i = 1, \dots, r$).

Every prelinear function g defined on \mathfrak{X} admits a unique extension to a linear function on $\text{span}(\mathfrak{X})$, the linear span of \mathfrak{X} (see Dudley [17], Lemma 2.30, page 88). This extension is given by

$$\tilde{g}\left(\sum_{i=1}^r \lambda_i x_i\right) = \sum_{i=1}^r \lambda_i g(x_i), \quad \text{with } x_i \in \mathfrak{X} \text{ and } \lambda_i \in \mathbb{R} \text{ (} i = 1, \dots, r \text{)}. \tag{2.23}$$

In the following, if (\mathfrak{X}, d) is a metric space contained in a vector space, we denote by $\mathcal{C}_{\text{pl}}(\mathfrak{X}, d)$ the subset of $\mathcal{C}(\mathfrak{X}, d)$ formed by prelinear functions on \mathfrak{X} . Further, if \mathcal{B} is a Banach space with norm $\|\cdot\|$, $d_{\mathcal{B}}$ stands for the metric on \mathcal{B} , that is, $d_{\mathcal{B}}(x, y) = \|x - y\|$ ($x, y \in \mathcal{B}$).

Corollary 2.6. Let \mathcal{B} be a Banach space and let δ, σ, ι and α be the maps in (1.1). Let us assume that the set $\mathfrak{X} \subset \mathcal{B}$ satisfies the following two conditions:

- (i) \mathfrak{X} is a weakly compact subset of \mathcal{B} .
- (ii) For each $g \in \mathcal{C}_{\text{pl}}(\mathfrak{X}, d_{\mathcal{B}})$, its linear extension \tilde{g} in (2.23) is continuous on $\text{span}(\mathfrak{X})$.

Then, the maps σ, ι and α are Hadamard directionally differentiable at $f \in \ell^\infty(\mathfrak{X})$ tangentially to $\mathcal{C}_{\text{pl}}(\mathfrak{X}, d_{\mathcal{B}})$ with derivatives, for $g \in \mathcal{C}_{\text{pl}}(\mathfrak{X}, d_{\mathcal{B}})$,

$$\sigma'_f(g) = \sup_{A_0^w(f)} g, \quad \iota'_f(g) = \inf_{B_0^w(f)} g \quad \text{and} \quad \alpha'_f(g) = \sup_{A_0^w(f)} g - \inf_{B_0^w(f)} g,$$

where $A_0^w(\cdot)$ and $B_0^w(\cdot)$ are defined in (2.21).

If additionally $f \in \mathcal{C}_{\text{pl}}(\mathfrak{X}, d_{\mathcal{B}}) \setminus \{0\}$, then the derivatives of δ, σ, ι and α are as in (2.18).

Proof. As in the previous proofs, we only discuss the map σ . Let us consider $x \in A_0^w(f)$ (defined in (2.21)) and $g \in \mathcal{C}_{\text{pl}}(\mathfrak{X}, d_{\mathcal{B}})$. We consider a sequence $\{x_n\} \subset \mathfrak{X}$ such that $x_n \rightarrow x$ and $f(x_n) \rightarrow \sup_{\mathfrak{X}} f$ (the existence of such a sequence is guaranteed by condition (i) and (2.7)). Condition (ii) and Hahn-Banach theorem imply that there exists a linear and continuous map, say \bar{g} , defined on \mathcal{B} such that $\bar{g} = \tilde{g}$ on $\text{span}(\mathfrak{X})$, and hence $\bar{g} = g$ on \mathfrak{X} . As $\bar{g} \in \mathcal{B}^*$ and $x_n \rightarrow x$, we conclude that $\lim_{n \rightarrow \infty} g(x_n) = g(x)$. This shows that $g_f^{\blacktriangle, w}(x) = g(x)$, with $g_f^{\blacktriangle, w}(x)$ defined as in (2.22), and the conclusion follows from the observation at the beginning of this section.

Finally, if $f \in \mathcal{C}_{\text{pl}}(\mathfrak{X}, d_{\mathcal{B}})$, the same argument used before shows that $A_0^w(f) = M^+(f)$, where the set $M^+(\cdot)$ is defined in (2.16). □

We observe that hypothesis (i) in the previous corollary is essential to extract a weakly convergent subsequence in \mathfrak{X} . We also observe that condition (ii) cannot be dropped as, in general,

the linear extension \tilde{g} of a function $g \in \mathcal{C}_{\text{pl}}(\mathfrak{X}, d_{\mathcal{B}})$ is not necessarily continuous in $\text{span}(\mathfrak{X})$ as the following example shows: Let \mathcal{B} be an infinite-dimensional Banach space with norm $\|\cdot\|$. We consider $\mathfrak{X} = \{x_n\}_{n=0}^{\infty} \subset \mathcal{B}$, where $x_0 = 0$ and $\{x_n\}_{n=1}^{\infty}$ is a linearly independent subset of \mathcal{B} such that $\|x_n\| = 1/n$ ($n \in \mathbb{N}$). It is easy to check that the function defined by $g(0) = 0$ and $g(x_n) = 1/\sqrt{n}$ ($n \in \mathbb{N}$) belongs to $\mathcal{C}_{\text{pl}}(\mathfrak{X}, d_{\mathcal{B}})$, but its linear extension \tilde{g} is *not* continuous because it is not bounded on the unit sphere since $\tilde{g}(x_n/\|x_n\|) = \sqrt{n}$ ($n \in \mathbb{N}$).

The following proposition provides easy to check conditions guaranteeing that Corollary 2.6 (ii) is fulfilled.

Proposition 2.2. *Let \mathcal{B} be a Banach space with norm $\|\cdot\|$ and $\mathfrak{X} \subset \mathcal{B}$. Let us assume that one of the following two conditions is satisfied:*

- (a) *There exists $x \in \mathfrak{X}$ and $\delta > 0$ such that $B(x, \delta) := \{y \in \mathcal{B} : \|y - x\| \leq \delta\} \subset \mathfrak{X}$.*
- (b) *\mathcal{B} is a Hilbert space and there exists $\{x_i\}_{i \in I} \subset \mathfrak{X}$, where I is an arbitrary index set such that $\text{span}(\mathfrak{X}) = \text{span}(\{x_i\}_{i \in I})$, $\{x_i\}_{i \in I}$ are pairwise orthogonal and $c := \inf_{i \in I} \|x_i\| > 0$.*

Then, for each $g \in \mathcal{C}_{\text{pl}}(\mathfrak{X}, d_{\mathcal{B}})$, its linear extension \tilde{g} in (2.23) is continuous on $\text{span}(\mathfrak{X})$.

Proof. Let us assume that (a) holds. As $g \in \mathcal{C}(\mathfrak{X}, d_{\mathcal{B}})$, the condition $B(x, \delta) \subset \mathfrak{X}$ ensures that \tilde{g} in (2.23) is continuous at x , and, by linearity, continuous on $\text{span}(\mathfrak{X})$.

Assume now that (b) is satisfied. For $x \in \text{span}(\mathfrak{X})$, we can write $x = \sum_{i=1}^r \lambda_i x_i$, with $\lambda_i \in \mathbb{R}$ ($i = 1, \dots, r$). Taking into account that $\|x\| = \sum_{i=1}^r |\lambda_i| \|x_i\| \geq c \sum_{i=1}^r |\lambda_i|$, we finally obtain that

$$|\tilde{g}(x)| \leq \|g\|_{\infty} \sum_{i=1}^r |\lambda_i| \leq \|g\|_{\infty} \|x\| / c.$$

The previous inequalities show that \tilde{g} is continuous on $\text{span}(\mathfrak{X})$ and the proof is complete. \square

Closed bounded convex subsets of a reflexive Banach space are weakly compact (see Brezis [10], Corollary 3.22). Therefore, the hypotheses of Corollary 2.6 are general enough to include many infinite-dimensional sets. Thanks to Proposition 2.2, an important example covered by Corollary 2.6 is when \mathfrak{X} is the closed unit ball of a reflexive Banach space, and, in particular, the closed unit ball of a Hilbert space. On the other hand, working with prelinear functions could seem to be too restrictive. However, we point out that if \mathbb{P} is a probability measure and a set \mathfrak{X} is \mathbb{P} -pre-Gaussian (Giné and Nickl [25], Definition 3.7.26, page 251), there is a version of the \mathbb{P} -bridge whose sample paths are prelinear (see Giné and Nickl [25], Theorem 3.7.28, page 252). Such a version is usually called *suitable*.

Remark 2.5. Corollary 2.4 still holds with the obvious modifications if \mathfrak{X} is in the conditions of Corollary 2.6. It is enough to replace convergence with weak convergence and $\mathcal{C}(\mathfrak{X}, d)$, $A_0(\cdot)$ and $B_0(\cdot)$ with $\mathcal{C}_{\text{pl}}(\mathfrak{X}, d_{\mathcal{B}})$, $A_0^w(\cdot)$ and $B_0^w(\cdot)$, respectively.

2.6. The case $\mathfrak{X} = \bar{\mathbb{R}}^d$ and the Skorohod space $\mathcal{D}(\bar{\mathbb{R}}^d)$

Throughout this section $\mathfrak{X} = \bar{\mathbb{R}}^d$ ($d \geq 1$) endowed with d_e , the metric corresponding to the Euclidean norm on $[0, 1]^d$ through a given homeomorphism. Hence, $(\bar{\mathbb{R}}^d, d_e)$ is a compact metric space and we can apply Corollaries 2.2 and 2.3 in Section 2.3.

Many important stochastic processes take values in the one-dimensional Skorohod space, $\mathcal{D}(\bar{\mathbb{R}})$, consisting of all the càdlàg functions, that is, right-continuous functions having limit from the left at every point. This space provides a natural and convenient setting to analyze the behaviour of processes with unidimensional time parameter and jumps in their paths such as Poisson processes, Lévy processes, empirical processes or discretizations of stochastic processes, among others. Skorohod-type spaces are usually equipped with different norms to make them separable. However, we are only interested in a multidimensional extension of the Skorohod space viewed as a subset of $\ell^\infty(\bar{\mathbb{R}}^d)$ with the supremum norm. The final aim of this section is providing alternative expressions for the directional derivatives in (2.11) when the involved functions belong to the d -dimensional Skorohod space.

