General asymptotic power results: Directional differentiability of maps related to the sup-norm

Luis Alberto Rodríguez Ramírez

Máster en Matemáticas y Aplicaciones

MÁSTERES DE LA UAM

2017 - 2018

Facultad de Ciencias

Universidad Autónoma de Madrid

Máster en
Matemáticas y
Aplicaciones

General asymptotic power results: Directional differentiability of maps related to the sup-norm

Luis Alberto Rodríguez Ramírez

Director: Javier Cárcamo Urtiaga
Lugar: UAM - Facultad de Ciencias
Dpto. de Matemáticas

Trabajo Fin de Máster. Curso 2017-2018

Universidad Autónoma de Madrid

Abstract

We show the directional Hadamard differentiability of various functionals related to the supremum norm. Additionally, we provide simple expressions for the derivatives in the space of *càdlàg* functions with the supremum norm (Skorohod space). The interest of these results lies in the fact that the (functional) Delta method can be used to give asymptotic results about the asociated statistics. As an application, we improve the results in [Raghavachari, 1973] and solve an open problem in [Jager and Wellner, 2004] about Berk-Jones type statistics.

Contents

1 Summary

Let us suppose that we want to estimate the Kolmogorov distance between two random variables X and Y

$$d(X,Y) = \|F - G\|_\infty,$$

where F and G are respectively the distribution functions of X and Y. Let us assume that we are given a sample $X_1, \dots, X_n$ of X. So, we use the statistic

$$\Delta_n = \sqrt{n}\left(\|F_n - G\|_\infty - \|F - G\|_\infty\right), \tag{1}$$

where $F_n(x) = \frac{1}{n}\sum_{k=1}^n \mathbf{1}_{\{X_k \le x\}}$ is the empirical distribution function associated to the sample and $\mathbf{1}_A$ stands for the indicator function of the set A.

To express it in a different way, in (1) we are comparing the supremum distance from the empirical distribution function F_n to G and the supremum distance from F to G. This idea is behind the Kolmogorov-Smirnov test. In fact, when $X =_{\text{st}} Y$ (X and Y have the same distribution, that is, $\mathbb{P}(X \le x) = \mathbb{P}(Y \le x)$, for all $x \in \mathbb{R}$) (2) becomes

$$\Delta_n = \sqrt{n}\,\|F_n - F\|_\infty. \tag{2}$$

In other words, (2) is the so called Kolmogorov-Smirnorv statistic. The limiting distribution of (2) under the null hypothesis ($H_0 : F = G$) was first derived by Kolmogorov:

$$\Delta_n \rightsquigarrow \sup_{t\in\mathbb{R}}\left(\left|\mathbb{B}_{F(t)}\right|\right),$$

where we use the arrow '$\rightsquigarrow$' to denote the weak convergence of probability measures and $\mathbb{B}_F$ is a F-Brownian bridge (see Subsection 4.1). From the modern perspective of the theory, it is only an application of the extended continuous mapping theorem (see Theorem 2.4).

On the contrary, if $F \neq G$, it means we are testing 'how different' X and Y are. Under some regularity requirements (continuity of F and G), [Raghavachari, 1973] derived the asymptotic distribution of (1). Hence, two questions arise: Can the regularity assumptions be weakened? Can the results be extended to higher dimensions?

Additionally, it is known that Kolmogorov-Smirnov test does not provide satisfactory results when the differences between F and G are concentrated on the tails of the functions. Some alternative tests have been proposed in the literature (see for instance [Berk and Jones, 1979]) and questions about the asymptotic behaviour appear (see [Jager and Wellner, 2004]).

Our ambition was to answer all these questions. The seed of the theory we present consists of separating the study of the functionals asociated to the statistics and the convergence of the underlying processes (in the spirit of [Pollard, 1984]). The most important tool for our purposes in this setting is the functional Delta method, which can be sucintly summarized as analysis of weak limit of random elements (possibly in an infinite dimensional vector space) via first order expansions.

The aim of the Section 2 is to discuss the (directional) differentiability of the supremum norm and various related maps, viewed as real functionals from the space of bounded functions defined on an arbitrary set or a measure space. We consider the supremum norm, the supremum, and the total amplitude of a real function. As an application, we use an extended version of the functional Delta method to derive the asymptotic distribution of various 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.

Throughout this work, $\mathfrak{X}$ is a nonempty set and $\ell^\infty(\mathfrak{X})$ is the real Banach space of bounded functions $f : \mathfrak{X} \longrightarrow \mathbb{R}$, 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-field and μ a positive measure, we denote by $L^\infty(\mathfrak{X}, \mathcal{A}, \mu)$ the set of classes of equivalence of measurable and essentially bounded functions $f : \mathfrak{X} \longrightarrow \mathbb{R}$ with the norm $\|f\|_{L^\infty(\mathfrak{X},\mathcal{A},\mu)} = \operatorname*{esssup}_{x\in\mathfrak{X}}(|f(x)|)$, where

$$\operatorname*{esssup}_{x\in\mathfrak{X}}(f(x)) = \sup(\{C \in \mathbb{R} : \ \mu(\{x \in \mathfrak{X} : \ f(x) \leq C\}) > 0\}.$$

An important example of this general setting is the case in which $\mathfrak{X} = \mathbb{R}^d$ or $\overline{\mathbb{R}}^d$ $(d \geq 1)$, with $\overline{\mathbb{R}} = [-\infty, +\infty]$ the extended real line. To avoid unnecessary repetitions, unless specifically mentioned, from now on we will only consider the supremum without specifying the posible underlying measure.

Let us assume $\{\mathbb{Q}_n\}$ are random elements taking values in $\ell^\infty(\mathfrak{X})$, $q \in \ell^\infty(\mathfrak{X})$, and

$$r_n\ (\mathbb{Q}_n - q) \rightsquigarrow \mathbb{Q} \quad \text{in } \ \ell^\infty(\mathfrak{X}), \tag{3}$$

where r_n is a sequence of real numbers such that $r_n \to \infty$ (usually $r_n = \sqrt{n}$ but it might be different), and $\mathbb{Q}$ is a Borel random element in $\ell^\infty(\mathfrak{X})$ (see [van der Vaart and Wellner, 1996]). Given $\varphi : \ell^\infty(\mathfrak{X}) \longrightarrow \mathbb{R}$ we are interested in the assymptotics of

$$D_n(\varphi) = D_\varphi(q, \mathbb{Q}_n, r_n) = r_n\ (\varphi(\mathbb{Q}_n) - \varphi(q))\,. \tag{4}$$

In the following, we denote the functional defined by the supremum norm by δ, that is,

$$\delta(f) = \|f\|_\infty, \quad \text{for} \quad f \in \ell^\infty(\mathfrak{X}). \tag{5}$$

In light of (4)-(5), a direct and intuitive approach to find the asymptotic distribution of $D_n(\varphi)$ could be analyzing the differentiability of δ in (5) 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.

In addition to the map δ in (5), we will also consider

$$\sigma(f) = \sup_{x\in\mathfrak{X}}(f(x)) \qquad \text{and} \qquad \alpha(f) = \operatorname*{amp}_{x\in\mathfrak{X}}(f(x)), \quad f \in \ell^\infty(\mathfrak{X}), \tag{6}$$

where $\operatorname*{amp}_{x\in\mathfrak{X}}(f(x)) = \sup_{x\in\mathfrak{X}}(f(x)) - \inf_{x\in\mathfrak{X}}(f(x))$ is the total amplitude of the function f. Observe that, from (4), when $\mathfrak{X} = \mathbb{R}^d$ and the target function $q = F - G$ is the difference between two distribution functions, σ and α are associated with the (multidimensional) *one-sided Kolmogorov statistic* and the so-called *Kuiper statistic* (see Subsection 4.1 for details). The asymptotic behaviour of these two statistics in dimension 1 was also discussed in [Raghavachari, 1973] (see also [DasGupta, 2008, Chapter 26]).

In Section 3 we deepen in the theory of differentiability focusing in the Skorohod space. The results about the derivative of the cited functionals in this space provides the necessary machinery to improve in Section 4 the results of [Raghavachari, 1973]. Finally, in Subsection 4.2 we answer an open question asked in [Jager and Wellner, 2004].

2 Directional differentiability and the Delta method

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.

In many situations it is common to face the problem of estimating a transformation, $\varphi(\theta)$, 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 certain 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 $\varphi(T_n)$ can be determined by expanding φ around θ (von Mises calculus, see [van der Vaart, 1998, Chapter 20]) 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 [van der Vaart and Wellner, 1996, 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 ([Giné and Nickl, 2016]).

