Statistical Science
2016, Vol. 31, No. 1, 61–79
DOI: 10.1214/15-STS532
© Institute of Mathematical Statistics, 2016

A Topologically Valid Definition of Depth for Functional Data

Alicia Nieto-Reyes and Heather Battey

Abstract. The main focus of this work is on providing a formal definition of statistical depth for functional data on the basis of six properties, recognising topological features such as continuity, smoothness and contiguity. Amongst our depth defining properties is one that addresses the delicate challenge of inherent partial observability of functional data, with fulfillment giving rise to a minimal guarantee on the performance of the empirical depth beyond the idealised and practically infeasible case of full observability. As an incidental product, functional depths satisfying our definition achieve a robustness that is commonly ascribed to depth, despite the absence of a formal guarantee in the multivariate definition of depth. We demonstrate the fulfillment or otherwise of our properties for six widely used functional depth proposals, thereby providing a systematic basis for selection of a depth function.

Key words and phrases: Functional data, multivariate statistics, partial observability, robustness, statistical depth.

1. INTRODUCTION

This work intersects the areas of functional data analysis (FDA) and statistical depth. FDA provides an alternative way of studying traditional data objects, recognising that it is sometimes more natural and more fruitful to view a collection of measurements as partially observed realisations of random functions. Prototypical examples of functional data objects include growth trajectories, handwriting data and brain imaging data. On the other hand, statistical depth (henceforth referred to as depth) is a powerful data analytic and inferential tool, able to reveal diverse features of the underlying distribution such as spread, shape and symmetry (Liu, Parelius and Singh, 1999). The ability of depth to reveal distributional features has been exploited in novel ways to define, inter alia, depth-based classifiers (e.g., Li, Cuesta-Albertos and Liu, 2012; Paindaveine and Van Bever, 2015).

The main focus of this work is on providing a formal definition of depth for functional data, justified on the basis of several properties. The definition fills an important void in the existing literature because naïve extensions of multivariate depth constructions, designed to satisfy the properties deemed suitable in multivariate space, neglect the topological features of functional data and often give rise to absurd depth computations (Dutta, Ghosh and Chaudhuri, 2011; Chakraborty and Chaudhuri, 2014b). The need for such a definition was first pointed out in the conference proceedings Nieto-Reyes (2011), where a crude first attempt to address the problem was made. Undesirable behaviour is also evident for specific constructions of functional depth examples that have been proposed without suitable reflection on the properties sought (López-Pintado and Jornsten, 2007).

The properties that constitute our definition not only provide a sophisticated extension of those defining the multivariate depth, recognising topological features such as continuity, contiguity and smoothness, but also implicitly address several common or inherent difficulties associated with functional data. Amongst our six depth defining properties is one that tackles the

Alicia Nieto-Reyes is Associate Professor, Departamento de Matemáticas, Estadística y Computación, Universidad de Cantabria, Spain (e-mail: alicia.nieto@unican.es). Heather Battey is Research Associate, ORFE, Princeton University, Sherrerd Hall, 98 Charlton Street, Princeton, NJ 08540, USA (e-mail: hbattey@princeton.edu); and Imperial College London, Department of Mathematics, South Kensington Campus, London, England SW7 2AZ; (e-mail: h.battey@imperial.ac.uk).

delicate challenge of inherent partial observability of functional data, providing a minimal guarantee on the performance of the empirical depth beyond the idealised and practically infeasible case of full observability. Robustness to the presence of outliers is often cited as one of the defining features of empirical depth (López-Pintado and Jornsten, 2007). Our definition of functional depth automatically yields a robust estimator of the population depth in the sense of qualitative robustness (Hampel, 1971). As we elucidate in Section 3.1, none of the properties constituting the multivariate definition of depth (Zuo and Serfling, 2000b; Liu, 1990) give rise to this property, thus a further contribution of our work is the insight that the existing definition for the multivariate framework is insufficient to guarantee robustness of the multivariate empirical depth. A further challenge, automatically addressed (if present) by our definition, pertains to functional data exhibiting little variability over a subset of the domain and significantly overlapping one another on this set. Intuitively, functional observations over such a domain ought to play a reduced role in the assignment of depth (Claeskens et al., 2014), especially in light of the partial observability and the convention to preprocess the partial observations.

We demonstrate the fulfilment or otherwise of our depth defining properties for six widely used functional depth functions, from which we conclude that the *h-depth* (Cuevas, Febrero and Fraiman, 2007) is the most well-reasoned in terms of number of properties satisfied.

The remainder of this paper is organised as follows. Section 2 provides an explanation of the notion of depth at the heuristic level, tracking its chronological development, before providing the formal definition of depth in $\mathbb{R}^p$, $p \geq 1$, as set forth in Zuo and Serfling (2000b), Liu (1990). Section 2 also formalises the functional data setting and defines the notation used in the paper. A formal definition of depth in function space appears in Section 3, together with a justification of the properties upon which it is based and a thorough discussion of their implications. Section 4 analyses existing constructions of functional depth, establishing the fulfillment or otherwise of each property appearing in the definition of functional depth. All the proofs appear in Section 5.

2. BACKGROUND AND NOTATION

2.1 Historical Development and a Heuristic Explanation of Depth

Unlike the univariate case in which there is no ambiguity in the definition of order, when data provide coordinates in a higher dimensional space the notion of order is ill-defined; for instance, in $\mathbb{R}^2$ it is not clear whether $(3, 6)$ is larger or smaller than $(5, 4)$. This fact led to a body of work in the 1970s, proposing new exploratory data analysis tools for assigning ranks to points in a data set. The method of *convex hull peeling*, credited to J. W. Tukey (Huber, 1972; Barnett, 1976) is a particularly intuitive example. A pedagogical description of the procedure for the bivariate case is provided in Green (1981), where readers are encouraged to envisage the data points as pins on a board. A large elastic band is looped around the pins forming the convex hull of the data points. The data points touching the elastic band are the extremes of the empirical distribution and are assigned rank one and discarded. The procedure is repeated to identify the next most extreme points, which are assigned rank two, and so on. Clearly, in this example, the empirical distribution plays an important role in the assignment of rank, where, roughly speaking, data points closer to the centre of the empirical distribution receive higher rank(s), giving rise to a centre-outward ordering.

J. W. Tukey coined the term *depth* in Tukey (1975) as the collection of exploratory procedures for assigning ranks to points in a data set. There, he proposed the celebrated *halfspace depth*, or *Tukey depth*, of a data point in $\mathbb{R}^p$ with respect to (henceforth w.r.t.) a multidimensional sample. Rousseeuw and Ruts (1999) later defined the halfspace depth w.r.t. a generic measure as opposed to the empirical measure, broadening the purely data analytic perspective. Thus, modern usage of the term depth refers to a much more general class of objects. The underlying mathematical idea behind these depth constructions and others is that a probability measure maps events in the Borel σ-algebra to $[0, 1]$, a space on which the assignment of order poses no concern.

Since Tukey's seminal work, many alternative examples of depth have been proposed. It was, however, the *simplicial depth* (Liu, 1990) that sparked a resurgence of research on the topic throughout the 1990s and 2000s. Simplicial depth is shown in Liu (1990) to possess several desirable properties, on the basis of which the definition of depth is formalised in Zuo and Serfling (2000b), reproduced in Definition 2.1 for ease of reference. In Definition 2.1, $\mathcal{P}$ denotes the class of distributions on the Borel sets of $\mathbb{R}^p$, and $P = P_X$ denotes the distribution of a general random vector X; the subscript X is suppressed when there is no need to be explicit.

DEFINITION 2.1 (Zuo and Serfling, 2000b; Liu, 1990). The bounded and non-negative mapping $D(\cdot,\cdot) : \mathbb{R}^p \times \mathcal{P} \longrightarrow \mathbb{R}$ is called a *statistical depth function* if it satisfies the following properties:

1. *Affine invariance.* $D(Ax + b, P_{AX+b}) = D(x, P_X)$ holds for any $\mathbb{R}^p$-valued random vector X, any $p \times p$ nonsingular matrix A and any $b \in \mathbb{R}^p$.
2. *Maximality at centre.* $D(\theta, P) = \sup_{x\in\mathbb{R}^p} D(x, P)$ holds for any $P \in \mathcal{P}$ having a unique centre of symmetry θ w.r.t. some notion of symmetry.
3. *Monotonicity relative to the deepest point.* For any $P \in \mathcal{P}$ having deepest point θ, $D(x, P) \leq D(\theta + \alpha(x - \theta), P)$ holds for all $\alpha \in [0, 1]$.
4. *Vanishing at infinity.* $D(x, P) \to 0$ as $\|x\| \to \infty$, for each $P \in \mathcal{P}$, where $\|\cdot\|$ is the Euclidean norm.

For a discussion of centre of symmetry in $\mathbb{R}^p$, see Zuo and Serfling (2000b); a more general discussion, applicable to function spaces, is provided in Section 3.1.2. Four further properties purported in Serfling (2006) as desirable but not necessary are reproduced in (i)–(iv) below:

(i) *Symmetry.* If P is symmetric about θ in some sense, then so is $D(x, P)$.

(ii) *Continuity of $D(x, P)$ as a function of x.* Or merely *upper semi-continuity.*

(iii) *Continuity of $D(x, P)$ as a function of P.*

(iv) *Quasi-concavity as a function of x.* The set $\{x : D(x, P) \geq c\}$ is convex for each real c.

Upper semicontinuity is a weaker requirement than continuity. In $\mathbb{R}^d$, it is natural to obligate the depth function to preserve the upper semicontinuity property of the distribution function. This statement has a straightforward extension to function spaces, which is addressed in Section 3. (iii), although not required to provide an order, is indispensable in view of the fact that statisticians do not have access to the true P but rather an empirical counterpart, which converges almost surely to P uniformly over the Borel sets of the domain on which P is defined. It has yet further important implications, explained in Section 3.1.6. Suitable modifications of properties 1–4, as well as (ii) and (iii), are thus amongst our defining properties of functional depth and are thoroughly justified and discussed in the functional framework in Section 3.1. No further attention is dedicated to properties (i) and (iv), which are not deemed necessary, neither in the multivariate nor functional framework. The designation of D as a bounded and non-negative mapping is also unnecessary to provide an order, and thus does not appear in our definition of functional depth in Section 3.

2.2 The FDA Framework

To formalise the FDA framework, a data point is thought of as a realisation of the random function $\{X(v) : v \in \mathcal{V}\}$, where $\mathcal{V}$ is a compact subset of $\mathbb{R}^d$ for $d \geq 1$. Letting Ω denote the underlying sample space, $\{X(v) : v \in \mathcal{V}\} := \{X(\omega, v) : \omega \in \Omega, v \in \mathcal{V}\}$ is the map $X : \Omega \to \mathfrak{F}$, where $\mathfrak{F}$ is a function space, whilst for a fixed $\omega \in \Omega$, $X(\omega, \cdot)$ maps from $\mathcal{V}$ to a vector space $\mathbb{F}$. There is a rich body of work concerning $\mathfrak{F} = \mathbb{L}_2(\mathcal{V}, \lambda)$, the space of Lebesgue square integrable functions from $\mathcal{V}$ to $\mathbb{F} = \mathbb{R}$ (here and henceforth, λ denotes Lebesgue measure on $\mathcal{V}$). Nonstandard choices of $\mathfrak{F}$ will undoubtedly become more prevalent in the FDA literature, which currently accommodates functional manifolds (Chen and Müller, 2012) and multivariate functional spaces $\mathfrak{F} = \bigotimes_{k=1}^{K} \mathbb{L}_2(\mathcal{V}_k, \lambda)$ (Chiou and Müller, 2014) as well as a variety of smoothness classes embedded in $\mathbb{L}_2(\mathcal{V}, \lambda)$. In the interest of generality, for the definition of functional depth, we do not restrict $\mathfrak{F}$ beyond the assumption that there exists a metric d on $\mathfrak{F}$ such that $(\mathfrak{F}, d)$ is a separable metric space.

A further distinguishing feature of functional data is that they are inherently partially observed. Although theoretically infinite dimensional data objects, due to the limitations of the data collection instruments or the experimental design, each functional data object is only ever recorded at a finite set of discretisation points, which we denote by $\mathcal{V}' \subset \mathcal{V}$.

The following notation is henceforth used throughout. $(\mathfrak{F}, d)$ is a separable metric space and $\mathcal{A}$ is the σ-algebra on $\mathfrak{F}$ generated by the open d metric balls. Separability of $(\mathfrak{F}, d)$ guarantees that $\mathcal{A}$ coincides with the Borel σ-algebra on $\mathfrak{F}$ (see, e.g., van der Vaart and Wellner, 1996, Chapter 1.7). $(\mathfrak{F}, \mathcal{A}, P)$ is a probability space with $P \in \mathcal{P}$, the space of all probability measures on the Borel sets of $\mathcal{A}$. Particular instances of $\mathfrak{F}$ to which reference is made are as follows: $\mathbb{H}(\mathcal{V})$, an infinite dimensional Hilbert space on $\mathcal{V}$; $\mathcal{C}(\mathcal{V})$, the space of continuous functions on $\mathcal{V}$; $\mathbb{L}_p(\mathcal{V}, \lambda)$, the space of Lebesgue p-integrable functions on $\mathcal{V}$, where $1 \leq p < \infty$; $\mathbb{L}_\infty(\mathcal{V})$, the space of uniformly bounded functions on $\mathcal{V}$; and $\mathfrak{W}^{k,p}(\mathcal{V}, \lambda)$, the Sobolev space of Lebesgue p-integrable functions on $\mathcal{V}$ whose weak derivatives up to order $k \geq 1$ are Lebesgue p-integrable on $\mathcal{V}$, where $1 \leq p < \infty$. To avoid excessive notation, unless otherwise stated, the argument(s) $\mathcal{V}$ and λ (if applicable) are tacit when we write $\mathcal{C}$, $\mathbb{H}$, $\mathbb{L}_p$, $\mathbb{L}_\infty$ and $\mathfrak{W}^{k,p}$. Similarly, $\|x\|_{\mathbb{L}_p(\mathcal{V},\lambda)} = (\int_{\mathcal{V}} x(v)^p \lambda(dv))^{1/p}$ is henceforth referred to in the more compact form $\|x\|_{\mathbb{L}_p}$. In normed spaces, the metric d will most naturally be a norm;

in this case $d = \|\cdot\|_{\mathbb{L}_p}$ is used to mean $d(x, y) = \|x - y\|_{\mathbb{L}_p}$. $\mathbb{H}$ is most naturally endowed with its inner product norm $\|x - y\|_{\mathbb{L}_2} = \sqrt{\langle x - y, x - y\rangle}$ for $x, y \in \mathbb{H}$, whilst $\mathbb{L}_\infty$ is most naturally endowed with the supremum norm $\|x - y\|_\infty = \sup_{v\in\mathcal{V}} |x(v) - y(v)|$ for $x, y \in \mathbb{L}_\infty$. Recall from the above introduction to the FDA framework that for any $\omega \in \Omega$, $X(\omega, \cdot) : \mathcal{V} \to \mathbb{F}$, where $\mathbb{F}$ is a vector space; unless otherwise stated, $\|\cdot\|$ will be used to denote an arbitrary norm on $\mathbb{F}$. For any $x \in \mathfrak{F}$, $x(H) := \{x(v) : v \in H \subseteq \mathcal{V}\}$, whilst $x(\mathcal{V})$ is tacitly implied by x. Finally, a sample $X_1, \ldots, X_n$ of random draws from P gives rise to the empirical measure P_n, a collection of $\frac{1}{n}$-weighted point masses at $X_1, \ldots, X_n$. $\widetilde{P}_n$ is used to denote the empirical measure of $\widetilde{X}_1, \ldots, \widetilde{X}_n$, which is a sample of reconstructed functional data objects based on the random sample $\{X_i(\mathcal{V}_i') : i = 1, \ldots, n\}$ of partially observed functional data objects, where $\mathcal{V}_i' \subset \mathcal{V}$ is a finite set that may be different for every $i \in \{1, \ldots, n\}$.