The d -dimensional Skorohod space, introduced in Neuhaus [36] (see also Bickel and Wichura [9]) and more recently considered in Seijo and Sen [45]), is usually defined in compact rectangles of $\bar{\mathbb{R}}^d$. We will firstly extend this space to functions defined in $\bar{\mathbb{R}}^d$.

For $v \in \{-1, 1\}$ and $x \in \bar{\mathbb{R}}$, let

$$I_v(x) := \begin{cases} \emptyset, & \text{if } v = -1, x = -\infty, \\ [-\infty, x), & \text{if } v = -1, x \in (-\infty, +\infty], \\ (x, +\infty], & \text{if } v = +1, x \in [-\infty, +\infty), \\ \emptyset, & \text{if } v = +1, x = +\infty, \end{cases}$$

and

$$\tilde{I}_v(x) := \begin{cases} [-\infty, x), & \text{if } v = -1, x < \infty, \\ \bar{\mathbb{R}}, & \text{if } v = -1, x = +\infty, \\ \emptyset, & \text{if } v = +1, x = +\infty, \\ [x, +\infty], & \text{if } v = +1, x < \infty. \end{cases}$$

We consider $\mathcal{V} := \{-1, 1\}^d$ the set of 2^d vertices of $[-1, 1]^d$. For $\mathbf{v} = (v_1, \dots, v_d) \in \mathcal{V}$ and $\mathbf{x} = (x_1, \dots, x_d) \in \bar{\mathbb{R}}^d$, we define the \mathbf{v} -quadrants of \mathbf{x} by

$$Q_{\mathbf{v}}(\mathbf{x}) := I_{v_1}(x_1) \times \dots \times I_{v_d}(x_d) \quad \text{and} \quad \tilde{Q}_{\mathbf{v}}(\mathbf{x}) := \tilde{I}_{v_1}(x_1) \times \dots \times \tilde{I}_{v_d}(x_d).$$

Observe that $Q_{\mathbf{v}}(\mathbf{x}) \subset \tilde{Q}_{\mathbf{v}}(\mathbf{x})$, $\tilde{Q}_{\mathbf{v}}(\mathbf{x}) \cap \tilde{Q}_{\mathbf{v}'}(\mathbf{x}) = \emptyset$ whenever $\mathbf{v}, \mathbf{v}' \in \mathcal{V}$ with $\mathbf{v} \neq \mathbf{v}'$, and $\bigcup_{\mathbf{v} \in \mathcal{V}} \tilde{Q}_{\mathbf{v}}(\mathbf{x}) = \bar{\mathbb{R}}^d$, for all $\mathbf{x} \in \bar{\mathbb{R}}^d$. Additionally, for each $\mathbf{x} \in \bar{\mathbb{R}}^d$, there exists a unique $\mathbf{v}_{\mathbf{x}} \in \mathcal{V}$ such that $\mathbf{x} \in \tilde{Q}_{\mathbf{v}_{\mathbf{x}}}(\mathbf{x})$. For instance, if $\mathbf{x} \in \bar{\mathbb{R}}^d$, we have that $\mathbf{v}_{\mathbf{x}} = \mathbf{1}$, where $\mathbf{1} := (1, \dots, 1)$.

With the previous concepts we can define the quadrant limits. Let us consider a function $f : \bar{\mathbb{R}}^d \rightarrow \mathbb{R}$, $\mathbf{v} \in \mathcal{V}$ and $\mathbf{x} \in \bar{\mathbb{R}}^d$. We say that $l \in \mathbb{R}$ is the \mathbf{v} -limit of f at \mathbf{x} if $Q_{\mathbf{v}}(\mathbf{x}) \neq \emptyset$ and for every sequence $\{\mathbf{x}_n\} \subset Q_{\mathbf{v}}(\mathbf{x})$ such that $\mathbf{x}_n \rightarrow \mathbf{x}$, we have that $f(\mathbf{x}_n) \rightarrow l$. In such a case, we denote $l \equiv f_{\mathbf{v}}(\mathbf{x})$. Additionally, it is said that f is continuous from above at $\mathbf{x} \in \bar{\mathbb{R}}^d$ if $f_{\mathbf{v}_{\mathbf{x}}}(\mathbf{x})$

exists and $f_{\mathbf{v}_x}(\mathbf{x}) = f(\mathbf{x})$. We say that f is continuous from above if it is continuous from above at every $\mathbf{x} \in \bar{\mathbb{R}}^d$.

Definition 2.5. The Skorohod space on $\bar{\mathbb{R}}^d$, denoted by $\mathcal{D}(\bar{\mathbb{R}}^d)$, is the collection of all continuous from above real functions f defined in $\bar{\mathbb{R}}^d$ for which the \mathbf{v} -limit of f exists for every $\mathbf{v} \in \mathcal{V}$ and $\mathbf{x} \in \bar{\mathbb{R}}^d$ such that $Q_{\mathbf{v}}(\mathbf{x}) \neq \emptyset$.

When $d = 1$, $\mathcal{D}(\bar{\mathbb{R}})$ is usual Skorohod space on $\bar{\mathbb{R}}$. The properties of the multidimensional Skorohod space in $[0, 1]^d$ shown in Neuhaus [36] can be extended with no difficulty to $\mathcal{D}(\bar{\mathbb{R}}^d)$. For instance, the elements in $\mathcal{D}(\bar{\mathbb{R}}^d)$ belong to $\mathcal{D}(\bar{\mathbb{R}})$ in each coordinate, have at most countably many discontinuities and all of them are of the “first class”. The fact that $\mathcal{D}(\bar{\mathbb{R}}^d) \subset \ell^\infty(\bar{\mathbb{R}}^d)$ follows from Neuhaus [36], Corollary 1.6, by noting that functions in $\mathcal{D}(\bar{\mathbb{R}}^d)$ have finite quadrant limits at infinity points.

Remark 2.6. We observe that if $f \in \mathcal{D}(\bar{\mathbb{R}}^d)$ and $\{\mathbf{x}_n\} \subset \tilde{Q}_{\mathbf{v}}(\mathbf{x})$ such that $\mathbf{x}_n \rightarrow \mathbf{x}$, then $f(\mathbf{x}_n) \rightarrow f_{\mathbf{v}}(\mathbf{x})$. This follows from the fact that

$$\tilde{Q}_{\mathbf{v}}(\mathbf{x}) = \{\mathbf{y} \in \bar{\mathbb{R}}^d : \mathbf{y} \in \overline{Q_{\mathbf{v}_y}(\mathbf{y}) \cap Q_{\mathbf{v}}(\mathbf{x})}\},$$

where \bar{A} denotes the closure of the set A . In other words, the functions in $\mathcal{D}(\bar{\mathbb{R}}^d)$ have quadrant limits in $\tilde{Q}_{\mathbf{v}}(\mathbf{x})$.

We are now in position to see how the derivatives in (2.11) look like when $\mathfrak{X} = \bar{\mathbb{R}}^d$ and the functions on which they act belong to $\mathcal{D}(\bar{\mathbb{R}}^d)$.

Corollary 2.7. For any $f \in \mathcal{D}(\bar{\mathbb{R}}^d) \setminus \{0\}$, the maps δ , σ , ι and α in (1.1) are Hadamard directionally differentiable at f tangentially to $\mathcal{D}(\bar{\mathbb{R}}^d)$. For $g \in \mathcal{D}(\bar{\mathbb{R}}^d)$, their derivatives are given by

$$\begin{aligned} \delta'_f(g) &= \max_{\mathbf{v} \in \mathcal{V}} \sup_{M_{\mathbf{v}}^+(|f|)} (g_{\mathbf{v}} \cdot \text{sgn}(f_{\mathbf{v}})), & \sigma'_f(g) &= \max_{\mathbf{v} \in \mathcal{V}} \sup_{M_{\mathbf{v}}^+(f)} g_{\mathbf{v}}, \\ \iota'_f(g) &= \min_{\mathbf{v} \in \mathcal{V}} \inf_{M_{\mathbf{v}}^-(f)} g_{\mathbf{v}}, & \alpha'_f(g) &= \max_{\mathbf{v} \in \mathcal{V}} \sup_{M_{\mathbf{v}}^+(f)} g_{\mathbf{v}} - \min_{\mathbf{v} \in \mathcal{V}} \inf_{M_{\mathbf{v}}^-(f)} g_{\mathbf{v}}, \end{aligned} \quad (2.24)$$

where for $h \in \mathcal{D}(\bar{\mathbb{R}}^d)$,

$$\begin{aligned} M_{\mathbf{v}}^+(h) &:= \{\mathbf{x} \in \bar{\mathbb{R}}^d : Q_{\mathbf{v}}(\mathbf{x}) \neq \emptyset \text{ and } h_{\mathbf{v}}(\mathbf{x}) = \sup h\}, \\ M_{\mathbf{v}}^-(h) &:= \{\mathbf{x} \in \bar{\mathbb{R}}^d : Q_{\mathbf{v}}(\mathbf{x}) \neq \emptyset \text{ and } h_{\mathbf{v}}(\mathbf{x}) = \inf h\}. \end{aligned} \quad (2.25)$$

Proof. This corollary can be proved as Corollary 2.2 by taking into account Remark 2.6 and the following fact: As the number of non-empty quadrants of each point in $\bar{\mathbb{R}}^d$ is finite, each sequence converging to a point $\mathbf{x} \in \bar{\mathbb{R}}^d$ has a subsequence contained in $\tilde{Q}_{\mathbf{v}}(\mathbf{x})$, for some $\mathbf{v} \in \mathcal{V}$. In particular, for every $h \in \mathcal{D}(\bar{\mathbb{R}}^d)$, it holds that $A_0(h) = \bigcup_{\mathbf{v} \in \mathcal{V}} M_{\mathbf{v}}^+(h)$ and $B_0(h) = \bigcup_{\mathbf{v} \in \mathcal{V}} M_{\mathbf{v}}^-(h)$, where $A_0(h)$ and $B_0(h)$ are defined in (2.12). \square

The sets $M_{\mathbf{v}}^+(h)$ (respectively, $M_{\mathbf{v}}^-(h)$) in (2.25) might coincide for different $\mathbf{v} \in \mathcal{V}$. For instance, when f is continuous, $M_{\mathbf{v}}^+(|f|) = M^+(|f|)$, $M_{\mathbf{v}}^+(f) = M^+(f)$, and $M_{\mathbf{v}}^-(f) = M^-(f)$, for all $\mathbf{v} \in \mathcal{V}$, where $M^+(\cdot)$ and $M^-(\cdot)$ are defined in (2.16).