As a reminder we are going to review the basics of weak convergence. Let $(\mathcal{D}, d)$ be a metric space, $\mathcal{B}_{\mathcal{D}}$ the Borel σ-field, $\{\mathbb{P}_n\}$ and $\mathbb{P}$ be Borel probability measures on $\mathcal{D}$, and $\mathcal{C}_b(\mathcal{D})$ denotes the set of all bounded continuous functions on $\mathcal{D}$. The classical definition of weak convergence is: *the sequence* $\{\mathbb{P}_n\}$ *converges weakly to* $\mathbb{P}$, $\mathbb{P}_n \rightsquigarrow \mathbb{P}$, *if for all* $f \in \mathcal{C}_b(\mathcal{D})$

$$\int_{\mathcal{D}} f(\omega)\, d\mathbb{P}_n(\omega) \to \int_{\mathcal{D}} f(\omega)\, d\mathbb{P}(\omega). \tag{7}$$

Given $\{(\Omega_n, \mathcal{A}_n, \mathbb{P}_n)\}$ and $(\Omega, \mathcal{A}, \mathbb{P})$ probability spaces and measurable mappings $\{X_n\}$ and X, $X_n : \Omega_n \longrightarrow \mathcal{D}$, and $X : \Omega \longrightarrow \mathcal{D}$ with respect to the Borel σ-field $\mathcal{B}_{\mathcal{D}}$. We say that the sequence $\{X_n\}$ converges weakly to X, written $X_n \rightsquigarrow X$, if $\mathbb{P}_{X_n} \rightsquigarrow \mathbb{P}_X$, we mean, the sequence of measures induced by $\{X_n\}$ converge weakly to the measure induced by X. Classical theory about this mode of convergence includes results such that Portmanteau theorem, continuous mapping theorem and, Prohorov's theorem, tools for establishing tightness and weak convergence results for product spaces.

However, the clasical theory requires $\{\mathbb{P}_n\}$ to be Borel measure for each n. The empirical process does not even satisfies this condition (see [Billingsley, 1999, Chapter 15, 156-158] or [van der Vaart and Wellner, 1996, Problem 1.7.3]). The key idea to solve this problem, due to Hoffman and Jørgensen, is to drop the requirement of Borel measurability of each X_n upholding the requirement (7), where the integrals of the left-hand side are now outer expectations with respect to $\mathbb{P}_X$, provided X is still Borel measurable.

Definition 2.1. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be an arbitrary probability space and $T : \Omega \longrightarrow \overline{\mathbb{R}}$ an arbitrary map. The *outer integral* of T with respecto to $\mathbb{P}$ is defined as

$$\mathbb{E}^*(T) = \inf\left(\left\{\mathbb{E}(U) :\ U \geq T,\ U : \Omega \longrightarrow \overline{\mathbb{R}} \text{ measurable},\ \mathbb{E}(U) \text{ exists}\right\}\right).$$

Here $\mathbb{E}(U)$ is understood to exists if $\mathbb{E}\left(U^{+}\right)$ or $\mathbb{E}\left(U^{-}\right)$ is finite, where $U^{+}(\omega)=\max(\{U(\omega),0\})$ and $U^{-}(\omega)=\max(\{-U(\omega),0\})$.

The *outer probability* of an arbitrary $B \subseteq \Omega$ is

$$\mathbb{P}^{*}(B)=\inf(\{\mathbb{P}(A):\ A \supseteq B \text{ and } A \in \mathcal{A}\})=\mathbb{E}^{*}\left(\mathbf{1}_{B}\right).$$

Note that the functions U in the definition of outer integral are allowed to take the value ∞, so that the set where we take the infimum is never empty.

Definition 2.2. Let $\{(\Omega_n,\mathcal{A}_n,\mathbb{P}_n)\}$ be a sequence of probability spaces and $X_n:\Omega_n \longrightarrow \mathcal{D}$ arbitrary maps. The sequence $\{X_n\}$ *converges weakly* to a Borel measure L if for all $f \in \mathcal{C}_b(\mathcal{D})$

$$\mathbb{E}^{*}\left(f\left(X_n\right)\right) \longrightarrow \int_{\mathcal{D}} f(\omega)\, dL(\omega).$$

This is denoted by $X_n \rightsquigarrow L$. If X has a law L we also say that $\{X_n\}$ converges weakly to X and write $X_n \rightsquigarrow X$.

The names *convergence in law* and *convergence in distribution* are used too.

Remark 2.3. If $\{X_n\}$ if a sequence of Borel measurable maps, then this definition is equivalent to the classical one because de outer expectation coincides with the (usual) expectation.

Most of the classical results can be stated and extended in this framework (see [van der Vaart and Wellner, 1996, Part 1]). One of the most important theorems for our purposes is the so called *extended continuous mapping theorem.*

Teorema 2.4. *Let $(\mathcal{D},d_{\mathcal{D}})$ and $(\mathcal{E},d_{\mathcal{E}})$ be metric spaces and let $\{\mathcal{D}_n\} \subseteq \mathcal{P}(\mathcal{D})$, $g_n:\mathcal{D}_n \longrightarrow \mathcal{E}$ satisfy*

asymptotic equicontinuity condition*: if $x_n \to x$ with $x_n \in \mathcal{D}_n$ for every n and $x \in \mathcal{D}_0$ then $g_n\left(x_n\right) \to g(x)$ where $\mathcal{D}_0 \subseteq \mathcal{D}$ and $g:\mathcal{D}_0 \longrightarrow \mathcal{E}$.*

Let X_n be maps with values in $\mathcal{D}_n$, let X be Borel measurable, separable and take values in $\mathcal{D}_0$. Then $X_n \rightsquigarrow X$ implies that $g_n\left(X_n\right) \rightsquigarrow g(X)$.

Additionally, the extended continuous mapping theorem is one of the necessary ingredients to deduce the functional delta method. The other one is an appropiate notion of differentiability.

Definition 2.5. Let $(\mathcal{D},\|\cdot\|_{\mathcal{D}})$ and $(\mathcal{E},\|\cdot\|_{\mathcal{E}})$ be Banach spaces. A map $\varphi:\mathcal{D}\longrightarrow\mathcal{E}$ 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 $\varphi_\theta':\mathcal{D}_0 \longrightarrow \mathcal{E}$ such that for all $h \in \mathcal{D}_0$ and all sequences $\{t_n\} \subset \mathbb{R}$ such that $t_n \searrow 0$.

$$\left\|\frac{\varphi\left(\theta+t_n h\right)-\varphi(\theta)}{t_n}-\varphi_\theta'(h)\right\|_{\mathcal{E}} \to 0. \tag{8}$$

Gâteaux differentiability is too weak for the Delta method to hold. To solve this problem, the directions along which we approach to $\varphi(\theta)$ in (8) have to be allowed to change with n. This naturally leads to the concept of Hadamard directional differentiability. We follow [Shapiro, 1990] for the next definition.

Definition 2.6. In the context of the previous definition, we say that $\varphi : \mathcal{D} \longrightarrow \mathcal{E}$ is *Hadamard directionally differentiable* at $\theta \in \mathcal{D}$ tangentially to a set $\mathcal{D}_0 \subset \mathcal{D}$ if there exists a map $\varphi'_\theta : \mathcal{D}_0 \longrightarrow \mathcal{E}$ such that for all $h \in \mathcal{D}_0$ and all sequences $\{h_n\} \subset \mathcal{D}$, $\{t_n\} \subset \mathbb{R}$ such that $t_n \searrow 0$ and $\|h_n - h\|_{\mathcal{D}} \to 0$

$$\left\| \frac{\varphi\left(\theta + t_n\, h_n\right) - \varphi(\theta)}{t_n} - \varphi'_\theta(h) \right\|_{\mathcal{E}} \to 0. \tag{9}$$

Now we will give some example to get more familiar with this new notion of differentiability provided in [Fang and Santos, 2015].

Example (Absolute value of mean). Let X be a real random variable and suppose we want to estimate de parameter

$$\varphi\left(\theta_0\right) = |\mathbb{E}(X)|.$$

Here $\theta_0 = \mathbb{E}(X)$, $\mathcal{D} = \mathcal{E} = \mathbb{R}$ and, $\varphi : \mathbb{R} \longrightarrow \mathbb{R}$ is defined by for all $\theta \in \mathbb{R} \quad \varphi(\theta) = |\theta|$. The Hadamard directional derivative $\varphi'_\theta : \mathbb{R} \longrightarrow \mathbb{R}$ equals

$$\varphi'_\theta(h) = \begin{cases} h & \text{if } \theta > 0 \\ |h| & \text{if } \theta = 0 \\ -h & \text{if } \theta < 0 \end{cases}.$$

Note that φ is actually (fully) Hadamard differentiable everywhere except at $\theta = 0$ but that it is still Hadamard directionally differentiable at that point.