3. FORMAL DEFINITION OF FUNCTIONAL DEPTH

The definition of functional depth provided in this section refers to the concept of centre of symmetry, which is elucidated in Section 3.1.2, and relies on the following preliminary definition.

DEFINITION 3.1. Let $(\mathfrak{F}, \mathcal{A}, P)$ be a probability space as in Section 2.2. Define $\mathcal{E}$ to be the smallest set in the σ-algebra $\mathcal{A}$ such that $P(\mathcal{E}) = P(\mathfrak{F})$. Then the convex hull of $\mathfrak{F}$ with respect to P is defined as

$$\mathfrak{C}(\mathfrak{F}, P) := \big\{x \in \mathfrak{F} : x(v) = \alpha L(v) + (1 - \alpha)U(v) : v \in \mathcal{V}, \alpha \in [0, 1]\big\},$$

where

$$U := \Big\{\sup_{x\in\mathcal{E}} x(v) : v \in \mathcal{V}\Big\} \quad \text{and}$$

$$L := \Big\{\inf_{x\in\mathcal{E}} x(v) : v \in \mathcal{V}\Big\}.$$

DEFINITION 3.2. Let $(\mathfrak{F}, \mathcal{A}, P)$ be a probability space as in Section 2.2. Let $\mathcal{P}$ be the space of all probability measures on $\mathfrak{F}$. The mapping $D(\cdot, \cdot) : \mathfrak{F} \times \mathcal{P} \longrightarrow \mathbb{R}$ is a *statistical functional depth* if it satisfies properties P-1 to P-6, below.

P-1. *Distance invariance.* $D(f(x), P_{f(X)}) = D(x, P_X)$ for any $x \in \mathfrak{F}$ and $f : \mathfrak{F} \to \mathfrak{F}$ such that for any $y \in \mathfrak{F}$, $d(f(x), f(y)) = a_f \cdot d(x, y)$, with $a_f \in \mathbb{R} \setminus \{0\}$.

P-2. *Maximality at centre.* For any $P \in \mathcal{P}$ possessing a unique centre of symmetry $\theta \in \mathfrak{F}$ w.r.t. some notion of functional symmetry, $D(\theta, P) = \sup_{x\in\mathfrak{F}} D(x, P)$.

P-3. *Strictly decreasing with respect to the deepest point.* For any $P \in \mathcal{P}$ such that $D(z, P) = \max_{x\in\mathfrak{F}} D(x, P)$ exists, $D(x, P) < D(y, P) < D(z, P)$ holds for any $x, y \in \mathfrak{F}$ such that $\min\{d(y, z), d(y, x)\} > 0$ and $\max\{d(y, z), d(y, x)\} < d(x, z)$.

P-4. *Upper semi-continuity in x.* $D(x, P)$ is upper semi-continuous as a function of x, that is, for all $x \in \mathfrak{F}$ and for all $\varepsilon > 0$, there exists a $\delta > 0$ such that

$$\sup_{y:d(x,y)<\delta} D(y, P) \le D(x, P) + \varepsilon. \tag{3.1}$$

P-5. *Receptivity to convex hull width across the domain.* $D(x, P_X) < D(f(x), P_{f(X)})$ for any $x \in \mathfrak{C}(\mathfrak{F}, P)$ with $D(x, P) < \sup_{y\in\mathfrak{F}} D(y, P)$ and $f : \mathfrak{F} \to \mathfrak{F}$ such that $f(y(v)) = \alpha(v)y(v)$ with $\alpha(v) \in (0, 1)$ for all $v \in L_\delta$ and $\alpha(v) = 1$ for all $v \in L_\delta^c$:

$$L_\delta := \operatorname*{argsup}_{H\subseteq\mathcal{V}} \Big\{\sup_{x,y\in\mathfrak{C}(\mathfrak{F},P)} d\big(x(H), y(H)\big) \le \delta\Big\}$$

for any $\delta \in [\inf_{v\in\mathcal{V}} d(L(v), U(v)), d(L, U))$ such that $\lambda(L_\delta) > 0$ and $\lambda(L_\delta^c) > 0$.

P-6. *Continuity in P.* For all $x \in \mathfrak{F}$, for all $P \in \mathcal{P}$ and for every $\varepsilon > 0$, there exists a $\delta(\varepsilon) > 0$ such that $|D(x, Q) - D(x, P)| < \varepsilon$ P-almost surely for all $Q \in \mathcal{P}$ with $d_{\mathcal{P}}(Q, P) < \delta$ P-almost surely, where $d_{\mathcal{P}}$ metricises the topology of weak convergence.

3.1 Discussion of the Functional Depth Defining Properties

3.1.1 *Discussion of* P-1, *distance invariance.* property P-1 is the generalisation from $\mathbb{R}^d$ to $\mathfrak{F}$ of property 1 of Zuo and Serfling (2000b), also considered in Theorem 3 of Liu (1990). It states that any mapping from $\mathfrak{F}$ to $\mathfrak{F}$ that preserves, up to a scaling factor, the relative distances between elements in the d metric also preserves the depth in the transformed space. As an example, consider $(\mathfrak{F}, d) = (\mathbb{L}_2, \|\cdot\|_{\mathbb{L}_2})$ and suppose $\mu := \mathbb{E}(X) = \int x P(dx)$ is known. Then property P-1 ensures that the depth is unaffected by recentring around the zero function because $\|x - y\|_{\mathbb{L}_2} = \|(x - \mu) - (y - \mu)\|_{\mathbb{L}_2}$ for all $x, y \in \mathbb{L}_2$.

3.1.2 *Discussion of* P-2, *maximality at centre.* P-2 is the most logically contentious of the properties listed. The reason is that, even for distributions on $\mathbb{R}^d$, there is no unique notion of symmetry, a fact that is a fortiori true in function spaces. Indeed, since depth itself was originally conceived as a way to give meaning to the concept of centre of symmetry, the deepest element is no less valid as a centre of symmetry than

any other definition, giving rise to the somewhat paradoxical conclusion that P-2 is always achieved with θ equal to the deepest point, as long as $\sup_{x\in\mathfrak{F}} D(x, P) = \max_{x\in\mathfrak{F}} D(x, P)$. It is more meaningful to consider the behaviour of D for a particular P for which many notions of centre of symmetry coincide at θ. In $\mathbb{R}$ such a P is the Gaussian distribution, for which the median is equal to the mean and is a centre of symmetry with respect to many notions of symmetry including central symmetry and halfspace symmetry (e.g., Zuo and Serfling, 2000a). In the setting of $\mathfrak{F} = \mathbb{H}$, the analogue of the Gaussian distribution is the Gaussian process. With this in mind, verification of the following property is insightful:

P-2G. *Maximality at Gaussian process mean*. For P a zero-mean, stationary, almost surely continuous Gaussian process on $\mathcal{V}$, $D(\theta, P) = \sup_{x\in\mathfrak{F}} D(x, P) \neq \inf_{x\in\mathfrak{F}} D(x, P)$, where θ is the zero mean function.

REMARK 3.3. Existence of $\mathbb{E}X$ is guaranteed when $X \sim P$ with P a Gaussian process.

The zero function of property P2-G is the centre of symmetry of the mean zero Gaussian process with respect to all notions of functional symmetry that have been tacitly introduced via existing depth constructions, for instance, pointwise angular symmetry in Fraiman and Muniz (2001) and López-Pintado and Romo (2009), and pointwise halfspace symmetry in Claeskens et al. (2014). If a distribution P_X on $\mathfrak{F}$ is pointwise halfspace symmetric about z, then for every $v \in \mathcal{V}$, the corresponding distribution of $X(v)$ is halfspace symmetric around $z(v)$.

Property P-2, in partnership with P-3, leads to the centre-outward ordering for which depth was originally conceived. Outward orderings from local centres of symmetry are also possible (see Paindaveine and Van Bever, 2013), and are induced by constructions that attach greater importance to probabilities $P(A)$ for Borel sets A to which the evaluation points x have close proximity, where proximity is measured by a suitable metric. The relative weighting depends on the features of P that one would like to detect through the use of the local depth function. As the weighting rule becomes close to uniform, the local features are blurred, resulting in global behaviour of any local depth construction. Local centre-outward orderings are not induced by our definition.

3.1.3 *Discussion of P-3, strictly decreasing with respect to the deepest point*. For some function spaces $\mathfrak{F}$, there is more than one natural metric d. For instance, if $\mathfrak{F} = \mathbb{L}_\infty \cap \mathfrak{W}^{k,2}$, $(\mathfrak{F}, d)$ is separable with respect to the supremum norm, the standard Sobolev inner product norm (Adams, 1975) or its slight generalisation, as employed in Silverman (1996). With this example in mind, setting $d = \|\cdot\|_\infty$ and $\mathcal{V} \subset \mathbb{R}$, property P-3 ensures that the depth prescribes successively lower depths to functions that only belong to successively larger envelopes around the deepest point z. However, when d is the standard Sobolev inner product norm, the depth prescribes successively lower depths to functions which lie in successively larger Sobolev balls around z, that is, its prescription takes account of the distance of x from z in derivative space as well as in $\mathbb{L}_2$ norm, assigning low depth to functions much rougher than z.

P-3 has two further implications. The first is that

$$\lim_{x:d(x,z)\to\infty} D(x, P) = \inf_{x\in\mathfrak{F}} D(x, P), \tag{3.2}$$

where z is such that $D(z, P) = \max_{x\in\mathfrak{F}} D(x, P)$ exists and where the convention in current literature is to construct $D(\cdot, P)$ such that $\inf_{x\in\mathfrak{F}} D(x, P) = 0$ for any $P \in \mathcal{P}$. Equation (3.2) itself leads to the conclusion of Lemma 3.4.

LEMMA 3.4. *Let $(\mathfrak{F}, d)$ be a functional metric space such that $d = \|\cdot\|_{\mathbb{L}_p}$, then for each $P \in \mathcal{P}$, (3.2) implies that $D(x, P) \to \inf_{x\in\mathfrak{F}} D(x, P)$ as $\|x(v)\| \to \infty$ for Lebesgue almost every $v \in \mathcal{V}$, where $\|\cdot\|$ is a norm on $\mathbb{F}$ (cf. Section 2.2).*

Requiring that $D(x, P) \to \inf_{x\in\mathfrak{F}} D(x, P)$ as $\|x(v)\| \to \infty$ for Lebesgue almost every $v \in \mathcal{V}$ is one natural analogue of property 4 of Zuo and Serfling (2000b), Liu (1990) and was suggested in Nieto-Reyes (2011), but we view property P-3 as more suitable in view of the arguments already set forth in this discussion. The second implication of P-3 is Lemma 3.5.

LEMMA 3.5. *Let $D(\cdot, \cdot) : \mathfrak{F} \times \mathcal{P} \to \mathbb{R}$ satisfy property* P-3 *and let z be as in* P-3. *Then*

$$z = \operatorname*{argmax}_{x\in\mathfrak{F}} D(x, P).$$

The direct analogue of property 3 of Zuo and Serfling (2000b), Liu (1990) is to relax the strict inequality in property P-3. The strict inequality in P-3 yields fewer ties in depth computations, which enables us to better differentiate amongst the different elements of $\mathfrak{F}$. Moreover, strict inequality in P-3 automatically implies nondegeneracy of functional depth because it prevents all the points in $\mathfrak{F}$ having the same depth. Degenerate behaviour of several depth constructions is observed in Chakraborty and Chaudhuri (2014b). They

show that, inter alia, the band depth and half-region depth constructions result in zero depth of every function in $\mathfrak{F}$ with probability one for common distributions such as continuous Gaussian processes.

3.1.4 *Discussion of* P-4, *upper semi-continuity in x*. In $\mathbb{R}$, there is a clear correspondence between the definition of depth and the cumulative distribution function $F(x) = P(X \leq x)$. The two natural ways of defining the depth at a point $x \in \mathbb{R}$ are $D(x, P) = P(X \leq x) \cdot P(X \geq x)$ and $D(x, P) = \min\{P(X \leq x), P(X \geq x)\}$, thus, from the càdlàg property of the cumulative distribution function, it is clear that, in $\mathbb{R}$, the depth is upper semicontinuous in the distance from the deepest point. The point we wish to make here is that, in order for the depth to reveal the features of the underlying distribution, it should, as a minimal requirement, satisfy the same properties as a cumulative distribution function, namely, being nondecreasing (P-3) and upper-semicontinuous (P-4).

3.1.5 *Discussion of* P-5, *receptivity to convex hull width across the domain*. Many functional data sets encountered in practice contain functional data points that exhibit little variability over a particular subset of the domain $L \subset \mathcal{V}$, and significantly overlap with one another on L. The phenomenon described arises, inter alia, in functional microarray data sets (Amaratunga and Cabrera, 2003) and in chemometric data sets (see, e.g., the yarn data set in the R package pls Swierenga et al., 1999). Although the instinct is to draw parallels with the notion of heteroskedasticity in linear regression, this is in fact an entirely different phenomenon, as it is usually still appropriate to view functional data as i.i.d. copies of a random function X; X simply possesses a variance function that is close to zero over L and a correlation function close to one over $L \times L$. P-5 obligates the depth to take heed of the values of $x \in \mathfrak{C}(\mathfrak{F}, P)$ over $\mathcal{V} \setminus L$ to a greater extent than over L. Heuristically, the order of the curves does not matter much over L. Property P-5 is particularly important in view of the discussion of P-6 because, over L, small measurement error can conceivably lead to reconstructed functions that overlap in a drastically different way to the same functions observed without measurement error. A simple solution available for integrated depth constructions is to integrate the pointwise depths over a weight function depending on the convex hull of the data. This solution, proposed in Claeskens et al. (2014), effectively reduces the influence of regions over which all functions nearly coincide.

3.1.6 *Discussion of* P-6, *continuity in P*. Examples of $d_{\mathcal{P}}(\cdot, \cdot)$ are the Prohorov and bounded Lipschitz metrics, which both metricise the topology of weak convergence in the sense that $d_{\mathcal{P}}(Q, P) \to 0$ P-almost surely is equivalent to $Q \to P$ P-almost surely (e.g., Dudley, 2002, Theorem 11.3.3).