We emphasize that $g_{\mathbf{v}} \equiv g$, for all $\mathbf{v} \in \mathcal{V}$, whenever $g \in \mathcal{C}(\bar{\mathbb{R}}^d, d_e)$. The following corollary is important for applications because many stochastic processes that commonly appear as weak limits of other processes have continuous paths a.s.

Corollary 2.8. *For any $f \in \mathcal{D}(\bar{\mathbb{R}}^d) \setminus \{0\}$, the maps δ , σ , ι and α in (1.1) are Hadamard directionally differentiable at f tangentially to $\mathcal{C}(\bar{\mathbb{R}}^d, d_e)$. For $g \in \mathcal{C}(\bar{\mathbb{R}}^d, d_e)$, their derivatives are given by*

$$\begin{aligned} \delta'_f(g) &= \max_{\mathbf{v} \in \mathcal{V}} \sup_{M_{\mathbf{v}}^+(|f|)} (g \cdot \text{sgn}(f_{\mathbf{v}})), & \sigma'_f(g) &= \max_{\mathbf{v} \in \mathcal{V}} \sup_{M_{\mathbf{v}}^+(f)} g, \\ \iota'_f(g) &= \min_{\mathbf{v} \in \mathcal{V}} \inf_{M_{\mathbf{v}}^-(f)} g, & \alpha'_f(g) &= \max_{\mathbf{v} \in \mathcal{V}} \sup_{M_{\mathbf{v}}^+(f)} g - \min_{\mathbf{v} \in \mathcal{V}} \inf_{M_{\mathbf{v}}^-(f)} g, \end{aligned}$$

with $M_{\mathbf{v}}^+(\cdot)$ and $M_{\mathbf{v}}^-(\cdot)$ defined in (2.25).

If additionally $f \in \mathcal{C}(\bar{\mathbb{R}}^d, d_e)$, the derivatives are as in (2.18).

2.7. Statistical applications

In a wide variety of situations Theorem 2.1 and its subsequent corollaries, joint with the extended Delta method in Proposition 2.1, provide the right framework to obtain a number of significant examples in which the asymptotic distribution of a statistic of interest can be determined with ease. The combination of these results is summarized in the following theorem.

Theorem 2.2. *Let $q \in \ell^\infty(\mathfrak{X}) \setminus \{0\}$ and assume that there exists \mathbb{Q}_n taking values in $\ell^\infty(\mathfrak{X})$ a.s. such that $r_n(\mathbb{Q}_n - q) \rightsquigarrow \mathbb{Q}$, for a sequence of real numbers satisfying that $r_n \rightarrow \infty$ and a Borel random element \mathbb{Q} in $\ell^\infty(\mathfrak{X})$. Then, for $\phi \in \{\delta, \sigma, \iota, \alpha\}$ in (1.1), we have that*

$$r_n(\phi(\mathbb{Q}_n) - \phi(q)) \rightsquigarrow \phi'_q(\mathbb{Q}), \tag{2.26}$$

where the derivatives ϕ'_q are given in (2.3). Moreover, we have that $r_n(\phi(\mathbb{Q}_n) - \phi(q)) = \phi'_q(r_n(\mathbb{Q}_n - q)) + o_{\mathbb{P}}(1)$.

Theorem 2.2 is still valid for the maps σ , ι and α when $q \equiv 0$ as $\sigma'_0(g) \equiv \sup_{\mathfrak{X}} g$, $\iota'_0(g) \equiv \inf_{\mathfrak{X}} g$ and $\alpha'_0(g) \equiv \text{amp}_{\mathfrak{X}}(g)$ are continuous maps. Further, for those $q \in \ell^\infty(\mathfrak{X})$ such that ϕ'_q is linear, i.e., ϕ is fully Hadamard differentiable at q (see Corollary 2.4 and Remarks 2.4 and 2.5), and when \mathbb{Q} is Gaussian, we conclude that $\phi'_q(\mathbb{Q})$ is normally distributed.

In this setting, despite $\phi \in \{\delta, \sigma, \iota, \alpha\}$ in (1.1) not being uniformly Hadamard differentiable (see Example 2.1), if we know that $q_n \rightarrow q$ in $\ell^\infty(\mathfrak{X})$ and $r_n(\mathbb{Q}_n - q_n) \rightsquigarrow \mathbb{Q}$, we can still conclude that $r_n(\phi(\mathbb{Q}_n) - \phi(q_n)) \rightsquigarrow \phi'_q(\mathbb{Q})$ if we assume: (i) ϕ is fully Hadamard differentiable; (ii) the sequence $r_n(q_n - q)$ is relatively compact. Further, if (i) is not satisfied and $r_n(q_n - q) \rightarrow h$, we still have that $r_n(\phi(\mathbb{Q}_n) - \phi(q_n)) \rightarrow_d \phi'_q(\mathbb{Q} + h) - \phi'_q(h)$. See van der Vaart and Wellner [54], page 375, for details.

In what follows, we will apply Theorem 2.2 in different contexts to obtain the asymptotic distribution of several statistics.

3. Distribution functions

Let \mathbf{X} and \mathbf{Y} be two non-degenerate random vectors taking values on \mathbb{R}^d ($d \geq 1$) with joint cumulative distribution functions $F(\mathbf{x}) := P(\mathbf{X} \leq \mathbf{x})$ and $G(\mathbf{x}) := P(\mathbf{Y} \leq \mathbf{x})$, $\mathbf{x} \in \mathbb{R}^d$, where ‘ \leq ’ stands for the coordinatewise order in \mathbb{R}^d . The goal in this section is to estimate $\phi(F - G)$, where $\phi \in \{\delta, \sigma, \alpha\}$ are defined in (1.1).

One-sample case. In this situation, we have at our disposal a random sample $\mathbf{X}_1, \dots, \mathbf{X}_n$ from \mathbf{X} . We estimate $F - G$ with $F_n - G$, where F_n is the empirical distribution function of the observed sample, that is,

$$F_n(\mathbf{x}) := \frac{1}{n} \sum_{i=1}^n 1_{[\mathbf{X}_i \leq \mathbf{x}]}, \quad \mathbf{x} \in \mathbb{R}^d,$$

and 1_A stands for the indicator function of the set A .

The problem consists in finding the behaviour, as $n \rightarrow \infty$, of

$$\begin{aligned} D_n(\delta) &= \sqrt{n}(\|F_n - G\|_\infty - \|F - G\|_\infty), \\ D_n(\sigma) &= \sqrt{n}(\sup(F_n - G) - \sup(F - G)), \\ D_n(\alpha) &= \sqrt{n}(\text{amp}(F_n - G) - \text{amp}(F - G)). \end{aligned} \tag{3.1}$$

When $F \neq G$, the asymptotic distribution of the statistics $D_n(\delta)$, $D_n(\sigma)$ and $D_n(\alpha)$ in (3.1) can be viewed as the limit under the alternative hypothesis of the corresponding two-sided and one-sided Kolmogorov–Smirnov test statistics and Kuiper statistic, respectively.

In this example, for $\phi \in \{\delta, \sigma, \alpha\}$, the statistics in (3.1) are $D_n(\phi) \equiv D_\phi(q, \mathbb{Q}_n, r_n)$ in (1.3) with $q = F - G$, $\mathbb{Q}_n = F_n - G$, and $r_n = \sqrt{n}$. The underlying normalized process, i.e., $r_n(\mathbb{Q}_n - q)$, is nothing but the multivariate *empirical process* (indexed by points),

$$\mathbb{E}_{n,F}(\mathbf{x}) := \sqrt{n}(F_n(\mathbf{x}) - F(\mathbf{x})), \quad n \in \mathbb{N}, \mathbf{x} \in \mathbb{R}^d. \tag{3.2}$$

When there is no confusion with respect to the underlying distribution, we simply use the notation \mathbb{E}_n for the empirical process in (3.2). As the collection of all indicator functions of lower (hyper)rectangles of $\bar{\mathbb{R}}^d$, $\{1_{(-\infty, x_1] \times \dots \times (-\infty, x_d]} : (x_1, \dots, x_d) \in \bar{\mathbb{R}}^d\}$, is Donsker (see van der Vaart and Wellner [54], Example 2.1.3, page 82), the empirical process converges in law in $\ell^\infty(\bar{\mathbb{R}}^d)$. The weak limit of \mathbb{E}_n , denoted in the following by \mathbb{B}_F , is a *F-Brownian bridge*, that is, a centered Gaussian process with covariance function $E(\mathbb{B}_F(\mathbf{x})\mathbb{B}_F(\mathbf{y})) = F(\mathbf{x} \wedge \mathbf{y}) - F(\mathbf{x})F(\mathbf{y})$. (Here $\mathbf{x} \wedge \mathbf{y} \equiv (x_1 \wedge y_1, \dots, x_d \wedge y_d)$ if $\mathbf{x} = (x_1, \dots, x_d)$ and $\mathbf{y} = (y_1, \dots, y_d)$.) If $d = 1$, the assertion “ $\mathbb{E}_n \rightsquigarrow \mathbb{B}_F$ in $\ell^\infty(\bar{\mathbb{R}})$ ” is nothing but the celebrated Donsker’s theorem (Kolmogorov–Doob–Donsker–Dudley central limit theorem). In such a case, $\mathbb{B}_F = \mathbb{B} \circ F$, where \mathbb{B} is a standard

Brownian bridge on $[0, 1]$. When $d \geq 2$, \mathbb{B}_F is also called a tied-down or pinned Brownian sheet based on the measure with distribution function F .

In this particular case we have that $F - G \in \mathcal{D}(\bar{\mathbb{R}}^d)$, $\mathbb{E}_n \in \mathcal{D}(\bar{\mathbb{R}}^d)$ a.s., and $\mathbb{E}_n \rightsquigarrow \mathbb{B}_F$ in $\ell^\infty(\bar{\mathbb{R}}^d)$. Therefore, as a direct consequence of Theorem 2.2 and Corollary 2.7 we obtain the following result.