Example (Intersection bounds). Let $\mathbb{X} = \left(X^{(1)}, X^{(2)}\right)$ be a bivariate real random vector and consider the problem of estimating the parameter

$$\varphi\left(\theta_0\right) = \max\left\{\mathbb{E}\left(X^{(1)}\right), \mathbb{E}\left(X^{(2)}\right)\right\}.$$

In this context, $\theta_0 = \left(\mathbb{E}\left(X^{(1)}\right), \mathbb{E}\left(X^{(2)}\right)\right)$, $\mathcal{D} = \mathbb{R}^2$, $\mathcal{E} = \mathbb{R}$, and $\varphi : \mathbb{R}^2 \longrightarrow \mathbb{R}$ is given by $\varphi(\theta) = \max\left\{\theta^{(1)}, \theta^{(2)}\right\}$ for $\theta = \left(\theta^{(1)}, \theta^{(2)}\right) \in \mathbb{R}^2$. Let $j^* = \underset{j \in \{1,2\}}{\operatorname{argmax}} \left(\theta^{(j)}\right)$. For any $h = \left(h^{(1)}, h^{(2)}\right) \in \mathbb{R}^2$ it is straightforward to verify that $\varphi'_\theta : \mathbb{R}^2 \longrightarrow \mathbb{R}$ is given by

$$\varphi'_\theta(h) = \begin{cases} h^{(j^*)} & \text{if } \theta^{(1)} \neq \theta^{(2)} \\ \max\left(\left\{h^{(1)}, h^{(2)}\right\}\right) & \text{if } \theta^{(1)} = \theta^{(2)} \end{cases}.$$

As in the previous example, φ'_θ is nonlinear precisely when Hadamard differentiability is not satisfied.

Example (Conditional moment inequalities). Let $X = (Y, Z)$ with $Y \in \mathbb{R}$ and $Z \in \mathbb{R}^d$ with $d \in \mathbb{N} \setminus \{0\}$. Fo a suitable set of functions $\mathcal{F} \subseteq \ell^\infty\left(\mathbb{R}^d\right)$ we are going to test wether $\mathbb{E}(Y|Z) \leq 0$ almost surely by estimating the parameter

$$\varphi\left(\theta_0\right) = \sup_{f \in \mathcal{F}} (\mathbb{E}(Y\, f(Z))),$$

where $\theta_0 \in \ell^\infty(\mathcal{F})$ satisfies $\theta_0(f) = \mathbb{E}(Y\, f(Z))$ for all $f \in \mathcal{F}$, $\mathcal{D} = \ell^\infty(\mathcal{F})$, $\mathcal{E} = \mathbb{R}$, and the map $\varphi : \mathcal{D} \longrightarrow \mathcal{E}$ is given by $\varphi(\theta) = \sup_{f\in\mathcal{F}}(\theta(f))$. This example is infinite dimensional so we are going to derive tangentially to $\mathcal{C}(\mathcal{F}) \subsetneq \ell^\infty(\mathcal{F})$, the space of continuous functions over $\mathcal{F}$. Additionally suppose $\mathbb{E}\left(Y^2\right) < \infty$ and that $\mathcal{F}$ is compact when endowed with the metric: $\|f\|_{L^2(Z)} = \left(\mathbb{E}\left(f(Z)^2\right)\right)^{1/2}$ for $f \in \mathcal{F}$. Then $\theta_0 \in \mathcal{C}(\mathcal{F})$ and φ is Hadamard directionally differentiable at any $\theta \in \mathcal{C}(\mathcal{F})$ tangentially to $\mathcal{C}(\mathcal{F})$. In particular, for $\Psi_{\mathcal{F}}(\theta) = \operatorname{argmax}_{f\in\mathcal{F}}(\theta(f))$ the directional derivative is

$$\varphi'_\theta(h) = \sup_{f\in\Psi_{\mathcal{F}}(\theta)}(h(f)).$$

Obviously, Hadamard directional differentiability implies the Gâteaux one. 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 2.5 and 2.6. Nevertheless, if equation (9) is satisfied, then φ'_θ is continuous and homogeneous of degree 1 [Shapiro, 1990, Proposition 3.1]. The important fact about Hadamard directional differentiability is that it allows the application of what we call the *extended (functional) Delta method*.

Proposition 2.7. *Let $(\mathcal{D}, \|\cdot\|_{\mathcal{D}})$ and $(\mathcal{E}, \|\cdot\|_{\mathcal{E}})$ be Banach spaces. spaces and $\varphi : \mathcal{D}_\varphi \subset \mathcal{D} \longrightarrow \mathcal{E}$, where $\mathcal{D}_\varphi$ is the domain of φ. Assume that φ is Hadamard directionally differentiable at $\theta \in \mathcal{D}_\varphi$ tangentially to a set $\mathcal{D}_0 \subset \mathcal{D}$. Let $T_n : \Omega_n \longrightarrow \mathcal{D}_\varphi$ be maps such that $r_n\,(T_n - \theta) \rightsquigarrow T$, for some sequence of numbers $r_n \to \infty$ and a random element T that takes values in $\mathcal{D}_0$. Then, $r_n\,(\varphi\,(T_n) - \varphi(\theta)) \rightsquigarrow \varphi'_\theta(T)$. If additionally φ'_θ can be continuously extended to $\mathcal{D}$, then we have that $r_n\,(\varphi\,(T_n) - \varphi(\theta)) = \varphi'_\theta\,(r_n\,(T_n - \theta)) + o_{\mathbb{P}}(1)$.*

Remark 2.8. The detailed proof of this result can be found in [Shapiro, 1991, Theorem 2.1] (see also [Römisch, 2006, Theorem 1] or [Fang and Santos, 2015, Theorem 2.1]), but it is essentially the same one as for the traditional Delta method in [van der Vaart, 1998, Theorem 20.8]. The key idea is to apply Theorem 2.4 to the sequence of functionals defined by $\varphi_n(h) = r_n\,\left(\varphi\left(\theta + r_n^{-1}\,h\right) - \varphi(\theta)\right)$, $n \in \mathbb{N}$ because the linearity of the derivative is not used in the proof.

Remark 2.9. The maps $\delta, \sigma, \alpha : \ell^\infty(\mathfrak{X}) \longrightarrow \mathbb{R}$ defined in (5) and (6) are Hadamard directionally differentiable at $f \in \ell^\infty(\mathfrak{X})$ if and only if they are Gâteaux directionally differentiable at f. This follows from the fact that for $\varphi \in \{\delta, \sigma, \alpha\}$, we have that

$$\left|\frac{\varphi\,(f + t_n\,h_n) - \varphi(f)}{t_n} - \varphi'_f(h)\right| \leq \left|\varphi\left({}^{f}/_{t_n} + h_n\right) - \varphi\left({}^{f}/_{t_n} + h\right)\right| + \left|\frac{\varphi\,(f + t_n\,h) - \varphi(f)}{t_n} - \varphi'_f(h)\right|$$

and $\left|\varphi\left({}^{f}/_{t_n} + h_n\right) - \varphi\left({}^{f}/_{t_n} + h\right)\right| \leq \|h_n - h\|_\infty$ for $\varphi \in \{\delta, \sigma\}$ while $\left|\alpha\left({}^{f}/_{t_n} + h_n\right) - \alpha\left({}^{f}/_{t_n} + h\right)\right| \leq 2\,\|h_n - h\|_\infty$.

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

Teorema 2.10. *For any $f \in \ell^\infty(\mathfrak{X})\backslash\{0\}$, the maps δ, σ and α in (5) and (6) are Hadamard directionally differentiable at f. For $g \in \ell^\infty(\mathfrak{X})$, their derivatives are respectively given by*

$$\delta'_f(g) = \lim_{\varepsilon \searrow 0} \left(\sup_{x \in A_\varepsilon(|f|)} (g(x)\ \mathrm{sgn}(f(x))) \right), \quad \sigma'_f(g) = \lim_{\varepsilon \searrow 0} \left(\sup_{x \in A_\varepsilon(f)} (g(x)) \right), \tag{10}$$

and

$$\alpha'_f(g) = \lim_{\varepsilon \searrow 0} \left(\sup_{x \in A_{\varepsilon(f)}} (g(x)) - \inf_{x \in B_{\varepsilon(f)}} (g(x)) \right), \tag{11}$$

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

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

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 easily. 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 $s_n \nearrow \infty$, we consider $\sigma_n(f) : \ell^\infty(\mathfrak{X}) \longrightarrow \mathbb{R}$ defined by

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

From remarks 2.8 and 2.9, it suffices to show that $\sigma_n(f, g) \to \sigma'_f(g)$, as $n \to \infty$, with $\sigma'_f(g)$ defined in (10). For $\varepsilon > 0$ and $x \notin A_\varepsilon(f)$, we have that

$$s_n\, f(x) + g(x) - s_n \sup_{y \in \mathfrak{X}} (f(y)) \leq \sup_{y \in \mathfrak{X}} (g(y)) - s_n\, \varepsilon.$$

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

$$\begin{aligned} \limsup_{n \to \infty} \sigma_n(f, g) &= \limsup_{n \to \infty} \left(\sup_{y \in A_\varepsilon(f)} (s_n\, f(y) + g(y)) - s_n \sup_{y \in \mathfrak{X}} (f(y)) \right) \\ &\leq \sup_{y \in A_\varepsilon(f)} (g(y)). \end{aligned} \tag{12}$$

Conversely, let us define

$$h(\epsilon) = \sup_{y \in A_\varepsilon(f)} (g(y)), \quad \varepsilon > 0. \tag{13}$$

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

$$g(x_m) \geq h\left({}^1/_m\right) - {}^1/_m \quad \text{and} \quad f(x_m) \geq \sup_{y \in \mathfrak{X}} (f(y)) - {}^1/_m. \tag{14}$$

From (14), for each s_n, we have that

$$\begin{aligned} h\left({}^1/_m\right) &\leq g\left(x_m\right) +^1 /_m \\ &= s_n\, f\left(x_m\right) + g\left(x_m\right) - s_n\, f\left(x_m\right) +^1 /_m \\ &\leq \sigma_n(f,g) +^{(s_n+1)} /_m. \end{aligned} \tag{15}$$

Now (15) implies that, for all $n \in \mathbb{N}$,

$$\lim_{\varepsilon \searrow 0} \left(\sup_{y \in A_\varepsilon(f)} (g(y)) \right) = \lim_{m \to \infty} h\left({}^1/_m\right) \leq \sigma_n(f,g). \tag{16}$$

The proof corresponding to σ follows from (12) and (16).

