On data depth in infinite dimensional spaces

Anirvan Chakraborty* and Probal Chaudhuri

Theoretical Statistics and Mathematics Unit,
Indian Statistical Institute
203, B. T. Road, Kolkata - 700108, INDIA.
emails: anirvan_r@isical.ac.in, probal@isical.ac.in

Abstract

The concept of data depth leads to a center-outward ordering of multivariate data, and it has been effectively used for developing various data analytic tools. While different notions of depth were originally developed for finite dimensional data, there have been some recent attempts to develop depth functions for data in infinite dimensional spaces. In this paper, we consider some notions of depth in infinite dimensional spaces and study their properties under various stochastic models. Our analysis shows that some of the depth functions available in the literature have degenerate behaviour for some commonly used probability distributions in infinite dimensional spaces of sequences and functions. As a consequence, they are not very useful for the analysis of data satisfying such infinite dimensional probability models. However, some modified versions of those depth functions as well as an infinite dimensional extension of the spatial depth do not suffer from such degeneracy, and can be conveniently used for analyzing infinite dimensional data.

Keywords: α-mixing sequences, band depth, fractional Brownian motions, geometric Brownian motions, half-region depth, half-space depth, integrated data depth, projection depth, spatial depth

1 Introduction

In finite dimensional spaces, depth functions provide a center-outward ordering of the points in the sample space relative to a given probability distribution, and various depth functions for probability distributions in $\mathbb{R}^d$ have been proposed in the literature (see, e.g., Liu et al. (1999) and Zuo and Serfling (2000) for some extensive review). Several desirable properties of depth functions have been listed in Zuo and Serfling (2000), and these properties

*Research of A. Chakraborty is partially supported by CSIR SPM Fellowship

have been utilized in developing several statistical procedures. Depth-weighted L-type location estimators like trimmed means have been considered in Donoho and Gasko (1992), Fraiman and Muniz (2001), Mosler (2002) and Zuo (2006). Depth functions have also been used to construct statistical classifiers (see, e.g., Jörnsten (2004), Ghosh and Chaudhuri (2005), Mosler and Hoberg (2006), Dutta and Ghosh (2012) and Li, Cuesta-Albertos and Liu (2012)). Another useful application of depths is in constructing depth contours (see, e.g., Donoho and Gasko (1992) and Mosler (2002)), which determine central and outlying regions of a probability distribution. These contours and regions are useful in outlier detection.

With the recent advancement of scientific techniques and measurement devices, we increasingly come across data that have dimensions much larger than the sample sizes. Such data cannot be handled using standard multivariate techniques due to their high dimensionalities and low sample sizes. A common approach for handling such data is to embed them into suitable infinite dimensional spaces (e.g., data lying in function spaces). Half-space depth (HD) (see, e.g., Donoho and Gasko (1992)), projection depth (PD) (see, e.g., Zuo and Serfling (2000)) and spatial depth (SD) (see, e.g., Vardi and Zhang (2000) and Serfling (2002)), which were originally defined for data in finite dimensional spaces, can have natural extensions into infinite dimensional spaces as we shall see in subsequent sections.

Fraiman and Muniz (2001) defined a notion of depth, which is called integrated data depth (ID), in function spaces. Fraiman and Muniz (2001) used this depth function to construct trimmed means, and they showed that the empirical ID is a strongly and uniformly consistent estimator of its population counterpart. These authors used ID to categorize extremal and central curves in the data consisting of 100 curves used to build the NASDAQ 100 index. Recently, López-Pintado and Romo (2009, 2011) introduced two different notions of data depth for functional data, and they called them band depth (BD) and half-region depth (HRD). These authors have used these depth functions for detecting the central and the peripheral sample curves of some real datasets including daily temperature curves for Canadian weather stations and gene expression data for lymphoblastic leukemia. Trimmed means based on BD have been discussed in López-Pintado and Romo (2006), and they used it to construct classifiers based on certain distance measures. The distance of an observation from a class was defined either as the distance from the trimmed mean of the class or as a trimmed weighted average of the distances from observations in the class. The procedure was implemented to classify the well-known Berkeley growth data (see Ramsay and Silverman (2005)). López-Pintado and Romo (2009) also proposed a rank based test for two-population problems using BD, and they used the procedure to test the equality of curves obtained by plotting relative diameters along the y-axis against relative heights along the x-axis for two groups of trees as well as the Berkeley growth data. These authors proved

that the empirical versions of both of BD and HRD converge uniformly almost surely to their population counterparts. However, it was observed by them that both the depth functions tend to take small values if the sample consists of irregular (non-smooth) curves that cross one another often. To overcome this problem, López-Pintado and Romo (2009, 2011) proposed modified versions of these depth functions, called modified band depth (MBD) and modified half-region depth (MHRD), using the "proportion of time" a sample curve spends inside a band or a half-region, respectively.

It was observed in Liu (1990) that the maximum value of simplicial depth of a point in $\mathbb{R}^d$ with respect to any angularly symmetric absolutely continuous probability distribution is 2^{-d}. Consequently, the simplicial depth of any point in $\mathbb{R}^d$ under such a distribution converges to zero as d grows to infinity. This observation motivated us to critically investigate the behaviour of the above-mentioned depth functions for some standard probability models that are widely used for data in infinite dimensional spaces. It will be shown that infinite dimensional extensions of HD and PD have degenerate behaviour in infinite dimensional spaces. Moreover, both BD and HRD suffer from degeneracy for some standard probability models in function spaces. However, their modified versions as well as ID do not suffer from any such degenerate behaviour for similar probability distributions in function spaces. We also extend the notion of SD into infinite dimensional spaces, and it is shown that such an extension leads to a well-behaved and statistically useful depth function for many infinite dimensional probability distributions.

2 Depths using linear projections

In this section, we shall consider depth functions that are defined using linear projections of a random element $\mathbf{X}$. We begin by recalling that in finite dimensional spaces, the definitions of both of HD and PD involve distributions of linear projections of $\mathbf{X}$. An extension of HD into Banach spaces has been considered in Dutta et al. (2011). Consider a Banach space $\mathcal{X}$, the associated Borel σ-field, a random element $\mathbf{X} \in \mathcal{X}$ and a fixed point $\mathbf{x} \in \mathcal{X}$. The HD of $\mathbf{x}$ with respect to the distribution of $\mathbf{X}$ is defined as $HD(\mathbf{x}) = \inf\{P(\mathbf{u}(\mathbf{X}-\mathbf{x}) \geq 0) : \mathbf{u} \in \mathcal{X}^*\}$, where $\mathcal{X}^*$ denotes the dual space of $\mathcal{X}$. The PD of $\mathbf{x}$ with respect to the distribution of $\mathbf{X}$ is defined as

$$PD(\mathbf{x}) = \left[1 + \sup_{\mathbf{u}\in\mathcal{X}^*} \frac{|\mathbf{u}(\mathbf{x}) - \theta(\mathbf{u}(\mathbf{X}))|}{\sigma(\mathbf{u}(\mathbf{X}))}\right]^{-1},$$

where $\theta(.)$ and $\sigma(.)$ are some measures of location and scatter of the distribution of $\mathbf{u}(\mathbf{X})$.

If $\mathcal{X}$ is a separable Hilbert space, $\mathcal{X}$ is isometrically isomorphic to l_2, the space of all square summable sequences. In that case, $\mathcal{X} = \mathcal{X}^* = l_2$, and $\mathbf{u}(\mathbf{X})$ and $\mathbf{u}(\mathbf{x})$ in the definitions of HD and PD given above are same as $\langle \mathbf{u}, \mathbf{X}\rangle$ and $\langle \mathbf{u}, \mathbf{x}\rangle$, respectively. Here

$\langle .,. \rangle$ denotes the usual inner product in l_2. We shall first consider the space l_2 equipped with its usual norm and the associated Borel σ-field. Consider a random sequence $\mathbf{X} = (X_1, X_2, \ldots) \in l_2$ such that $\sum_{k=1}^{\infty} E(X_k^2) < \infty$, which implies $E(\mathbf{X}) = (E(X_1), E(X_2), \ldots) \in l_2$. Let us set $Y_1 = X_1 - E(X_1)$, and denote by Y_k the residual of linear regression of X_k on $(X_1, X_2, \ldots, X_{k-1})$ for $k \geq 2$. In other words, for $k \geq 2$, $Y_k = X_k - \beta_{0k} - \sum_{j=1}^{k-1} \beta_{jk} X_j$, where $\beta_{0k} + \sum_{j=1}^{k-1} \beta_{jk} X_j$ is the linear regression of X_k on $(X_1, X_2, \ldots, X_{k-1})$. Thus, $\mathbf{Y} = (Y_1, Y_2, \ldots)$ is a sequence of uncorrelated random variables with zero means. Further, since $\tau_k^2 = E(Y_k^2) \leq E(X_k^2)$ for all $k \geq 1$, we have $\sum_{k=1}^{\infty} \tau_k^2 < \infty$, and hence, $\mathbf{Y} \in l_2$ with probability one. We now state a theorem that establishes a degeneracy result for both of HD and PD under appropriate conditions on the distribution of $\mathbf{Y}$.

Theorem 2.1. *Let μ denote the probability distribution of $\mathbf{X}$ in l_2. Assume that the residual sequence $\mathbf{Y}$ obtained from $\mathbf{X}$ is α-mixing with the mixing coefficients $\{\alpha_k\}$ satisfying $\sum_{k=1}^{\infty} \alpha_k^{1-1/2p} < \infty$ for some $p \geq 1$. Further, assume that $\tau_k > 0$ for all $k \geq 1$, and $\sup_{k \geq 1} E\{(Y_k/\tau_k)^{2r}\} < \infty$ for some $r > p$. Then, $HD(\mathbf{x}) = PD(\mathbf{x}) = 0$ for all $\mathbf{x}$ in a subset of l_2 with μ-measure one. Here $HD(\mathbf{x})$ and $PD(\mathbf{x})$ denote the half-space and the projection depths of $\mathbf{x}$ with respect to μ, respectively, and in the definition of $PD(\mathbf{x})$, we choose $\theta(.)$ and $\sigma(.)$ to be the mean and the standard deviation, respectively.*

It is obvious that for any Gaussian probability measure μ, the assumptions in the preceding theorem hold. Recently, it has been shown in Dutta et al. (2011) that HD has degenerate behaviour when the probability distribution of $\mathbf{X} = (X_1, X_2, \ldots)$ is such that $X_1, X_2, \ldots$ are independent random variables satisfying suitable moment conditions. Note that if $\mathbf{X} = (X_1, X_2, \ldots)$ is a sequence of independent random variables with zero means, we have $\mathbf{Y} = \mathbf{X}$. In that case, if we choose $p = 1$ and $r = 2$, the moment assumption in the above theorem implies that $\sum_{k=1}^{\infty} E\{(X_k/\sigma_k)^4\}/k^2 < \infty$, which is the condition assumed in Theorem 3 in Dutta et al. (2011). It is worth mentioning here that the above result is actually true whenever $\sum_{k=1}^{\infty} (Y_k/\tau_k)^2 = \infty$ with probability one (see the proof in Section 6). This, for instance, holds whenever $\mathbf{Y}$ is a sequence of independent non-degenerate random variables. The moment and the mixing assumptions on $\mathbf{Y}$ stated in the theorem are only sufficient to ensure $\sum_{k=1}^{\infty} (Y_k/\tau_k)^2 = \infty$ with probability one, but by no means they are necessary.