Almost sure convergence of empirical depth to population depth. The importance of property P-6 is evident when replacing Q with P_n. In this case, fulfillment of P-6 implies that the depth based on the empirical distribution converges almost surely to its population counterpart, that is, the estimator $D(\cdot, P_n) \to D(\cdot, P)$ P-almost surely. This is particularly important when the depth is to be used for statistical inference. In this case, the objective is to gain understanding of population truths based on a random sample from that population. By contrast, in data analysis problems, the statistician typically has access to the whole population. Functional data analysis is, however, slightly different in view of the inherent partial observability of functional data.

Partial observability of functional data. A second fundamental observation pertaining to P-6 is that it tacitly addresses the inherent partial observability problem of functional data analysis. The latter gives rise to the delicate challenge of P_n being inaccessible in its entirety. More specifically, whilst P_n is a collection of weighted point masses at $X_1, \ldots, X_n$, each valued in $\mathfrak{F}$, the practitioner only has access to P_n', a collection of weighted point masses on $\{X_i(\mathcal{V}_i') : i = 1, \ldots, n\}$, where $\mathcal{V}_i' \subset \mathcal{V}$ is a finite set that may be different for every $i \in \{1, \ldots, n\}$. The issue of partial observability of functional data is usually addressed through a preliminary interpolation or smoothing step to obtain an approximate reconstruction of the functional data object. Let $\widetilde{X}_1, \ldots, \widetilde{X}_n$ be a sample of reconstructed functional data objects obtained from the random sample $\{X_i(\mathcal{V}_i') : i = 1, \ldots, n\}$ of partially observed functional data objects or even from $\{X_i^*(\mathcal{V}_i') : i = 1, \ldots, n\}$, where $X_i^*(\mathcal{V}_i') = \{X_i(v) + \varepsilon_i, v \in \mathcal{V}_i'\}$ with $\{\varepsilon_i : i = 1, \ldots, n\}$ independent mean zero noise variables. Let $\widetilde{P}_n$ be the empirical probability measure over $\widetilde{X}_1, \ldots, \widetilde{X}_n$. Then provided the reconstruction is such that $\widetilde{P}_n \to P$ P-almost surely, property P-6 delivers the desired convergence of the functional depth.

Qualitative robustness. Importantly, fulfillment of P-6 produces an embodiment of the empirical depth with the quintessential feature of robustness (cf. Theorem 3.7 below). The following definition of qualitative robustness is a restatement of Definition (A) in

Hampel (1971) in the more specific terms of the empirical depth. Here, we subscript the empirical depth by P and Q to emphasise that P_n and Q_n are random draws from P and Q respectively. With this notation, $\mathcal{L}(D_P(\cdot, P_n))$ is the probability measure on $\mathbb{R}$ induced by the mapping $D_P(\cdot, P_n)$. The theorem and definition are stated in terms of P_n, but it applies analogously when P_n is replaced by $\widetilde{P}_n$.

DEFINITION 3.6 (Qualitative robustness). Let P_n and Q_n be the empirical measures corresponding to the n random draws from P and Q respectively. For any $x \in \mathfrak{F}$, $D_P(x, P_n)$ is *robust* at $P \in \mathcal{P}$ if and only if for all $\varepsilon > 0$ there exists a $\delta > 0$ such that for any $Q \in \mathcal{P}$ satisfying $d_{\mathcal{P}}(Q, P) < \delta$, $d_{\mathcal{P}}(\mathcal{L}(D_P(x, P_n)), \mathcal{L}(D_Q(x, Q_n))) < \varepsilon$ for all n.

THEOREM 3.7 (An application of Hampel et al. (1986), Section 2.2, Theorem 2). *If D_P satisfies property* P-6, *then $D_P(\cdot, P_n)$ is robust at P for any $P \in \mathcal{P}$.*

Qualitative robustness of the empirical depth is a desirable feature, as it ensures that conclusions are not inordinately affected by outliers.

3.2 Implications for Applications

In this section we emphasise the roles played by P-1 to P-6 for different kinds of applications.

Regarding P-1, in many applications, one would like the conclusions of statistical analysis or inference to be invariant to changes in the units of measurement. Nevertheless, for applications in which the ranking amongst the functions is the object of interest rather than the precise value of the depth, a weaker requirement may be sought: invariance of the ordering rather than invariance of the depth values. This requirement would be suitable for constructing trimmed sample statistics by discarding the most extreme order statistics. There are applications in which the value of the depths themselves are of interest, and thus invariance in the precise sense of P-1 is important. For instance, in certain model systems, systemic stability is related to diversity of a population and distance of the population centre of symmetry from a point, p, that is independent of the population. This situation arises in the model of the financial system considered by Beale et al. (2011), where $P = P_n$, that is, the whole population is available. One may construct a measure, R, of systemic risk from $d(z, p)$ and $\sum_{i=1}^{n} D(X_i, P)$, where $z = \text{argmax}_{x \in \mathfrak{F}} D(x, P)$, $P = P_n$ and X_i is the relevant functional observation on individual i. The systemic risk contribution of individual i is then $R_i = D(X_i, P)/R$. This hints at the possibility of regulatory mechanisms designed to incentivise high systemic risk individuals towards a more systemically stable configuration in $\mathfrak{F}$ space. P-3 ensures the diversity information is captured in the prescribed depths whilst P-3 and P-4 together ensure that the depth is not simply a ranking but captures the relative proximities of each individual to the centre of symmetry.

The centre-outward ordering induced by P-2 and P-3 and the information on relative proximities induced by P-3 and P-4 are qualities that enhance the ability of functional DD classifiers (Li, Cuesta-Albertos and Liu, 2012) to differentiate between samples drawn from two different distributions. Moreover, the centre-outward ordering guaranteed by P-3 provides the necessary and sufficient conditions for defining nearest neighbours (Paindaveine and Van Bever, 2015). Depth-based nearest neighbours have been effectively exploited (in the same reference) to define new classifiers, but they also offer prospects for nearest neighbour-based nonparametric regression (e.g., Devroye, Györfi and Lugosi, 1996).

Property P-5 is also important for functional classification. If curves are from two different populations, both possessing covariance function close to zero over a subset $L \subset \mathcal{V}$, a functional DD plot classifier based on a depth violating P-5, ceteris paribus, has less power to discriminate between the two samples than one based on a depth satisfying P-5. Since classification is an inference (supervised learning) problem, P-6 is important for ensuring that the sample depths of each $x \in \mathcal{F}$ converge to the corresponding population depths as $n \to \infty$. This assumption underpins the success of the DD classifier.

Regardless of the precise nature of the application, P-6 is important for all of them, with its precise role depending on whether the application concerns inference or data analysis. For inference problems, the requirement is that the empirical depth converges to the population depth. Moreover, we require that the empirical depth based on the discretised functional data converges to the population depth. For data analysis problems, the aim is for the empirical depth based on the discretised functional data to converge to the empirical depth.

4. A COMPARATIVE STUDY OF EXISTING FUNCTIONAL DEPTH PROPOSALS

In this section we explore several popular constructions that have been proposed as functional depths in

the literature. As we will see in due course, there is no single construction that satisfies all six properties in our definition of functional depth, which emphasises the necessity for further work in the area. Only functional depth constructions that have been proposed at the population level rather than simply at the sample level are explored, which rules out the construction based on distances that appears in Nieto-Reyes (2011) and the one based on tilting that appears in Genton and Hall (2015).

4.1 Existing Functional Depth Constructions

In each of the depth constructions outlined below, X is a functional random variable defined on the probability space $(\mathfrak{F}, \mathcal{A}, P)$ (cf. Section 2.2) and, where relevant, expectation $\mathbb{E}$ is taken with respect to P unless otherwise stated. Sample analogues are obtained by replacing P by P_n for the idealised case and by $\widetilde{P}_n$ for the practically relevant case in which functional data objects are only observed at a finite set of evaluation points (cf. Section 3.1.6). For completeness, the sample versions of each depth construction in the idealised case are included after their population counterparts. The nonidealised sample versions, $D(\cdot, \widetilde{P}_n)$, are obtained by replacing $\{X_1, \ldots, X_n\}$ by $\{\widetilde{X}_1, \ldots, \widetilde{X}_n\}$, in $D(\cdot, P_n)$. The constructions below need not uniquely prescribe a choice of metric d; however, in most cases, there is a natural choice of d with which to assess the fulfillment of properties P-1 to P-6 in Definition 3.2. In each construction, $(\mathfrak{F}, d)$ is as stated, $\mathcal{A}$ is the Borel sigma algebra (also the d-ball σ-algebra; cf. Section 2.2), and P is a probability measure on the Borel sets of $\mathcal{A}$.

4.1.1 *The h-depth.* Let $(\mathfrak{F}, d) = (\mathbb{H}, \|\cdot\|_{\mathbb{L}_2})$. The h-depth (Cuevas, Febrero and Fraiman, 2007) at $x \in \mathbb{H}$ w.r.t. P is defined as

$$D_h(x, P) := \mathbb{E}K_h\big(\|x - X\|_{\mathbb{L}_2}\big), \tag{4.1}$$

where, for fixed $h > 0$, $K_h(\cdot) = (1/h)K(\cdot/h)$, with $K(\cdot)$ the Gaussian kernel. The sample analogue of (4.1) is $D_h(x, P_n) := \frac{1}{n}\sum_{i=1}^n K_h(\|x - X_i\|_{\mathbb{L}_2})$. The h-depth is the only example we consider that can be described as local (cf. Section 3.1.2), a feature that is dispelled when the parameter h is sufficiently large.

4.1.2 *The random Tukey depth.* Let $(\mathfrak{F}, d) = (\mathbb{H}, \|\cdot\|_{\mathbb{L}_2})$. Defining $\mathfrak{U} := \{u_1, \ldots, u_k\}$, where u_j $j = 1, \ldots, k$ are realisations of U_j $j = 1, \ldots, k$, each drawn independently from a nondegenerate probability measure μ on $\mathbb{H}$, the random Tukey depth (Cuesta-Albertos and Nieto-Reyes, 2008) at $x \in \mathbb{H}$ w.r.t. P is

$$\begin{aligned} D_{\mathrm{RT}}(x, P) &= D_{\mathfrak{U}}(x, P) \\ &:= \min_{u \in \mathfrak{U}} D_1\big(\langle u, x\rangle, P_u\big), \end{aligned} \tag{4.2}$$

where, for any probability measure Q on the Borel sets of $\mathbb{R}$, $D_1(t, Q) = \min\{Q(-\infty, t], Q[t, -\infty)\}$, P_u is the marginal of P on $\{\langle u, x\rangle : x \in \mathbb{H}\}$. μ is taken as a nondegenerate stationary Gaussian measure on $\mathbb{H}$. For a discussion of the choice of k, see Cuesta-Albertos and Nieto-Reyes (2008). The sample analogue of (4.2) is simply obtained by replacing P with P_n.

4.1.3 *The band depth.* Let $(\mathfrak{F}, d) = (\mathcal{C}, \|\cdot\|_\infty)$ and let $\mathcal{V} \subset \mathbb{R}$. For $j \geq 2$, introduce the random j-simplex in $\mathfrak{F}$, $\mathcal{S}_j(P) = \{y \in \mathfrak{F} : y(v) = \alpha_1 X_1(v) + \cdots + \alpha_j X_j(v) : (\alpha_k)_{k=1}^j \in \Delta^j \ \forall v \in \mathcal{V}, (X_k)_{k=1}^j \sim P\}$, where $\Delta^j \subset \mathbb{R}^{j-1}$ is the unit j-simplex. The band depth (López-Pintado and Romo, 2009) at $x \in \mathfrak{F}$ is defined as

$$D_J(x, P) = \sum_{j=2}^J P_{\mathcal{S}_j}\big(x \in \mathcal{S}_j(P)\big), \tag{4.3}$$

where $P_{\mathcal{S}_j}$ is the probability measure over the random simplices constructed from the random j-tuple $X_1, \ldots, X_j$.

When P is replaced by P_n, there are n choose j distinct sets in the set of all random j-simplices on $\mathfrak{F}$ giving rise to the sample analogue of equation (4.3), $D_J(x, P_n) = \sum_{j=2}^J \binom{n}{j}^{-1} \sum_{1 \leq i_1 < \cdots < i_j \leq n} \mathbb{1}\{x \in B_{i_j}\}$, where $B_{i_j} := \{y \in \mathfrak{F} : y(v) = \alpha_1 X_{i_1}(v) + \cdots + \alpha_j X_{i_j}(v) : (\alpha_k)_{k=1}^j \in \Delta^j \ \forall v \in \mathcal{V}\}$ and $\{(i_1, \ldots, i_j) : i = 1, \ldots, n\}$ defines the set of all possible j-tuples from $X_1, \ldots, X_n$.

4.1.4 *The modified band depth.* Let $(\mathfrak{F}, d) = (\mathcal{C}, \|\cdot\|_\infty)$ and let $\mathcal{V} \subset \mathbb{R}$. For $j \geq 2$, define a random j-simplex in $\mathbb{R}$ to be of the form $\mathcal{S}_j(v, P) = \{y(v) \in \mathbb{R} : y(v) = \alpha_1 X_1(v) + \cdots + \alpha_j X_j(v) : (\alpha_k)_{k=1}^j \in \Delta^j, (X_k)_{k=1}^j \sim P\}$, where $\Delta^j \subset \mathbb{R}^{j-1}$ is the unit j-simplex. The modified band depth (López-Pintado and Romo, 2009) at $x \in \mathfrak{F}$ is

$$\begin{aligned} &D_{\mathrm{MJ}}(x, P) \\ &\quad = \sum_{j=2}^J \mathbb{E}\big[\lambda\big\{v \in \mathcal{V} : x \in \mathcal{S}_j(v, P)\big\}\big]/\lambda(\mathcal{V}), \end{aligned} \tag{4.4}$$

where expectation is with respect to the measure $P_{\mathcal{S}_j}$, as defined above in the definition of the band depth. In Section 5 it will sometimes be convenient to re-

fer to $S_J(v, P) = [L_J(v), U_J(v)]$, where $L_J(v) := \min_{y \in \mathfrak{X}_j} y(v)$ and $U_J(v) := \max_{y \in \mathfrak{X}_j} y(v)$, where $\mathfrak{X}_j = (X_1, \ldots, X_j)$ and $X_1, \ldots, X_j \sim P$.

When P is replaced by P_n, there are n choose j distinct sets in the set of all random j-simplices on $\mathfrak{F}$ giving rise to the sample analogue of equation (4.4),

$$D_{\mathrm{MJ}}(x, P_n) := \sum_{j=2}^{J} \binom{n}{j}^{-1} \cdot \sum_{1 \le i_1 < \cdots < i_j \le n} \lambda\{v \in \mathcal{V} : x(v) \in B_{i,j}(v)\}/\lambda(\mathcal{V}),$$

where $B_{i_j}(v) := \{y(v) \in \mathbb{R} : y(v) = \alpha_1 X_{i_1}(v) + \cdots + \alpha_j X_{i_j}(v) : (\alpha_k)_{k=1}^{j} \in \Delta^j\}$ and $\{(i_1, \ldots, i_j) : i = 1, \ldots, n\}$ defines the set of all possible j-tuples from $X_1, \ldots, X_n$.