Now, we consider the map δ in (5). Assume that $f \in \ell^\infty(\mathfrak{X})$ with $\|f\|_\infty > 0$.

For $g \in \ell^\infty(\mathfrak{X})$, we have to show that $\delta_n(f,g) \to \delta'_f(g)$, as $n \to \infty$, where $\delta_n(f,g) = \|s_n\, f + g\|_\infty - s_n\,\|f\|_\infty$ and $s_n \nearrow \infty$. First, for $\varepsilon <^{\|f\|_\infty} /_2$ and $s_n >^{2\,\|g\|_\infty} /_{\|f\|_\infty}$, it is readily checked that $s_n\,|f| + \operatorname{sgn}(f)\, g \geq 0$ globally on $A_\varepsilon(|f|)$. We hence conclude that

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

The proof for α follows from the duality between the 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. □

If $\mathfrak{X}$ is a compact metric space, the derivatives in (10) and (11) can be characterized by means of convergent sequences in $\mathfrak{X}$ as the following corollary shows.

Corollary 2.11. *In the context of Theorem 2.10, let us further assume that $\mathfrak{X}$ is a compact metric space. The derivatives in (10)–(11) can be expressed as*

$$\begin{aligned} \delta'_f(g) &= \sup_{y \in A_0(|f|)} \left((g\ \operatorname{sgn}(f))^{\blacktriangle}_{|f|}(y) \right), \\ \sigma'_f(g) &= \sup_{y \in A_0(f)} \left(g^{\blacktriangle}_f(y) \right), \\ \alpha'_f(g) &= \sup_{y \in A_0(f)} \left(g^{\blacktriangle}_f(y) \right) - \sup_{y \in B_0(f)} \left(g^{\blacktriangledown}_f(y) \right), \end{aligned} \tag{17}$$

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

$$\begin{aligned} A_0(h) = \Big\{ x \in \mathfrak{X} :\ &\text{there exists } \{x_n\} \subset \mathfrak{X} \text{ with } x_n \to x \text{ and} \\ &h\left(x_n\right) \to \sup_{y \in \mathfrak{X}} (h(y)) \Big\}, \\ B_0(h) = \Big\{ x \in \mathfrak{X} :\ &\text{there exists } \{x_n\} \subset \mathfrak{X} \text{ with } x_n \to x \text{ and} \\ &h\left(x_n\right) \to \inf_{y \in \mathfrak{X}} (h(y)) \Big\}, \end{aligned} \tag{18}$$

and

$$\begin{aligned} h_l^{\blacktriangle}(x) &= \sup\left(\left\{\limsup_{n\to\infty} h\left(x_n\right): \ x_n \to x \ \text{ and } \ l\left(x_n\right) \to \sup_{y\in\mathfrak{X}}(l(y))\right\}\right), \quad x \in A_0(l), \\ h_l^{\blacktriangledown}(x) &= \inf\left(\left\{\liminf_{n\to\infty} h\left(x_n\right): \ x_n \to x \ \text{ and } \ l\left(x_n\right) \to \inf_{y\in\mathfrak{X}}(l(y))\right\}\right), \quad x \in B_0(l). \end{aligned} \tag{19}$$

Proof. We only give a detailed proof for σ because the rest of the cases are analogous. We consider the sequence $\{x_m\}$ satisfying (14) obtained in the proof of Theorem 2.10. As $\mathfrak{X}$ is compact, we can extract a convergent subsequence $x_{m_k} \to x$ in $\mathfrak{X}$, as $k \to \infty$. From (14), we have that $x \in A_0(f)$ and, recalling (13), from Theorem 2.10, we obtain that

$$\sigma'_f(g) = \lim_{k\to\infty} h\left({}^1/_{m_k}\right) \leq \limsup_{k\to\infty} g\left(x_{m_k}\right) \leq g_f^{\blacktriangle}(x) \leq \sup_{y\in A_0(f)} \left(g_f^{\blacktriangle}(y)\right). \tag{20}$$

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

$$\limsup_{n\to\infty}\left(g\left(x_n\right)\right) \leq \sup_{y\in A_\varepsilon(f)}(g(y)), \quad \text{for all } \varepsilon > 0. \tag{21}$$

The conclusion follows from (20), (21) and Theorem 2.10. □

In the following, if $\mathfrak{X}$ is a metric space we denote by $\mathcal{C}(\mathfrak{X})$ the subset of $\ell^\infty(\mathfrak{X})$ constituted by continuous functions. We observe that if $g \in \mathcal{C}(\mathfrak{X})$, 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 (19). This observation yields the following corollary.

Corollary 2.12. *Let $\mathfrak{X}$ be a compact metric space. For any $f \in \ell^\infty(\mathfrak{X})$, the maps δ, σ and α in (5) and (6) are Hadamard directionally differentiable at f tangentially to the set $\mathcal{C}(\mathfrak{X})$ with derivatives, for $g \in \mathcal{C}(\mathfrak{X})$,*

$$\begin{aligned} \delta'_f(g) &= \sup_{y\in A_0(|f|)}(g(y)\ \mathrm{sgn}(f(y))), \\ \sigma'_f(g) &= \sup_{y\in A_0(f)}(g(y)), \\ \alpha'_f(g) &= \sup_{y\in A_0(f)}(g(y)) - \inf_{y\in B_0(f)}(g(y)), \end{aligned}$$

where $A_0(\cdot)$ and $B_0(\cdot)$ are defined in (18).

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 (see [Fang and Santos, 2015, Proposition 2.1]). This kind of results can be traced back to [Banach, 1932] (see also [Leonard and Taylor, 1983], [Leonard and Taylor, 1985], and the references therein). In these works the supremum norm differentiability was investigated from the point of view of functional analysis within the space $\mathcal{C}(\mathfrak{X})$, with $\mathfrak{X}$ a compact metric space. The following result, a direct consequence of Corollary 2.12, provides similar outcomes by using a different approach. We denote by $\mathrm{card}(A)$ the cardinal of the set A.

Corollary 2.13. *Assume that $\mathfrak{X}$ 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 (18). We have that:*

(a) The map δ in (5) is (fully) Hadamard differentiable at f tangentially to the set $\mathcal{C}(\mathfrak{X})$ if and only if $\operatorname{card}(A_0(|f|)) = 1$. *In such a case, $\delta'_f(g) = g(x^*) \operatorname{sgn}(f(x^*))$, where $A_0(|f|) = \{x^*\}$.*

(b) The map σ in (6) is (fully) Hadamard differentiable at f tangentially to the set $\mathcal{C}(\mathfrak{X})$ if and only if $\operatorname{card}(A_0(f)) = 1$. *In such case, $\sigma'_f(g) = g(x^+)$, where $A_0(f) = \{x^+\}$.*

(c) The map α in (6) is (fully) Hadamard differentiable at f tangentially to the set $\mathcal{C}(\mathfrak{X})$ if and only if $\operatorname{card}(A_0(f)) = \operatorname{card}(B_0(f)) = 1$. *In such case, $\alpha'_f(g) = g(x^+) - g(x^-)$, where $A_0(f) = \{x^+\}$ and $B_0(f) = \{x^-\}$.*

When $f \in \mathcal{C}(\mathfrak{X})$, the condition $\operatorname{card}(A_0(|f|)) = 1$ 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 is important too. It is a sufficient condition for the asymptotic normality provided the limit in (3) is Gaussian.

3 The Skorohod space $\mathcal{D}\big(\overline{\mathbb{R}}^d\big)$

Many important stochastic processes take values in the 1-dimensional Skorohod space, $\mathcal{D}\big(\overline{\mathbb{R}}\big)$, 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, and many others. The d-dimensional Skorohod space, introduced in [Neuhaus, 1971] (see also [Bickel and Wichura, 1971]) and more recently considered in [Seijo and Sen, 2011]), is usually defined in compact rectangles of $\mathbb{R}^d$. Starting from the definition of the Skorohod space on $[0,1]^d$ in [Neuhaus, 1971], in this section we extend this space to functions defined in $\overline{\mathbb{R}}^d$. We also provide alternative expressions for the directional derivatives in (17) when the involved functions belong to the d-dimensional Skorohod space.