The degeneracy of HD and PD stated in the previous theorem is not restricted to separable Hilbert spaces only. Let us consider the space $C[0,1]$ of continuous functions defined on $[0,1]$ along with its supremum norm and the associated Borel σ-field. Recall that the dual space of $C[0,1]$ is the space of finite signed Borel measures on $[0,1]$ equipped with its total variation norm. The following result shows the degeneracy of HD and PD for a class of probability measures in $C[0,1]$.

Theorem 2.2. *Consider a random element* $\mathbf{X}$ *in* $C[0,1]$ *having a Gaussian distribution with a positive definite covariance kernel, and let* μ *denote the distribution of* $\mathbf{X}$*. Then,* $HD(\mathbf{x}) = PD(\mathbf{x}) = 0$ *for all* $\mathbf{x}$ *in a subset of* $C[0,1]$ *with* μ*-measure one. Here we denote the half-space and the projection depths of* $\mathbf{x}$ *with respect to* μ *by* $HD(\mathbf{x})$ *and* $PD(\mathbf{x})$*, respectively, and we choose* $\theta(.)$ *as the mean and* $\sigma(.)$ *as the standard deviation in the definition of* $PD(\mathbf{x})$*.*

The degeneracy of HD stated in Theorems 2.1 and 2.2 can be interpreted as follows. Let $\mathcal{X}$ be either l_2 or $C[0,1]$. Then, for any $\mathbf{x} \in \mathcal{X}$, we can choose a hyperplane in $\mathcal{X}$ through $\mathbf{x}$ in such a way that the probability content of one of the half-spaces is as small as we want. On the other hand, the degeneracy result about PD in the above theorems implies that one can find an element $\mathbf{u} \in \mathcal{X}^*$ so that the distance of $\mathbf{u}(\mathbf{x})$ from the mean of $\mathbf{u}(\mathbf{X})$ relative to the standard deviation of $\mathbf{u}(\mathbf{X})$ will be as large as desired. Such degenerate behaviour of HD and PD clearly implies that they are not suitable for center-outward ordering of the points in $\mathcal{X}$, and these depth functions cannot be used to determine the central and the outlying regions for many probability measures including Gaussian distributions in $\mathcal{X}$. One reason for such degeneracy is that the dual space $\mathcal{X}^*$ is too large, and its unit ball is not compact (see also Mosler and Polyakova (2012)). It will be appropriate to note here that unlike what we have mentioned about simplicial depth in the Introduction, it is easy to verify that the maximum values of HD and PD for any symmetric probability distribution in $\mathcal{X}$ such that any linear function has a continuous distribution are 1/2 (see, e.g., Dutta et al. (2011)) and 1, respectively, and these maximum values are achieved at the centre of symmetry of the probability distribution. In other words, although HD and PD have degenerate behaviour in $\mathcal{X}$, the half-space median and the projection median remain well-defined for symmetric distributions in $\mathcal{X}$.

Let us now consider a simple classification problem, which involves class distributions in $\mathcal{X}$ ($\mathcal{X} = l_2$ or $C[0,1]$ as in the preceding paragraph), where the two classes differ only by a shift in the location. Let $\mathbf{X}$ and $\mathbf{Z}$ denote random elements from the two class distributions, where $\mathbf{Z}$ has the same distribution as $\mathbf{X} + \mathbf{c}$ for some fixed $\mathbf{c} \in \mathcal{X}$. Let us denote by $HD_{\mathbf{X}}$ and $HD_{\mathbf{Z}}$ the half-space depth functions computed using the distributions of $\mathbf{X}$ and $\mathbf{Z}$, respectively. Similarly, let $PD_{\mathbf{X}}$ and $PD_{\mathbf{Z}}$ be the projection depth functions based on the distributions of $\mathbf{X}$ and $\mathbf{Z}$, respectively. Then, under the assumptions of Theorem 2.1 or 2.2, it is easy to verify using the arguments in the proofs of those theorems (see Section 6) that $HD_{\mathbf{X}}(\mathbf{w}) = HD_{\mathbf{Z}}(\mathbf{w}) = PD_{\mathbf{X}}(\mathbf{w}) = PD_{\mathbf{Z}}(\mathbf{w}) = 0$ for almost every realization $\mathbf{w}$ of $\mathbf{X}$ and $\mathbf{Z}$. This implies that neither HD nor PD is suitable for classification purpose in the space $\mathcal{X}$ for such class distributions.

A modified version of Tukey depth, called the random Tukey depth (RTD), was proposed in Cuesta-Albertos and Nieto-Reyes (2008) for probability distributions in l_2. It is defined

as $RTD(\mathbf{x}) = \min_{1\le j\le N} \min\{P(\langle \mathbf{U}_j, \mathbf{X}\rangle \le \langle \mathbf{U}_j, \mathbf{x}\rangle), P(\langle \mathbf{U}_j, \mathbf{X}\rangle \ge \langle \mathbf{U}_j, \mathbf{x}\rangle)\}$, where $\mathbf{U}_j$'s are N i.i.d. observations from some probability distribution in l_2 independent of $\mathbf{X}$, and the probability in the definition of RTD is conditional on them. It is easy to see that the support of the distribution of $RTD(\widetilde{\mathbf{X}})$ is the whole of $[0, 1/2]$ for Gaussian and many other distributions in l_2, where $\widetilde{\mathbf{X}}$ denotes an independent copy of $\mathbf{X}$. However, Cuesta-Albertos and Nieto-Reyes (2008) mentioned some theoretical and practical difficulties with RTD including the problem of choosing N and the distribution of $\mathbf{U}_j$'s. A depth function for probability distributions in Banach spaces was introduced in Cuevas and Fraiman (2009), which is called Integrated dual depth (IDD). It is defined as $IDD(\mathbf{x}) = \int_{\mathcal{X}^*} D_{\mathbf{u}}(\mathbf{u}(\mathbf{x}))Q(d\mathbf{u})$, where $\mathbf{x} \in \mathcal{X}$, Q is a probability measure in $\mathcal{X}^*$, and $D_{\mathbf{u}}$ is a depth function defined on $\mathbb{R}$. Cuevas and Fraiman (2009) recommended that one can choose a finite number of i.i.d. random elements $\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_N$ from a probability distribution in $\mathcal{X}^*$, which will be independent of $\mathbf{X}$ and compute IDD using $N^{-1}\sum_{k=1}^{N} D_{\mathbf{U}_k}(\mathbf{U}_k(\mathbf{x}))$. It can be easily shown that if $D_{\mathbf{u}}$ is any standard depth function (e.g., HD, SD or simplicial depth) that maps $\mathbb{R}$ onto a non-degenerate interval, then for Gaussian and many other distributions of $\mathbf{X}$ in $\mathcal{X}$, $IDD(\mathbf{X})$ will have a non-degenerate distribution with an appropriate interval as its support. However, like RTD, there are no natural guidelines available in practice for choosing the probability distribution Q in the dual space $\mathcal{X}^*$ and the number N of the random directions $\mathbf{U}_j$'s.

3 Depths based on coordinate random variables

In this section, we shall discuss depths that use the underlying coordinate system of the sample space. We begin by considering BD and HRD that were discussed in the Introduction. BD and HRD of any $\mathbf{x} = \{x_t\}_{t\in[0,1]} \in C[0,1]$ with respect to the probability distribution of a random element $\mathbf{X} = \{X_t\}_{t\in[0,1]}$ in $C[0,1]$ are defined as

$$BD(\mathbf{x}) = \sum_{j=2}^{J} P\left(\min_{i=1,\ldots,j} X_{i,t} \le x_t \le \max_{i=1,\ldots,j} X_{i,t},\ \forall\, t \in [0,1]\right) \quad \text{and} \tag{1}$$

$$HRD(\mathbf{x}) = \min\{P(X_t \le x_t,\ \forall\, t \in [0,1]),\ P(X_t \ge x_t,\ \forall\, t \in [0,1])\}, \tag{2}$$

respectively. Here $\mathbf{X}_i = \{X_{i,t}\}_{t\in[0,1]}$, $i = 1, 2, \ldots, J$, denote independent copies of $\mathbf{X}$. López-Pintado and Romo (2009, 2011) defined finite dimensional versions of these two depth functions as follows. For J independent copies $\mathbf{X}_i = (X_{i,1}, X_{i,2}, \ldots, X_{i,d})$, $i = 1, 2, \ldots, J$, of $\mathbf{X} = (X_1, X_2, \ldots, X_d)$ and a fixed $\mathbf{x} = (x_1, x_2, \ldots, x_d)$,

$$BD(\mathbf{x}) = \sum_{j=2}^{J} P\left(\min_{1\le i\le j} X_{i,k} \le x_k \le \max_{1\le i\le j} X_{i,k},\ \forall\, k = 1, 2 \ldots, d\right) \quad \text{and}$$

$$HRD(\mathbf{x}) = \min\{P(X_k \le x_k,\ \forall\ \ k = 1, 2 \ldots, d),\ P(X_k \ge x_k,\ \forall\ \ k = 1, 2 \ldots, d)\},$$

respectively. The above definitions of BD and HRD in function spaces and finite dimensional Euclidean spaces lead to a natural definition of these depth functions in a sequence space. For J i.i.d. copies $\mathbf{X}_i = (X_{i,1}, X_{i,2}, \ldots)$ of an infinite random sequence $\mathbf{X} = (X_1, X_2, \ldots)$ and a fixed sequence $\mathbf{x} = (x_1, x_2, \ldots)$, we can define

$$\begin{aligned} BD(\mathbf{x}) &= \sum_{j=2}^{J} P\left(\min_{1\le i\le j} X_{i,k} \le x_k \le \max_{1\le i\le j} X_{i,k},\ \forall\ k \ge 1\right) \quad \text{and} \\ HRD(\mathbf{x}) &= \min\{P(X_k \le x_k,\ \forall\ k \ge 1),\ P(X_k \ge x_k,\ \forall\ k \ge 1)\}, \end{aligned}$$

respectively. However, as the following theorem shows, such versions of BD and HRD in sequence spaces will have degenerate behaviour for certain α-mixing sequences.

Theorem 3.1. *Let $\mathbf{X} = (X_1, X_2, \ldots)$ be an α-mixing sequence of random variables and denote the distribution of $\mathbf{X}$ by μ. Also, assume that the mixing coefficients $\{\alpha_k\}$ satisfy $\sum_{k=1}^{\infty} \alpha_k^{1-1/2p} < \infty$ for some $p \ge 1$, and the X_k's are non-atomic for each $k \ge 1$. Then, $BD(\mathbf{x}) = HRD(\mathbf{x}) = 0$ for all $\mathbf{x}$ with μ-measure one, where $BD(\mathbf{x})$ and $HRD(\mathbf{x})$ denote the band and the half-region depths of $\mathbf{x}$ with respect to μ, respectively.*

The preceding theorem implies that for i.i.d. copies of a random sequence satisfying appropriate α-mixing conditions, any given sample sequence will not lie in a band or a half-region formed by the other sample sequences with probability one. A question that now arises is whether a similar phenomenon holds for probability distributions in function spaces like $C[0,1]$. Unfortunately, as the next theorem shows, BD and HRD continue to exhibit degenerate behaviour for a well-known class of probability measures in $C[0,1]$.