Proposition 3.1. *Assume that $F \neq G$ and let \mathbb{B}_F be an F -Brownian bridge. For $\phi \in \{\delta, \sigma, \alpha\}$, we consider the statistics $D_n(\phi)$ defined in (3.1). We have that $D_n(\phi) \rightsquigarrow \phi'_{F-G}(\mathbb{B}_F)$, where the derivatives ϕ'_{F-G} are given as in (2.24).*

When $d = 1$, Proposition 3.1 improves Raghavachari [39], Theorems 1, 2 and 3, as here F and G are not assumed to be continuous. If F is continuous, then $\mathbb{B}_F \in \mathcal{C}(\bar{\mathbb{R}}^d, d_e)$ a.s., and the limiting distributions in Proposition 3.1 have simpler expressions (see (2.18)). The following corollary provides a multidimensional extension of the results in Raghavachari [39].

Corollary 3.1. *In the conditions of Proposition 3.1, let us further assume that $F, G \in \mathcal{C}(\bar{\mathbb{R}}^d, d_e)$ and we consider the sets $M^+(\cdot)$ and $M^-(\cdot)$ defined in (2.16). We have that:*

- (i) $D_n(\delta) \rightsquigarrow \sup_{M^+(|F-G|)}(\mathbb{B}_F \cdot \text{sgn}(F - G));$
- (ii) $D_n(\sigma) \rightsquigarrow \sup_{M^+(F-G)} \mathbb{B}_F;$
- (iii) $D_n(\alpha) \rightsquigarrow \sup_{M^+(F-G)} \mathbb{B}_F - \inf_{M^-(F-G)} \mathbb{B}_F.$

Remark 3.1. In the setting of the previous corollary, when $M^+(|F - G|)$ (respectively, $M^+(F - G)$, and $M^+(F - G)$ and $M^-(F - G)$) contains only one point, the mapping δ (respectively, σ and α) is fully Hadamard differentiable at $F - G$ (see Corollary 2.4). In particular, the asymptotic distribution of $D_n(\delta)$ (respectively, $D_n(\sigma)$ and $D_n(\alpha)$) is a zero mean Gaussian distribution. The asymptotic variance can be directly computed from the covariances of \mathbb{B}_F .

Remark 3.2. In Dette, Kokot and Aue [15], Theorem 6.1, the authors obtained a similar version of the results in Raghavachari [39] for convergence of suprema of non-centered processes indexed by directed sets. Using the results in Section 2.3 we can state the following slightly more general result: Let (T, d) be a compact metric space and $\mu \in \mathcal{C}(T, d) \setminus \{0\}$. Let $\{X_a : a \in A\}$ be a net of random variables taking values in $\ell^\infty(T)$ and $r : A \rightarrow [0, \infty)$ satisfying that $\lim_a r_a = \infty$ (with $r_a = r(a)$). Assume that $Z_a := r_a(X_a - \mu) \rightsquigarrow Z$ in $\ell^\infty(T)$, where Z is a Gaussian random variable with paths in $\mathcal{C}(T, d)$ a.s., then

$$D_a(\delta) = r_a(\|X_a\|_\infty - \|\mu\|_\infty) \rightsquigarrow \delta'_\mu(Z) = \sup_{M^+(|\mu|)} \text{sgn}(\mu)Z.$$

It is worth noting that we can drop the assumption on the normalizing sequence r_a in Dette, Kokot and Aue [15], Theorem 6.1. A similar result can be provided when (T, d) is a totally bounded metric space by using the results in Section 2.4 (see Corollary 2.5(b)).

Remark 3.3. The results in the paper can be used to make inferences on the quantity $\delta(F - F_0) = \|F - F_0\|_\infty$, where F_0 is a fixed and known distribution function. It should be taken into

account that in general the corresponding limiting distribution, $\delta'_{F-F_0}(\mathbb{B}_F)$, cannot be approximated by a standard bootstrap approach. It is known that the standard bootstrap fails when the mapping is not fully Hadamard differentiable and the limit of the underlying process is Gaussian; see Fang and Santos [20], Theorem 3.1. Observe that the map δ is fully differentiable at $F - F_0$ if and only if $F - F_0$ is a peaking function (see Corollary 2.4). Therefore, an alternative approach has to be used if we do not assume this “peaking condition” on $F - F_0$. In Fang and Santos [20], Theorem 3.2, a method to consistently estimate this type of asymptotic distributions is proposed. The key idea is estimating in a suitable way the directional derivative. A detailed study of these topics is beyond the scope and space limitations of the present paper.

Two-sample case. Here, two (mutually independent) random samples are available, one of size n from F and another one of size m from G . Let F_n and G_m be the empirical distribution functions of the two samples, respectively, and set $N \equiv \frac{nm}{n+m}$. The two-sided, and one-sided Kolmogorov–Smirnov and Kuiper statistics in the two sample case are given by

$$\begin{aligned} D_{n,m}(\delta) &:= \sqrt{N}(\|F_n - G_m\|_\infty - \|F - G\|_\infty), \\ D_{n,m}(\sigma) &:= \sqrt{N}(\sup(F_n - G_m) - \sup(F - G)), \\ D_{n,m}(\alpha) &:= \sqrt{N}(\text{amp}(F_n - G_m) - \text{amp}(F - G)). \end{aligned} \tag{3.3}$$

In the general setting specified in (1.3), this situation corresponds to the case $q = F - G$, $\mathbb{Q}_{n,m} = F_n - G_m$ and $r_{n,m} = \sqrt{N}$. Hence, we have that

$$r_{n,m}(\mathbb{Q}_{n,m} - q) = \sqrt{\frac{m}{n+m}} \mathbb{E}_{n,F} - \sqrt{\frac{n}{n+m}} \tilde{\mathbb{E}}_{m,G}$$

with $\mathbb{E}_{n,F}$ and $\tilde{\mathbb{E}}_{m,G}$ independent empirical processes. We further observe that if the sampling scheme is balanced, that is, $n/(n+m) \rightarrow \lambda$, with $0 < \lambda < 1$ as $n, m \rightarrow \infty$, then $r_{n,m}(\mathbb{Q}_{n,m} - q) \rightsquigarrow \sqrt{1-\lambda}\mathbb{B}_F - \sqrt{\lambda}\tilde{\mathbb{B}}_G$ in $\ell^\infty(\mathbb{R}^d)$, where \mathbb{B}_F and $\tilde{\mathbb{B}}_G$ are two independent Brownian bridges associated with F and G , respectively. Hence, Theorem 2.2 and Corollary 2.8 directly imply the following result which improves and generalizes Raghavachari [39], Theorems 4 and 5.

Proposition 3.2. *Let us consider a sampling scheme such that as $n, m \rightarrow \infty$, $n/(n+m) \rightarrow \lambda$, with $0 < \lambda < 1$ and let \mathbb{B}_F and $\tilde{\mathbb{B}}_G$ be two independent Brownian bridges associated with F and G , respectively. For $\phi \in \{\delta, \sigma, \alpha\}$, we consider the statistics $D_{n,m}(\phi)$ defined in (3.3). We have that $D_{n,m}(\phi) \rightsquigarrow \phi'_{F-G}(\sqrt{1-\lambda}\mathbb{B}_F - \sqrt{\lambda}\tilde{\mathbb{B}}_G)$, where the derivatives ϕ'_{F-G} are given in (2.24). If we further have that $F, G \in \mathcal{C}(\mathbb{R}^d, d_e)$, then the derivatives can be expressed as in (2.18).*

4. Copulas

In this section, for simplicity, we will assume that the involved distribution functions are continuous. Let us assume that the d -dimensional distribution function F has copula C and continuous marginal distribution functions F_1, \dots, F_d . In other words, $F(\mathbf{x}) = C(F_1(x_1), \dots, F_d(x_d))$, for

$\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$. Let F_n and $F_{n,i}$ ($i = 1, \dots, d$) be the empirical joint and i th marginal distribution functions of a random sample of size n from F . The *empirical copula* is

$$C_n(\mathbf{u}) := F_n(F_{n,1}^{-1}(u_1), \dots, F_{n,d}^{-1}(u_d)), \quad \mathbf{u} := (u_1, \dots, u_d) \in [0, 1]^d, \tag{4.1}$$

where $F_{n,i}^{-1}$ stands for the generalized inverse of $F_{n,i}$, that is, the marginal quantile function of the i th coordinate sample. The *empirical copula process* is defined by

$$\mathbb{C}_n(\mathbf{u}) := \sqrt{n}(C_n(\mathbf{u}) - C(\mathbf{u})), \quad n \in \mathbb{N}, \mathbf{u} \in [0, 1]^d. \tag{4.2}$$

Empirical copula processes play the same role for copulas as empirical processes for distribution functions and they have been extensively used in goodness-of-fit testing problems for copulas (see Fermanian [21] for an overview about this subject).

Several works have been devoted to discuss the asymptotic behaviour of \mathbb{C}_n in (4.2). For instance, in Segers [44] (see also the references therein) it is shown that, under certain not very restrictive smoothness assumptions on the underlying copula C , \mathbb{C}_n converges weakly in $\ell^\infty([0, 1]^d)$. Specifically, let us assume that C satisfies the following regularity condition:

Condition 1. For each $i \in \{1, \dots, d\}$, the i th first order partial derivative of C , $\partial_i C$, exists and is continuous on the set $\{\mathbf{u} = (u_1, \dots, u_d) \in [0, 1]^d : 0 < u_i < 1\}$.

If Condition 1 is satisfied, $\mathbb{C}_n \rightsquigarrow \mathbb{C}$ in $\ell^\infty([0, 1]^d)$ (see Segers [44], Proposition 3.1), where \mathbb{C} is a Gaussian process that can be represented as

$$\mathbb{C}(\mathbf{u}) = \mathbb{B}_C(\mathbf{u}) - \sum_{i=1}^d \partial_i C(\mathbf{u}) \mathbb{B}_C^{(i)}(u_i), \quad \mathbf{u} = (u_1, \dots, u_d) \in [0, 1]^d, \tag{4.3}$$

with \mathbb{B}_C a C -Brownian bridge (see Section 3) and $\mathbb{B}_C^{(i)}(u_i) := \mathbb{B}_C(1, \dots, 1, u_i, 1, \dots, 1)$, the variable u_i appearing at the i th entry.