First, for $v \in \{-1, 1\}$ and $x \in \overline{\mathbb{R}}$, let

$$I_v(x) = \begin{cases} [-\infty, x[, & \text{if } \ v = -1, \ x \in \overline{\mathbb{R}}, \\ ]x, \infty], & \text{if } \ v = 1, \ x \in \overline{\mathbb{R}}, \end{cases}$$

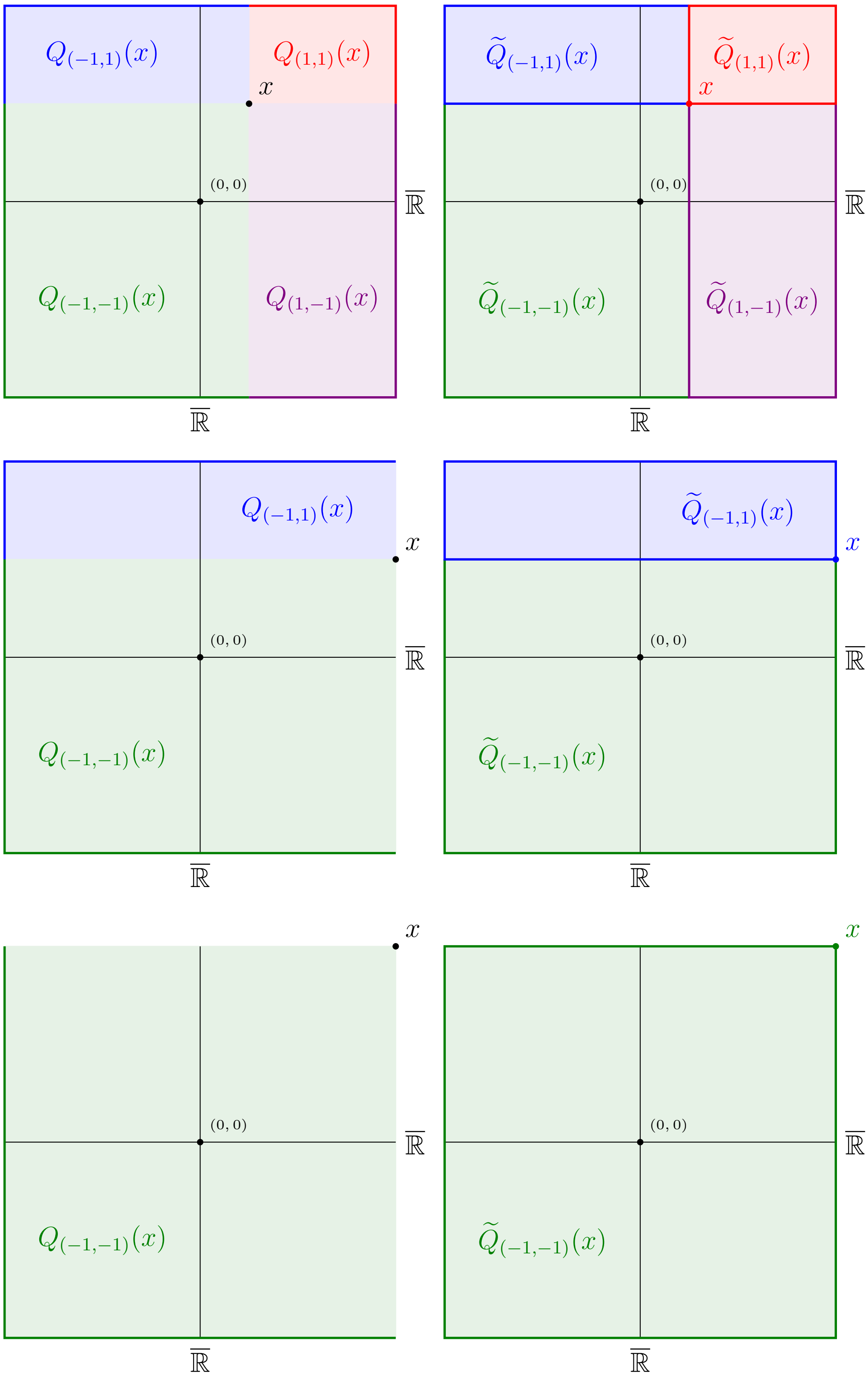
and

$$\widetilde{I}_v(x) = \begin{cases} [-\infty, x[, & \text{if } \ v = -1, \ x < \infty, \\ \overline{\mathbb{R}}, & \text{if } \ v = -1, \ x = \infty, \\ \varnothing & \text{if } v = 1, \ x = \infty, \\ [x, \infty], & \text{if } \ v = 1, \ x < \infty. \end{cases}$$

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

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

Observe that $Q_{\mathbf{v}}(\mathbf{x}) \subset \widetilde{Q}_{\mathbf{v}}(\mathbf{x})$, $\widetilde{Q}_{\mathbf{v}}(\mathbf{x}) \cap \widetilde{Q}_{\mathbf{v}'}(\mathbf{x}) = \varnothing$ whenever $\mathbf{v}$, $\mathbf{v}' \in \mathcal{V}$ with $\mathbf{v} \neq \mathbf{v}'$, and $\bigcup_{\mathbf{v} \in \mathcal{V}} \widetilde{Q}_{\mathbf{v}}(\mathbf{x}) = \overline{\mathbb{R}}^d$, for all $\mathbf{x} \in \overline{\mathbb{R}}^d$. Additionally, for each $\mathbf{x} \in \overline{\mathbb{R}}^d$, there exists a unique $\mathbf{v}_{\mathbf{x}} \in \mathcal{V}$ such that $\mathbf{x} \in \widetilde{Q}_{\mathbf{v}_{\mathbf{x}}}(\mathbf{x})$. For instance, if $\mathbf{x} \in \mathbb{R}^d$, we have that $\mathbf{v}_{\mathbf{x}} = \mathbf{1}$, where $\mathbf{1} \equiv (1, \overset{d)}{\ldots}, 1)$. When $d = 2$ the situation can be summarized in the following figures:



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

$\mathbf{x} \in \overline{\mathbb{R}}^d$ if $f_{\mathbf{v_x}}(\mathbf{x})$ 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 $\mathbf{x} \in \overline{\mathbb{R}}^d$.

Definition 3.1. The *Skorohod space* on $\overline{\mathbb{R}}^d$, denoted by $\mathcal{D}(\overline{\mathbb{R}}^d)$, is the collection of all continuous from above real functions f defined in $\overline{\mathbb{R}}^d$ for which the $\mathbf{v}$-limit of f exists for every $\mathbf{v} \in \mathcal{V}$ and $\mathbf{x} \in \overline{\mathbb{R}}^d$ such that $Q_{\mathbf{v}}(\mathbf{x}) \neq \varnothing$.

When $d = 1$, $\mathcal{D}(\overline{\mathbb{R}})$ is usual Skorohod space on $\overline{\mathbb{R}}$. The properties of the multidimensional Skorohod space in $[0,1]^d$ shown in [Neuhaus, 1971] can be extended with no difficulty to $\mathcal{D}(\overline{\mathbb{R}}^d)$. For instance, the elements in $\mathcal{D}(\overline{\mathbb{R}}^d)$ belong to $\mathcal{D}(\overline{\mathbb{R}})$ in each coordinate, have at most countably many discontinuities and all of them are finite jump discontinuities. The fact that $\mathcal{D}(\overline{\mathbb{R}}^d) \subset \ell^\infty(\overline{\mathbb{R}}^d)$ follows from [Neuhaus, 1971, Corollary 1.6] by noting that functions in $\mathcal{D}(\overline{\mathbb{R}}^d)$ have finite quadrant limits at infinity points.

Remark 3.2. We observe that if $f \in \mathcal{D}(\overline{\mathbb{R}}^d)$ and $\{\mathbf{x}_n\} \subset \widetilde{Q}_{\mathbf{v}}(\mathbf{x})$ such that $\mathbf{x}_n \to \mathbf{x}$, then $f(\mathbf{x}_n) \to f_{\mathbf{v}}(\mathbf{x})$. This follows from the fact that

$$\widetilde{Q}_{\mathbf{v}}(\mathbf{x}) = \left\{\mathbf{y} \in \overline{\mathbb{R}}^d : \ \mathbf{y} \in \overline{Q_{\mathbf{v_y}}(\mathbf{y}) \cap Q_{\mathbf{v}}(\mathbf{x})}\right\},$$

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

We are now in position to see how the derivatives in (17) look like when $\mathfrak{X} = \overline{\mathbb{R}}^d$ and the functions on which they act belong to $\mathcal{D}(\overline{\mathbb{R}}^d)$.