Theorem 3.2. *Let $\{X_t\}_{t\in[0,1]}$ be a Feller process having continuous sample paths. Assume that for some $x_0 \in \mathbb{R}$, $P(X_0 = x_0) = 1$, and the distribution of X_t is non-atomic and symmetric about x_0 for each $t \in (0,1]$. Then, $BD(\mathbf{x}) = HRD(\mathbf{x}) = 0$ for all $\mathbf{x}$ in a set of μ-measure one, where μ denotes the probability distribution of $\mathbf{X}$, and the depth functions BD and HRD are obtained using μ.*

We refer to Revuz and Yor (1991) for an exposition on Feller processes that include Brownian motions, geometric Brownian motions and Brownian bridges. The above theorem implies that for many well-known stochastic processes, BD and HRD will be degenerate at zero. Consequently, BD and HRD will not be suitable for depth-based statistical procedures like trimming, identification of central and outlying data points, etc. for such distributions in $C[0,1]$ like HD and PD. Consider next distinct Feller processes $\mathbf{X}$ and $\mathbf{Y}$ on $C[0,1]$, and let $BD_{\mathbf{X}}$, $BD_{\mathbf{Y}}$, $HRD_{\mathbf{X}}$ and $HRD_{\mathbf{Y}}$ denote the BD's and the HRD's obtained using the distributions of $\mathbf{X}$ and $\mathbf{Y}$, respectively. Then, if both of $\mathbf{X}$ and $\mathbf{Y}$ satisfy the conditions of Theorem 3.2, using the arguments in the proofs of Lemma 6.1 and 6.2 (see Section 6), it follows that $BD_{\mathbf{X}}(\mathbf{z}) = BD_{\mathbf{Y}}(\mathbf{z}) = HRD_{\mathbf{X}}(\mathbf{z}) = HRD_{\mathbf{Y}}(\mathbf{z}) = 0$ for almost every realization

$\mathbf{z}$ of $\mathbf{X}$ and $\mathbf{Y}$. This implies that neither BD nor HRD will be able to discriminate between the distributions of $\mathbf{X}$ and $\mathbf{Y}$.

As mentioned in the Introduction, it was observed by López-Pintado and Romo (2009, 2011) that the depth functions BD and HRD tend to take small values if the sample curves intersect each other often; and this observation motivated them to consider modified versions of BD and HRD, namely MBD and MHRD, respectively. MBD and MHRD for probability distributions in $C[0,1]$, as defined by López-Pintado and Romo (2009, 2011), are given below. For a fixed $\mathbf{x} = \{x_t\}_{t\in[0,1]} \in C[0,1]$ and J i.i.d. copies $\mathbf{X}_i = \{X_{i,t}\}_{t\in[0,1]}$ of a random element $\mathbf{X} = \{X_t\}_{t\in[0,1]} \in C[0,1]$,

$$\begin{aligned} MBD(\mathbf{x}) &= \sum_{j=2}^{J} E\left[\lambda\left(\left\{t\in[0,1] : \min_{i=1,\ldots,j} X_{it} \le x_t \le \max_{i=1,\ldots,j} X_{it}\right\}\right)\right] \text{ and} \\ MHRD(\mathbf{x}) &= \min\{E[\lambda(\{t\in[0,1] : X_t \le x_t\})],\ E[\lambda(\{t\in[0,1] : X_t \ge x_t\})]\}, \end{aligned}$$

where $\lambda(.)$ is the Lebesgue measure on $[0,1]$. Fraiman and Muniz (2001) defined ID for probability measures on $C[0,1]$ as follows. For $\mathbf{x} = \{x_t\}_{t\in[a,b]} \in C[0,1]$ and a random element $\mathbf{X} = \{X_t\}_{t\in[0,1]} \in C[0,1]$, $ID(\mathbf{x}) = \int_0^1 D_t(x_t)dt$, where for every t, $D_t(.)$ denotes a univariate depth function on the real line obtained using the distribution of X_t. As observed in López-Pintado and Romo (2009), if we choose $J = 2$ in the definition of MBD, then $MBD(\mathbf{x}) = \int_0^1 2F_t(x_t)(1 - F_t(x_t))dt$, which is $ID(\mathbf{x})$ defined using the simplicial depth for each coordinate variable. Here F_t denotes the distribution of X_t for each $t \in [0,1]$. Indeed, we have the following equivalent representations of MBD and MHRD by Fubini's theorem. For any $\mathbf{x} = \{x_t\}_{t\in[0,1]} \in C[0,1]$,

$$\begin{aligned} MBD(\mathbf{x}) &= \sum_{j=2}^{J} E\left[\int_0^1 I\left(\min_{i=1,\ldots,j} X_{it} \le x_t \le \max_{i=1,\ldots,j} X_{it}\right) dt\right] \\ &= \sum_{j=2}^{J} \int_0^1 \left[1 - F_t^j(x_t-) - (1 - F_t(x_t))^j\right] dt \text{ and} \qquad (3) \\ MHRD(\mathbf{x}) &= \min\left\{E\left[\int_0^1 I(X_t \le x_t)dt\right],\ E\left[\int_0^1 I(X_t \ge x_t)dt\right]\right\} \\ &= \min\left\{\int_0^1 P(X_t \le x_t)dt,\ \int_0^1 P(X_t \ge x_t)dt\right\}. \qquad (4) \end{aligned}$$

It is easy to see from (3) that if $\mathbf{X} = \{X_t\}_{t\in[0,1]} \in C[0,1]$ is symmetrically distributed about $\mathbf{a} = \{a_t\}_{t\in[0,1]} \in C[0,1]$, i.e., $\mathbf{X} - \mathbf{a}$ and $\mathbf{a} - \mathbf{X}$ have the same distribution, then MBD has a unique maximum at $\mathbf{a}$. The same is true for ID provided that for all $t \in [0,1]$, the univariate depth D_t in the definition of ID has a unique maximum at a_t (cf. the property "FD4center" in Mosler and Polyakova (2012, p. 10), Theorems 3 and 4 in Liu (1990) and property "P2" in Zuo and Serfling (2000, p. 463)). Consider next $\mathbf{x} = \{x_t\}_{t\in[0,1]} \in C[0,1]$

and $\mathbf{y} = \{y_t\}_{t\in[0,1]} \in C[0,1]$ satisfying either $a_t \leq x_t \leq y_t$ or $y_t \leq x_t \leq a_t$ for all $t \in [0,1]$, i.e., $\mathbf{y}$ is farther away from $\mathbf{a}$ than $\mathbf{x}$. Then, $MHRD(\mathbf{y}) \leq MHRD(\mathbf{x})$ and $MBD(\mathbf{y}) \leq MBD(\mathbf{x})$. Further, if $D_t(x_t)$ is a decreasing function of $|x_t - a_t|$ for all $t \in [0,1]$, we have $ID(\mathbf{y}) \leq ID(\mathbf{x})$ (cf. the "FD4pw Monotone" property in Mosler and Polyakova (2012, p. 9)). Consider next any $\mathbf{x} = \{x_t\}_{t\in[0,1]} \in C[0,1]$ satisfying $x_t \neq 0$ for all t in a subset of $[0,1]$ with Lebesgue measure one. It follows from representations (3) and (4) for MBD and MHRD that both $MBD(\mathbf{a} + n\mathbf{x})$ and $MHRD(\mathbf{a} + n\mathbf{x})$ converge to zero as $n \to \infty$. Further, if $D_t(s) \to 0$ as $|s - a_t| \to \infty$ for all $t \in [0,1]$, then $ID(\mathbf{a} + n\mathbf{x}) \to 0$ as $n \to \infty$. So, all these depth functions tend to zero as one moves away from the center of symmetry along suitable lines. This can be viewed as a weaker version of the "FD3" property in Mosler and Polyakova (2012) (see also Theorem 1 in Liu (1990) and property "P4" in Zuo and Serfling (2000, p. 464)).

The following theorem shows that MBD, MHRD and ID have non-degenerate distributions with adequate spread for a class of probability distributions in $C[0,1]$ that includes many popular stochastic models. The properties of these depth functions discussed above and the theorem stated below show that these depth functions are suitable choices for a center-outward ordering of elements of $C[0,1]$ with respect to the distributions of a large class of stochastic processes, and can be used for constructing central and outlying regions, trimmed estimators, and also for outlier detection. Moreover, due to the continuity of ID and MBD, and the fact that they attain their unique maximum at the centre of symmetry of any probability distribution, both of these depth functions will be able to discriminate between two distributions with distinct centres of symmetry.

For the next theorem, in the definition of ID, we shall assume $D_t(.) = \psi(F_t(.))$ for all $t \in [0,1]$, where ψ is a bounded continuous positive function satisfying $\psi(0+) = \psi(1-) = 0$, and F_t denotes the distribution of Y_t.

Theorem 3.3. *Consider the process $\mathbf{X} = \{X_t\}_{t\in[0,1]} = \{g(t, Y_t)\}_{t\in[0,1]}$, where $\{Y_t\}_{t\in[0,1]} \in C[0,1]$ is a fractional Brownian motion starting at some $y_0 \in \mathbb{R}$. Assume that the function $g : [0,1] \times \mathbb{R}$ is continuous, and $g(t,.)$ is strictly increasing with $g(t,s) \to \infty$ as $s \to \infty$ for each $t \in [0,1]$. Then the following hold.*

(a) The depth functions $MBD(\mathbf{x})$, $MHRD(\mathbf{x})$ and $ID(\mathbf{x})$ take all values in $(0, A_J]$, $(0, 1/2]$ and $\psi((0,1))$, respectively, as $\mathbf{x}$ varies in $C[0,1]$, where MBD, MHRD and ID are obtained using the distribution of $\mathbf{X}$, and $A_J = J - 2 + 2^{-J+1}$ for any $J \geq 2$ with J as in the definitions of BD and MBD.

(b) The supports of the distributions of $MBD(\widetilde{\mathbf{X}})$, $MHRD(\widetilde{\mathbf{X}})$ and $ID(\widetilde{\mathbf{X}})$ are $[0, A_J]$, $[0, 1/2]$ and the closure of $\psi((0,1))$, respectively. Here $\widetilde{\mathbf{X}}$ denotes an independent copy of $\mathbf{X}$.

(c) The above conclusions hold if $\{Y_t\}_{t\in[0,1]}$ is a fractional Brownian bridge starting at

$y_0 \in \mathbb{R}$.

Note that since ψ is a continuous non-constant function, the support of the distribution of $ID(\widetilde{\mathbf{X}})$ is actually a closed non-degenerate interval. Here, by the support of a probability distribution in any metric space, we mean the smallest closed set with probability one. Let us also observe that in the above theorem, the depths are computed based on the entire process $\mathbf{X} = \{X_t\}_{t\in[0,1]}$ starting from time $t = 0$. But in practice, it might very often be the case that we observe the process from some time point $t_0 > 0$, and then the depths are to be computed based on the observed path $\{X_t\}_{t\in[t_0,1]}$. Even in that case, the conclusions of the above theorem hold (see Remark 6.4 in Section 6).