4.1.5 *The half-region depth.* In the same setting as for the band depth, the half-region depth (López-Pintado and Romo, 2011) w.r.t. P at $x \in \mathfrak{F}$ is

$$D_{\mathrm{HR}}(x, P) := \min\{P(X \in H_x), P(X \in E_x)\}, \tag{4.5}$$

where H_x is the hypograph of x, that is, $H_x := \{y \in \mathfrak{F} : y(v) \le x(v)\ v \in \mathcal{V}\}$, and E_x is the epigraph of x, that is, $E_x := \{y \in \mathfrak{F} : y(v) \ge x(v)\ v \in \mathcal{V}\}$. Thus, the halfspace depth is the minimum between the proportion of curves in the epigraph and hypograph of x. The sample analogue of (4.5) is obtained by replacing $P(X \in H_x)$ in (4.5) by $\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{X_i \in H_x\}$ and analogously for $P(X \in E_x)$.

4.1.6 *The modified half-region depth.* In the same setting as for the band depth, the half-region depth (López-Pintado and Romo, 2011) w.r.t. P at $x \in \mathfrak{F}$ is

$$D_{\mathrm{MHR}}(x, P) = \min\{\mathbb{E}[\lambda\{v \in \mathcal{V} : X(v) \le x(v)\}], \mathbb{E}[\lambda\{v \in \mathcal{V} : X(v) \ge x(v)\}]\}/\lambda(\mathcal{V}),$$

with sample analogue

$$D_{\mathrm{MHR}}(x, P_n) = \min\left\{\frac{1}{n}\sum_{i=1}^{n} \lambda\{v \in \mathcal{V} : X_i(v) \le x(v)\}, \frac{1}{n}\sum_{i=1}^{n} \lambda\{v \in \mathcal{V} : X_i(v) \ge x(v)\}\right\} \Big/ \lambda(\mathcal{V}).$$

In Table 1, we summarise the depth constructions presented in detail above.

4.1.7 *Other existing functional depth proposals.* In addition to the six functional depth proposals exposed above, there are several other constructions that have appeared in the literature. The integrated depth is proposed in Fraiman and Muniz (2001) as the first depth for functional data. It is defined by integrating over the continuum of one-dimensional pointwise depths at each point $x(v)$, $v \in \mathcal{V}$. As noted in Claeskens et al. (2014), the integrated depth is related to the modified band depth of López-Pintado and Romo (2009). More specifically, the modified band depth with $J = 2$, the recommended value in López-Pintado and Romo (2009), coincides with the integrated depth when computed w.r.t. a probability distribution with absolutely continuous marginals. This correspondence is due to the use of the simplicial depth for the one-dimensional pointwise depth, as initially proposed in Fraiman and Muniz (2001). Other one-dimensional pointwise depths are equally valid, but do not give rise to this same link with the modified band depth. The multivariate functional halfspace depth of Claeskens et al. (2014) generalises the integrated depth, allowing multivariate functions through the use of the multidimensional pointwise Tukey depth, and through the inclusion of a weight function to downweight the influence of the pointwise depth values over regions where

TABLE 1
Summary of existing depth constructions

Depth	$(\mathfrak{F}, d)$	$\mathcal{V}$	Construction
$D_h(x, P)$	$\mathbb{H}, \|\cdot\|_{\mathbb{L}_2}$	$\mathcal{V} \subset \mathbb{R}^d, d \ge 1$	$\mathbb{E}K_h(\|x - X\|_{\mathbb{L}_2})$
$D_{\mathrm{RT}}(x, P)$	$\mathbb{H}, \|\cdot\|_{\mathbb{L}_2}$	$\mathcal{V} \subset \mathbb{R}^d, d \ge 1$	$\min\{D_1(\langle u, x\rangle, P_u) : u \in \mathfrak{U}\}$
$D_J(x, P)$	$\mathcal{C}, \|\cdot\|_\infty$	$\mathcal{V} \subset \mathbb{R}$	$\sum_{j=2}^{J} P_{\mathcal{S}_j}(x \in \mathcal{S}_j(P))$
$D_{\mathrm{MJ}}(x, P)$	$\mathcal{C}, \|\cdot\|_\infty$	$\mathcal{V} \subset \mathbb{R}$	$\sum_{j=2}^{J} \mathbb{E}[\lambda\{v \in \mathcal{V} : x(v) \in \mathcal{S}_j(v, P)\}]/\lambda(\mathcal{V})$
$D_{\mathrm{HR}}(x, P)$	$\mathcal{C}, \|\cdot\|_\infty$	$\mathcal{V} \subset \mathbb{R}$	$\min\{P(X \in H_x), P(X \in E_x)\}$
$D_{\mathrm{MHR}}(x, P)$	$\mathcal{C}, \|\cdot\|_\infty$	$\mathcal{V} \subset \mathbb{R}$	$\min\{\mathbb{E}[\lambda\{v \in \mathcal{V} : X(v) \le x(v)\}]/\lambda(\mathcal{V}), \mathbb{E}[\lambda\{v \in \mathcal{V} : X(v) \ge x(v)\}]/\lambda(\mathcal{V})\}$

the convex hull width is small. Another approach to generalise the integrated depth to multivariate functions that was proposed in the recent literature is in Hlubinka et al. (2015). Other functional depth proposals include the integrated dual depth of Cuevas and Fraiman (2009), proposed as the population analogue of the random projection depth (Cuevas, Febrero and Fraiman, 2007). There, the double random projection depth was also proposed as the first example of depth suitable for multivariate functional data. Additionally, Chakraborty and Chaudhuri (2014a, 2014b) study from a functional depth perspective the spatial depth of Chaudhuri (1996), Vardi and Zhang (2000) and Serfling (2002) to the functional setting, which has proved to be a useful construction. For a generalisation of some of these depths, see Mosler (2013).

4.2 A Property-Wise Analysis of Existing Functional Depths

In the theoretical results that follow, D_h, D_{RT}, D_J, D_{MJ}, D_{HR}, D_{MHR} and their respective $(\mathfrak{F}, d)$ are as in Table 1, and $\mathcal{D} := \{D_h, D_{\mathrm{RT}}, D_J, D_{\mathrm{MJ}}, D_{\mathrm{HR}}, D_{\mathrm{MHR}}\}$. The conclusions of the following theorems are summarised in Table 2. We comment here on reasons for which the different examples of depth satisfy, or fail to satisfy, the corresponding properties. For a deeper insight, see the proofs in Section 5.

THEOREM 4.1 (Property P-1. Distance invariance). *All elements of $\mathcal{D}$ satisfy property P-1 with the exception of D_h.*

The part of the proof of Theorem 4.1 concerning the h depth assumes that the same h is used in $D_h(x, P_X)$ and $D_h(f(x), P_{f(X)})$, but the conclusion remains valid if we allow for h to depend on f. To see it, simply observe that $\frac{1}{h}\exp\{-\|x - X\|^2/2h^2\} \neq \frac{1}{h_f}\exp\{-a\|x - X\|^2/2h_f^2\}$ for any $h > 0, h_f > 0$.

Recall from our discussion of P-2 that, since there is no unique measure of centre of symmetry, θ, in general, it is more meaningful to consider the behaviour of D for a particular case of P in which all standard notions of centre of symmetry coincide at θ. We thus consider here adherence to P-2G.

TABLE 2
Adherence of existing depth constructions to depth defining properties

	P-1	P-2G	P-3	P-4	P-5	P-6
D_h	✗	✓	✓	✓	✓	✓
D_{RT}	✓	✓	✗	✓	✗	✓
D_J	✓	✓	✗	✓	✗	✓
D_{MJ}	✓	✓	✗	✓	✗	✓
D_{HR}	✓	✗	✗	✓	✗	✓
D_{MHR}	✓	✓	✗	✓	✗	✓

THEOREM 4.2 (Property P-2G, maximality at Gaussian process mean). *With the exception of D_{HR}, all elements of $\mathcal{D}$ satisfy property P-2G, where $J \geq 3$ in D_J.*

The intuitive explanation for D_{HR} failing to satisfy P2-G is that the expected number of upcrossings of a mean zero Gaussian process above a level a is strictly decreasing in $|a|$. Hence, the probability that a Gaussian process is either entirely above or entirely below a is strictly increasing in $|a|$. The modified version of D_{HR} does not suffer this drawback, as it takes account of the duration of excursions above $|a|$.

For sufficiently small h, the h-depth becomes a local depth rather than a global depth and, hence, as alluded to in the discussion in Section 3.1.3, one would not expect a centre outward ordering from a unique centre of symmetry, but rather an outward ordering from points of high local depth. As such, verification of P-3 is only achievable when h is sufficiently large for D_h to constitute a global depth. We implicitly impose this assumption in Lemma 4.3 below by imposing that the deepest element (as measured by D_h) exists and coincides with the mean.

LEMMA 4.3. *Provided that $\mathbb{E}X$ exists and $D_h(\mathbb{E}X, P) = \sup_{x\in\mathfrak{F}} D_h(x, P)$, $D = D_h$ satisfies* P-3.

Lemma 4.3 works for any type of distribution, including both continuous and discrete. However, the counterexamples in the proof of Theorem 4.4 demonstrate that noncontinuous distributions preclude adherence to P-3 for elements of $\mathcal{D}\setminus\{D_h\}$. The constructions of these depths are based more directly on terms of the form $P(B_x)$ for B_x a Borel set that depends on $x \in \mathfrak{F}$. For noncontinuous distributions and the constructions we consider, there exist $x, y \in \mathfrak{F}$ with $x \neq y$ that yield $P(B_x) = P(B_y)$, resulting in the assignment of equal depths to x and y.

THEOREM 4.4 (Property P-3, strictly decreasing w.r.t. the deepest point). *The elements of $\mathcal{D} \setminus \{D_h\}$ do not satisfy property* P-3.

Lemma 4.5, as well as being of independent interest, is used in the proof of Theorem 4.6.

LEMMA 4.5. *For any $P \in \mathcal{P}$, $D_h(x, P)$ is continuous in x.*

THEOREM 4.6 (Property P-4, upper semi-continuity in x). *All elements of $\mathcal{D}$ satisfy property* P-4.

Upper semi-continuity of the elements of $\mathcal{D}$ arises naturally because all depth constructions preserve the upper semi-continuity of the distribution function induced by P. A stricter requirement of continuity would, in most cases, rule out the possibility of P with finite support.

THEOREM 4.7 (Property P-5, receptivity to convex hull width across the domain). *Provided that* $\mathbb{E}X$ *exists*, $D = D_h$ *satisfies* P-5. *The elements of* $\mathcal{D} \setminus \{D_h\}$ *do not satisfy property* P-5.

The intuition behind the nonadherence of the elements of $\mathcal{D} \setminus D_h$ to P-5 is that their constructions all result in an assignment of rank, neglecting the relative distances (as measured in some suitable metric, d, with respect to P) between elements of $\mathfrak{F}$. By contrast, the h-depth is essentially a weighted $\mathbb{L}_2(\mathcal{V}, \lambda)$, where the weights depend on P. As such, it is able to appropriately exploit the information contained in P such that the influence of variations in X over L_δ is commensurate with δ.

THEOREM 4.8 (Property P-6, continuity in P). *All elements of* $\mathcal{D} \setminus \{D_J, D_{\mathrm{RT}}\}$ *satisfy property* P-6. D_{RT} *satisfies* P-6 *when the limiting distribution is continuous or the sequence of distributions is the sequence of empirical distributions*. D_J *satisfies* P-6 *when* $\mathfrak{F}$ *is restricted to be the space of equicontinuous functions on* $\mathcal{V} \subset \mathbb{R}$.

All elements of $\mathcal{D} \setminus \{D_J, D_{\mathrm{MJ}}\}$ are either constructed from sets of the form $P(B_x)$ for B_x a Borel set that depends on $x \in \mathfrak{F}$ or as an integral of a bounded Lipschitz function with respect to P, which yields adherence to P-6 by the well-known Portmanteau theorem for weak convergence (cf. Section 5 for details). The construction of D_J and D_{MJ} results in a stochastic process whose behaviour is governed by P. As is shown in Section 5, convergence of Q to P guarantees weak convergence of the respective stochastic processes, which in turn results in pointwise P-a.s. convergence of depths.

Amongst the six constructions we consider, the h-depth satisfies 5 of the 6 properties we seek. This should not be interpreted as a recommendation to favour the h-depth. As discussed in Section 3.2, each property has different implications for different application areas and a depth construction should thus be chosen with the application in mind. As the h-depth fails to satisfy P-1, a proposal is to substitute the proposed kernel. As a simple illustration, if the kernel function resulted in $D_h(x, P) := \frac{1}{\sqrt{2\pi}} \exp\{-\|x - X\|^2/2h^2\}$, property P-1 would be satisfied when allowing h to depend on f, where f is defined in Definition 3.2.

5. PROOFS

PROOF OF LEMMA 3.4. For any $x, z \in \mathfrak{F}$, $(d(x, z))^p \leq (\sup_{v\in\mathcal{V}} \|x(v) - z(v)\|)^p \lambda(\mathcal{V})$. Fixing z, as $\lambda(\mathcal{V})$ is finite, $d(x, z) \to \infty$ implies that

$$\sup_{v\in\mathcal{V}} \|x(v)\| \to \infty.$$

Thus, $D(x, P) \to \inf_{x\in\mathfrak{F}} D(x, \mathfrak{F})$ as $\sup_{v\in\mathcal{V}} \|x(v)\| \to \infty$ and, a fortiori, as $\|x(v)\| \to \infty$ for Lebesgue almost every $v \in \mathcal{V}$. □

PROOF OF LEMMA 3.5. Suppose for a contradiction that there exist $z_1, z_2 \in \mathfrak{F}$ $z_1 \neq z_2$ such that $D(z_1, P) = D(z_2, P) = \max_{x\in\mathfrak{F}} D(x, P)$. As $z_1 \neq z_2$ implies $d(z_1, z_2) > 0$, we may take in the statement of P-3 $x = z_1$ and $z = z_2$, which yields by P-3 $D(z_1, P) < D(z_2, P)$, a contradiction. □

PROOF OF THEOREM 4.1 (PROPERTY P-1). *h-depth*. When $(\mathfrak{F}, d) = (\mathbb{H}, \|\cdot\|_{\mathbb{L}_2})$, the set of functions that satisfy $d(f(x), f(y)) = a_f \cdot d(x, y)$ for any $x, y \in \mathfrak{F}$ is given by

$$\{f : f(x(v)) = \sqrt{a(v)}x(v), a(v) = a_f > 0 \quad \forall v \in \mathcal{V}\}. \tag{5.1}$$

Since $K_h(a\|x - X\|) \neq K_h(\|x - X\|)$ for all $a \neq 1$, there exist functions in the set (5.1) for which $D_h(x, P_X) \neq D_h(f(x), P_{f(X)})$. *Random Tukey depth*: Let $(\mathfrak{F}, d) = (\mathbb{H}, |\langle\cdot, \cdot\rangle|)$, then the set of functions that satisfy $d(f(x), f(y)) = a_f \cdot d(x, y)$ for any $x, y \in \mathfrak{F}$ is given by equation (5.1). The result follows since $\{y : \langle u, x - y\rangle \geq 0\} = \{y : \langle u, \sqrt{a}y - \sqrt{a}x\rangle \geq 0\}$ for all $v \in \mathbb{H}$.