Using Theorem 2.2 and Corollary 2.8, we immediately obtain the following result. Though details are omitted, similar results can be stated for the unilateral Kolmogorov–Smirnov and Kuiper statistics and the associated two sample problems. Therefore, we obtain analogous outcomes to those of Raghavachari [39] for copulas instead of distribution functions.

Proposition 4.1. *Let C be a copula satisfying Condition 1 and let C_n be as in (4.1). For any copula $D \neq C$, the statistic*

$$T_n(C, D) := \sqrt{n}(\|C_n - D\|_\infty - \|C - D\|_\infty)$$

converges in distribution to $\delta'_{C-D}(\mathbb{C}) = \sup_{M^+(\{C-D\})}(\mathbb{C} \cdot \text{sgn}(C - D))$, with \mathbb{C} defined in (4.3) and the set $M^+(\cdot)$ is given in (2.16).

For any bivariate copula C , we consider the survival copula \bar{C} defined by

$$\bar{C}(u, v) := u + v - 1 + C(1 - u, 1 - v), \quad (u, v) \in [0, 1]^2.$$

The statistic

$$\bar{T}_n(C) := \sqrt{n}(\|C_n - \bar{C}_n\|_\infty - \|C - \bar{C}\|_\infty), \tag{4.4}$$

where C_n is given in (4.1), has been used in Genest and Nešlehová [24] to derive a test of radial symmetry for bivariate copulas. The next proposition provides the asymptotic distribution of such statistic.

Proposition 4.2. *Let C be a bivariate copula satisfying Condition 1 (for $d = 2$). The statistic $\bar{T}_n(C)$ in (4.4) converges in distribution to*

$$\delta'_{C-\bar{C}}(\mathbb{C}^*) = \sup_{M^+(|C-\bar{C}|)} (\mathbb{C}^* \cdot \text{sgn}(C - \bar{C})),$$

where $\mathbb{C}^*(u, v) := \mathbb{C}(u, v) - \mathbb{C}(1 - u, 1 - v)$, $(u, v) \in [0, 1]^2$, and \mathbb{C} and the set $M^+(\cdot)$ are defined in (4.3) and (2.16), respectively.

Proof. From Theorem 2.2, it will suffice to show that

$$C_n^* := \sqrt{n}(C_n - \bar{C}_n - (C - \bar{C})) \rightsquigarrow \mathbb{C}^* \quad \text{in } \ell^\infty([0, 1]^2).$$

Observe that $C_n^*(u, v) = \mathbb{C}_n(u, v) - \mathbb{C}_n(1 - u, 1 - v)$, $(u, v) \in [0, 1]^2$, with \mathbb{C}_n being the empirical copula process defined in (4.2). Therefore, from Condition 1 together with Segers [44], Proposition 3.1, and the continuous mapping theorem, we have that $C_n^* \rightsquigarrow \mathbb{C}^*$ in $\ell^\infty([0, 1]^2)$ and the proof is complete. \square

5. On a question by Jager and Wellner related to the Berk–Jones statistic

Let F_n be the empirical distribution function of a sample of size n from a univariate random variable with continuous distribution function F . Suppose that we want to test the null hypothesis $H_0 : F = G$ versus the alternative $H_1 : F \neq G$, where G is a fixed (and usually known) continuous distribution function. Berk and Jones [8] (see also DasGupta [13], Chapter 26.7) introduced the test statistic

$$R(F_n, G) := \sup_{x \in \mathbb{R}} K(F_n(x), G(x)), \tag{5.1}$$

where

$$K(x, y) := x \log\left(\frac{x}{y}\right) + (1 - x) \log\left(\frac{1 - x}{1 - y}\right),$$

for $x \in [0, 1]$ and $y \in (0, 1)$. (The values of $K(x, y)$ when $x = 0$ and $x = 1$ are taken by continuity.)

For each $x \in \mathbb{R}$, $nK(F_n(x), G(x))$ is the log-likelihood ratio statistic for testing $H_0 : F(x) = G(x)$ against $H_1 : F(x) \neq G(x)$. Hence, $R(F_n, G)$ in (5.1) is nothing but the supremum of these

pointwise likelihood ratio tests statistics. Additionally, $K(x, y)$ is the Kullback–Leibler divergence between two Bernoulli distributions with means x and y . Hence, $K(x, y) \geq 0$ with equality if and only if $x = y$. In particular, $R(F_n, G) = \|K(F_n, G)\|_\infty$.

Berk and Jones [8] computed the asymptotic distribution of (the normalized version of) $R(F_n, F)$, i.e., the distribution of the statistic under the null hypothesis $F = G$. For a detailed proof, see Wellner and Koltchinskii [56], Theorem 1.1, or Jager and Wellner [29], Theorem 3.1. It holds that

$$nR(F_n, F) - d_n \rightsquigarrow Y_4, \quad \text{as } n \rightarrow \infty, \tag{5.2}$$

where $P(Y_4 \leq x) = \exp(-4 \exp(-x))$ for $x \in \mathbb{R}$, i.e., Y_4 has double-exponential extreme value distribution, and

$$d_n := \log_2 n - \frac{1}{2} \log_3 n - \frac{1}{2} \log(4\pi),$$

with $\log_2 n := \log(\log n)$ and $\log_3 n := \log(\log_2 n)$.

In Jager and Wellner [28], Question 2, page 329, it was set out the open problem of finding the asymptotic behaviour of the Berk–Jones statistic under the alternative hypothesis. In other words, assuming that $F \neq G$, the question consists in finding conditions on F and G for which the statistic

$$B_n := \sqrt{n}(R(F_n, G) - R(F, G)), \tag{5.3}$$

converges in distribution and, in such a case, identifying its weak limit, where $R(F_n, G)$ is given in (5.1) and $R(F, G) := \sup_{x \in \mathbb{R}} K(F(x), G(x))$.

Here we give a precise answer for the previous question. First, we note that B_n in (5.3) has the general form of (1.3). In other words,

$$B_n = D_\sigma(q = K(F, G), \mathbb{Q}_n = K(F_n, G), r_n = \sqrt{n}), \tag{5.4}$$

where σ is defined in (1.1). Therefore, from (5.4) and Theorem 2.2, to obtain the asymptotic distribution of B_n in (5.3) it is enough to find the weak limit of the process \mathbb{W}_n given by

$$\mathbb{W}_n := \sqrt{n}(K(F_n, G) - K(F, G)). \tag{5.5}$$

This result is stated in the following theorem.

Theorem 5.1. *Let us assume that the function $\log(\frac{F(1-G)}{G(1-F)})$ is monotone around $\pm\infty$ and*

$$\int_{\mathbb{R}} \log^2\left(\frac{F(t)(1-G(t))}{G(t)(1-F(t))}\right) dF(t) < \infty.$$

The process \mathbb{W}_n defined in (5.5) satisfies that $\mathbb{W}_n \rightsquigarrow \mathbb{W}$ in $\ell^\infty(\bar{\mathbb{R}})$, where

$$\mathbb{W} := \mathbb{B}_F \log \frac{F(1-G)}{G(1-F)}, \tag{5.6}$$

and \mathbb{B}_F is an F -Brownian bridge.

Proof. Using Taylor’s theorem, we have that

$$K(F_n, G) - K(F, G) = (F_n - F) \log \frac{F(1 - G)}{G(1 - F)} + \frac{1}{2} \frac{(F_n - F)^2}{F_n^*(1 - F_n^*)}, \tag{5.7}$$

where F_n^* is between F and F_n . We set

$$\tilde{\mathbb{W}}_n := \sqrt{n}(F_n - F) \log \frac{F(1 - G)}{G(1 - F)}. \tag{5.8}$$

From (5.5) and (5.7), we have that

$$\|\mathbb{W}_n - \tilde{\mathbb{W}}_n\|_\infty = \frac{\sqrt{n}}{2} \left\| \frac{(F_n - F)^2}{F_n^*(1 - F_n^*)} \right\|_\infty. \tag{5.9}$$

Now, from (5.9) and Wellner and Koltchinskii [56], equation (2.2) (see also Jager and Wellner [29], equation (9)), we obtain that

$$\begin{aligned} \|\mathbb{W}_n - \tilde{\mathbb{W}}_n\|_\infty &=_{\text{st}} \sqrt{n}R(F_n, F) \\ &= \frac{1}{\sqrt{n}}(nR(F_n, F) - d_n) + \frac{d_n}{\sqrt{n}}, \end{aligned} \tag{5.10}$$

where ‘ $=_{\text{st}}$ ’ stands for equality in distribution. From (5.2) and (5.10), we conclude that $\|\mathbb{W}_n - \tilde{\mathbb{W}}_n\|_\infty \rightsquigarrow 0$. Hence, the processes \mathbb{W}_n and $\tilde{\mathbb{W}}_n$ have the same asymptotic behaviour (see van der Vaart [53], Theorem 18.10). Finally, the conclusion follows from van der Vaart [53], Example 19.12, page 273. \square

Remark 5.1. As it follows from the proof of Theorem 5.1, the process \mathbb{W}_n behaves asymptotically as $\tilde{\mathbb{W}}_n$ in (5.8), which is a *weighted empirical process*. Therefore, necessary and sufficient conditions for the convergence of the process \mathbb{W}_n defined in (5.5) are given by the Chibisov-O’Reilly theorem (see Shorack and Wellner [49], page 462).

We are now in position to solve the question proposed in Jager and Wellner [28].

Corollary 5.1. *In the conditions of Theorem 5.1, the statistic B_n in (5.3) satisfies that*

$$B_n \rightsquigarrow \sigma'_{K(F,G)}(\mathbb{W}) = \sup_{M^+(K(F,G))} \mathbb{W}, \quad \text{as } n \rightarrow \infty,$$

where \mathbb{W} is given in (5.6) and the set $M^+(\cdot)$ is defined in (2.16).

Remark 5.2. Similar results can be stated for the family of test statistics $S_n(s)$ based on ϕ -divergences introduced by Jager and Wellner [29]. Details are omitted.