4 Spatial depth in infinite dimensional spaces

In this section, we shall consider an extension of the notion of spatial depth from $\mathbb{R}^d$ into infinite dimensional spaces. Spatial depth of $\mathbf{x} \in \mathbb{R}^d$ with respect to the probability distribution of a random vector $\mathbf{X} \in \mathbb{R}^d$ is defined as $SD(\mathbf{x}) = 1 - ||E\{(\mathbf{x} - \mathbf{X})/||\mathbf{x} - \mathbf{X}||\}||$ (see, e.g., Vardi and Zhang (2000) and Serfling (2002)). It has been widely used for various statistical procedures including clustering and classification (see, e.g., Jörnsten (2004) and Ghosh and Chaudhuri (2005)), construction of depth-based central and outlying regions and depth-based trimming (see Serfling (2006)). This depth function extends naturally to any Hilbert space $\mathcal{X}$. For an $\mathbf{x} \in \mathcal{X}$ and a random element $\mathbf{X} \in \mathcal{X}$, we can define $SD(\mathbf{x})$ using the same expression as above, where $||.||$ is to be taken as the usual norm in $\mathcal{X}$, and the expectation is in the Bochner sense (see, e.g., Araujo and Giné (1980, p. 100)). Spatial depth function inherits many of its interesting properties from finite dimensions. For instance, $SD(\mathbf{x})$ is a continuous function in $\mathbf{x}$ if the distribution of $\mathbf{X}$ is non-atomic, which is a direct consequence of dominated convergence theorem. Further, it follows from Kemperman (1987) that if $\mathcal{X}$ is strictly convex, and the distribution of $\mathbf{X}$ is non-atomic and is not supported on a line in $\mathcal{X}$, then the function SD has a unique maximum at the spatial median $\mathbf{m}$ (say) of $\mathbf{X}$, and its maximum value is 1 (cf. the property "FD4center" in Mosler and Polyakova (2012, p. 10), Theorems 3 and 4 in Liu (1990) and property "P2" in Zuo and Serfling (2000, p. 463)). Further, if we consider the sequence $\{\mathbf{m} + n\mathbf{x}\}_{n\geq 1}$ for any fixed non-zero $\mathbf{x} \in \mathcal{X}$, it follows by a simple application of dominated convergence theorem that $SD(\mathbf{m} + n\mathbf{x}) \to 0$ as $n \to \infty$ (cf. the "FD3" property in Mosler and Polyakova (2012), Theorem 1 in Liu (1990) and property "P4" in Zuo and Serfling (2000, p. 464)).

A natural question that arises now is whether SD suffers from degeneracy similar to what was observed in the case of some of the depth functions discussed earlier or its distribution is well spread out. As the next theorem shows, the distribution of SD is actually supported on the entire unit interval for a large class of probability measures in a separable

Hilbert space $\mathcal{X}$ including Gaussian probabilities.

Theorem 4.1. *Let $\mathcal{X}$ be a separable Hilbert space and consider a random element $\mathbf{X} = \sum_{k=1}^{\infty} X_k \phi_k$, where $\{\phi_k\}_{k\geq 1}$ is an orthonormal basis of $\mathcal{X}$. Assume that $\mathbf{X}$ has a nonatomic probability distribution μ with $\sum_{k=1}^{\infty} E(X_k^2) < \infty$, and the support of the conditional distribution of $(X_1, X_2, \ldots, X_d)$ given $(X_{d+1}, X_{d+2}, \ldots)$ is the whole of $\mathbb{R}^d$ for each $d \geq 1$. Then, the function $SD(\mathbf{x})$ defined using the distribution μ takes all the values in $(0, 1]$ as $\mathbf{x}$ varies in $\mathcal{X}$. Further, if $\widetilde{\mathbf{X}}$ denotes an independent copy of $\mathbf{X}$, the support of the distribution of $SD(\widetilde{\mathbf{X}})$ will be the whole of $[0, 1]$.*

Since $C[0, 1] \subseteq L_2[0, 1]$, for any probability distribution on $C[0, 1]$, SD is defined in the same way as in the case of the separable Hilbert space $L_2[0, 1]$. Thus, for a random element $\mathbf{X} \in C[0, 1]$, if the sequence $(X_1, X_2, \ldots)$ obtained from the orthogonal decomposition of $\mathbf{X}$ in $L_2[0, 1]$ satisfies the conditions of Theorem 4.1, then the support of the distribution of $SD(\widetilde{\mathbf{X}})$ will be the whole of $[0, 1]$. In particular, for any Gaussian process having a continuous mean function and a continuous positive definite covariance kernel, we can have $(X_1, X_2, \ldots)$ to be the coefficients of the Karhunen-Loeve expansion of $\mathbf{X}$, which will then be a sequence of independent Gaussian random variables, and consequently, the conditions of Theorem 4.1 will hold. Those assumptions, however, need not hold when $\mathbf{X}$ is a function of some Gaussian process in $C[0, 1]$ like what we have considered in Theorem 3.3. Indeed, even if $\mathbf{X}$ admits a Karhunen-Loeve type expansion in such a case, the sequence of coefficients need not satisfy the conditions of Theorem 4.1. However, as the next theorem shows, the distribution of SD has full support on the unit interval in some of these situations as well.

Theorem 4.2. *Consider the process $\mathbf{X} = \{X_t\}_{t\in[0,1]} = \{g(t, Y_t)\}_{t\in[0,1]}$ as in Theorem 3.3. Then, the function $SD(\mathbf{x})$ defined using the distribution of $\mathbf{X}$ takes all values in $(0, 1)$ as $\mathbf{x}$ varies in $C[0, 1]$. Moreover, the support of the distribution of $SD(\widetilde{\mathbf{X}})$ is the whole of $[0, 1]$, where $\widetilde{\mathbf{X}}$ is an independent copy of $\mathbf{X}$.*

It follows from arguments that are very similar to those in Remark 6.4 in Section 6 that the above result holds even if SD is computed based on the process $\{X_t\}_{t\in[t_0,1]}$, where $t_0 > 0$. The properties of SD stated at the beginning of this section along with the results in Theorems 4.1 and 4.2 imply that like ID, MBD and MHRD, SD can also be used for various depth-based statistical procedures. The spatial depth function can also be used to discriminate between two probability measures in a separable Hilbert space or $C[0, 1]$. For instance, for any two non-atomic probability measures having distinct and unique spatial medians, the associated spatial depth functions will be continuous, each having a unique maximum at the corresponding spatial median. In that case, spatial depth will be able to distinguish between the two distributions.

We conclude this section with the discussion of another notion of depth, which has a

somewhat similar nature as that of SD. For a random element $\mathbf{X}$ and a fixed element $\mathbf{x}$ in $L_2[0,1]$, the h-depth introduced in Cuevas et al. (2007) is defined as $E\{K_h(||\mathbf{x}-\mathbf{X}||)\}$, where $K_h(t) = h^{-1}K(t/h)$ for some fixed kernel K and $h > 0$ is a tuning parameter. Suppose that K is a bounded continuous probability density function supported on the whole of $[0,\infty)$ with $K(s) \to 0$ as $s \to \infty$, and $(X_1, X_2, \ldots)$ satisfies the conditions in Theorem 4.1, which ensures that the support of the distribution of $\mathbf{X}$ is the whole of $L_2[0,1]$. Then, in view of the continuity of $E\{K_h(||\mathbf{x}-\mathbf{X}||)\}$ as a function of $\mathbf{x}$, which is a consequence of the dominated convergence theorem, it is not difficult to see that the support of the distribution of the h-depth evaluated at an independent copy $\widetilde{\mathbf{X}}$ of $\mathbf{X}$ will be the whole of $[0, A]$. Here $A = \sup_{\mathbf{x}\in L_2[0,1]} E\{K_h(||\mathbf{x}-\mathbf{X}||)\}$. However, no specific guidelines are available for choosing K and h in practice.

5 Demonstration using real and simulated data

In the three preceding sections, we have investigated the behaviour of several depth functions in infinite dimensional spaces. The results derived in those sections are all about the population versions of different depth functions. In this section, we try to investigate to what extent those results are reflected in the empirical versions of the corresponding depth functions computed using some simulated and real datasets. First, we shall consider some simulated and real sequence data. The simulated dataset consists of 50 i.i.d. observations on a d-dimensional Gaussian random vector $\mathbf{X} = (X_1, X_2, \ldots, X_d)$ that satisfies $Cov(X_k, X_l) = r^{-|k-l|}/(kl)^2$, where $r = 0.1$, $k, l = 1, 2, \ldots, d$, and $d = 4000$. The real dataset that is considered next is obtained from http://datam. i2r.a-star.edu.sg/datasets/krbd/ColonTumor/ColonTumor.zip, and it contains expressions of $d = 2000$ genes in tumor tissue biopsies corresponding to 40 colon tumor patients and 22 normal samples of colon tissue. For both these datasets, we can view each sample point as the first d coordinates of an infinite sequence.

In all our samples, since the dimension is much larger than the sample size, the empirical versions of both of HD and PD turn out to be zero (see Figure 1). This is a consequence of the fact that when the dimension is larger than the sample size, and no sample point lies in the subspace spanned by the remaining sample points, the HD and the PD of any data point with respect to the empirical distribution of the remaining data points is zero (see, e.g., remarks at the beginning of Section 4 in Dutta et al. (2011)). It is also observed from the dotplots in Figure 1 that empirical BD and HRD are both degenerate at zero for the two datasets. However, the distribution of empirical SD is well spread out in the corresponding dotplots in Figure 1.

For the colon data, we have prepared another dotplot (see Fig. 2), which shows the

difference between the two empirical SD values for each data point, where one depth value is obtained with respect to the empirical distribution of the tumor tissue sample, and the other one is obtained using that of the normal sample. The value of this difference for a data point corresponding to the tumor tissue is plotted in the panel with heading "Tumor tissue", where all the values are positive. This implies that each data point in the sample of tumor tissue has higher depth value with respect to the empirical distribution of the tumor sample than its depth value with respect to the empirical distribution of the normal tissue sample. On the other hand, a data point corresponding to the normal tissue is plotted in the panel with heading "Normal tissue", where all the values, except only two, are negative. In other words, except for those two cases, each data point in the sample of normal tissue has higher depth value with respect to the normal tissue sample. Thus, SD adequately discriminates between the two samples, and maximum depth or other depth-based classifiers (see, e.g., Ghosh and Chaudhuri (2005) and Li, Cuesta-Albertos and Liu (2012)) constructed using SD will yield good results for this dataset.