Corollary 3.3. *For any $f \in \mathcal{D}(\overline{\mathbb{R}}^d) \setminus \{0\}$, the maps δ, σ and α in (5) and (6) are Hadamard directionally differentiable at f tangentially to $\mathcal{D}(\overline{\mathbb{R}}^d)$. For $g \in \mathcal{D}(\overline{\mathbb{R}}^d)$, their derivatives are given by*

$$\delta'_f(g) = \max_{\mathbf{v}\in\mathcal{V}} \left(\sup_{\mathbf{y}\in M^+_{\mathbf{v}}(|f|)} ((g \ \mathrm{sgn}(f))_{\mathbf{v}}(\mathbf{y})) \right), \qquad \sigma'_f(g) = \max_{\mathbf{v}\in\mathcal{V}} \left(\sup_{\mathbf{y}\in M^+_{\mathbf{v}}(f)} (g_{\mathbf{v}}(\mathbf{y})) \right),$$

and

$$\alpha'_f(g) = \max_{\mathbf{v}\in\mathcal{V}} \left(\sup_{\mathbf{y}\in M^+_{\mathbf{v}}(f)} (g_{\mathbf{v}}(\mathbf{y})) \right) - \min_{\mathbf{v}\in\mathcal{V}} \left(\inf_{\mathbf{y}\in M^-_{\mathbf{v}}(f)} (g_{\mathbf{v}}(\mathbf{y})) \right),$$

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

$$\begin{aligned} M^+_{\mathbf{v}}(h) &= \left\{\mathbf{x} \in \overline{\mathbb{R}}^d : \ Q_{\mathbf{v}}(\mathbf{x}) \neq \varnothing \ \text{ and } \ h_{\mathbf{v}}(\mathbf{x}) = \sup_{\mathbf{y}\in\overline{\mathbb{R}}^d} (h(\mathbf{y}))\right\}, \\ M^-_{\mathbf{v}}(h) &= \left\{\mathbf{x} \in \overline{\mathbb{R}}^d : \ Q_{\mathbf{v}}(\mathbf{x}) \neq \varnothing \ \text{ and } \ h_{\mathbf{v}}(\mathbf{x}) = \inf_{\mathbf{y}\in\overline{\mathbb{R}}^d} (h(\mathbf{y}))\right\}. \end{aligned} \tag{22}$$

Proof. This corollary can be proved as Corollary 2.11 by taking into account Remark 3.2 and the following fact: As the number of non-empty quadrants of each point in $\overline{\mathbb{R}}^d$ is finite, each sequence converging to a point $\mathbf{x} \in \overline{\mathbb{R}}^d$ has a subsequence contained in $\widetilde{Q}_{\mathbf{v}}(\mathbf{x})$, for some $\mathbf{v} \in \mathcal{V}$. In particular, for every $h \in \mathcal{D}(\overline{\mathbb{R}}^d)$, 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 (18). $\square$

The sets $M_{\mathbf{v}}^{+}(h)$ (respectively, $M_{\mathbf{v}}^{-}(h)$) in (22) 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 for $h \in \ell^{\infty}\big(\overline{\mathbb{R}}^d\big)$,

$$
\begin{aligned}
M^{+}(h) &= \left\{ \mathbf{x} \in \overline{\mathbb{R}}^d : \ h(\mathbf{x}) = \sup_{\mathbf{y} \in \overline{\mathbb{R}}^d} (h(\mathbf{y})) \right\}, \\
M^{-}(h) &= \left\{ \mathbf{x} \in \overline{\mathbb{R}}^d : \ h(\mathbf{x}) = \inf_{\mathbf{y} \in \overline{\mathbb{R}}^d} (h(\mathbf{y})) \right\}.
\end{aligned}
\tag{23}
$$

Observe that $M^{+}(|f|)$ (respectively, $M^{+}(f)$ and $M^{-}(f)$) is the set of extremal points corresponding to the sup-norm (respectively, the supremum and infimum) of f.

We emphasize that $g_{\mathbf{v}} \equiv g$, for all $\mathbf{v} \in \mathcal{V}$, whenever $g \in \mathcal{C}\big(\overline{\mathbb{R}}^d\big)$. 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 3.4. *For any $f \in \mathcal{D}\big(\overline{\mathbb{R}}^d\big) \setminus \{0\}$, the maps δ, σ and α in (5) and (6) are Hadamard directionally differentiable at f tangentially to $\mathcal{C}\big(\overline{\mathbb{R}}^d\big)$. For $g \in \mathcal{C}\big(\overline{\mathbb{R}}^d\big)$, their derivatives are given by*

$$
\delta'_f(g) = \max_{\mathbf{v} \in \mathcal{V}} \left(\sup_{\mathbf{y} \in M_{\mathbf{v}}^{+}(|f|)} (g(\mathbf{y}) \ \mathrm{sgn}(f)_{\mathbf{v}}(\mathbf{y})) \right), \quad \sigma'_f(g) = \max_{\mathbf{v} \in \mathcal{V}} \left(\sup_{\mathbf{y} \in M_{\mathbf{v}}^{+}(f)} (g(\mathbf{y})) \right), \tag{24}
$$

and

$$
\alpha'_f(g) = \max_{\mathbf{v} \in \mathcal{V}} \left(\sup_{\mathbf{y} \in M_{\mathbf{v}}^{+}(f)} (g(\mathbf{y})) \right) - \min_{\mathbf{v} \in \mathcal{V}} \left(\inf_{\mathbf{y} \in M_{\mathbf{v}}^{-}(f)} (g(\mathbf{y})) \right), \tag{25}
$$

with $M_{\mathbf{v}}^{+}(\cdot)$ and $M_{\mathbf{v}}^{-}(\cdot)$ defined in (22). If additionally $f \in \mathcal{C}\big(\overline{\mathbb{R}}^d\big)$, we have that

$$
\begin{aligned}
\delta'_f(g) &= \sup_{\mathbf{y} \in M^{+}(|f|)} (g(\mathbf{y}) \ \mathrm{sgn}(f(\mathbf{y}))), \\
\sigma'_f(g) &= \sup_{\mathbf{y} \in M^{+}(f)} (g(\mathbf{y})), \\
\alpha'_f(g) &= \sup_{\mathbf{y} \in M^{+}(f)} (g(\mathbf{y})) - \inf_{\mathbf{y} \in M^{-}(f)} (g(\mathbf{y})),
\end{aligned}
\tag{26}
$$

where $M^{+}(\cdot)$ and $M^{-}(\cdot)$ are defined in (23).

4 Applications

In a wide variety of situations Theorem 2.10 and its subsequent corollaries, joint with the extended Delta method in Proposition 2.7, 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.

Teorema 4.1. *Let $q \in \ell^\infty(\mathfrak{X})$ 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 $r_n \to \infty$ and a random element $\mathbb{Q}$ in $\ell^\infty(\mathfrak{X})$. Then, for $\varphi \in \{\delta, \sigma, \alpha\}$ (in (5) and (6)), we have that*

$$r_n(\varphi(\mathbb{Q}_n) - \varphi(q)) \rightsquigarrow \varphi_q'(\mathbb{Q}), \tag{27}$$

where the derivatives φ_q' are given in (10)-(11). Moreover, we have that $r_n(\varphi(\mathbb{Q}_n) - \varphi(q)) = \varphi_q'(r_n(\mathbb{Q}_n - q)) + o_{\mathbb{P}}(1)$.

This theorem is still valid for the maps σ and α when $q = 0$ as $\sigma_0'(g) = \sup_{y\in\mathfrak{X}}(g(y))$ and $\alpha_0'(g) = \operatorname{amp}_{y\in\mathfrak{X}}(g(y))$ 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 2.13), and when $\mathbb{Q}$ is Gaussian, we have that $\varphi_q'(\mathbb{Q})$ is normally distributed.

In what follows we will apply the previous general result in two different contexts to obtain the asymptotic distribution of several statistics.

4.1 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 $F(\mathbf{x}) = \mathbb{P}(\mathbf{X} \leq \mathbf{x})$ and $G(\mathbf{x}) = \mathbf{P}(\mathbf{Y} \leq \mathbf{x})$, $\mathbf{x} \in \mathbb{R}^d$, where '$\leq$' stands for the coordinatewise order in $\mathbb{R}^d$, that is, given $\mathbf{x} = (x_1, \ldots, x_d), \mathbf{y} = (y_1, \ldots, y_d) \in \mathbb{R}^d$ we say $\mathbf{x} \leq \mathbf{y}$ if and only if for every $i \in \{1, \ldots, d\}$ $x_i \leq y_i$. The goal in this section is to estimate $\phi(F - G)$, where $\phi \in \{\delta, \sigma, \alpha\}$ is any of the functionals introduced in Section 1.

One-sample case: In this situation we have at our disposal a random sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$ 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,

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

and $\mathbf{1}_A$ stands for the indicator function of the set A.