For D_J, D_{MJ}, D_{HR} and D_{MHR}, letting $(\mathfrak{F}, d) = (\mathcal{C}(\mathcal{V}), \|\cdot\|_\infty)$, the set of functions satisfying $d(f(x), f(y)) = a_f \cdot d(x, y)$ for any $x, y \in \mathfrak{F}$ is given by

$$\{f : f(x(v)) = a(v)x(v) + b(v), |a(v)| = a_f > 0 \quad \forall v \in \mathcal{V}\}.$$

Then, $D(x, P_X) = D(f(x), P_{f(X)})$ for those instances of depth listed above by the following observations. *Band depth*: the result is Theorem 3 of López-Pintado and Romo (2009). *Modified band depth*: for $a_f > 0$, $x(v) \in [L_j(v), U_j(v)]$ if and only if $a_f x(v) \in [a_f L_j(v), a_f U_j(v)]$. *Half-region depth*: we have $P[X(v) \leq x(v), v \in \mathcal{V}] = P[a_f X(v) \leq a_f x(v), v \in$

$\mathcal{V}]$ and analogously for $P[X(v) \geq x(t), v \in \mathcal{V}]$. *Modified half-region depth*: we have $E[\lambda\{v \in \mathcal{V} : x(v) \leq X(v)\}] = E[\lambda\{v \in \mathcal{V} : a_f x(v) \leq a_f X(v)\}]$ and analogously for $E[\lambda\{v \in \mathcal{V} : x(v) \geq X(v)\}]$. □

PROOF OF THEOREM 4.2 (PROPERTY P-2G). *h-depth*. Suppose for a contradiction that

$$z := \operatorname*{argsup}_{x \in \mathfrak{F}} D_h(x, P)$$

is such that $D_h(z, P_X) > D_h(E[X], P_X)$. Since

$$\begin{aligned} &\operatorname*{argsup}_x D_h(x, P_X) \\ &\quad = \operatorname*{argsup}_x \mathbb{E}\left[\exp\left\{-\frac{\|x - X\|^2}{2h^2}\right\}\right] \\ &\quad = \operatorname*{arginf}_x \mathbb{E}\big[\|x - X\|^2\big], \end{aligned} \tag{5.2}$$

the previous supposition is equivalent to $\mathbb{E}[\|z - X\|^2] < \mathbb{E}[\|\mathbb{E}[X] - X\|^2] = \mathbb{E}[\|X\|^2]$. After some algebra we obtain $\|x\|^2 < 2\int x(v)\mathbb{E}[X(v)]\,dv = 0$, a contradiction.

Random Tukey depth. For any $u \in \mathfrak{F} = \mathbb{H}$, we have that $\langle u, \mathbb{E}X\rangle$ is the mean of P_u because $\mathbb{E}[\langle X, u\rangle] = \mathbb{E}\int X(v)u(v)\,dv = \int \mathbb{E}X(v)u(v)\,dv$. Since, for P a Gaussian process, the mean of P_u coincides with the median of P_u, we have $D_1(\langle u, \mathbb{E}X\rangle, P_u) = \frac{1}{2}$. Then, by the definition of random Tukey depth, $D_{\mathrm{RT}}(\mathbb{E}X, P) = \min_{u \in \mathfrak{U}} \frac{1}{2} = \frac{1}{2}$, the maximum attainable value for the random Tukey depth, hence, $D_{\mathrm{RT}}(\mathbb{E}X, P) = \sup_{x \in \mathfrak{F}} D_{\mathrm{RT}}(x, P)$.

Band depth and modified band depth. By the definition of the band depth and the modified band depth,

$$\sup_{x \in \mathfrak{F}} D_J(x, P) \leq \sum_{j=2}^{J} \sup_{x \in \mathfrak{F}} P_{\mathcal{S}_j}\big(x \in \mathcal{S}_j(P)\big)$$

and

$$\begin{aligned} &\sup_{x \in \mathfrak{F}} D_{\mathrm{MJ}}(x, P) \\ &\quad \leq \sum_{j=2}^{J} \sup_{x \in \mathfrak{F}} \mathbb{E}\big[\lambda\big\{v \in \mathcal{V} : x(v) \in \mathcal{S}_j(v, P)\big\}/\lambda(\mathcal{V})\big]. \end{aligned}$$

Since each of $X_1, \ldots, X_J$ is a random draw from P, whose mean is $\theta = \mathbb{E}X$, and since $P_{\mathcal{S}_j}$ is a continuous distribution over simplices (because P is continuous), the x which maximises the probability of a random j-simplex enveloping it is clearly $x = \theta$, yielding $\sup_{x \in \mathfrak{F}} D_J(x, P) = D_J(\theta, P)$. Similarly, the x for which the expected duration spent in any simplex is largest is also $x = \theta$, yielding $\sup_{x \in \mathfrak{F}} D_{\mathrm{MJ}}(x, P) = D_{\mathrm{MJ}}(\theta, P)$.

Half-region depth. By Adler (1981), Theorem 4.1.1, the expected number of upcrossings of a level $\bar{u}$ of a zero-mean, stationary, almost surely continuous random process on $\mathcal{V}$ is

$$\mathbb{E}[N_{\bar{u}}] = \sqrt{\frac{-R''(0)}{R(0)}}\frac{\lambda(\mathcal{V})}{2\pi}\exp\left\{-\frac{\bar{u}^2}{2R(0)}\right\}, \tag{5.3}$$

where $R(0) = \mathbb{E}[|X(v)|^2]$ and $-R''(0)$ is the variance of $X(v)$, which is constant by stationarity of X. Equation (5.3) is maximised at $\bar{u} = 0$, hence, for any $\bar{u}$ such that $0 < |\bar{u}| < \infty$, $\min\{P(X(v) \leq \bar{u}\ \forall v \in \mathcal{V}), P(X(v) \geq \bar{u}\ \forall v \in \mathcal{V})\} > \min\{P(X(v) \leq 0\ \forall v \in \mathcal{V}), P(X(v) \geq 0\ \forall v \in \mathcal{V})\}$.

Modified half-region depth. Demonstrating that $D_{\mathrm{HR}}(x, P)$ achieves its maximum value at the zero mean function of the Gaussian process P entails a proof that the expected measure of the level zero excursion set is $1/2$, where the level zero excursion set is defined as

$$A_0 := A_0(X, \mathcal{V}) := \big\{v \in \mathcal{V} : X(v) \geq 0\big\}.$$

By Rice (1945), from which equation (5.3) also originally derived, the expected length of an excursion above zero is $\pi\sqrt{R(0)/[-R''(0)]}$. Recalling that $\mathcal{V}$ is a compact subset of $\mathbb{R}$ and assuming an excursion starts at $\min\{v \in \mathcal{V}\}$, we thus have, using equation (5.3),

$$\mathbb{E}\big[\lambda(A_0)\big] = \frac{\lambda(\mathcal{V})}{2}\sqrt{\frac{-R''(0)}{R(0)}}\sqrt{\frac{R(0)}{-R''(0)}} = \frac{\lambda(\mathcal{V})}{2}.$$

Hence, $D_{\mathrm{MHR}}(\mathbb{E}X, P) = 1/2$, which coincides with $\sup_{x \in \mathfrak{F}} D(x, P)$. □

PROOF OF LEMMA 4.3 (PROPERTY P-3, *h*-DEPTH). Observe that D_h is translation invariant, that is, for any $x, b \in \mathfrak{F}$ and $P_X \in \mathcal{P}$

$$\begin{aligned} &D_h(x, P_X) \\ &\quad = \mathbb{E}\left[\frac{1}{h\sqrt{2\pi}}\exp\left\{-\frac{\|x - X\|^2}{2h^2}\right\}\right] \\ &\quad = \mathbb{E}\left[\frac{1}{h\sqrt{2\pi}}\exp\left\{-\frac{\|(x - b) - (X - b)\|^2}{2h^2}\right\}\right]. \end{aligned}$$

Thus, set $\mathbb{E}[X] = 0$ without loss of generality.

Suppose for a contradiction $D_h(x, P) \geq D_h(y, P)$. Substituting $\|x - X\|^2 = \|x\|^2 + \|X\|^2 - 2\int x(v) \times X(v)\,dv$ in the expression for D_h gives the inequality

$$\begin{aligned} &\exp\left\{-\frac{\|x\|^2 - \|y\|^2}{2h^2}\right\} \\ &\quad \geq \mathbb{E}\left[\exp\left\{\frac{\int (y(v) - x(v))X(v)\,dv}{h^2}\right\}\right]. \end{aligned}$$

By the statement of P-3 and the fact that $\mathbb{E}[X] = 0$, we have $\|x\| > \|y\|$ and so

$$1 > \exp\left\{-\frac{\|x\|^2 - \|y\|^2}{2h^2}\right\}. \tag{5.4}$$

On the other hand, by Jensen's inequality,

$$\mathbb{E}\left[\exp\left\{\frac{\int (y(v) - x(v))X(v)\,dv}{h^2}\right\}\right] \geq \exp\left\{\frac{\int (y(v) - x(v))\mathbb{E}[X](v)\,dv}{h^2}\right\},$$

which is equal to 1 because $\mathbb{E}[X] = 0$. This together with (5.4) yields the contradiction

$$1 > \exp\left\{-\frac{\|x\|^2 - \|y\|^2}{2h^2}\right\} \geq 1. \qquad \square$$

PROOF OF THEOREM 4.4 (PROPERTY P-3). *Random Tukey depth*. The proof is by counterexample. Let $P \in \mathcal{P}$ be a discrete distribution with support $\{x_1, x_2\}$ with $x_1(v) = 2$ for all $v \in \mathcal{V}$ and $x_2(v) = -1$ for all $v \in \mathcal{V}$. Let $u \in \mathbb{H}$ be an arbitrary realisation of the random variable U whose distribution is μ. The inner product with u of any $y \in \mathcal{Y} := \{y(v) = c \ \forall v \in \mathcal{V} \text{ with } c \in (-1, 2)\}$ gives rise to $\langle u, y\rangle \in (\min\{\langle u, x_1\rangle, \langle u, x_2\rangle\}, \max\{\langle u, x_1\rangle, \langle u, x_2\rangle\})$. It follows that $D_{\mathrm{RT}}(y, P) = \max_{x\in\mathfrak{F}} D_{\mathrm{RT}}(x, P)$ for any y in the closure of $\mathcal{Y}$, which contradicts Lemma 3.5.

Band depth. The proof is by counterexample. Take $P \in \mathcal{P}$ discrete with $P(\{x_1\}) = P(\{x_2\}) = 1/2$, where $x_1(v) = -c$ for all $v \in \mathcal{V}$, $x_2(v) = c$ for all $v \in \mathcal{V}$. Then $P_{\mathcal{S}_j}$ $j = J = 2$ is discrete with $P_{\mathcal{S}_j}(\mathcal{S}_{j,1}) = P_{\mathcal{S}_j}(\mathcal{S}_{j,2}) = 1/4$ and $P_{\mathcal{S}_j}(\mathcal{S}_{j,3}) = 1/2$, where $\mathcal{S}_{j,1} = \{x_1\}$, $\mathcal{S}_{j,2} = \{x_2\}$ and $\mathcal{S}_{j,3} = \{[x_1(v), x_2(v)] : v \in \mathcal{V}\}$. Then $D_J(z, P)$ has two global maxima, at $z = x_1$ and at $z = x_2$, with $D_J(z, P) = 3/4$. Without loss of generality, set $z = x_1$. For any $x, y \in \mathfrak{F} = \mathcal{C}(\mathcal{V})$ such that $\max\{d(y, z), d(y, x)\} < d(x, z)$ and $x_2(v) < x(v) < x_1(v)$, $x_2(v) < y(v) < x_1(v)$ for all $v \in \mathcal{V}$. Then $D_J(x, P) = D_J(y, P) = 1/2$, violating P-3.

Modified band depth. The proof uses the same counterexample as in the proof for the band depth. We have

$$\begin{aligned} D_{\mathrm{MJ}}(z, P) &= \lambda\{v \in \mathcal{V} : z(v) \in \mathcal{S}_{j,1}(v, P)\} P_{\mathcal{S}_j}(\mathcal{S}_{j,1})/\lambda(\mathcal{V}) \\ &\quad + \lambda\{v \in \mathcal{V} : z(v) \in \mathcal{S}_{j,2}(v, P)\} P_{\mathcal{S}_j}(\mathcal{S}_{j,2})/\lambda(\mathcal{V}) \\ &\quad + \lambda\{v \in \mathcal{V} : z(v) \in \mathcal{S}_{j,3}(v, P)\} P_{\mathcal{S}_j}(\mathcal{S}_{j,3})/\lambda(\mathcal{V}), \end{aligned}$$

and $D_{\mathrm{MJ}}(z, P)$ is maximised at $z = x_1$ and $z = x_2$, giving $D_{\mathrm{MJ}}(z, P) = 3/4$. Without loss of generality, set $z = x_1$. For any $x, y \in \mathfrak{F} = \mathcal{C}(\mathcal{V})$ such that $\max\{d(y, z), d(y, x)\} < d(x, z)$ and $x_2(v) < x(v) < x_1(v)$, $x_2(v) < y(v) < x_1(v)$ for all $v \in \mathcal{V}$. Then $D_{\mathrm{MJ}}(x, P) = D_{\mathrm{MJ}}(y, P) = 1/2$, violating P-3.

Half-region depth. Let P, x and y be as for the (modified) band depth. Then $D(z, P) = P(X(v) \geq z(v), v \in \mathcal{V}) = P(X(v) \leq z(v), v \in \mathcal{V})$. But $P(X(v) \geq x(v), v \in \mathcal{V}) = P(X(v) \geq z(v), v \in \mathcal{V}) = P(X(v) \geq y(v), v \in \mathcal{V})$, hence $D_{\mathrm{HR}}(x, P) = D_{\mathrm{HR}}(y, P) = D_{\mathrm{HR}}(z, P)$ despite the fact that $d(y, z) < d(x, z)$.

