NONPARAMETRIC METHODS FOR DATA IN INFINITE DIMENSIONAL SPACES

ANIRVAN CHAKRABORTY

Indian Statistical Institute, Kolkata

December, 2014

ANIRVAN CHAKRABORTY

Thesis submitted to the Indian Statistical Institute
in partial fulfillment of the requirements
for the award of the degree of
Doctor of Philosophy.

December, 2014

Thesis Advisor : Prof. Probal Chaudhuri

STATISTICAL INDIAN INSTITUTE

UNITY IN DIVERSITY

Indian Statistical Institute
203, B. T. Road, Kolkata, India.

To my parents and teachers

Acknowledgement

First of all, I pay my obeisance and thank the Almighty for everything.

Next, I would like to express my sincere and heartfelt gratitude to my supervisor Prof. Probal Chaudhuri. Without his constant guidance, it would not have been possible for me to build up my thesis in the form as it stands today. His endless enthusiasm has always served as an impetus for me throughout my research tenure. Words will not be able to voice the impact of his scholarly abilities on my research work. I am indebted to him for whatever I have learnt during my years as a Ph.D. student.

I express my deepest respect for my parents, who have given their all to mould me into a good human being. Their unending love, care and support have been an inspiration for me to carry on with my dissertation work. I also thank my family members, who have always supported me in my life and work. I apologize to all of them for not being able to spend enough time with them during my tenure as a Ph.D. student.

Next, I thank my teachers for their constant support and guidance, and I am obliged to them for imparting their knowledge to me. I am thankful to my teachers during my days as an undergraduate student at the St. Xavier's College, Kolkata and as a postgraduate student in the Indian Statistical Institute at Delhi and Kolkata. I am very grateful to Dr. Anil Kr. Ghosh, Prof. Arup Bose and Prof. Alok Goswami of Stat. Math. Unit of the Indian Statistical Institute, Kolkata, for teaching me various courses during the first two years of my research tenure and also introducing me to many frontline areas of research. I have also benefitted from several interactions with various other faculty members of different departments of the Indian Statistical Institute.

I thank the Council of Scientific and Industrial Research, Government of India, for supporting me with the Shyama Prasad Mukherjee Fellowship. I am thankful to the Indian Statistical Institute for providing me with excellent infrastructure to carry out my research work. I also thank various administrative departments of the Indian Statistical Institute for different matters during my research tenure. Special thanks to the Office of the Stat. Math. Unit for being extremely helpful in various ways.

It would be very selfish on my part not to mention those who brought a gust of fresh air (both academically as well as non-academically) to my sometimes tiresome and monochromatic life as a Ph.D. student. It was a joy to have fellow research scholars like Subhajit da, Shubhra da, Soumalya, Arnab, Tapas da, Shubhabrata, Zinna, Munmun, Suman and Sabyasachi around me. I will surely remember some of the quality time that I have spent with them.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

For univariate as well as finite dimensional multivariate data, there is an extensive literature on nonparametric statistical methods. One of the reasons for the popularity of nonparametric methods is that it is often difficult to justify the assumptions (e.g., Gaussian distribution of the data) made in the models used in parametric methods. Nonparametric procedures use more flexible models, which involve less assumptions. So, they are more robust against possible departures from the model assumptions, and are applicable to a wide variety of data. Nonparametric methods outperform their parametric competitors in many situations, where the assumptions required for the parametric methods are not satisfied.

Nowadays, with the advancement in the technology and measurement apparatus, statisticians often have to analyze data, which are curves or functions observed over a domain. Such data are increasingly becoming common in various fields of science like biomedical sciences (ECG and EEG curves of patients observed over a time period, MRI and other image data obtained from patients), cognitive sciences (data on hand-writing and speech patterns of subjects), chemical science (spectrometric data observed over a range of wavelengths), environmental science (air pollutant levels at different places over a period of time), meteorology (temperature curves at different places over a year, precipitation levels at different locations during a year) etc (see Ramsay and Silverman (2005) for a detailed exposition). A major difference between this type of data and standard multivariate data is that the set of points in the domain, where one sample observation is recorded may be different than those for the other sample observations. Further, the number of such points is often very large compared to the number of samples making the dimension of the data larger than the sample size. As a result, standard multivariate techniques cannot be used for analyzing such data. However, this type of data can be conveniently handled by viewing them as observations from some infinite dimensional space, e.g., the space of functions defined on an interval in the real line. Due to the advantages in using nonparametric methods for multivariate data lying in finite dimensional spaces, one may expect that such procedures will also be useful for analyzing data, which lie in infinite dimensional spaces. In this thesis, we will address this issue by investigating various nonparametric procedures for such data.