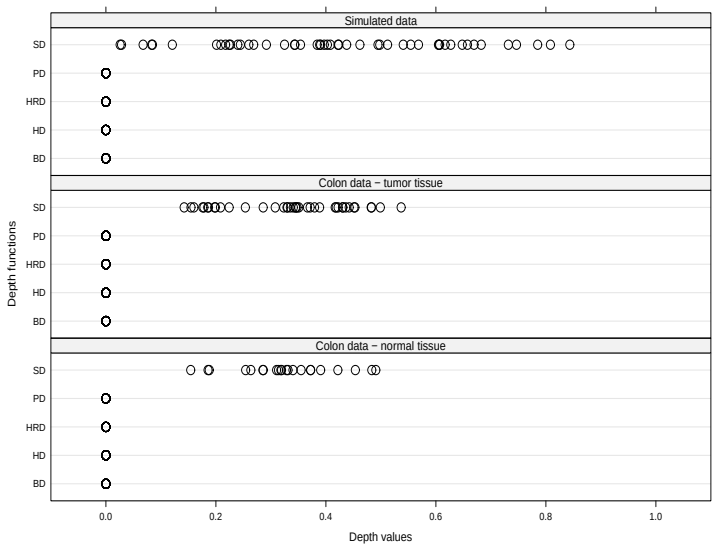

Figure 1: *Dotplots of SD, PD, HRD, HD and BD for simulated data and colon data.*

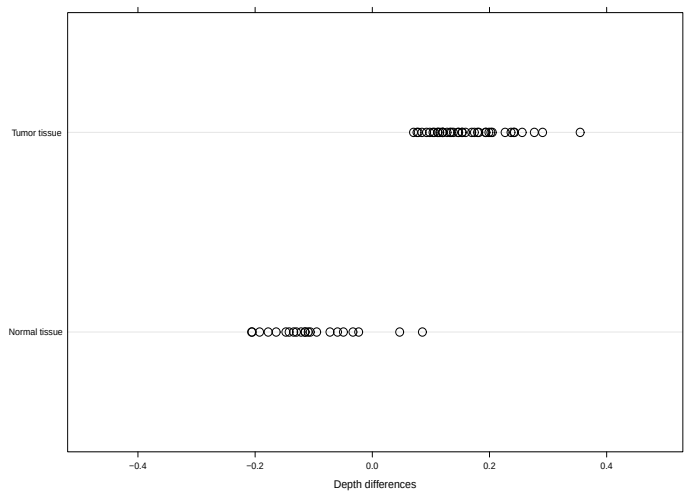

Figure 2: *Dotplots of depth differences based on SD for colon data. The horizantal axis corresponds to the difference between empirical SD values of each data point with respect to the tumor tissue sample and the normal tissue sample.*

We shall next consider some simulated and real functional data. Each of the three simulated datasets consists of 50 observations from (i) a standard Brownian motion on $[0, 1]$, (ii) a zero mean fractional Brownian motion on $[0, 1]$ with covariance function $K(t, s) = (1/2)[t^{2H} + s^{2H} - |t - s|^{2H}]$, where $t, s \in [0, 1]$, and we choose the Hurst index $H = 0.75$, and (iii) a geometric Brownian motion defined as $X_t = exp((r - \sigma^2/2)t + \sigma B_t)$, where $t \in [0, 1]$ and $r = \sigma = 0.5$. Here $\{B_t\}_{t\in[0,1]}$ denotes the standard Brownian motion on $[0, 1]$. For all three simulated datasets, the sample functions were observed at $d = 2000$ equispaced points in $(0, 1)$. We have also considered two real datasets, the first one being the lip movement data, which is available at www.stats.ox.ac.uk/∼silverma/fdacasebook/LipPos.dat, and contains 32 sample observations on the movement of the lower lip. The curves are the trajectories traced by the lower lip while pronouncing the word "bob". The measurements are taken at $d = 501$ time points in a time interval of 700 milliseconds. The second real dataset is the growth acceleration dataset derived from the well-known Berkeley growth data (see Ramsay and Silverman (2005)), which contains two subclasses, namely, the boys and the girls. Heights of 39 boys and 54 girls were measured at 31 time points between ages 1 and 18 years. The growth acceleration curves are obtained through monotone spline smoothing available in the R package "fda", and those are recorded at $d = 101$ equispaced ages in the interval $[1, 18]$. For these functional datasets, we calculated MBD by taking $J = 2$ as suggested in López-Pintado and Romo (2009), and D_t in the definition of ID was taken to be SD for each t, which is equivalent to the depth function used in Fraiman and Muniz (2001).

As shown in the dotplots in Figures 3 and 4, for all of the above simulated and real data, the distributions of empirical ID, MBD, MHRD and SD are well spread out. Empirical BD and HRD are both degenerate at zero for the Brownian motion and the geometric Brownian motion (see Figure 3). For the fractional Brownian motion, the maximum value of empirical BD was 0.024, with its median = 0 and the third quartile = 0.004, whereas the maximum value of empirical HRD was 0.020 with its third quartile = 0 (see Figure 3). For the lip movement data, the empirical HRD is degenerate at zero, while the maximum value of empirical BD is 0.006 with its third quartile = 0 (see Figure 4). For the growth acceleration data, the HRD again turns out to be degenerate at zero, while BD takes a maximum value of 0.004 for boys and 0.008 for girls, and the third quartile for BD = 0 for boys as well as girls (see Figure 4).

For the growth acceleration data, Fig. 5 shows the dotplots for the differences between the two depth values with respect to the empirical distributions of the boys and the girls based on SD, MHRD, MBD and ID. The value of this difference for a data point corresponding to a boy (respectively, a girl) is plotted in the panel with heading "Boys" (respectively, "Girls"). For SD, MBD and ID, most of the data points corresponding to the boys have higher depth values with respect to the empirical distribution of the boys than with respect to the empirical distribution of the girls. On the other hand, most of the data points corresponding to the girls have higher depth values with respect to the empirical distribution of the girls. This implies that each of ID, MBD and SD adequately discriminates between the two samples, and depth-based classifiers (see, e.g., Ghosh and Chaudhuri (2005) and Li, Cuesta-Albertos and Liu (2012)) constructed using ID, MBD or SD will perform well for this dataset. However, the plot corresponding to MHRD shows that a large number of data points in the sample of boys have higher depth values with respect to the empirical distribution of the girls, and almost half of the data points in the sample of girls have higher depth values with respect to the empirical distribution of the boys. This indicates that MHRD does not discriminate well between the two samples.

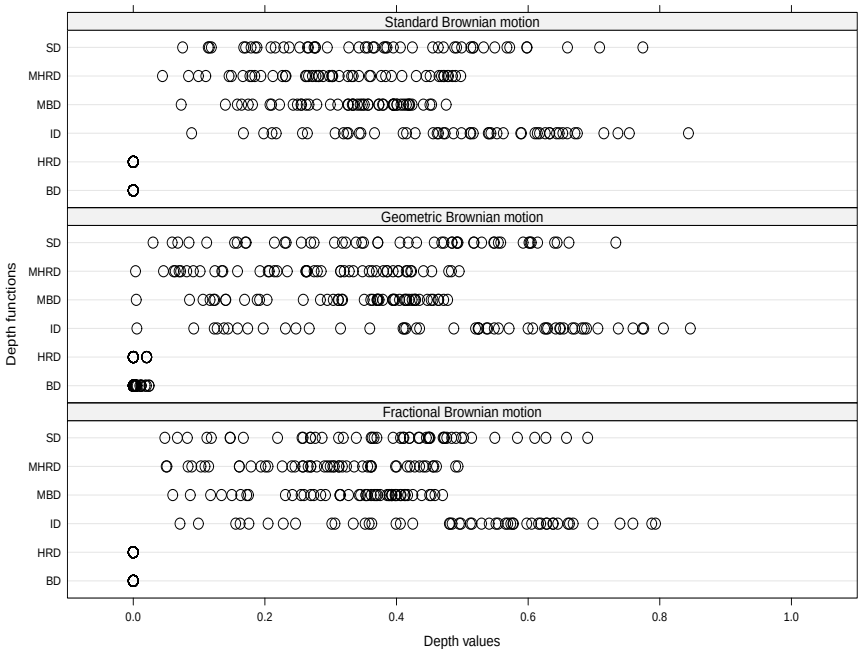

Figure 3: *Dotplots of SD, MHRD, MBD, ID, HRD and BD for simulated standard Brownian motion, geometric Brownian motion and fractional Brownian motion.*

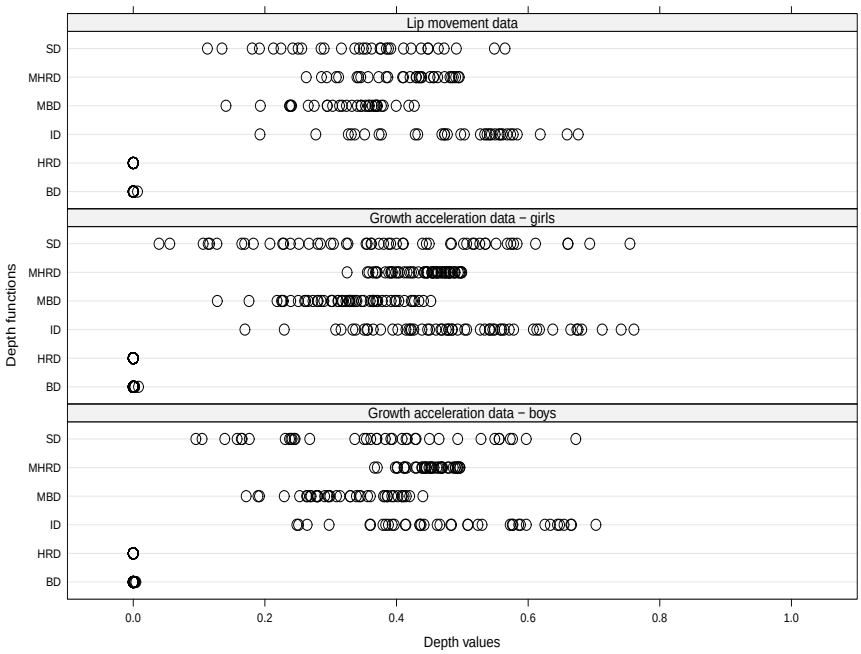

Figure 4: *Dotplots of SD, MHRD, MBD, ID, HRD and BD for lip movement data and growth acceleration data.*

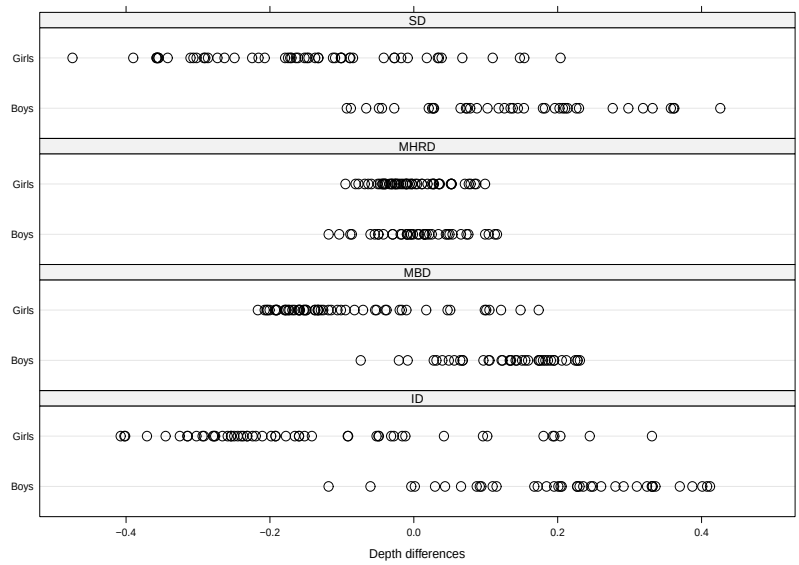

Figure 5: *Dotplots for depth differences based on SD, MHRD, MBD and ID for growth accceleration data. The horizantal axis corresponds to the difference between empirical depth values of each data point with respect to the boys and the girls.*