Modified half-region depth. Let P, x and y be as for the (modified) band depth. Then for any $\omega \in \Omega$, $\lambda\{v \in \mathcal{V} : X(\omega, v) \leq x(v)\} = \lambda\{v \in \mathcal{V} : X(\omega, v) \leq y(v)\}$ and likewise for the converse inequality. Hence, $D_{\mathrm{MHR}}(x, P) = D_{\mathrm{MHR}}(y, P)$ despite the fact that $d(y, z) < d(x, z)$. $\square$

PROOF OF LEMMA 4.5. Write $\exp\{-\|x - X(\omega)\|/2h\}/\sqrt{2\pi h} =: F(x, \omega)$. Then for P-almost every $\omega \in \Omega$, $F(\cdot, \omega)$ is continuous at x. Moreover, since $\exp\{-z\}$ is bounded on $z \in \mathbb{R}^+$, there exists a P-integrable function $g(\omega)$ such that $F(y, \omega) \leq g(\omega)$ for P-almost every $\omega \in \Omega$ and all y in a neighbourhood of x. Since the above holds for all $x \in \mathfrak{F}$, it follows by Theorem 7.43 of Shapiro, Dentcheva and Ruszczyński (2009) that $\mathbb{E}[\exp\{-\| \cdot - X(\omega)\|/2h\}/\sqrt{2\pi h}]$ is continuous at x for all $x \in \mathfrak{F}$. $\square$

PROOF OF THEOREM 4.6 (PROPERTY P-4). *h-depth*. By Lemma 4.5, D_h is continuous in x so a fortiori, it is upper semicontinuous.

Random Tukey depth. The case of $D_{\mathrm{RT}}(y, P) \leq D_{\mathrm{RT}}(x, P)$ is trivial. When $D_{\mathrm{RT}}(y, P) > D_{\mathrm{RT}}(x, P)$, the condition in (3.1) is

$$\begin{aligned} &\sup_{y:\|y-x\|<\delta} \min_{u\in\mathfrak{U}} D_1(\langle u, y\rangle, P_v) \\ &\quad \leq \min_{u\in\mathfrak{U}} D_1(\langle u, x\rangle, P_u) + \varepsilon. \end{aligned} \tag{5.5}$$

We verify the existence of a δ satisfying (5.5) for all

$$0 < \varepsilon \leq 1/2 - D_{\mathrm{RT}}(x, P). \tag{5.6}$$

Note that if $D_{\mathrm{RT}}(x, P) \geq 1/2$, we are in the case of $D_{\mathrm{RT}}(y, P) \leq D_{\mathrm{RT}}(x, P)$. For the less interesting scenario in which $\varepsilon > 1/2 - D_{\mathrm{RT}}(x, P)$, the construction of δ satisfying (5.5) is more involved. Let $u \in \mathfrak{U}$ such that $D_{\mathrm{RT}}(x, P) = D_1(\langle u, x\rangle, P_u)$, and notice that $D_{\mathrm{RT}}(y, P) \leq D_1(\langle u, y\rangle, P_u)$ for all $u \in \mathfrak{U}$. Additionally, $D_{\mathrm{RT}}(y, P) > D_{\mathrm{RT}}(x, P) = D_1(\langle u, x\rangle, P_u)$ implies $D_1(\langle u, y\rangle, P_u) > D_{\mathrm{RT}}(x, P)$. For ε satisfying (5.6), $D_{\mathrm{RT}}(x, P) = P_u(-\infty, \langle u, x\rangle]$ implies $D_1(\langle u, y\rangle, P_u) = P_u(-\infty, \langle u, y\rangle]$ and, analogously, $D_{\mathrm{RT}}(x, P) = P_u[\langle u, x\rangle, -\infty)$ that $D_1(\langle u, y\rangle, P_u) = [\langle u, y\rangle, -\infty)$.

With these observations, we see that (5.5) is achieved with $\delta < \sup\{\eta > 0 : P(B(\eta)) \le \varepsilon\}$, where

$$B(\eta) := \big\{y \in \mathfrak{F} : D_{\mathrm{RT}}(y, P) > D_{\mathrm{RT}}(x, P) = D_1\big(\langle u, x\rangle, P_u\big), \big|\langle u, y - x\rangle\big| < \eta\big\}.$$

Band depth and half-region depth. López-Pintado and Romo (2009) (Theorem 3) and López-Pintado and Romo (2011) (Proposition 6) prove that for all $x \in \mathfrak{F}$ and for all $\varepsilon > 0$, there exists a $\delta > 0$ such that

$$\sup_{y:|\|y\|_\infty - \|x\|_\infty| < \delta} D(y, P) \le D(x, P) + \varepsilon$$

for the respective depth constructions, $D = D_J$ and $D = D_{\mathrm{HR}}$. Since $|\|y\|_\infty - \|x\|_\infty| \le d(y, x)$, the proof is complete.

Modified band depth. The case of $D_{\mathrm{MJ}}(y, P) \le D_{\mathrm{MJ}}(x, P)$ is trivial. When $D_{\mathrm{MJ}}(y, P) > D_{\mathrm{MJ}}(x, P)$, the condition in (3.1) is

$$\sup_{y:\|x-y\|_\infty < \delta} \sum_{j=2}^{J} \mathbb{E}\big[\lambda\big\{v \in \mathcal{V} : y(v) \in \big[L_j(v), U_j(v)\big], x(v) \notin \big[L_j(v), U_j(v)\big]\big\}/\lambda(\mathcal{V})\big] \le \varepsilon, \quad J \ge 2. \tag{5.7}$$

Taking $\delta < \sup\{\eta > 0 : \sum_{j=2}^{J} \mathbb{E}[\lambda\{v \in \mathcal{V} : x(v) \notin B_j(v), \min(|x(v) - L_j(v)|, |x(v) - U_j(v)|) < \eta\}] \le \varepsilon\lambda(\mathcal{V})\}$ ensures (5.7) is satisfied.

Modified half-region depth. The case of $D_{\mathrm{MHR}}(y, P) \le D_{\mathrm{MHR}}(x, P)$ is trivial. When $D_{\mathrm{MHR}}(y, P) > D_{\mathrm{MHR}}(x, P)$, the condition in (3.1) is

$$\sup_{y:\|x-y\|_\infty < \delta} \mathbb{E}\big[\lambda\big\{v \in \mathcal{V} : y(v) \le X(v) \le x(v)\big\}\big] \le \varepsilon. \tag{5.8}$$

We verify the existence of a δ satisfying (5.8) for all $0 < \varepsilon \le 1/2 - D(x, P)$. For the less interesting case of $\varepsilon > 1/2 - D(x, P)$, the construction of δ satisfying (5.8) is more involved. Let

$$\Gamma := \big\{\eta > 0 : \mathbb{E}\big[\lambda\big\{v \in \mathcal{V} : \big(X(v) \le x(v)\big)\mathbb{1}\{x \in \mathcal{A}\}, \big(X(v) \ge x(v)\big)\mathbb{1}\{x \in \mathcal{B}\}, \big|x(v) - X(v)\big| < \eta\big\}\big] < \varepsilon\lambda(\mathcal{V})\big\},$$

where $\mathcal{A} := \{x \in \mathfrak{F} : D(x, P) = \mathbb{E}[\lambda\{v \in \mathcal{V} : x(v) \le X(v)\}]\}$ and $\mathcal{B} := \{x \in \mathfrak{F} : D(x, P) = \mathbb{E}[\lambda\{v \in \mathcal{V} : x(v) \ge X(v)\}]\}$. Then taking $\delta < \sup\{\eta \in \Gamma\}$ ensures (5.8) is satisfied. □

PROOF OF THEOREM 4.7 (PROPERTY P-5). *h-depth.* We obtain $D(f(x), P) > D(x, P)$ by simple calculation: $(\alpha(v))^2(x(v) - X(v))^2 < (x(v) - X(v))^2$ for all $v \in L_\delta$ with $\lambda(L_\delta) > 0$, hence,

$$D\big(f(x), P_{f(X)}\big) = \frac{1}{h\sqrt{2\pi}}\mathbb{E}\bigg[\exp\bigg\{-\frac{1}{2h^2}\bigg(\int_{L_\delta^c}\big(x(v) - X(v)\big)^2 dv + \int_{L_\delta}\big(\alpha(v)\big)^2\big(x(v) - X(v)\big)^2 dv\bigg)\bigg\}\bigg]$$

whilst

$$D(x, P) = \frac{1}{h\sqrt{2\pi}}\mathbb{E}\bigg[\exp\bigg\{-\frac{1}{2h^2}\bigg(\int_{L_\delta^c}\big(x(v) - X(v)\big)^2 dv + \int_{L_\delta}\big(x(v) - X(v)\big)^2 dv\bigg)\bigg\}\bigg].$$

Random Tukey depth. The proof is by counterexample. Let P be a discrete probability with $P[x_i] = 1/3$ for $i = 1, 2, 3$ and $x_1(v) > 0$, $x_2(v) = 0$ and $x_3(v) < 0$ for all $v \in \mathcal{V}$, with x_1 and x_3 nonconstant functions. Suppose for a contradiction that the following inequality is satisfied for $a = x_1$ and $a = x_3$,

$$D_{\mathrm{RT}}(a, P_X) < D_{\mathrm{RT}}\big(f(a), P_{f(X)}\big). \tag{5.9}$$

If $a = x_1$, let's denote $b = x_3$ and else, if $a = x_3$, $b = x_1$. In general, as $\langle u, x_2\rangle = \langle u, f(x_2)\rangle = 0$, in order for the inequality (5.9) to be satisfied, any given $u \in \mathfrak{U}$ has to fulfil either

$$\min\{0, \langle u, f(b)\rangle\} < \langle u, f(a)\rangle < \max\{0, \langle u, f(b)\rangle\}$$

with $\langle u, f(b)\rangle \ne 0$ or

$$\langle u, f(a)\rangle = 0 \ne \langle u, f(b)\rangle \quad \text{or} \quad \langle u, f(a)\rangle = \langle u, f(b)\rangle. \tag{5.10}$$

However, in order for the inequality (5.9) to be simultaneously satisfied by $a = x_1$ and $a = x_3$, only (5.10) can apply for each $u \in \mathfrak{U}$; but $\mu\{u : \langle u, f(x_1)\rangle = \langle u, f(x_3)\rangle\} = 0$ because, as $\alpha(v) > 0$ for all $v \in \mathcal{V}$, $f(x_1)(v) > 0$ and $f(x_3)(v) < 0$ for all $v \in \mathcal{V}$. Thus, (5.9) cannot be simultaneously satisfied by $a = x_1$ and $a = x_3$, which leads to contradiction.

Band depth, modified band depth, half-region depth and modified half-region depth. The proof is by counterexample. We follow the counterexample of the random Tukey depth but state it here for the sake of completeness. Let P be a discrete probability with $P[x_i] = 1/3$ for $i = 1, 2, 3$ and $x_1(v) > 0$, $x_2(v) = 0$ and $x_3(v) < 0$ for all $v \in \mathcal{V}$, with x_1 and x_3 nonconstant functions. As $\alpha(v) > 0$ for all $v \in \mathcal{V}$, $f(x_1)(v) > 0$,

$f(x_2)(v) = 0$ and $f(x_3)(v) < 0$ for all $v \in \mathcal{V}$. In the case of the band depth and the modified band depth, for $j \in \{2, 3\}$, the transformation simply shrinks the convex hull of any simplex over the L_δ region, whilst the probability of any simplex based on the transformation is the same as that of the original simplex to which it corresponds. It is thus immediate that $D(x_1, P_X) = D(f(x_1), P_{f(X)})$ for any $D \in \{D_J, D_{\mathrm{MJ}}, D_{\mathrm{HR}}, D_{\mathrm{MHR}}\}$. $\square$

The proof of Theorem 4.8 relies on the following definition.

DEFINITION 5.1 (e.g., van der Vaart and Wellner, 1996). For any map $\Psi : \mathbb{D} \mapsto \mathbb{K}$, with $\mathbb{D}$ and $\mathbb{K}$ normed spaces endowed with norms $\|\cdot\|_{\mathbb{D}}$ and $\|\cdot\|_{\mathbb{K}}$ respectively, the Fréchet derivative of Ψ (if it exists) is the linear continuous map $D\Psi_a : \mathbb{D} \mapsto \mathbb{K}$ such that

$$\|\Psi(a+b) - \Psi(a) - D\Psi_a(b)\|_{\mathbb{K}} = o(\|b\|_{\mathbb{D}}).$$

PROOF OF THEOREM 4.8 (PROPERTY P-6). *h-depth.* Let $d_{\mathcal{P}}$ of property P-6 be the Prohorov metric or the bounded Lipschitz metric (see, e.g., Dudley, 2002, page 394). Then by the Portmanteau theorem (see, e.g., Dudley, 2002, Theorem 11.3.3), $d_{\mathcal{P}}(Q, P) \to 0$ implies $|\int f(y)(P - Q)(dy)| \to 0$ for all $f \in BL(\mathfrak{F}, d)$, where $BL(\mathfrak{F}, d) := \{f : \mathfrak{F} \to \mathbb{R} : \|f\|_{\mathrm{BL}} < \infty\}$, $\|f\|_{\mathrm{BL}} = \|f\|_L + \|f\|_\infty$, and

$$\|f\|_L := \sup_{z \neq y} \frac{|f(y) - f(z)|}{d(y, z)}.$$

Given that, for any $x \in \mathfrak{F}$,

$$\begin{aligned}
&\big|D_h(x, P) - D_h(x, Q)\big| \\
&\quad = \left| \int K_h\big(\|x - y\|_{\mathbb{L}_2}\big) P(dy) \right. \\
&\qquad \left. - \int K_h\big(\|x - y\|_{\mathbb{L}_2}\big) Q(dy) \right|,
\end{aligned}$$

it suffices by the previous observations to show that $K_h(\|x - \cdot\|_{\mathbb{L}_2}) \in BL(\mathfrak{F}, d)$. First note

$$\begin{aligned}
&\big\|K_h\big(\|x - \cdot\|_{\mathbb{L}_2}\big)\big\|_\infty \\
&\quad = \sup_{y \in \mathfrak{F}} \big|K_h\big(\|x - y\|_{\mathbb{L}_2}\big)\big| \\
&\quad = \sup_{y \in \mathfrak{F}} \left| \frac{1}{h} \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{\|x - y\|_{\mathbb{L}_2}^2}{2h^2} \right\} \right| \\
&\quad = (h\sqrt{2\pi})^{-1} < \infty.
\end{aligned}$$

Thus, it only remains to show $\|K_h(\|x - \cdot\|_{\mathbb{L}_2})\|_L < \infty$. Taking $\Psi = K_h$, $a = x - z$, and $b = z - y$ in Definition 5.1 yields

$$\begin{aligned}
&\big(\big|K_h\big(\|x - y\|_{\mathbb{L}_2}\big) - K_h\big(\|x - z\|_{\mathbb{L}_2}\big) \\
&\qquad - DK_{h,(x-z)}(z - y)\big|\big)/\big(\|y - z\|_{\mathbb{L}_2}\big) = o(1),
\end{aligned}$$

hence, to establish

$$\sup_{z \neq y} \frac{|K_h(\|x - y\|_{\mathbb{L}_2}) - K_h(\|x - z\|_{\mathbb{L}_2})|}{\|y - z\|_{\mathbb{L}_2}} < \infty,$$

it is sufficient to show

$$\sup_{z \neq y} \frac{|DK_{h,a}(z - y)|}{\|y - z\|_{\mathbb{L}_2}} < \infty.$$

Let $\psi(\cdot) = \|\cdot\|_{\mathbb{L}_2}^2$ and $\varphi(\cdot) = \frac{1}{h} \frac{1}{\sqrt{2\pi}} \exp\{-\frac{(\cdot)}{2h^2}\}$. We can thus write $DK_{h,a}(z - y) = D_a(\varphi \circ \psi)(z - y)$, and by the chain rule of Fréchet derivatives, $D_a(\varphi \circ \psi)(b) = D_a\varphi((\psi)(b)) \circ D_a\psi(b)$. We start by computing $D_a\psi(b)$. Setting $\Psi = \psi$ in Definition 5.1 gives $|\langle a+b, a+b\rangle - \langle a, a\rangle - D\psi_a(b)| = o(\|b\|)$ and noticing that $|\langle a + b, a + b\rangle - \langle a, a\rangle - 2\langle a, b\rangle| = \langle b, b\rangle = \|b\|_{\mathbb{L}_2}^2 = o(\|b\|_{\mathbb{L}_2})$, we conclude $D\psi_a(b) = 2\langle a, b\rangle = 2\langle x - z, z - y\rangle$.