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

$$\begin{aligned} D_n(\delta) &= \sqrt{n}\, \left(\|F_n - G\|_\infty - \|F - G\|_\infty \right), \\ D_n(\sigma) &= \sqrt{n} \left(\sup_{\mathbf{y}\in\mathbb{R}^d} \left(F_n(\mathbf{y}) - G(\mathbf{y})\right) - \sup_{\mathbf{y}\in\mathbb{R}^d} \left(F(\mathbf{y}) - G(\mathbf{y})\right) \right) \\ D_n(\alpha) &= \sqrt{n} \left(\operatorname*{amp}_{\mathbf{y}\in\mathbb{R}^d} \left(F_n(\mathbf{y}) - G(\mathbf{y})\right) - \operatorname*{amp}_{\mathbf{y}\in\mathbb{R}^d} (F(\mathbf{y}) - G(\mathbf{y})) \right). \end{aligned} \tag{28}$$

When $F \neq G$, the asymptotic distribution of the statistics in (28) 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 $\varphi \in \{\delta, \sigma, \alpha\}$, the statistics in (28) are $D_n(\varphi) = D_\varphi\left(q, \mathbb{Q}_n, r_n\right)$ in (4) with $q = F - G$, $\mathbb{Q}_n = F_n - G$, and $r_n = \sqrt{n}$. The underlying normalized process, i.e., $r_n\left(\mathbb{Q}_n - q\right)$, is nothing but the multivariate *empirical process* (indexed by points),

$$\mathbb{E}_{n,F}(\mathbf{x}) = \sqrt{n}\, \left(F_n(\mathbf{x}) - F(\mathbf{x})\right), \quad n \in \mathbb{N}, \quad \mathbf{x} \in \mathbb{R}^d. \tag{29}$$

When there is no confusion with respect to the underlying distribution, we simply use the notation $\mathbb{E}_n$ for the empirical process in (29). As the collection of all indicator functions of lower (hyper)rectangles of $\overline{\mathbb{R}}^d$, $\left\{\mathbf{1}_{]-\infty,x_1]\times\ldots\times]-\infty,x_d]} : (x_1, \ldots, x_d) \in \overline{\mathbb{R}}^d\right\}$, is Donsker (see [van der Vaart and Wellner, 1996, Example 2.1.3, p.82]), the empirical process converges in law in $\ell^\infty\left(\overline{\mathbb{R}}^d\right)$. 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 $\mathbb{E}\left(\mathbb{B}_F(\mathbf{x})\, \mathbb{B}_F(\mathbf{y})\right) = F(\mathbf{x} \wedge \mathbf{y}) - F(\mathbf{x})\, F(\mathbf{y})$. (Here $\mathbf{x} \wedge \mathbf{y} \equiv (x_1 \wedge y_1, \ldots, x_d \wedge y_d)$ if $\mathbf{x} = (x_1, \ldots, x_d)$ and $\mathbf{y} = (y_1, \ldots, y_d)$.) If $d = 1$, the assertion "$\mathbb{E}_n \rightsquigarrow \mathbb{B}_F$ in $\ell^\infty\left(\overline{\mathbb{R}}\right)$" 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 $F - G \in \mathcal{D}\left(\overline{\mathbb{R}}^d\right)$, $\mathbb{E}_n \in \mathcal{D}\left(\overline{\mathbb{R}}^d\right)$ a.s., $\mathbb{E}_n \rightsquigarrow \mathbb{B}_F$ in $\ell^\infty\left(\overline{\mathbb{R}}^d\right)$, and $\mathbb{B}_F \in \mathcal{C}\left(\overline{\mathbb{R}}^d\right)$ a.s. Therefore, as a direct consequence of Theorem 4.1 and Corollary 3.4 we obtain the following result.

Proposition 4.2. *Assume that $F \neq G$ and let $\mathbb{B}_F$ be an F-Brownian bridge. For $\varphi \in \{\delta, \sigma, \alpha\}$, we consider the statistics $D_n(\varphi)$ defined in (28). We have that $D_n(\varphi) \rightsquigarrow \varphi'_{F-G}\left(\mathbb{B}_F\right)$, where the derivatives φ'_{F-G} are given as in (24)–(25).*

When $d = 1$, Proposition 4.2 improves [Raghavachari, 1973, Theorems 1, 2 and 3] as here F and G are not assumed to be continuous. We also remark has been obtained with no additional effort. If $F - G$ is continuous, the limiting distributions in Proposition 4.2 have simpler expressions. The following corollary provides a multidimensional extension of the results in [Raghavachari, 1973].

Corollary 4.3. *In the conditions of Proposition 4.2, let us further assume that* $F - G \in \mathcal{C}(\overline{\mathbb{R}}^d)$. *We have that:*

$$D_n(\delta) \rightsquigarrow \sup_{\mathbf{y}\in M^+(|F-G|)} (\mathbb{B}_F(\mathbf{y}) \operatorname{sgn}(F(\mathbf{y}) - G(\mathbf{y}))), \quad D_n(\sigma) \rightsquigarrow \sup_{\mathbf{y}\in M^+(F-G)} (\mathbb{B}_F(\mathbf{y})),$$
$$D_n(\alpha) \rightsquigarrow \sup_{\mathbf{y}\in M^+(F-G)} (\mathbb{B}_F(\mathbf{y})) - \inf_{\mathbf{y}\in M^-(F-G)} (\mathbb{B}_F(\mathbf{y})),$$

where the sets $M^+(\cdot)$ *and* $M^-(\cdot)$ *are defined in (23).*

Remark 4.4. In the setting of the previous corollary, when $M^+(|F-G|)$ (respectively, $M^+(F-G)$, and $M^-(F-G)$) contains only one point, the mapping δ (respectively, σ and α) is fully Hadamard differentiable at $F - G$ (see Corollary 2.13). 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 $\mathbb{B}_F$.

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 $N = \frac{nm}{n+m}$. The two-sided and one-sided Kolmogorov-Smirnov and Kuiper statistics in the two sample case are given by

$$\begin{aligned}
D_{n,m}(\delta) &= \sqrt{N}\left(\|F_n - G_m\|_\infty - \|F - G\|_\infty\right), \\
D_{n,m}(\sigma) &= \sqrt{N}\left(\sup_{\mathbf{y}\in\overline{\mathbb{R}}^d} (F_n(\mathbf{y}) - G_m(\mathbf{y})) - \sup_{\mathbf{y}\in\overline{\mathbb{R}}^d} (F(\mathbf{y}) - G(\mathbf{y}))\right) \\
D_{n,m}(\alpha) &= \sqrt{N}\left(\operatorname*{amp}_{\mathbf{y}\in\overline{\mathbb{R}}^d} (F_n(\mathbf{y}) - G_m(\mathbf{y})) - \operatorname*{amp}_{\mathbf{y}\in\overline{\mathbb{R}}^d} (F(\mathbf{y}) - G(\mathbf{y}))\right).
\end{aligned} \tag{30}$$

In the general setting specified in (4), 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

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

with $\mathbb{E}_{n,F}$ and $\widetilde{\mathbb{E}}_{m,G}$ independent empirical processes. We further observe that if the sampling scheme is balanced, that is, $^n/_{(n+m)} \to \lambda$, with $0 < \lambda < 1$ as $n, m \to \infty$, then $r_{n,m}\,(\mathbb{Q}_{n,m} - q) \rightsquigarrow \sqrt{1-\lambda}\,\mathbb{B}_F - \sqrt{\lambda}\,\widetilde{\mathbb{B}}_G$ in $\ell^\infty(\overline{\mathbb{R}}^d)$, where $\mathbb{B}_F$ and $\widetilde{\mathbb{B}}_G$ are two independent Brownian bridges associated with F and G, respectively. Hence, Theorem 4.1 and Corollary 3.4 directly imply the following result which improves and generalizes [Raghavachari, 1973, Theorems 4 and 5].

Proposition 4.5. *Let us consider a sampling scheme such that as* n, $m \to \infty$, $^n/_{(n+m)} \to \lambda$, *with* $0 < \lambda < 1$ *and let* $\mathbb{B}_F$ *and* $\widetilde{\mathbb{B}}_G$ *be two independent Brownian bridges associated with* F *and* G, *respectively. For* $\varphi \in \{\delta, \sigma, \alpha\}$, *we consider the statistics* $D_{n,m}(\varphi)$ *defined in (30). We have that* $D_{n,m}(\varphi) \rightsquigarrow \varphi'_{F-G}\left(\sqrt{1-\lambda}\,\mathbb{B}_F - \sqrt{\lambda}\,\widetilde{\mathbb{B}}_G\right)$, *where the derivatives* φ'_{F-G} *are given in (24)–(25). If we further have that* $F - G \in \mathcal{C}(\overline{\mathbb{R}}^d)$, *then the derivatives can be expressed as in (26).*

4.2 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 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, 1979] (see also [DasGupta, 2008, Chapter 26.7]) introduced the test statistic

$$R\left(F_n, G\right) = \sup_{x\in\mathbb{R}} \left(K\left(F_n(x), G(x)\right)\right), \tag{31}$$

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 $x = 0$ are taken by continuity.)