6 Technical details

Proof. of Theorem 2.1 Let $\mathbf{X}(d) = (X_1, X_2, \ldots, X_d)'$ and $\mathbf{Y}(d) = (Y_1, Y_2, \ldots, Y_d)'$ be d-dimensional column vectors that consist of the first d coordinates of the sequences $\mathbf{X}$ and $\mathbf{Y}$. Observe that $\mathbf{Y}(d) = T_d(\mathbf{X}(d))$, where $T_d : \mathbb{R}^d \to \mathbb{R}^d$ is a bijective affine map. By definition, the half-space depth of a point $\mathbf{x} \in l_2$ relative to the distribution of $\mathbf{X}$ will satisfy

$$\begin{aligned} HD(\mathbf{x}) &= \inf_{\mathbf{u}\in l_2} P(\langle \mathbf{u}, \mathbf{X} - \mathbf{x}\rangle \geq 0) \leq \inf_{d\geq 1} \inf_{\mathbf{v}\in\mathbb{R}^d} P(\mathbf{v}'\mathbf{X}(d) \geq \mathbf{v}'\mathbf{x}(d)) \\ &= \inf_{d\geq 1} \inf_{\mathbf{v}\in\mathbb{R}^d} P(\mathbf{v}'\mathbf{Y}(d) \geq \mathbf{v}'\mathbf{y}(d)) \\ &\leq \inf_{d\geq 1} \inf_{\mathbf{v}\in\mathbb{R}^d:\mathbf{v}'\mathbf{y}(d)>0} P(\mathbf{v}'\mathbf{Y}(d) \geq \mathbf{v}'\mathbf{y}(d)), \end{aligned} \tag{5}$$

where $\mathbf{x}(d) = (x_1, x_2, \ldots, x_d)'$ is the vector of first d coordinates of $\mathbf{x}$, $\mathbf{y}(d) = (y_1, y_2, \ldots, y_d)'$ $= T_d(\mathbf{x}(d))$ and $\mathbf{v} = (v_1, v_2, \ldots, v_d)'$. Throughout this section, any finite dimensional vector will be a column vector, and $'$ will denote its transpose. Since $Y_1, Y_2, \ldots, Y_d$ are uncorrelated, it follows from (5) and Chebyshev inequality that

$$HD(\mathbf{x}) \leq \inf_{d\geq 1} \inf_{\mathbf{v}:\mathbf{v}'\mathbf{y}(d)>0} \frac{Var(\mathbf{v}'\mathbf{Y}(d))}{(\mathbf{v}'\mathbf{y}(d))^2} = \inf_{d\geq 1} \inf_{\mathbf{v}:\mathbf{v}'\mathbf{y}(d)>0} \frac{\sum_{k=1}^d v_k^2 \tau_k^2}{\left[\sum_{k=1}^d v_k y_k\right]^2}. \tag{6}$$

(6) implies, by an application of Cauchy-Schwarz inequality, that

$$HD(\mathbf{x}) \leq \inf_{d\geq 1} \left[\sum_{k=1}^{d} y_k^2/\tau_k^2\right]^{-1}. \tag{7}$$

In view of the moment and the mixing conditions assumed on the Y_k's in the theorem, it follows from Corollary 4 in Hansen (1991) that

$$d^{-1}\sum_{k=1}^{d} Y_k^2/\tau_k^2 \to 1 \text{ a.s. } \Rightarrow \inf_{d\geq 1} \left[\sum_{k=1}^{d} Y_k^2/\tau_k^2\right]^{-1} = 0 \text{ a.s.} \tag{8}$$

(7) and (8) imply that $HD(\mathbf{x}) = 0$ for all $\mathbf{x}$ in a subset of l_2 with μ-measure one.

Next, using the definition of PD and arguments similar to those used above, we get that for any $\mathbf{x} \in l_2$,

$$\begin{aligned}
\frac{1-PD(\mathbf{x})}{PD(\mathbf{x})} &= \sup_{\mathbf{u}\in l_2} \frac{|\langle \mathbf{u},\mathbf{x}\rangle - E(\langle \mathbf{u},\mathbf{X}\rangle)|}{\sqrt{Var(\langle \mathbf{u},\mathbf{X}\rangle)}} \geq \sup_{d\geq 1}\sup_{\mathbf{v}\in\mathbb{R}^d} \frac{|\mathbf{v}'\mathbf{x}(d) - E(\mathbf{v}'\mathbf{X}(d))|}{\sqrt{Var(\mathbf{v}'\mathbf{X}(d))}} \\
&\geq \sup_{d\geq 1}\sup_{\mathbf{v}\in\mathbb{R}^d} \frac{|\mathbf{v}'\mathbf{y}(d)|}{\sqrt{Var(\mathbf{v}'\mathbf{Y}(d))}} \geq \sup_{d\geq 1}\sup_{\mathbf{v}\in\mathbb{R}^d} \frac{\left|\sum_{k=1}^{d} v_k y_k\right|}{\sqrt{\sum_{k=1}^{d} v_k^2\tau_k^2}} \\
&= \sup_{d\geq 1}\sum_{k=1}^{d} \frac{y_k^2}{\tau_k^2}.
\end{aligned} \tag{9}$$

As in the case of HD, in view of the moment and the mixing conditions on the Y_k's assumed in the theorem, (8) and (9) now imply that $PD(\mathbf{x}) = 0$ for all $\mathbf{x}$ in a subset of l_2 with μ-measure one. □

Proof. of Theorem 2.2 Let us denote the dual space of $C[0,1]$ by $\mathcal{M}[0,1]$. Consider the measure $\mathbf{u}_d \in \mathcal{M}[0,1]$, which assigns point mass v_k at k/d, $k = 1,2,\ldots,d$. So, we have $\mathbf{u}_d(\mathbf{x}) = \sum_{k=1}^{d} v_k x_{k/d}$ for any $\mathbf{x} = \{x_t\}_{t\in[0,1]} \in C[0,1]$. Let $\mathbf{v} = (v_1, v_2, \ldots, v_d)'$, $\mathbf{X}_d = (X_{1/d}, X_{2/d}, \ldots, X_{d/d})'$ and $\mathbf{x}_d = (x_{1/d}, x_{2/d}, \ldots, x_{d/d})'$. For each $d \geq 1$, define $Y_{d,1} = X_{1/d} - E(X_{1/d})$, and let $Y_{d,k}$ denote the residual of linear regression of $X_{k/d}$ on $(X_{1/d}, X_{2/d}, \ldots, X_{(k-1)/d})$ for $k = 2,3,\ldots,d$. Then, $\mathbf{Y}_d = (Y_{d,1}, Y_{d,2}, \ldots, Y_{d,k})'$ has a multivariate Gaussian distribution with independent components in view of the Gaussian distribution of $\mathbf{X}$. The proof now follows by straightforward modification of the arguments used in the proof of Theorem 2.1 and using $\mathbf{Y}_d$ in place of $\mathbf{Y}(d)$. □

Proof. of Theorem 3.1 Let $\widetilde{\mathbf{X}} = (\widetilde{X}_1, \widetilde{X}_2, \ldots)$ and $\mathbf{X}_i = (X_{i,1}, X_{i,2}, \ldots)$, $i = 1,2,\ldots,J$, be independent copies of $\mathbf{X}$. We first note that $BD(\mathbf{x}) = HRD(\mathbf{x}) = 0$ with probability one iff $E\{BD(\widetilde{\mathbf{X}})\} = E\{HRD(\widetilde{\mathbf{X}})\} = 0$. Let us first consider the case of BD. Note that $E\{BD(\widetilde{\mathbf{X}})\} = \sum_{j=2}^{J} P(\min_{1\leq i\leq j} X_{i,k} \leq \widetilde{X}_k \leq \max_{1\leq i\leq j} X_{i,k},\ \forall\ k \geq 1)$. So, $E\{BD(\widetilde{\mathbf{X}})\} = 0$ iff $P(\min_{1\leq i\leq j} X_{i,k} \leq \widetilde{X}_k \leq \max_{1\leq i\leq j} X_{i,k},\ \forall\ k \geq 1) = 0$ for all $2 \leq j \leq J$.

Consequently, it is enough to show that for any $2 \leq j \leq J$, the event $\{\min_{1\leq i\leq j} X_{i,k} > \widetilde{X}_k\} \cup \{\max_{1\leq i\leq j} X_{i,k} < \widetilde{X}_k\}$ occurs for some $k \geq 1$ with probability one. Now, the sequence $(\min_{1\leq i\leq j} X_{i,1} - \widetilde{X}_1, \min_{1\leq i\leq j} X_{i,2} - \widetilde{X}_2, \ldots)$ is α-mixing for any $1 \leq j \leq J$, and its mixing coefficients satisfy the conditions assumed in the theorem. On the other hand, $P(\min_{1\leq i\leq j} X_{i,k} > \widetilde{X}_k) = 2^{-j}$ for all $k \geq 1$, by the continuity of the distributions of the X_k's. So, using Corollary 4 in Hansen (1991), we have $d^{-1}\sum_{k=1}^{d} I(\min_{1\leq i\leq j} X_{i,k} > \widetilde{X}_k) \to 2^{-j}$ as $d \to \infty$ with probability one for all $1 \leq j \leq J$. So, the event $\{\min_{1\leq i\leq j} X_{i,k} > \widetilde{X}_k\}$ actually occurs for infinitely many $k \geq 1$ with probability one. Thus, $BD(\mathbf{x}) = 0$ for all $\mathbf{x}$ in a subset of l_2 with μ-measure one.

The proof for HRD follows by taking $j = 1$, and we skip further details. □

Lemma 6.1. *Let $\{X_t\}_{t\in[0,1]}$ be a Feller processes in $C[0,1]$ satisfying the conditions of Theorem 3.2. Let $\mathbf{X}_i = \{X_{i,t}\}_{t\in[0,1]}$, $i = 1, 2, \ldots, J$, denote independent copies of $\mathbf{X}$, and define $T_j = \inf\{t > 0 : \min_{1\leq i\leq j} X_{i,t} > x_0\}$ and $S_j = \inf\{t > 0 : \max_{1\leq i\leq j} X_{i,t} < x_0\}$ for $1 \leq j \leq J$. Then, $P(T_j = 0) = P(S_j = 0) = 1$ for all $1 \leq j \leq J$.*

Proof. Consider the multivariate Feller process $\{(X_{1,t}, X_{2,t}, \ldots, X_{j,t})\}_{t\in[0,1]}$, where $1 \leq j \leq J$. Since, $P(T_j \leq t) \geq P(\min_{1\leq i\leq j} X_{i,t} > x_0) = 2^{-j}$ and $P(S_j \leq t) \geq P(\max_{1\leq i\leq j} X_{i,t} < x_0) = 2^{-j}$ for every $t > 0$, we have

$$P(T_j = 0) = \lim_{s\downarrow 0} P(T_j \leq s) \geq 2^{-j} \quad \text{and} \quad P(S_j = 0) = \lim_{s\downarrow 0} P(S_j \leq s) \geq 2^{-j}. \tag{10}$$

From the continuity of the sample paths of the processes, and using Propositions 2.16 and 2.17 in Revuz and Yor (1991), it follows that $P(T_j = 0) = 0$ or 1 and $P(S_j = 0) = 0$ or 1 for all $1 \leq j \leq J$. The proof is now complete using (10). □

Lemma 6.2. *Let $\{X_t\}_{t\in[0,1]}$ be a Feller process on $C[0,1]$ satisfying the conditions of Theorem 3.2. Also, let $\mathbf{f} = \{f_t\}_{t\in[0,1]} \in C[0,1]$ be such that $f_0 = x_0$ and $f_t - x_0$ changes sign infinitely often in any right neighbourhood of zero. Then, $P(T = 0) = P(S = 0) = 1$, where $T = \inf\{t > 0 : X_t - f_t > 0\}$ and $S = \inf\{t > 0 : X_t - f_t < 0\}$.*

Proof. For any $t > 0$, let $0 < r < t$ be such that $f_r < x_0$. Then, $P(T \leq t) \geq P(T \leq r) \geq P(X_r > f_r) \geq P(X_r > x_0) = 1/2$. Now, arguing as in the proof of Lemma 6.1, we get that $P(T = 0) = 1$ since $\{X_t - f_t\}_{t\in[0,1]}$ is a Feller proces staring at 0. Next, let $0 < s < t$ be such that $f_s > x_0$. By similar arguments, we get that $P(S = 0) = 1$. □