6. Maximum mean discrepancies

6.1. Definition and examples

Let X and Y be two random variables taking values on a topological space (\mathcal{X}, τ) with Borel probability measures P and Q , respectively. Throughout this section we will use the notation $E_P(f)$ to denote the mathematical expectation of f with respect to the probability measure P . Further, the set \mathfrak{X} will be a class of real functions defined on \mathcal{X} that will be denoted by \mathcal{F} . We consider a statistic to measure the dissimilarity between P and Q (see Fortet and Mourier [22] and Müller [33]).

Definition 6.1. Let us consider a class \mathcal{F} of measurable functions $f : \mathcal{X} \rightarrow \mathbb{R}$. The *maximum mean discrepancy* (MMD in short) between P and Q with respect to the class \mathcal{F} is defined by

$$\text{MMD}[\mathcal{F}, P, Q] := \sup_{f \in \mathcal{F}} (E_P(f) - E_Q(f)). \tag{6.1}$$

To avoid indeterminate forms in the difference between expectations in (6.1), it is usually assumed that \mathcal{F} is a subset of $\mathcal{C}(\mathcal{X}, \tau)$, the class of bounded and continuous real functions on \mathcal{X} . The probability distribution of the variables is usually completely identified with the MMD with respect to $\mathcal{C}(\mathcal{X}, \tau)$. In fact, if (\mathcal{X}, d) is a metric space, then $P = Q$ if and only if $E_P(f) = E_Q(f)$, for all $f \in \mathcal{C}(\mathcal{X}, d)$ (see Dudley [18], Lemma 9.3.2). However, the class $\mathcal{C}(\mathcal{X}, d)$ is in general too large to deal with, so that suitable subsets are usually employed in practice. Another possibility is assuming that the functions $f \in \mathcal{F}$ satisfy that $\sup_{x \in \mathcal{X}} |f(x)|/b(x) < \infty$, for a measurable function $b : \mathcal{X} \rightarrow [1, \infty)$ such that $E_P(b) < \infty$ and $E_Q(b) < \infty$. For simplicity, in the following we will not mention these necessary integrability requirements and we will assume that $\sup_{f \in \mathcal{F}} E_P(f), \sup_{f \in \mathcal{F}} E_Q(f) < \infty$.

We observe that when \mathcal{F} is symmetric, that is, $-f \in \mathcal{F}$ whenever $f \in \mathcal{F}$, we have that $\text{MMD}[\mathcal{F}, P, Q] = \sup_{f \in \mathcal{F}} |E_P(f) - E_Q(f)|$. In other words, the MMD in (6.1) is the *integral probability metric* generated by \mathcal{F} (see Müller [33]). In Rachev et al. [38], Section 4.4, it is also said that the metric has a ζ -structure; see Zolotarev [57]. In this section we will also assume that \mathcal{F} is symmetric.

Some frequently used probability metrics can be expressed as $\text{MMD}[\mathcal{F}, P, Q]$, for a suitable choice of the set of functions \mathcal{F} . In the following examples X and Y are two random variables with distribution functions F and G and associated probability measures P and Q , respectively.

1. *Kolmogorov metric.* This distance is $\|F - G\|_\infty$, which is the integral probability metric generated by $\mathcal{F} = \{1_{(-\infty, x]} : x \in \mathbb{R}\}$. Further, it is also generated by the set of all functions of bounded variation 1 (see Müller [33], Theorem 5.2).
2. *L^p metrics.* For $1 \leq p < \infty$, this metric is defined by $d_p(F, G) := \|F - G\|_p$ ($\|\cdot\|_p$ being the usual L^p -norm). When X and Y are integrable, d_p admits the dual representation (see Rachev et al. [38], page 73) $d_p(F, G) = \text{MMD}[\mathcal{F}_p, P, Q]$, where \mathcal{F}_p is the class of all Lebesgue a.e. differentiable functions f such that the derivative f' satisfies $\|f'\|_q \leq 1$ (q being the conjugate of p , i.e., q is such that $1/p + 1/q = 1$).

3. *Wasserstein metric.* This distance is a particular and important case of the L^p -metric with $p = 1$. Its generator is also the class $\mathcal{F}_W \equiv$ the set of functions $f : \mathbb{R} \rightarrow \mathbb{R}$ satisfying the Lipschitz condition $|f(x) - f(y)| \leq |x - y|$, for all $(x, y) \in \mathbb{R}^2$. By the Kantorovich-Rubinstein theorem, $\|F - G\|_1 = \text{MMD}[\mathcal{F}_W, P, Q]$. In the context of image processing, this metric is called the *earth mover's distance* (see Rubner, Tomasi and Guibas [42]). The importance of the Wasserstein metric, as well as its relevance for optimal transport problems, has been summarized in Villani [55], Section 6.
4. *Bounded Lipschitz metric.* This metric (see Huber [27], page 29) is the integral probability metric generated by $\mathcal{F}_{BL} := \{f : \|f\|_{BL} \leq 1\}$, where $\|f\|_{BL} := \|f\|_L + \|f\|_\infty$ and $\|\cdot\|_L$ is the Lipschitz norm given by

$$\|f\|_L := \sup_{x \neq y \in \mathbb{R}} \frac{|f(x) - f(y)|}{|x - y|}.$$

5. *Zolotarev ideal metrics of order r .* For $r \in \mathbb{N}$, let \mathcal{Z}_r be the class of $(r - 1)$ -times continuously differentiable functions $f : \mathbb{R} \rightarrow \mathbb{R}$ satisfying the Lipschitz condition $|f^{(r-1)}(x) - f^{(r-1)}(y)| \leq |x - y|$, for all $(x, y) \in \mathbb{R}^2$. (Here we use the notation $f^{(0)} \equiv f$.) The class \mathcal{Z}_r can also be substituted by the set of functions f having r th derivative $f^{(r)}$ a.e. and such that $|f^{(r)}| \leq 1$ a.e. The metric $\zeta_r \equiv \text{MMD}[\mathcal{Z}_r, P, Q]$ is called the *Zolotarev metric of order r* (see Rachev et al. [38] for a general reference and properties of these distances). Convergence in ζ_r -metric implies weak convergence plus convergence of the r th absolute moment. Zolotarev metrics have been used in Rao [40] to obtain a CLT for independent, non-identically distributed random variables. As mentioned in Rachev et al. [38], Section 15, the case $r = 2$ is appropriate for investigating some ageing properties of lifetime distributions. In Baíllo, Cárcamo and Getman [3], ζ_2 has also been used to generate new distance measures for classifying X-ray astronomy data into stellar classes. The metric ζ_3 has been considered in the context of distributional recurrences (see Neininger and Rüschendorf [34] and Neininger and Rüschendorf [35]).
6. *Zolotarev metric of order r in L^p .* For $r \in \mathbb{N}$, and $1 \leq p \leq \infty$, the metric $\zeta_{r,p}$ is generated by $\mathcal{Z}_{r,p}$, the set of functions $f : \mathbb{R} \rightarrow \mathbb{R}$ for which $f^{(r+1)}$ exists and satisfies $\|f^{(r+1)}\|_q \leq 1$, where q is the conjugate of p . Note that $\zeta_{r,1} \equiv \zeta_{r+1}$ (the Zolotarev ideal metric of order $r + 1$). In risk theory, the metrics $\zeta_{1,\infty}$ and $\zeta_{1,1}$ are respectively called the *stop-loss distance* and the *integrated stop-loss distance* (see Denuit et al. [14]).
7. *Kernel distances.* When $\mathcal{F} = \{f : \|f\|_{\mathcal{H}} \leq 1\}$ is the unit ball in a reproducing kernel Hilbert space \mathcal{H} , the associated MMD is called *kernel distance*.

6.2. An asymptotic result for the MMD over Donsker classes

The use of the empirical counterpart of the MMD was already considered in Fortet and Mourier [22] and it has been extensively employed in machine learning when \mathcal{F} is the unit ball in a reproducing kernel Hilbert space (RKHS) (see Gretton et al. [26]). In Sriperumbudur et al. [52], the authors showed the consistency and rate of convergence of some estimators of various integral probability metrics. The asymptotic behaviour of an estimator of the Zolotarev metric of order r in L^p has been discussed in Cárcamo [11]. Here we provide a general result regarding the

estimation of the MMD. We only consider the two sample case as this situation is the most frequently considered in the literature, but similar results can be obtained in the one sample case.

Let X_1, \dots, X_n and Y_1, \dots, Y_m be two independent random samples from X and Y with probability measures P and Q , respectively. We denote by P_n and Q_m the empirical measures associated with these samples, that is, $P_n = n^{-1} \sum_{i=1}^n \delta_{X_i}$ and $Q_m = m^{-1} \sum_{j=1}^m \delta_{Y_j}$, where δ_a stands for the Dirac delta at the point a . Given a class of functions \mathcal{F} , the empirical counterpart of $\text{MMD}[\mathcal{F}, P, Q]$ in (6.1) is given by

$$\text{MMD}[\mathcal{F}, P_n, Q_m] = \sup_{f \in \mathcal{F}} \left(\frac{1}{n} \sum_{i=1}^n f(X_i) - \frac{1}{m} \sum_{j=1}^m f(Y_j) \right). \tag{6.2}$$

In this section, we are interested in the asymptotic behaviour of the quantity

$$M_{m,n} := \sqrt{N} (\text{MMD}[\mathcal{F}, P_n, Q_m] - \text{MMD}[\mathcal{F}, P, Q]), \quad \text{with } N \equiv \frac{nm}{n+m}. \tag{6.3}$$