For each $x \in \mathbb{R}$, $n\, K\left(F_n(x), G(x)\right)$ is the log-likelihood ratio statistic for testing $H_0 : F(x) = G(x)$ against $H_1 : F(x) \neq G(x)$. Hence, $R\left(F_n, G\right)$ in (31) 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, $R\left(F_n, G\right) = \left\|K\left(F_n, G\right)\right\|_\infty$.

[Berk and Jones, 1979] computed the asymptotic distribution of (the normalized version of) $R\left(F_n, F\right)$, i.e., the distribution of the statistic under the null hypothesis $F = G$. For a detailed proof, see [Wellner and Koltchinskii, 2003, Theorem 1.1] or [Jager and Wellner, 2007, Theorem 3.1]. It holds that

$$n\, R\left(F_n, F\right) - d_n \rightsquigarrow Y_4, \quad \text{as } n \to \infty, \tag{32}$$

where $\mathbb{P}\left(Y_4 \leq x\right) = \exp\left(-4 \exp(-x)\right)$ 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 $\log_3(n) = \log\left(\log_2(n)\right)$.

In [Jager and Wellner, 2004, Question 2, p.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

$$B_n = \sqrt{n}\, \left(R\left(F_n, G\right) - R(F, G)\right), \tag{33}$$

converges in distribution and, in such case, identifying its weak limit, where $R\left(F_n, G\right)$ is given in (31) 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 (33) can be rewritten in the general form of (4). In other words,

$$B_n = D_\sigma\left(q = K(F, G), \mathbb{Q}_n = K\left(F_n, G\right), r_n = \sqrt{n}\right), \tag{34}$$

where σ is defined in (6). As K is non-negative, it also holds that
$B_n = D_\delta\left(K(F,G), K\left(F_n, G\right), \sqrt{n}\right)$ with δ in (5). Therefore, from (34) and Theorem 4.1, to obtain the asymptotic distribution of B_n in (33) it is enough to find the weak limit of the process $\mathbb{W}_n$ given by

$$\mathbb{W}_n = \sqrt{n}\,\left(K\left(F_n, G\right) - K(F,G)\right). \tag{35}$$

This result is stated in the following theorem.

Teorema 4.6. *Let us assume that*

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

The process $\mathbb{W}_n$ defined in (35) satisfies that $\mathbb{W}_n \rightsquigarrow \mathbb{W}$ in $\ell^\infty\left(\overline{\mathbb{R}}\right)$, where

$$\mathbb{W} = \mathbb{B}_F\, \log\frac{F\,(1-G)}{G\,(1-F)}, \tag{36}$$

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

Proof. Using Taylor's theorem (von Mises calculus), we have that

$$K\left(F_n, G\right) - K(F,G) = \left(F_n - F\right)\,\log\frac{F\,(1-G)}{G\,(1-F)} + \frac{1}{2}\,\frac{\left(F_n - F\right)^2}{F_n^*\,\left(1 - F_n^*\right)}, \tag{37}$$

where F_n^* is between F and F_n (in a ball centered at F and radius $\left\|F_n - F\right\|_\infty$). We set

$$\widetilde{\mathbb{W}}_n = \sqrt{n}\left(F_n - F\right)\,\log\frac{F\,(1-G)}{G\,(1-F)}. \tag{38}$$

From (35) and (37), we have that

$$\left\|\mathbb{W}_n - \widetilde{\mathbb{W}}_n\right\|_\infty = \frac{\sqrt{n}}{2}\,\left\|\frac{\left(F_n - F\right)^2}{F_n^*\,\left(1 - F_n^*\right)}\right\|_\infty. \tag{39}$$

Now, from (39) and [Wellner and Koltchinskii, 2003, equation (2.2)] (see also [Jager and Wellner, 2007, equation (9)], we obtain that

$$\begin{aligned}\left\|\mathbb{W}_n - \widetilde{\mathbb{W}}_n\right\|_\infty &=_{\text{st}} \sqrt{n}\,R\left(F_n, F\right)\\ &= \frac{1}{\sqrt{n}}\,\left(n\,R\left(F_n, F\right) - d_n\right) + \frac{d_n}{\sqrt{n}},\end{aligned} \tag{40}$$

where '$=_{\text{st}}$' stands for equality in distribution. From (32) and (40), we conclude that $\left\|\mathbb{W}_n - \widetilde{\mathbb{W}}_n\right\|_\infty \rightsquigarrow 0$. Hence, the processes $\mathbb{W}_n$ and $\widetilde{\mathbb{W}}_n$ have the same asymptotic behaviour (see [van der Vaart, 1998, Theorem 18.10]). Finally, the conclusion follows from [van der Vaart, 1998, Example 19.12, p.273]. □

Remark 4.7. As it follows from the proof of Theorem 4.6, the process $\mathbb{W}_n$ behaves asymptotically as $\widetilde{\mathbb{W}}_n$ in (38), which is actually a *weighted empirical process.* Therefore, necessary and sufficient conditions for the convergence of the process $\mathbb{W}_n$ defined in (35) are given by the Chibisov-O'Reilly theorem.

We are now in position to solve the question proposed in [Jager and Wellner, 2004].

Corollary 4.8. *In the conditions of Theorem 4.6, the statistic B_n in (33) satisfies that*

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

where $\mathbb{W}$ is given in (36) and the set $M^+(\cdot)$ is defined in (23).

Remark 4.9. Similar results can be stated for the family of test statistics $S_n(s)$ based on ϕ-divergences introduced by [Jager and Wellner, 2007]. Details are omitted.

References

[Banach, 1932] Banach, S. (1932). *Théorie des opérations linéaires.* Instytut Matematyczny Polskiej Akademi Nauk.

[Berk and Jones, 1979] Berk, R. and Jones, D. (1979). Goodness-of-fit test statistics that dominate the kolmogorov statistics. *Probability Theory and Related Fields*, 47(1):47–59.

[Bickel and Wichura, 1971] Bickel, P. J. and Wichura, M. J. (1971). Convergence criteria for multiparameter stochastic processes and some applications. *The Annals of Mathematical Statistics*, 42(5):1656–1670.

[Billingsley, 1999] Billingsley, P. (1999). *Convergence of probability measures.* Wiley, New York.

[DasGupta, 2008] DasGupta, A. (2008). *Asymptotic theory of statistics and probability.* Springer, New York.

[Fang and Santos, 2015] Fang, Z. and Santos, A. (2015). Inference on directionally differentiable functions. *arXiv: 1404.3763v2.*

[Giné and Nickl, 2016] Giné, E. and Nickl, R. (2016). *Mathematical foundations of infinite-dimensional statistical models.* Cambridge University Press, New York.

[Jager and Wellner, 2004] Jager, L. and Wellner, J. A. (2004). *On the "Poisson boundaries" of the family of weighted Kolmogorov statistics*, volume 45 of *Lecture Notes–Monograph Series*, pages 319–331. Institute of Mathematical Statistics, Beachwood, Ohio, USA.

[Jager and Wellner, 2007] Jager, L. and Wellner, J. A. (2007). Goodness-of-fit tests via phi-divergences. *The Annals of Statistics*, 35(5):2018–2053.

[Leonard and Taylor, 1983] Leonard, I. E. and Taylor, K. F. (1983). Supremum norm differentiability. *International Journal of Mathematics and Mathematical Sciences*, 6(4):705–713.

[Leonard and Taylor, 1985] Leonard, I. E. and Taylor, K. F. (1985). Essential supremum norm differentiability. *International Journal of Mathematics and Mathematical Sciences*, 8(3):433–439.

[Neuhaus, 1971] Neuhaus, G. (1971). On weak convergence of stochastic processes with multidimensional time parameter. *The Annals of Mathematical Statistics*, 42(4):1285–1295.

[Pollard, 1984] Pollard, D. (1984). *Convergence of Stochastic Processes.* Springer New York, New York, NY.

[Raghavachari, 1973] Raghavachari, M. (1973). Limiting distributions of kolmogorov-smirnov type statistics under the alternative. *The Annals of Statistics*, 1(1):67–73.

[Römisch, 2006] Römisch, W. (2006). *Delta method, infinite dimensional.*, pages –. Wiley.

[Seijo and Sen, 2011] Seijo, E. and Sen, B. (2011). A continuous mapping theorem for the smallest argmax functional. *Electronic Journal of Statistics*, 5(1):421–439.

[Shapiro, 1990] Shapiro, A. (1990). On concepts of directional differentiability. *Journal of Optimization Theory and Applications*, 66(3):477–487.

[Shapiro, 1991] Shapiro, A. (1991). Asymptotic analysis of stochastic programs. *Annals of Operations Research*, 30(1):169–186.

[van der Vaart, 1998] van der Vaart, A. W. (1998). *Weak Convergence and Empirical Processes.* Cambridge University Press, Cambridge.

[van der Vaart and Wellner, 1996] van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer-Verlag New York, New York.

[Wellner and Koltchinskii, 2003] Wellner, J. and Koltchinskii, V. (2003). A note on the asymptotic distribution of berk—jones type statistics under the null hypothesis. *High Dimensional Probability III*, 55(1):321–332.