Proof. of Theorem 3.2 We first prove the result for BD using similar ideas as in the proof

of Theorem 3.1. From the definition of BD in (1), we have

$$\begin{aligned} E\{BD(\widetilde{\mathbf{X}})\} &= \sum_{j=2}^{J} P\left(\min_{1\le i\le j} X_{i,t} \le \widetilde{X}_t \le \max_{1\le i\le j} X_{i,t},\ \forall\, t\in[0,1]\right) \\ &\le \sum_{j=2}^{J} P\left(\min_{1\le i\le j} X_{i,t} \le \widetilde{X}_t,\ \forall\, t\in[0,1]\right) \\ &= \sum_{j=2}^{J} E\left\{P\left(\min_{1\le i\le j} X_{i,t} \le \widetilde{X}_t,\ \forall\, t\in[0,1] \;\middle|\; \mathbf{X}_1,\mathbf{X}_2,\ldots,\mathbf{X}_J\right)\right\}. \quad (11) \end{aligned}$$

For any fixed j, let $\mathbf{z} = \{z_t\}_{t\in[0,1]}$ be a realization of the process $\{\min_{1\le i\le j} X_{i,t}\}_{t\in[0,1]}$. Then, from Lemma 6.1, it follows that $\mathbf{z}$ satisfies, with probability one, the assumptions made on the function $\mathbf{f}$ in Lemma 6.2. So, using Lemma 6.2, we have $P(z_t \le \widetilde{X}_t,\ \forall\, t\in[0,1]) = 0$ for all $\mathbf{z}$ in a set of probability one. Hence, the expectation in (11) is zero, which implies that $E\{BD(\widetilde{\mathbf{X}})\} = 0$. Thus, $BD(\mathbf{x}) = 0$ on a set of μ-measure one.

The proof for HRD follows by taking $\mathbf{z}$ to be a realization of the process $\mathbf{X}$, and using Lemma 6.1 and similar arguments as above. □

Lemma 6.3. *Let* $\mathbf{G}$ *be the map on* $C[0,1]$ *defined as* $\mathbf{G}(\mathbf{f}) = \{g(t,f_t)\}_{t\in[0,1]}$*, where* $\mathbf{f} = \{f_t\}_{t\in[0,1]} \in C[0,1]$ *and* $g : [0,1]\times\mathbb{R}\to\mathbb{R}$ *is continuous. Then,* $\mathbf{G}$ *is a continuous map from* $C[0,1]$ *into* $C[0,1]$*.*

Proof. Let $t_n \to t$ in $[0,1]$ as $n\to\infty$. By the continuity of g, and the fact that $\mathbf{f} = \{f_t\}_{t\in[0,1]} \in C[0,1]$, we have $g(t_n, f_{t_n}) \to g(t, f_t)$ as $n\to\infty$. This shows that $\mathbf{G}$ maps $C[0,1]$ into $C[0,1]$. Let us now fix $\epsilon > 0$, $t\in[0,1]$ and $\mathbf{f}\in C[0,1]$. Consider a sequence of functions $\mathbf{f}_n = \{f_{n,t}\}_{t\in[0,1]}$ in $C[0,1]$ such that $||\mathbf{f}_n - \mathbf{f}|| \to 0$ as $n\to\infty$. Note that the function g is uniformly continuous on $[0,1]\times I$, where I is any compact interval of the real line. Thus, $\sup_{t\in[0,1]} |g(t,f_{n,t}) - g(t,f_t)| \to 0$, and this proves the continuity of $\mathbf{G}$. □

Proof. of Theorem 3.3 (a) Since the process $\mathbf{Y} = \{Y_t\}_{t\in[0,1]}$ has almost surely continuous sample paths, Lemma 6.3 implies that the sample paths of the process $\mathbf{X} = \mathbf{G}(\mathbf{Y})$ also lie in $C[0,1]$ almost surely. Consider now $\mathbf{x}_p = \mathbf{G}(\mathbf{y}_p)$, where $p\in(0,1)$ and $\mathbf{y}_p = \{F_t^{-1}(p)\}_{t\in[0,1]}$. Note that the distribution F_t of Y_t is Gaussian for all $t\in(0,1]$ with zero mean and variance σ_t^2 (say), which is a continuous function in t. So, $F_t^{-1}(p) = \sigma_t\Phi^{-1}(\zeta_p)$, where Φ and ζ_p denote the distribution function and the pth quantile of the standard normal variable, respectively. Hence, $\mathbf{y}_p \in C[0,1]$, and in view of Lemma 6.3, we have $\mathbf{x}_p = \mathbf{G}(\mathbf{y}_p) \in C[0,1]$.

Note that by strict monotonicity of $g(t,.)$ for all $t\in[0,1]$, we have $MBD(\mathbf{x}_p) = \sum_{j=2}^{J}[1-p^j-(1-p)^j]$, $MHRD(\mathbf{x}_p) = \min(p, 1-p)$ and $ID(\mathbf{x}_p) = \psi(p)$. These depth functions are bounded above by $A_J = J-2+2^{-J+1}$, $1/2$ and $\sup_{s\in(0,1)}\psi(s)$, respectively, where the upper bounds are attained in MBD and MHRD iff $p = 1/2$. Let us now write $C_{y_0}[0,1] = \{\mathbf{f} =$

$\{f_t\}_{t\in[0,1]} \in C[0,1] : f_0 = y_0\}$, and define $H_0 = \mathbf{G}(C_{y_0}[0,1]) = \{\mathbf{G}(\mathbf{f}) : \mathbf{f} \in C_{y_0}[0,1]\}$. Since $\mathbf{x}_p \in H_0$, we have $MBD(H_0) = \{MBD(\mathbf{x}) : \mathbf{x} \in H_0\} = (0, A_J]$, $MHRD(H_0) = (0, 1/2]$ and $ID(H_0) = \psi((0,1))$ by varying $p \in (0,1)$. This completes the proof of part (a).

(b) It follows from the proof of Proposition 5.1 in Guasoni (2006) that the support of a fractional Brownian motion, say $\{Z_t\}_{t\in[0,1]}$, starting at zero is the whole of $C_0[0,1]$. Since the distribution of $\{Y_t\}_{t\in[0,1]}$ is same as that of $\{Z_t+y_0\}_{t\in[0,1]}$, the support of the distribution of $\{Y_t\}_{t\in[0,1]}$ is the whole of $C_{y_0}[0,1]$. By continuity of $\mathbf{G}$ proved in Lemma 6.3, any point in H_0 is a support point of the distribution of $\widetilde{\mathbf{X}}$. On the other hand, for every fixed $t \in [0,1]$, since $g(t,.)$ is a continuous strictly monotone function, and the distribution of Y_t is continuous, it follows that the distribution of X_t is continuous. So, using the dominated convergence theorem, we get that MBD, MHRD and ID are continuous functions on $C[0,1]$. This and the fact that any point in H_0 is a support point of the distribution of $\widetilde{\mathbf{X}}$ completes the proof of part (b).

(c) If $\{Y_t\}_{t\in[0,1]}$ is a fractional Brownian bridge "tied" down to b_0 at $t = 1$ (say), then it has the same distribution as that of $\{Z_t - Cov(Z_t, Z_1)(Z_1 - b_0)\}_{t\in[0,1]}$. So, the support of $\{Y_t\}_{t\in[0,1]}$ is the set $\{\mathbf{f} = \{f_t\}_{t\in[0,1]} \in C_{y_0}[0,1] : f_1 = b_0\}$. The proof now follows from arguments similar to those in parts (a) and (b). □

Remark 6.4. It follows from the proof of Proposition 5.1 in Guasoni (2006) that a fractional Brownian motion $\{Y_t\}_{t\in[0,1]}$ starting at y_0 has as its support as the whole of $C_{y_0}[0,1]$, which implies that the support of $\{Y_t\}_{t\in[t_0,1]}$ is the whole of $C[t_0,1]$ for any $t_0 > 0$. Consequently, if MBD, MHRD and ID are computed based on the distribution of $\{X_t\}_{t\in[t_0,1]}$, the supports of the distributions of $MBD(\widetilde{\mathbf{X}})$, $MHRD(\widetilde{\mathbf{X}})$ and $ID(\widetilde{\mathbf{X}})$ will be $[0, A_J]$, $[0, 1/2]$ and the closure of $\psi((0,1))$, respectively.

Proof. of Theorem 4.1 First, we shall prove that the support of $\widetilde{\mathbf{X}}$ is the whole of l_2, where $\widetilde{\mathbf{X}} = (\widetilde{X}_1, \widetilde{X}_2, \ldots)$ is an independent copy of $\mathbf{X} = (X_1, X_2, \ldots)$. For this, let us fix $\mathbf{x} \in l_2$ and $\eta > 0$. Then, there exists $d \geq 1$ satisfying $||\mathbf{x} - \mathbf{x}[d]|| < \eta$, where $\mathbf{x}[d] = (x_1, x_2, \ldots, x_d, 0, 0, \ldots)$. Further, in view of the assumption on the second moments of the X_k's, we can choose $M > d$ such that $\sum_{k>M} E(\widetilde{X}_k^2) < \eta^2/4$. Then,

$$\begin{aligned} & P(||\widetilde{\mathbf{X}} - \mathbf{x}|| < 2\eta) > P(||\widetilde{\mathbf{X}} - \mathbf{x}[d]|| < \eta) \\ > \; & P\left(\sum_{k\leq M} (\widetilde{X}_k - x_k)^2 < \frac{\eta^2}{2} \;\middle|\; \sum_{k>M} \widetilde{X}_k^2 < \frac{\eta^2}{2}\right) \; P\left(\sum_{k>M} \widetilde{X}_k^2 < \frac{\eta^2}{2}\right). \end{aligned} \tag{12}$$

Using Markov inequality, we get

$$P\left(\sum_{k>M} \widetilde{X}_k^2 < \frac{\eta^2}{2}\right) \; > \; 1 - \frac{\sum_{k>M} E(\widetilde{X}_k^2)}{\eta^2/2} \; > \; 1/2. \tag{13}$$

(12) and (13) now imply that

$$P(||\widetilde{\mathbf{X}} - \mathbf{x}|| < 2\eta) \quad > \quad \frac{1}{2} P\left(\sum_{k \le M} (\widetilde{X}_k - x_k)^2 < \frac{\eta^2}{2} \;\middle|\; \sum_{k > M} \widetilde{X}_k^2 < \frac{\eta^2}{2} \right). \tag{14}$$

From the conditional full support assumption on the X_k's, it follows that the expression on the right hand side of the inequality (14) is positive for each $\eta > 0$. This implies that $\mathbf{x}$ lies in the support of $\widetilde{\mathbf{X}}$.