For an arbitrary $s \in \mathfrak{F}$, set $w = (\psi)(s)$, which belongs to $\mathbb{R}^+$, thus,

$$D_a\varphi(w) = -\frac{1}{2h^3} \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{w}{2h^2} \right\}.$$

The chain rule delivers

$$\begin{aligned}
&DK_{h,a}(z - y) \\
&\quad = D_a(\varphi \circ \psi)(z - y) \\
&\quad = -\frac{1}{h^3} \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{\|y - z\|_{\mathbb{L}_2}^2}{2h^2} \right\} \langle x - z, z - y\rangle,
\end{aligned}$$

hence,

$$\begin{aligned}
&\sup_{z \neq y} \frac{|DK_{h,a}(z - y)|}{\|y - z\|_{\mathbb{L}_2}} \\
&\quad = \sup_{z \neq y} \frac{|\frac{1}{h^3} \frac{1}{\sqrt{2\pi}} \exp\{-\frac{\|z-y\|_{\mathbb{L}_2}^2}{2h^2}\} \langle x - z, z - y\rangle|}{\|y - z\|_{\mathbb{L}_2}} \\
&\quad \leq \sup_{z \neq y} \frac{1}{h^3} \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{\|y - z\|_{\mathbb{L}_2}^2}{2h^2} \right\} \\
&\qquad \cdot \max\left\{ \|x\|_{\mathbb{L}_2}^2, \|y\|_{\mathbb{L}_2}^2, \|z\|_{\mathbb{L}_2}^2 \right\} < \infty
\end{aligned}$$

because $x, y, z \in \mathfrak{F} = \mathbb{L}_2$ implies they each have finite $\mathbb{L}_2$ norm.

Random Tukey depth. $d_{\mathcal{P}}(Q, P) \to 0$ P-a.s. for any metric $d_{\mathcal{P}}(\cdot,\cdot)$ metricising the topology of weak convergence is equivalent to $Q \to P$ P-a.s., which in turn implies $Q_u \to P_u$ P-a.s. for all $u \in \mathbb{H}$. As P is continuous and u is drawn with a nondegenerate stationary Gaussian measure, P_u is also continuous. It follows that

$$\max\{\big|P_u\big(-\infty, \langle u, x\rangle\big] - Q_u\big(-\infty, \langle u, x\rangle\big]\big|, \big|P_u\big[\langle u, x\rangle, \infty\big) - Q_u\big[\langle u, x\rangle, \infty\big)\big|\} \to 0 \quad P\text{-a.s.}$$

and, consequently, $|D_1(\langle u, x\rangle, P_u) - D_1(\langle u, x\rangle, Q_u)| \to 0$ P-a.s. for any $u \in \mathbb{H}$. Then

$$\begin{aligned}&\big|D_{\mathrm{RT}}(x, P) - D_{\mathrm{RT}}(x, Q)\big| \\ &\quad= \Big|\min_{u\in\mathfrak{U}} D_1\big(\langle u, x\rangle, P_u\big) - \min_{u\in\mathfrak{U}} D_1\big(\langle u, x\rangle, Q_u\big)\Big| \\ &\quad\le \max_{u\in\mathfrak{U}}\big|D_1\big(\langle u, x\rangle, P_u\big) - D_1\big(\langle u, x\rangle, Q_u\big)\big| \\ &\quad\to 0 \quad P\text{-a.s.},\end{aligned}$$

where the inequality follows because, for any $w \in \mathfrak{U}$, $\min_{u\in\mathfrak{U}} D_1(\langle u, x\rangle, P_u) \le D_1(\langle w, x\rangle, P_w)$, and likewise for Q. The empirical case follows from the proof of Theorem 2.10 in Cuesta-Albertos and Nieto-Reyes (2008).

Band depth. Since $d_{\mathcal{P}}(P, Q)$ metricises the weak topology, $d_{\mathcal{P}}(P, Q) < \delta \to 0$ is the same as writing $X_\delta \rightsquigarrow Y$ as $\delta \to 0$, where $\rightsquigarrow$ denotes weak convergence and X_δ and Y are random variables $X_\delta : \Omega \to \mathfrak{F}$ and $Y : \Omega \to \mathfrak{F}$ such that, for any $A \in \mathcal{A}$, $P(A) = \mathbb{P}(X_\delta^{-1}(A))$ and $Q(A) = \mathbb{P}(Y^{-1}(A))$, where $\mathbb{P}$ is a probability on the underlying sample space Ω. By the Portmanteau theorem (e.g., Dudley, 2002, Theorem 11.3.3), $V_N \to_d V$ if and only if $\mathbb{E}f(V_N) \to \mathbb{E}f(V)$ for all bounded Lipschitz functions f. Define $X_{\delta,1}, \ldots, X_{\delta,J}$ to be i.i.d. copies of X_δ and $Y_1, \ldots, Y_J$ to be i.i.d. copies of Y. Then, by the Portmanteau theorem, for any $\ell \in \{1, \ldots, j\}$ where $j \in \{2, \ldots, J\}$ and for any $(\alpha_1, \ldots, \alpha_j) \in \Delta^j$, since f is bounded and continuous, there exists a $\delta < \delta_\ell$ such that

$$\begin{aligned}&\left|\mathbb{E}\left[f\left(\left[\sum_{k\neq\ell}\alpha_k X_{\delta,k}\right] + \alpha_\ell X_{\delta,\ell}\right)\right]\right. \\ &\qquad\left. - \mathbb{E}\left[f\left(\left[\sum_{k\neq\ell}\alpha_k X_{\delta,k}\right] + \alpha_\ell Y_\ell\right)\right]\right| < \delta/j.\end{aligned}$$

Hence,

$$\begin{aligned}&\left|\mathbb{E}\left[f\left(\sum_{k=1}^{j}\alpha_k X_{\delta,k}\right)\right] - \mathbb{E}\left[f\left(\sum_{k=1}^{j}\alpha_k Y_k\right)\right]\right| \\ &\quad\le \sum_{\ell=1}^{j}\left|\mathbb{E}\left[f\left(\sum_{k\neq\ell}\alpha_k X_{\delta,k} + \alpha_\ell X_{\delta,\ell}\right)\right]\right. \\ &\qquad\left. - \mathbb{E}\left[f\left(\sum_{k\neq\ell}\alpha_k X_{\delta,k} + \alpha_\ell Y_\ell\right)\right]\right| < \delta\end{aligned}$$

for all $\delta < \min\{\delta_\ell : \ell \in \{1, \ldots, j\}\}$. Letting

$$Z_{X(\delta),j}(\boldsymbol{\alpha}) := \sum_{k=1}^{j}\alpha_k X_{\delta,k}$$

and

$$Z_{Y,j}(\boldsymbol{\alpha}) := \sum_{k=1}^{j}\alpha_k Y_k,$$

we conclude through a second application of the Portmanteau theorem that $Z_{X(\delta),j}(\boldsymbol{\alpha}) \to_d Z_{Y,j}(\boldsymbol{\alpha})$ for any $j \in \{2, \ldots, J\}$ and any $\boldsymbol{\alpha} \in \Delta^j$. Hence, for every finite collection $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_\ell$ where $\boldsymbol{\alpha}_k \in \Delta^j$ for each $k \in \{1, \ldots, \ell\}$, $(Z_{X(\delta),j}(\boldsymbol{\alpha}_1), \ldots, Z_{X(\delta),j}(\boldsymbol{\alpha}_\ell)) \rightsquigarrow (Z_{Y,j}(\boldsymbol{\alpha}_1), \ldots, Z_{Y,j}(\boldsymbol{\alpha}_\ell))$. Here $(Z_{X(\delta),j}(\boldsymbol{\alpha}_1), \ldots, Z_{X(\delta),j}(\boldsymbol{\alpha}_\ell))$ is an arbitrary finite set of marginals (in the $\boldsymbol{\alpha}$ index) $Z_{X(\delta),j}(\boldsymbol{\alpha}) : \Omega^j \to \mathfrak{F}$ of the stochastic process $Z_{X(\delta),j} := \{Z_{X(\delta),j}(\boldsymbol{\alpha}) : \boldsymbol{\alpha} \in \Delta^j\}$ which is the map $Z_{X(\delta),j} : \Omega^j \to \mathfrak{F}(\Delta^j) = C(\mathcal{V}, \Delta^j) \subset \mathbb{L}^\infty(\mathcal{V} \times \Delta^j)$, where $\mathbb{L}^\infty(\mathcal{V} \times \Delta^j)$ is the space of bounded functions from $(\mathcal{V} \times \Delta^j)$ to $\mathbb{R}$. Similarly, $(Z_{Y,j}(\boldsymbol{\alpha}_1), \ldots, Z_{Y,j}(\boldsymbol{\alpha}_\ell))$ is an arbitrary finite set of marginals of the stochastic process $Z_{Y,j} := \{Z_{Y,j}(\boldsymbol{\alpha}) : \boldsymbol{\alpha} \in \Delta^j\}$. Hence, in order to show that $Z_{X(\delta),j} \rightsquigarrow Z_{Y,j}$ for every $j \in \{2, \ldots, J\}$, it only remains by Theorem 1.5.4 of van der Vaart and Wellner (1996) to show that, for any $j \in \{2, \ldots, J\}$, $Z_{X(\delta),j}$ is asymptotically tight, that is, for every $\xi > 0$ there exists a compact set K such that $\liminf_{\delta\to 0} P_{Z(\delta),j}(Z_{X(\delta),j} \in K^\eta) \le 1 - \xi$ for every $\eta > 0$, where $P_{Z(\delta),j}$ is defined at every $A \in \mathcal{A}$ by $P_{Z(\delta),j}(A) = \mathbb{P}^j(Z_{X(\delta),j}^{-1}(A))$.

By Theorem 1.5.7 of van der Vaart and Wellner (1996), $Z_{X(\delta),j}$ is asymptotically tight if and only if $Z_{X(\delta),j}(v, \boldsymbol{\alpha})$ is tight in $\mathbb{R}$ for every $w = (v, \boldsymbol{\alpha})$, and there exists a semimetric d_w on $\mathcal{W} = (\mathcal{V} \times \Delta^j)$ such that $(\mathcal{W}, d_w)$ is totally bounded and $Z_{X(\delta),j}$ is uniformly d_w-equicontinuous in probability, that is, for every $\kappa, \varsigma > 0$ there exists a γ such that

$$\limsup_{\delta\to 0} P_{Z(\delta),j}\Big(\sup_{w,w':d_w(w,w')<\gamma}\big|Z_{X(\delta),j}(w) - Z_{X(\delta),j}(w')\big| > \kappa\Big) < \varsigma.$$

Tightness of $Z_{X(\delta),j}(v, \boldsymbol{\alpha})$ holds by completeness of $\mathfrak{F}$, which gives rise to tightness of X_δ and hence $Z_{X(\delta),j}$

because tightness is preserved under convex combinations. Since $\mathcal{V}$ is compact, so too is $\mathcal{W}$, hence, $(\mathcal{W}, d_w)$ is totally bounded with respect to the ℓ_1 norm. We have

$$\begin{aligned}
&\Pr\Big(\sup_{w,w':d_w(w,w')<\gamma} |Z_{X(\delta),j}(w) - Z_{X(\delta),j}(w')| > \kappa\Big) \\
&\quad\le \Pr\Big(\sup_{w,w':d_w(w,w')<\gamma} |Z_{X(\delta),j}(v,\boldsymbol{\alpha}) \\
&\qquad - Z_{X(\delta),j}(v',\boldsymbol{\alpha})| > \kappa/2\Big) \\
&\qquad + \Pr\Big(\sup_{w,w':d_w(w,w')<\gamma} |Z_{X(\delta),j}(v',\boldsymbol{\alpha}) \\
&\qquad - Z_{X(\delta),j}(v',\boldsymbol{\alpha}')| > \kappa/2\Big) = I + II.
\end{aligned}$$

By the statement of Theorem 4.8, $\mathfrak{F}$ is the space of d_w-equicontinuous functions over $\mathcal{V}$. Since convex combinations of d_w-equicontinuous functions are d_w-equicontinuous, $Z_{X(\delta),j}(\cdot,\boldsymbol{\alpha})$ is d_w-equicontinuous with probability 1. It follows that for every $\kappa, \varsigma > 0$, there exists a $\gamma > 0$ such that $I < \varsigma/2$. Noting that $v' \in \mathcal{V}$ is fixed in II, taking γ sufficiently small also gives rise to $II < \varsigma/2$, proving tightness. Asymptotic tightness is immediate because the bounds on I and II hold independently of δ.

From here we know $Z_{X(\delta),j} \rightsquigarrow Z_{Y,j}$ for every $j \in \{2,\dots,J\}$. It follows by Theorem 11.3.3 of Dudley (2002) that there exists a $\eta(\delta) \searrow 0$ as $\delta \searrow 0$ such that $\rho(P_{Z(\delta),j}, Q_{Z(Y),j}) = M < \eta(\delta)$, where $Q_{Z(Y),j}(A) = \mathbb{P}^j(Z_{Y,j}^{-1}(A))$, that is, for all $A \in \mathcal{A}$, $P_{Z(\delta),j}(A) \le Q_{Z(Y),j}(A^\xi) + \xi$ for all $\xi \in [M, \eta(\delta))$. Hence, letting $B(x) = \bigcup\{A \in \mathcal{A} : x \in A\}$, we have $P_{Z(\delta),j}(B(x)) \le Q_{Z(Y),j}(B(x)^\xi) + \xi$ for all $\xi \in [M, \eta(\delta))$ and by the symmetry of the Prohorov metric and the fact that $B(x) \subset B(x)^\xi$ for $\xi > 0$, we conclude that $|P_{Z(\delta),j}(B(x)) - Q_{Z(Y),j}(B(x))| \le \xi < \eta(\delta)$. We have

$$\begin{aligned}
&|D_J(x,P) - D_J(Q)| \\
&\quad\le \sum_{j=2}^{J} |P_{Z(\delta),j}(B(x)) - Q_{Z(Y),j}(B(x))| \\
&\quad< (J-1)\eta(\delta).
\end{aligned}$$

Setting $\varepsilon = (J-1)\eta(\delta)$, we see that the result follows by taking every δ_ℓ in the above derivations equal to $\delta = \eta^{-1}(\varepsilon/(J-1))$.