We observe that $M_{m,n}$ is precisely $D_{n,m}(\sigma) = D_\sigma(D, \mathbb{D}_{n,m}, r_{n,m})$ in (1.3), where the underlying space is $\mathfrak{X} = \mathcal{F}$; the target functional is $D \in \ell^\infty(\mathcal{F})$ given by

$$D(f) := E_P(f) - E_Q(f), \quad f \in \mathcal{F}; \tag{6.4}$$

its estimator is

$$\mathbb{D}_{n,m}(f) := E_{P_n}(f) - E_{Q_m}(f) = \frac{1}{n} \sum_{i=1}^n f(X_i) - \frac{1}{m} \sum_{j=1}^m f(Y_j), \quad f \in \mathcal{F};$$

and $r_{n,m} := \sqrt{N}$. Therefore, from Theorem 2.2, to derive the asymptotic distribution of $M_{m,n}$ in (6.3) we only need to study the weak convergence in $\ell^\infty(\mathcal{F})$ of the process $r_{n,m}(\mathbb{D}_{n,m} - D) := \mathbb{G}_{n,m}$ given by

$$\mathbb{G}_{n,m} := \sqrt{\frac{m}{n+m}} \mathbb{G}_{n,P} - \sqrt{\frac{n}{n+m}} \mathbb{G}_{m,Q}, \tag{6.5}$$

where

$$\mathbb{G}_{n,P} := \sqrt{n}(P_n - P) \quad \text{and} \quad \mathbb{G}_{m,Q} := \sqrt{m}(Q_m - Q)$$

are two independent \mathcal{F} -indexed empirical processes associated with P and Q , respectively. In other words, for $f \in \mathcal{F}$, we have that

$$\mathbb{G}_{n,P}(f) = n^{-1/2} \sum_{i=1}^n (f(X_i) - E_P(f)) \quad \text{and} \quad \mathbb{G}_{m,Q}(f) = m^{-1/2} \sum_{j=1}^m (f(Y_j) - E_Q(f)).$$

Given a probability measure P , we recall that a class of functions \mathcal{F} is said to be *P-Donsker* if $\mathbb{G}_{n,P} \rightsquigarrow \mathbb{G}_P$ in $\ell^\infty(\mathcal{F})$, where \mathbb{G}_P is a P -Brownian bridge, that is, $\{\mathbb{G}_P(f) : f \in \mathcal{F}\}$ is a zero-mean Gaussian process with covariance function

$$E[\mathbb{G}_P(f_1)\mathbb{G}_P(f_2)] = E_P(f_1 f_2) - E_P(f_1)E_P(f_2), \quad f_1, f_2 \in \mathcal{F}.$$

Additionally, \mathcal{F} is *universal Donsker* if it is *P-Donsker*, for every probability measure P on the sample space.

We observe that whenever \mathcal{F} is *P-Donsker*, the process \mathbb{G}_P can be uniquely extended to the d_P -closure of the symmetric convex hull generated by \mathcal{F} (see Giné and Nickl [25], Theorem 3.7.28), where d_P is the intrinsic pseudo-metric on \mathcal{F} defined by

$$d_P^2(f, g) := E(\mathbb{G}_P(f) - \mathbb{G}_P(g))^2 = E_P(f - g)^2 - (E_P(f - g))^2, \quad f, g \in \mathcal{F}.$$

To simplify the writing, we will not use a different notation for this extension of \mathbb{G}_P .

We are in position to state the main result in this section that determines the asymptotic distribution of the statistic $M_{n,m}$ in (6.3) over Donsker classes.

Theorem 6.1. *Let X and Y be two random variables with probability measures P and Q , respectively. Let us assume that*

- (a) *The sampling scheme is balanced, that is, $n/(n+m) \rightarrow \lambda$, with $0 < \lambda < 1$, as $n, m \rightarrow \infty$.*
- (b) *The class \mathcal{F} is simultaneously P and Q -Donsker.*

We consider the metric d on \mathcal{F} given by

$$d(f, g) := \sqrt{E_P(f - g)^2} + \sqrt{E_Q(f - g)^2}, \quad f, g \in \mathcal{F}. \quad (6.6)$$

We have that (\mathcal{F}, d) is a totally bounded metric space, the function D in (6.4) belongs to $\mathcal{C}_u(\mathcal{F}, d)$ and the statistic $M_{n,m}$ defined in (6.3) satisfies that

$$M_{n,m} \rightsquigarrow \sup_{\bar{M}^+(D, d)} \mathbb{G},$$

where $\mathbb{G} := \sqrt{1-\lambda}\mathbb{G}_P - \sqrt{\lambda}\mathbb{G}_Q$ is a zero-mean Gaussian process with \mathbb{G}_P and \mathbb{G}_Q two independent \mathcal{F} -indexed Brownian bridges associated with P and Q , respectively, and

$$\bar{M}^+(D, d) := \{f \in (\bar{\mathcal{F}}, d) : E_P(f) - E_Q(f) = \text{MMD}[\mathcal{F}, P, Q]\}$$

with $\bar{\mathcal{F}}$ being the d -completion of \mathcal{F} .

Proof. First, from (a) and (b) we have that $\mathbb{G}_{n,m} \rightsquigarrow \mathbb{G}$, where $\mathbb{G}_{n,m}$ is in (6.5). Hence, by Theorem 2.2, $M_{n,m} \rightsquigarrow \sigma'_D(\mathbb{G})$. Now, as \mathcal{F} is P and Q -Donsker, the pseudo-metric spaces (\mathcal{F}, d_P) and (\mathcal{F}, d_Q) are totally bounded, where d_P and d_Q are the natural pseudo-metrics given by $d_S^2(f, g) := E_S(f - g)^2 - (E_S(f - g))^2$, for $S \in \{P, Q\}$ and $f, g \in \mathcal{F}$ (see Giné and Nickl [25], Remark 3.7.27). Further, $\mathbb{G}_P \in \mathcal{C}_u(\mathcal{F}, d_P)$ and $\mathbb{G}_Q \in \mathcal{C}_u(\mathcal{F}, d_Q)$ a.s. Now, as the class \mathcal{F} is bounded in $L^1(P)$ and $L^1(Q)$ (i.e., $\sup_{f \in \mathcal{F}} |E_P(f)|, \sup_{f \in \mathcal{F}} |E_Q(f)| < \infty$) and (\mathcal{F}, d_P) , (\mathcal{F}, d_Q) are totally bounded, using the same ideas as in the proof of Giné and Nickl [25], Theorem 3.7.40, page 262, we conclude that $(\mathcal{F}, d_{L^2(P)})$ and $(\mathcal{F}, d_{L^2(Q)})$ are also totally bounded, where $d_{L^2(S)}^2(f, g) := E_S(f - g)^2$ ($f, g \in \mathcal{F}$ and $S \in \{P, Q\}$). It is easy to check that this implies that (\mathcal{F}, d) is totally bounded, where d is in (6.6). On the other hand, by Cauchy–Schwarz inequality, we have that $|D(f) - D(g)| \leq d(f, g)$ and hence $D \in \mathcal{C}_u(\mathcal{F}, d)$. Further, the paths of

\mathbb{G} are in $\mathcal{C}_n(\mathcal{F}, d)$ a.s. since $d_P, d_Q \leq d$. Therefore, the conclusion follows by applying Corollary 2.5(b). \square

Condition (b) in Theorem 6.1 is the key assumption that has to be checked to apply the previous result. In other words, we have to ensure that \mathcal{F} is P and Q-Donsker. There are many results in the literature on empirical processes guaranteeing that a class of functions is Donsker (see van der Vaart and Wellner [54]). For instance, it is well known that the set of indicators generating the Kolmogorov distance is universal Donsker. The unit ball for the Bounded Lipschitz metric is P-Donsker whenever P has some finite moments (see Nickl and Pötscher [37], Corollary 5 and Remark 2). In the same work, Nickl and Pötscher [37] showed that bounded subsets of general function spaces defined over \mathbb{R}^d are Donsker under some appropriate conditions on the underlying probability measure. Examples include (weighted) Besov, Sobolev, Hölder, and Triebel type spaces. Some of these results have been extended in Sriperumbudur [51].

Acknowledgements

This research has been supported by the Spanish MCyT grant MTM2016-78751-P. We thank Holger Dette for pointing out to us the reference Dette, Kokot and Aue [15]. The first author thanks Carlos Mora-Corral (Department of Mathematics, Universidad Autónoma de Madrid) for showing to him the counterexample in Section 2.5 that proves that the linear extension \tilde{g} of a function $g \in \mathcal{C}_{\text{pl}}(\mathcal{X}, d_B)$ is not necessarily continuous. We are also indebted to the reviewers, AE and Editor for their appropriate and constructive suggestions and the references Beare and Moon [6], Kaido [30], Seo [46] and Beare and Shi [7]. Their comments have led to a improved version of the original manuscript.

References

- [1] Álvarez-Esteban, P.C., del Barrio, E., Cuesta-Albertos, J.A. and Matrán, C. (2012). Similarity of samples and trimming. *Bernoulli* **18** 606–634. MR2922463 <https://doi.org/10.3150/11-BEJ351>
- [2] Álvarez-Esteban, P.C., del Barrio, E., Cuesta-Albertos, J.A. and Matrán, C. (2016). A contamination model for the stochastic order. *TEST* **25** 751–774. MR3554414 <https://doi.org/10.1007/s11749-016-0494-2>
- [3] Baíllo, A., Cárcamo, J. and Getman, K. (2019). New distance measures for classifying X-ray astronomy data into stellar classes. *Adv. Data Anal. Classif.* **13** 531–557. MR3954521 <https://doi.org/10.1007/s11634-018-0309-2>
- [4] Banach, S. (1936). *Théorie des Opérations Linéaires. Monografie Mat.* **1**. Warsaw.
- [5] Beare, B.K. and Fang, Z. (2017). Weak convergence of the least concave majorant of estimators for a concave distribution function. *Electron. J. Stat.* **11** 3841–3870. MR3714300 <https://doi.org/10.1214/17-EJS1349>
- [6] Beare, B.K. and Moon, J.-M. (2015). Nonparametric tests of density ratio ordering. *Econometric Theory* **31** 471–492. MR3348455 <https://doi.org/10.1017/S0266466614000401>
- [7] Beare, B.K. and Shi, X. (2019). An improved bootstrap test of density ratio ordering. *Écon. Stat.* **10** 9–26. MR3945178 <https://doi.org/10.1016/j.ecosta.2018.08.002>