Since the distribution of $\mathbf{X}$ is non-atomic, SD is a continuous function on l_2 as mentioned in Section 4. Thus, the set $\{SD(\mathbf{x}) : \mathbf{x} \in l_2\}$ is an interval in $[0, 1]$. Hence, from the properties of SD discussed in Section 4, we get that the function SD takes all values in $(0, 1]$. This and the continuity of SD together imply that the support of the distribution of $SD(\widetilde{\mathbf{X}})$ is the whole of $[0, 1]$. □

Lemma 6.5. *The set $H_0 = \mathbf{G}(C_{y_0}[0, 1])$ is convex. Here, $\mathbf{G}$ is as in Lemma 6.1 and $C_{y_0}[0, 1]$ is as in the proof of Theorem 3.3.*

Proof. Let us take $\mathbf{f} = \{f_t\}_{t \in [0,1]}$ and $\mathbf{h} = \{h_t\}_{t \in [0,1]} \in C_{y_0}[0, 1]$. Fix $\lambda \in (0, 1)$ and $t \in [0, 1]$. Let $L = \max(||\mathbf{f}||, ||\mathbf{h}||)$. By continuity of $g(t, .)$, the range of $g(t, s)$ for $s \in [-L, L]$ is a closed and bounded interval, say $[a, b]$. Thus, $\lambda g(t, f_t) + (1 - \lambda) g(t, h_t) \in [a, b]$. Since $g(t, .)$ is continuous and strictly increasing, there is a unique $q_t \in [-L, L]$ such that $g(t, q_t) = \lambda g(t, f_t) + (1 - \lambda) g(t, h_t)$. Now let $t_n \to t \in [0, 1]$ as $n \to \infty$. Since $g(t_n, q_{t_n}) = \lambda g(t_n, f_{t_n}) + (1 - \lambda) g(t_n, h_{t_n})$, by continuity of g, we have

$$g(t_n, q_{t_n}) \to \lambda g(t, f_t) + (1 - \lambda) g(t, h_t) = g(t, q_t) \tag{15}$$

as $n \to \infty$. Suppose now, if possible, $q_{t_n} \nrightarrow q_t$ as $n \to \infty$. Then, there exists $\epsilon_0 > 0$ and a subsequence $\{t_{n_j}\}_{j \ge 1}$ such that $|q_{t_{n_j}} - q_t| > \epsilon_0$ for all $j \ge 1$. A further subsequence of $\{t_{n_j}\}_{j \ge 1}$ will converge to some $b_t \in [-L, L]$, and hence, $|b_t - q_t| \ge \epsilon_0$. Along that latter subsequence, we have $g(t_{n_j}, q_{t_{n_j}})$ converging to $g(t, b_t)$. This and (15) together imply that $g(t, b_t) = g(t, q_t)$. So, by strict monotonicity of $g(t, .)$, we get that $b_t = q_t$, which yields a contradiction. Hence, $q_{t_n} \to q_t$ as $n \to \infty$, which implies that $\mathbf{q} = \{q_t\}_{t \in [0,1]} \in C_{y_0}[0, 1]$. This proves the convexity of H_0. □

Lemma 6.6. *Every point in H_0 is a support point of the distribution of $\widetilde{\mathbf{X}}$ in $L_2[0, 1]$. Here $\widetilde{\mathbf{X}}$ is as in Theorem 4.2.*

Proof. Fix $\mathbf{f} \in C_{y_0}[0, 1]$ and $\eta > 0$. Let $||.||$ denote the supremum norm on $C[0, 1]$ as before, and $||.||_2$ denote the usual norm on $L_2[0, 1]$. Since $||\mathbf{y}||_2 \le ||\mathbf{y}||$ for any $\mathbf{y} \in C[0, 1]$, we have $P(||\mathbf{G}(\mathbf{Y}) - \mathbf{G}(\mathbf{f})||_2 < \eta) > P(||\mathbf{G}(\mathbf{Y}) - \mathbf{G}(\mathbf{f})|| < \eta)$. By the continuity of $\mathbf{G}$ proved in

Lemma 6.3, there exists $\delta > 0$ depending on η and $\mathbf{f}$ such that $P(||\mathbf{G}(\mathbf{Y}) - \mathbf{G}(\mathbf{f})|| < \eta) > P(||\mathbf{Y} - \mathbf{f}|| < \delta)$. Since any element in $C_{y_0}[0,1]$ is a support point of the distribution of $\mathbf{Y}$ in $C[0,1]$, we have $P(||\mathbf{Y} - \mathbf{f}|| < \delta) > 0$. It now follows that $\mathbf{G}(\mathbf{f}) \in H_0$ is a support point of the distribution of $\widetilde{\mathbf{X}} = \mathbf{G}(\widetilde{\mathbf{Y}})$ in $L_2[0,1]$, where $\widetilde{\mathbf{Y}}$ denotes an independent copy of $\mathbf{Y}$. This completes the proof. □

Proof. of Theorem 4.2 We will first show that $SD(\mathbf{x})$ takes all values in $(0,1)$ as $\mathbf{x}$ varies in $C[0,1]$. As discussed in Section 4, the spatial depth function is continuous on $L_2[0,1]$. We have $H_0 \subseteq C[0,1] \subseteq L_2[0,1]$, and H_0 is convex by Lemma 6.5, which implies that the set $SD(H_0) = \{SD(\mathbf{f}) : \mathbf{f} \in H_0\}$ is an interval in $[0,1]$. It follows from the non-atomicity of $\mathbf{X}$ and Lemma 4.14 in Kemperman (1987) that $SD(\mathbf{m}) = 1$, where $\mathbf{m}$ is a spatial median of $\mathbf{X}$ in $L_2[0,1]$. Further, from Remark 4.20 in Kemperman (1987), it follows that $\mathbf{m}$ lies in the closure of H_0 in $L_2[0,1]$. Thus, there exists a sequence $\{\mathbf{m}_n\}_{n\geq 1}$ in $H_0 \subseteq C[0,1]$ such that $||\mathbf{m}_n - \mathbf{m}||_2 \to 0$ as $n \to \infty$, where $||.||_2$ is the usual norm in $L_2[0,1]$ as before. Hence, by continuity of the spatial depth function,, we have $SD(\mathbf{m}_n) \to 1$ as $n \to \infty$. We next consider the sequence of linear functions $\{\mathbf{r}_n\}_{n\geq 1}$, where $\mathbf{r}_n = \{g(0, y_0) + d_n t\}_{t\in[0,1]}$ and $d_n \to \infty$ as $n \to \infty$. Since $g(t.,)$ is a strictly increasing continuous function for each $t \in [0,1]$, there exists $f_{n,t}$ such that $g(t, f_{n,t}) = g(0, y_0) + d_n t$. Using the assumptions about g, it can be shown that for each $n \geq 1$, the function $\mathbf{f}_n = \{f_{n,t}\}_{t\in[0,1]} \in C_{y_0}[0,1]$, which implies that $\mathbf{r}_n = \mathbf{G}(\mathbf{f}_n) \in H_0$. Now, using dominated convergence theorem, we have $SD(\mathbf{r}_n) \to 0$ as $n \to \infty$ in view of the fact that $d_n \to \infty$, and $\mathbf{r}_n/d_n$ converges to the identity function $\{t\}_{t\in[0,1]} \in C[0,1]$ as $n \to \infty$. Hence, $SD(H_0) \supseteq (0,1)$. Note that we will have $SD(H_0) = (0,1]$ if the spatial median $\mathbf{m}$ actually lies in H_0. Using Lemma 6.6, and the continuity of SD along with the fact that $SD(H_0) \supseteq (0,1)$, we get that the support of the distribution of $SD(\widetilde{\mathbf{X}})$ is the whole of $[0,1]$. □

References

Araujo, A. and Giné, E. (1980) *The central limit theorem for real and Banach valued random variables.* New York-Chichester-Brisbane: John Wiley & Sons.

Cuesta-Albertos, J. A., Nieto-Reyes, A. (2008). The random Tukey depth. *Computational Statistics & Data Analysis, 52*, 4979–4988.

Cuevas, A., Febrero, M., Fraiman, R. (2007). Robust estimation and classification for functional data via projection-based depth notions. *Computational Statistics, 22*, 481–496.

Cuevas, A. and Fraiman, R. (2009). On depth measures and dual statistics. A methodology for dealing with general data. *Journal of Multivariate Analysis, 100*, 753–766.

Donoho, D. L. and Gasko, M. (1992). Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *The Annals of Statistics, 20*, 1803–1827.

Dutta, S. and Ghosh, A. K. (2012) On robust classification using projection depth. *Annals of the Institute of Statistical Mathematics, 64*, 657–676.

Dutta, S. , Ghosh, A. K. and Chaudhuri, P. (2011). Some intriguing properties of Tukey's half-space depth. *Bernoulli, 17*, 1420–1434.

Fraiman, R. and Muniz, G. (2001). Trimmed means for functional data. *Test, 10*, 419–440.

Ghosh, A. K. and Chaudhuri, P. (2005). On maximum depth and related classifiers. *Scandinavian Journal of Statistics, 32*, 327–350.

Guasoni, P. (2006). No arbitrage under transaction costs, with fractional Brownian motion and beyond. *Mathematical Finance, 16*, 569–582.

Hansen, B. E. (1991) Strong laws for dependent heterogeneous processes. *Econometric Theory, 7*, 213–221.

Jörnsten, R. (2004) Clustering and classification based on the L_1 data depth. *Journal of Multivariate Analysis, 90*, 67–89.

Kemperman, J. H. B. (1987) The median of a finite measure on a Banach space. In *Statistical data analysis based on the L_1-norm and related methods (Neuchâtel, 1987)* (pp. 217–230). Amsterdam: North-Holland.

Li, J., Cuesta-Albertos, J. A. and Liu, R. (2012) DD-classifier: Nonparametric classification procedure based on DD-plot. *Journal of the American Statistical Association, 107*, 737–753.

Liu, R. (1990) On a notion of data depth based on random simplices. *The Annals of Statistics, 18*, 405–414.

Liu, R. Y. , Parelius, J. M. and Singh, K. (1999) Multivariate analysis by data depth: descriptive statistics, graphics and inference. *The Annals of Statistics, 27*, 783–858.

López-Pintado, S. and Romo, J. (2006) Depth-based classification for functional data. *Data depth: robust multivariate analysis, computational geometry and applications, DIMACS Ser. Discrete Math. Theoret. Comput. Sci., 72*, 103–119.

López-Pintado, S. and Romo, J. (2009) On the concept of depth for functional data. *Journal of the American Statistical Association, 104*, 718–734.

López-Pintado, S. and Romo, J. (2011) A half-region depth for functional data. *Computational Statistics and Data Analysis, 55*, 1679–1695.

Mosler, K. (2002) *Multivariate dispersion, central regions and depth.* Berlin: Springer-Verlag.

Mosler, K. and Hoberg, R. (2006) Data analysis and classification with the zonoid depth. *Data depth: robust multivariate analysis, computational geometry and applications, DIMACS Ser. Discrete Math. Theoret. Comput. Sci., 72*, 49–59.

Mosler, K. and Polyakova, Y. (2012) General notions of depth for functional data. Technical Report. arXiv:1208.1981v1.

Ramsay, J. O. and Silverman, B. W. (2005) *Functional data analysis.* New York: Springer.

Revuz, D. and Yor, M. (1991) *Continuous martingales and Brownian motion.* Berlin: Springer-Verlag.

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)* (pp. 25–38). Basel: Birkhäuser.

Serfling, R. (2006) Depth functions in nonparametric multivariate inference. *Data depth: robust multivariate analysis, computational geometry and applications, DIMACS Ser. Discrete Math. Theoret. Comput. Sci., 72*, 1–16.

Vardi, Y. and Zhang, C-H. (2000) The multivariate L_1-median and associated data depth. *Proceedings of the National Academy of Sciences of the United States of America, 97*, 1423–1423.

Zuo, Y. (2006) Multidimensional trimming based on projection depth. *The Annals of Statistics, 34*, 2211–2251.

Zuo, Y. and Serfling, R. (2000) General notions of statistical depth function. *The Annals of Statistics, 28*, 461–482.