Modified band depth. Let $[z(\cdot,\boldsymbol{\alpha}) : \boldsymbol{\alpha} \in \Delta^j]$ be the set of all convex combinations of the elements of $\mathfrak{F}$, and $P_{Z(\delta),j}$ and $Q_{Y,j}$ be probability measures on that set, as defined in the proof of Theorem 4.8 for the band depth. We have

$$\begin{aligned}
&|D_{\mathrm{MJ}}(x,P) - D(x,Q)| \\
&\quad= \Big|\sum_{j=2}^{J} \frac{1}{\lambda(\mathcal{V})}\big(\mathbb{E}[\lambda\{v \in \mathcal{V} : x \in \mathcal{S}_j(v,P)\}] \\
&\qquad - \mathbb{E}[\lambda\{v \in \mathcal{V} : x \in \mathcal{S}_j(v,P)\}]\big)\Big| \\
&\quad\le \sum_{j=2}^{J} \Big|\frac{1}{\lambda(\mathcal{V})} \int \lambda\{v \in \mathcal{V} : x \in [z(v,\boldsymbol{\alpha}) : \boldsymbol{\alpha} \in \Delta^j]\} \\
&\qquad \cdot (P_{Z(\delta),j} - Q_{Y,j})(dz)\Big|.
\end{aligned}$$

But by compactness of $\mathcal{V}$, $\lambda\{v \in \mathcal{V} : x \in [z(v,\boldsymbol{\alpha}) : \boldsymbol{\alpha} \in \Delta^j]\}$ is bounded and continuous in z because $z \in \mathfrak{F}(\Delta^j) = C(\mathcal{V} \times \Delta^j)$. Hence, $|D_{\mathrm{MJ}}(x,P) - D(x,Q)| \to 0$ as $\delta \to 0$ by the Portmanteau theorem (Dudley, 2002, Theorem 11.3.3) and the fact that $P_{Z(\delta),j} \to Q_{Y,j}$ as $\delta \to 0$, as demonstrated in the proof for the band depth.

Half-region depth. Take $d_{\mathcal{P}}(P,Q) = \rho(P,Q)$, where $\rho(P,Q)$ is defined as in the proof for the band depth. Suppose $\rho(P,Q) = M < \delta$ P-a.s., where $\delta > 0$. Then for any $A \in \mathcal{A}$ and any $\eta \in [M,\delta)$, $P(A) - Q(A^\eta) \le \eta < \delta$. Let E_x denote the epigraph of x and let H_x denote the hypograph of x. $|P(E_x) - Q(E_x)| \le |P(E_x) - Q(E_x^\eta)| \le \eta < \delta$ P-a.s. and $|P(H_x) - Q(H_x)| \le |P(H_x) - Q(H_x^\eta)| \le \eta < \delta$ P-a.s., hence, $\max\{|P(E_x) - Q(E_x)|, |P(H_x) - Q(H_x)|\} < \delta$ P-a.s. It follows that, for all $\varepsilon > 0$, $|D_{\mathrm{HR}}(x,P) - D_{\mathrm{HR}}(x,Q)| < \varepsilon$ P-a.s. as long as $d_{\mathcal{P}}(P,Q) < \delta$ P-a.s. with $\delta = \varepsilon$.

Modified half-region depth. Since $(\mathfrak{F}, d) = (\mathcal{C}(\mathcal{V}), \|\cdot\|_\infty)$ is separable and complete, P and Q are tight and by Theorem 11.3.5 and Corollary 11.6.4 of Dudley (2002), $\rho(P,Q) = \alpha(X,Y)$, where X and Y are random variables with laws P and Q, respectively, ρ is the Prohorov metric defined and used throughout the proof of Theorem 4.8, and α is the Ky-Fan metric, defined by $\alpha(X,Y) := \inf\{\eta > 0 : \Pr(d(X,Y) > \eta) \le \eta\}$. Let L be an arbitrary subset of $\mathcal{V}$, and let X_L and Y_L be the random variables X and Y defined over the restricted space with corresponding probability laws P_L and Q_L respectively. Since $P \to Q$, there exists a $\delta_L > 0$ such that $\rho(P_L, Q_L) < \delta_L$, hence $\alpha(X_L, Y_L) < \delta_L$, hence $\Pr(d(X_L, Y_L) \ge \delta_L) < \delta_L$ and for any Borel set A_L of $C(L)$, if $X_L \in A_L$, then $Y_L \in A_L^{\delta_L}$, hence for any $L \subset \mathcal{V}$ and a sufficiently small δ_L, $\{X_L(v) < x(v), Y_L(v) > x(v) : v \in L\}$ and $\{X_L(v) > x(v), Y_L(v) < x(v) : v \in$

$L\}$ are events of probability zero under the joint law of X_L and Y_L. By this argument,

$$\begin{aligned}&\left|D_{\mathrm{MHR}}(x,P)-D_{\mathrm{MHR}}(x,Q)\right|\\&\quad\le\max\left\{\left|\int\lambda\{v\in\mathcal{V}:y(v)\le x(v)\}(P-Q)(dy)\right|,\right.\\&\qquad\left.\left|\int\lambda\{v\in\mathcal{V}:y(v)\ge x(v)\}(P-Q)(dy)\right|\right\}\end{aligned}$$

with probability 1. Both terms in this expression converge to zero as $\delta\to 0$ by Theorem 11.3.3 of Dudley (2002) because $\lambda\{v\in\mathcal{V}:y(v)\le x(v)\}$ and $\lambda\{v\in\mathcal{V}:y(v)\ge x(v)\}$ are continuous in y and bounded by compactness of $\mathcal{V}$. □

6. ACKNOWLEDGEMENTS

Alicia Nieto-Reyes is grateful to the School of Mathematics at the University of Bristol for kind hospitality during the 2013/2014 academic year, during which time this work was carried out. We thank Juan Cuesta-Albertos and Peter Green for their comments on a single-authored unpublished manuscript, from which the seeds of this paper originated, and Peter Hall for kindly reading the final draft. We also thank three anonymous referees for constructive suggestions.

Alicia Nieto-Reyes supported in part by the Spanish Ministerio de Ciencia e Innovación under grant MTM201128657-C0202. Heather Battey supported in part by the EPSRC under grant EP/D063485/1.

REFERENCES

ADAMS, R. A. (1975). *Sobolev Spaces*. Academic Press, New York. MR0450957

ADLER, R. J. (1981). *The Geometry of Random Fields*. Wiley, Chichester. MR0611857

AMARATUNGA, D. and CABRERA, J. (2003). *Exploration and Analysis of DNA Microarray and Protein Array Data*. Wiley, New York.

BARNETT, V. (1976). The ordering of multivariate data. *J. Roy. Statist. Soc. Ser. A* **139** 318–355. MR0445726

BEALE, N., RAND, D., BATTEY, H., CROXSON, K., NOWAK, M. A. and MAY, R. M. (2011). Individual versus systemic risk and the Regulator's Dilemma. *Proc. Natl. Acad. Sci.* **108** 12647–12652.

CHAKRABORTY, A. and CHAUDHURI, P. (2014a). The spatial distribution in infinite dimensional spaces and related quantiles and depths. *Ann. Statist.* **42** 1203–1231. MR3224286

CHAKRABORTY, A. and CHAUDHURI, P. (2014b). On data depth in infinite dimensional spaces. *Ann. Inst. Statist. Math.* **66** 303–324. MR3171407

CHAUDHURI, P. (1996). On a geometric notion of quantiles for multivariate data. *J. Amer. Statist. Assoc.* **91** 862–872. MR1395753

CHEN, D. and MÜLLER, H.-G. (2012). Nonlinear manifold representations for functional data. *Ann. Statist.* **40** 1–29. MR3013177

CHIOU, J.-M. and MÜLLER, H.-G. (2014). Linear manifold modelling of multivariate functional data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 605–626. MR3210730

CLAESKENS, G., HUBERT, M., SLAETS, L. and VAKILI, K. (2014). Multivariate functional halfspace depth. *J. Amer. Statist. Assoc.* **109** 411–423. MR3180573

CUESTA-ALBERTOS, J. A. and NIETO-REYES, A. (2008). The random Tukey depth. *Comput. Statist. Data Anal.* **52** 4979–4988. MR2526207

CUEVAS, A., FEBRERO, M. and FRAIMAN, R. (2007). Robust estimation and classification for functional data via projection-based depth notions. *Comput. Statist.* **22** 481–496. MR2336349

CUEVAS, A. and FRAIMAN, R. (2009). On depth measures and dual statistics. A methodology for dealing with general data. *J. Multivariate Anal.* **100** 753–766. MR2478196

DEVROYE, L., GYÖRFI, L. and LUGOSI, G. (1996). *A Probabilistic Theory of Pattern Recognition*. *Applications of Mathematics* (*New York*) **31**. Springer, New York. MR1383093

DUDLEY, R. M. (2002). *Real Analysis and Probability*. *Cambridge Studies in Advanced Mathematics* **74**. Cambridge Univ. Press, Cambridge. MR1932358

DUTTA, S., GHOSH, A. K. and CHAUDHURI, P. (2011). Some intriguing properties of Tukey's half-space depth. *Bernoulli* **17** 1420–1434. MR2854779

FRAIMAN, R. and MUNIZ, G. (2001). Trimmed means for functional data. *TEST* **10** 419–440. MR1881149

GENTON, M. G. and HALL, P. (2015). A tilting approach to ranking influence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* DOI:10.1111/rssb.12102. To appear.

GREEN, P. J. (1981). Peeling bivariate data. In *Interpretting Multivariate Data* (V. Barnett, ed.). Wiley, New York.

HAMPEL, F. R. (1971). A general qualitative definition of robustness. *Ann. Math. Statist.* **42** 1887–1896. MR0301858

HAMPEL, F. R., RONCHETTI, E. M., ROUSSEEUW, P. J. and STAHEL, W. A. (1986). *Robust Statistics*: *The Approach Based on Influence Functions*. Wiley, New York. MR0829458

HLUBINKA, D., GIJBELS, I., OMELKA, M. and NAGY, S. (2015). Integrated data depth for smooth functions and its application in supervised classification. *Comput. Statist.* DOI:10.1007/s00180-015-0566-x. To appear.

HUBER, P. J. (1972). The 1972 Wald lecture. Robust statistics: A review. *Ann. Math. Statist.* **43** 1041–1067. MR0314180

LI, J., CUESTA-ALBERTOS, J. A. and LIU, R. Y. (2012). *DD*-classifier: Nonparametric classification procedure based on *DD*-plot. *J. Amer. Statist. Assoc.* **107** 737–753. MR2980081

LIU, R. Y. (1990). On a notion of data depth based on random simplices. *Ann. Statist.* **18** 405–414. MR1041400

LIU, R. Y., PARELIUS, J. M. and SINGH, K. (1999). Multivariate analysis by data depth: Descriptive statistics, graphics and inference. *Ann. Statist.* **27** 783–858. MR1724033

LÓPEZ-PINTADO, S. and JORNSTEN, R. (2007). Functional analysis via extensions of the band depth. In *Complex Datasets and Inverse Problems*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **54** 103–120. IMS, Beachwood, OH. MR2459182

LÓPEZ-PINTADO, S. and ROMO, J. (2009). On the concept of depth for functional data. *J. Amer. Statist. Assoc.* **104** 718–734. MR2541590

LÓPEZ-PINTADO, S. and ROMO, J. (2011). A half-region depth for functional data. *Comput. Statist. Data Anal.* **55** 1679–1695. MR2748671

MOSLER, K. (2013). Depth statistics. In *Robustness and Complex Data Structures* 17–34. Springer, Heidelberg. MR3135871

NIETO-REYES, A. (2011). On the properties of functional depth. In *Recent Advances in Functional Data Analysis and Related Topics* (F. Ferraty, ed.). *Contrib. Statist.* 239–244. Physica-Verlag/Springer, Heidelberg. MR2815588

PAINDAVEINE, D. and VAN BEVER, G. (2013). From depth to local depth: A focus on centrality. *J. Amer. Statist. Assoc.* **108** 1105–1119. MR3174687

PAINDAVEINE, D. and VAN BEVER, G. (2015). Nonparametrically consistent depth-based classifiers. *Bernoulli* **21** 62–82. MR3322313

RICE, S. O. (1945). Mathematical analysis of random noise. *Bell Syst. Tech. J.* **24** 46–156. MR0011918

ROUSSEEUW, P. J. and RUTS, I. (1999). The depth function of a population distribution. *Metrika* **49** 213–244. MR1731769

SERFLING, R. (2002). A depth function and a scale curve based on spatial quantiles. In *Statistical Data Analysis Based on the L_1-Norm and Related Methods* (*Neuchâtel*, 2002). *Stat. Ind. Technol.* 25–38. Birkhäuser, Basel. MR2001302

SERFLING, R. (2006). Depth functions in nonparametric multivariate inference. In *Data Depth*: *Robust Multivariate Analysis, Computational Geometry and Applications*. *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.* **72** 1–16. Amer. Math. Soc., Providence, RI. MR2343109

SHAPIRO, A., DENTCHEVA, D. and RUSZCZYŃSKI, A. (2009). *Lectures on Stochastic Programming*: *Modeling and Theory*. *MPS/SIAM Series on Optimization* **9**. SIAM, Philadelphia, PA. MR2562798

SILVERMAN, B. W. (1996). Smoothed functional principal components analysis by choice of norm. *Ann. Statist.* **24** 1–24. MR1389877

SWIERENGA, H., DE WEIJER, A. P., VAN WIJK, R. J. and BUYDENS, L. M. C. (1999). Strategy for constructing robust multivariate calibration models. *Chemom. Intell. Lab. Syst.* **49** 1–17.

TUKEY, J. W. (1975). Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians* (*Vancouver, BC*, 1974), *Vol.* 2 523–531. Canad. Math. Congress, Montreal, QC. MR0426989

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York. MR1385671

VARDI, Y. and ZHANG, C.-H. (2000). The multivariate L_1-median and associated data depth. *Proc. Natl. Acad. Sci. USA* **97** 1423–1426 (electronic). MR1740461

ZUO, Y. and SERFLING, R. (2000a). On the performance of some robust nonparametric location measures relative to a general notion of multivariate symmetry. *J. Statist. Plann. Inference* **84** 55–79. MR1747497

ZUO, Y. and SERFLING, R. (2000b). General notions of statistical depth function. *Ann. Statist.* **28** 461–482. MR1790005