- [8] Berk, R.H. and Jones, D.H. (1979). Goodness-of-fit test statistics that dominate the Kolmogorov statistics. *Z. Wahrsch. Verw. Gebiete* **47** 47–59. MR0521531 <https://doi.org/10.1007/BF00533250>
- [9] Bickel, P.J. and Wichura, M.J. (1971). Convergence criteria for multiparameter stochastic processes and some applications. *Ann. Math. Stat.* **42** 1656–1670. MR0383482 <https://doi.org/10.1214/aoms/1177693164>
- [10] Brezis, H. (2011). *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Universitext. New York: Springer. MR2759829
- [11] Cárcamo, J. (2017). Integrated empirical processes in L^p with applications to estimate probability metrics. *Bernoulli* **23** 3412–3436. MR3654811 <https://doi.org/10.3150/16-BEJ851>
- [12] Conway, J.B. (1990). *A Course in Functional Analysis*, 2nd ed. *Graduate Texts in Mathematics* **96**. New York: Springer. MR1070713
- [13] DasGupta, A. (2008). *Asymptotic Theory of Statistics and Probability*. *Springer Texts in Statistics*. New York: Springer. MR2664452
- [14] Denuit, M., Dhaene, J., Goovaerts, M. and Kaas, R. (2005). *Actuarial Theory for Dependent Risks: Measures, Orders and Models*. New York: Wiley.
- [15] Dette, H., Kokot, K. and Aue, A. (2018). Functional data analysis in the Banach space of continuous functions. Preprint. Available at arXiv:1710.07781v2 [math.ST].
- [16] Dette, H., Möllenhoff, K., Volgushev, S. and Bretz, F. (2018). Equivalence of regression curves. *J. Amer. Statist. Assoc.* **113** 711–729. MR3832221 <https://doi.org/10.1080/01621459.2017.1281813>
- [17] Dudley, R.M. (1999). *Uniform Central Limit Theorems*. *Cambridge Studies in Advanced Mathematics* **63**. Cambridge: Cambridge Univ. Press. MR1720712 <https://doi.org/10.1017/CBO9780511665622>
- [18] Dudley, R.M. (2002). *Real Analysis and Probability*. *Cambridge Studies in Advanced Mathematics* **74**. Cambridge: Cambridge Univ. Press. MR1932358 <https://doi.org/10.1017/CBO9780511755347>
- [19] Dümbgen, L. (1993). On nondifferentiable functions and the bootstrap. *Probab. Theory Related Fields* **95** 125–140. MR1207311 <https://doi.org/10.1007/BF01197342>
- [20] Fang, Z. and Santos, A. (2019). Inference on directionally differentiable functions. *Rev. Econ. Stud.* **86** 377–412. MR3936869 <https://doi.org/10.1093/restud/rdy049>
- [21] Fermanian, J.-D. (2013). An overview of the goodness-of-fit test problem for copulas. In *Copulae in Mathematical and Quantitative Finance. Lect. Notes Stat.* **213** 61–89. Heidelberg: Springer. MR3288239 https://doi.org/10.1007/978-3-642-35407-6_4
- [22] Fortet, R. and Mourier, E. (1953). Convergence de la répartition empirique vers la répartition théorique. *Ann. Sci. Éc. Norm. Supér.* (3) **70** 267–285. MR0061325
- [23] Freitag, G., Lange, S. and Munk, A. (2006). Non-parametric assessment of non-inferiority with censored data. *Stat. Med.* **25** 1201–1217. MR2225587 <https://doi.org/10.1002/sim.2444>
- [24] Genest, C. and Nešlehová, J.G. (2014). On tests of radial symmetry for bivariate copulas. *Statist. Papers* **55** 1107–1119. MR3266384 <https://doi.org/10.1007/s00362-013-0556-4>
- [25] Giné, E. and Nickl, R. (2016). *Mathematical Foundations of Infinite-Dimensional Statistical Models*. *Cambridge Series in Statistical and Probabilistic Mathematics*. New York: Cambridge Univ. Press. MR3588285 <https://doi.org/10.1017/CBO9781107337862>
- [26] Gretton, A., Borgwardt, K.M., Rasch, M.J., Schölkopf, B. and Smola, A. (2012). A kernel two-sample test. *J. Mach. Learn. Res.* **13** 723–773. MR2913716
- [27] Huber, P.J. (1981). *Robust Statistics*. *Wiley Series in Probability and Mathematical Statistics*. New York: Wiley. MR0606374
- [28] Jager, L. and Wellner, J.A. (2004). On the “Poisson boundaries” of the family of weighted Kolmogorov statistics. In *A Festschrift for Herman Rubin. Institute of Mathematical Statistics Lecture Notes – Monograph Series* **45** 319–331. Beachwood, OH: IMS. MR2126907 <https://doi.org/10.1214/lnms/1196285400>

- [29] Jager, L. and Wellner, J.A. (2007). Goodness-of-fit tests via phi-divergences. *Ann. Statist.* **35** 2018–2053. MR2363962 <https://doi.org/10.1214/0009053607000000244>
- [30] Kaido, H. (2016). A dual approach to inference for partially identified econometric models. *J. Econometrics* **192** 269–290. MR3463675 <https://doi.org/10.1016/j.jeconom.2015.12.017>
- [31] Leonard, I.E. and Taylor, K.F. (1983). Supremum norm differentiability. *Int. J. Math. Math. Sci.* **6** 705–713. MR0729390 <https://doi.org/10.1155/S0161171283000605>
- [32] Leonard, I.E. and Taylor, K.F. (1985). Essential supremum norm differentiability. *Int. J. Math. Math. Sci.* **8** 433–439. MR0809064 <https://doi.org/10.1155/S0161171285000473>
- [33] Müller, A. (1997). Integral probability metrics and their generating classes of functions. *Adv. in Appl. Probab.* **29** 429–443. MR1450938 <https://doi.org/10.2307/1428011>
- [34] Neininger, R. and Rüschendorf, L. (2004). A general limit theorem for recursive algorithms and combinatorial structures. *Ann. Appl. Probab.* **14** 378–418. MR2023025 <https://doi.org/10.1214/aoap/1075828056>
- [35] Neininger, R. and Rüschendorf, L. (2004). On the contraction method with degenerate limit equation. *Ann. Probab.* **32** 2838–2856. MR2078559 <https://doi.org/10.1214/009117904000000171>
- [36] Neuhaus, G. (1971). On weak convergence of stochastic processes with multidimensional time parameter. *Ann. Math. Stat.* **42** 1285–1295. MR0293706 <https://doi.org/10.1214/aoms/1177693241>
- [37] Nickl, R. and Pötscher, B.M. (2007). Bracketing metric entropy rates and empirical central limit theorems for function classes of Besov- and Sobolev-type. *J. Theoret. Probab.* **20** 177–199. MR2324525 <https://doi.org/10.1007/s10959-007-0058-1>
- [38] Rachev, S.T., Klebanov, L.B., Stoyanov, S.V. and Fabozzi, F.J. (2013). *The Methods of Distances in the Theory of Probability and Statistics*. New York: Springer. MR3024835 <https://doi.org/10.1007/978-1-4614-4869-3>
- [39] Raghavachari, M. (1973). Limiting distributions of Kolmogorov–Smirnov type statistics under the alternative. *Ann. Statist.* **1** 67–73. MR0346976
- [40] Rao, M.M., ed. (1997). *Real and Stochastic Analysis: Recent Advances. Probability and Stochastics Series*. Boca Raton, FL: CRC Press. MR1464220 <https://doi.org/10.1080/17442508808833525>
- [41] Römisch, W. (2004). Delta method, infinite dimensional. In *Encyclopedia of Statistical Sciences*. New York: Wiley.
- [42] Rubner, Y., Tomasi, C. and Guibas, L.J. (2000). The Earth Mover’s distance as a metric for image retrieval. *Int. J. Comput. Vis.* **40** 99–121.
- [43] Schmoyer, R.L. (1988). Linear interpolation with a nonparametric accelerated failure-time model. *J. Amer. Statist. Assoc.* **83** 441–449. MR0971370
- [44] Segers, J. (2012). Asymptotics of empirical copula processes under non-restrictive smoothness assumptions. *Bernoulli* **18** 764–782. MR2948900 <https://doi.org/10.3150/11-BEJ387>
- [45] Seijo, E. and Sen, B. (2011). A continuous mapping theorem for the smallest argmax functional. *Electron. J. Stat.* **5** 421–439. MR2802050 <https://doi.org/10.1214/11-EJS613>
- [46] Seo, J. (2018). Tests of stochastic monotonicity with improved power. *J. Econometrics* **207** 53–70. MR3856760 <https://doi.org/10.1016/j.jeconom.2018.04.004>
- [47] Shapiro, A. (1990). On concepts of directional differentiability. *J. Optim. Theory Appl.* **66** 477–487. MR1080259 <https://doi.org/10.1007/BF00940933>
- [48] Shapiro, A. (1991). Asymptotic analysis of stochastic programs **30** 169–186. MR1118896 <https://doi.org/10.1007/BF02204815>
- [49] Shorack, G.R. and Wellner, J.A. (1986). *Empirical Processes with Applications to Statistics. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics*. New York: Wiley. MR0838963
- [50] Sommerfeld, M. and Munk, A. (2018). Inference for empirical Wasserstein distances on finite spaces. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 219–238. MR3744719 <https://doi.org/10.1111/rssb.12236>

- [51] Sriperumbudur, B. (2016). On the optimal estimation of probability measures in weak and strong topologies. *Bernoulli* **22** 1839–1893. MR3474835 <https://doi.org/10.3150/15-BEJ713>
- [52] Sriperumbudur, B.K., Fukumizu, K., Gretton, A., Schölkopf, B. and Lanckriet, G.R.G. (2012). On the empirical estimation of integral probability metrics. *Electron. J. Stat.* **6** 1550–1599. MR2988458 <https://doi.org/10.1214/12-EJS722>
- [53] van der Vaart, A.W. (1998). *Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge: Cambridge Univ. Press. MR1652247 <https://doi.org/10.1017/CBO9780511802256>
- [54] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics. Springer Series in Statistics*. New York: Springer. MR1385671 <https://doi.org/10.1007/978-1-4757-2545-2>
- [55] Villani, C. (2009). *Optimal Transport: Old and New. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **338**. Berlin: Springer. MR2459454 <https://doi.org/10.1007/978-3-540-71050-9>
- [56] Wellner, J.A. and Koltchinskii, V. (2003). A note on the asymptotic distribution of Berk–Jones type statistics under the null hypothesis. In *High Dimensional Probability, III (Sandjberg, 2002). Progress in Probability* **55** 321–332. Basel: Birkhäuser. MR2033896
- [57] Zolotarev, V.M. (1983). Probability metrics. *Theory Probab. Appl.* **28** 278–302.

Received December 2018 and revised July 2019