Ranks, distributions and quantiles have been used to develop various nonparametric procedures for univariate data (see, e.g., Lehmann (1975) and Hájek et al. (1999)). Various extensions of these notions are available in the literature for multivariate data lying in finite dimensional spaces, and several well-known nonparametric procedures have been developed based on them (see, e.g., Puri and Sen (1971) and Oja (2010)). My thesis will be mainly devoted to investigating similar notions in infinite dimensional spaces and studying nonparametric statistical methods based on them.

Perhaps, the most popular nonparametric test for univariate data is the Wilcoxon–Mann–Whitney rank sum test for two sample problems. For finite dimensional multivariate data, several extensions of the Wilcoxon–Mann–Whitney test have been studied, and some of these extensions have been shown to be asymptotically more efficient than the Hotelling's T^2 test for many non-Gaussian distributions (see, e.g., Puri and Sen (1971), Randles and Peters (1990), Liu and Singh (1993), Choi and Marden (1997), Chakraborty and Chaudhuri (1999) and Oja (1999)). There have been some work on developing two sample tests for means of functional data (see, e.g., Cox and Lee (2008), Zhang et al. (2010) and Horváth et al. (2013)). Cuevas et al. (2004) studied a functional analysis of variance test, while Shen and Faraway (2004) and Zhang and Chen (2007) investigated tests in the setup of a functional linear model. For testing the hypothesis of equality of two mean functions, each of the previous tests can be appropriately simplified and used. However, all these tests for functional data have low powers when the sample observations have non-Gaussian and heavy-tailed distributions or when the samples are contaminated by outliers. In `Chapter 2`, we have investigated a notion of spatial ranks for probability distributions in infinite dimensional spaces, and have studied a Wilcoxon–Mann–Whitney type test for two sample problems based on them. The proposed test statistic has a limiting Gaussian distribution under the null hypothesis. Further, it is shown this test is consistent against all those alternatives, where the spatial median of $\mathbf{Y} - \mathbf{X}$ is different from zero. In particular, the test is consistent for location shift alternatives. Here $\mathbf{X}$ and $\mathbf{Y}$ are random observations from the two probability distributions in some infinite dimensional space. A theoretical comparison of the asymptotic powers of the proposed test and some other consistent tests available in the literature under a sequence of shrinking location alternatives have been carried out. It is shown that the proposed test outperforms the competing tests for some processes with heavy-tailed distributions, and all these tests have comparable powers for some well-known Gaussian processes. An empirical study of the robustness and the finite sample powers of this test and some of the other tests in the literature have also been conducted, which further establishes the superiority of the proposed test over many of the two sample tests available in the literature. The contents of `Chapter 2` are partially based on Chakraborty and Chaudhuri (2014d).

Univariate medians and quantiles have been extended in a number of ways for data and distributions in finite dimensional spaces (see, e.g., Small (1990), Chaudhuri (1996) and Koltchinskii (1997) for reviews). However, many of these well-known multivariate medians do not have meaningful extensions into infinite dimensional spaces that can be used to analyze data lying in those spaces. On the other hand, the spatial median and the spatial quantiles extend easily into infinite dimensional spaces (see, e.g., Valadier (1984), Kemperman (1987) and Chaudhuri (1996)). Recently, there have been some attempts to use spatial median type estimators for analyzing various real functional datasets (see, e.g., Gervini (2008), Chaouch and Goga (2012) and Cardot et al. (2013)). However, not much is known about the properties of spatial quantiles in infinite dimensional spaces. In finite dimensional spaces, spatial distributions have been extensively studied (see, e.g., Koltchinskii (1997), Möttönen et al. (1997) and Oja (2010)), and they can be viewed as inverses of spatial quantiles. In `Chapter 3`, we have investigated the notion of spatial distributions in infinite dimensional spaces along with those of the associated quantiles. Glivenko-Cantelli and Donsker type results have also been established for empirical spatial distribution processes, which arise from data lying in such spaces. We have investigated a version of empirical spatial quantile in infinite dimensional spaces, and a Bahadur-type asymptotic linear representation of this estimator along with its weak convergence have been obtained. A study of the asymptotic efficiency of the sample spatial median relative to the sample mean is presented for some well-known probability models in infinite dimensional spaces. It is shown that the sample spatial median is asymptotically more efficient relative to the sample mean for a class of processes with heavy-tailed distributions. The contents of `Chapter 3` are partially based on Chakraborty and Chaudhuri (2014c).

Several nonparametric statistical procedures for finite dimensional data have been developed using different notions of depth functions (see, e.g., Donoho and Gasko (1992), Liu et al. (1999) and Ghosh and Chaudhuri (2005)). A depth function provides a center-outward ordering of the points in the sample space relative to a given probability distribution (see, e.g., Zuo and Serfling (2000) for some extensive review). There have been some work on developing depth functions for probability distributions in function spaces. Fraiman and Muniz (2001) studied a depth function, called the integrated data depth, for probability distributions in $C[0,1]$. Recently, López-Pintado and Romo (2009, 2011) studied the band depth, the half-region depth and their modified versions for probability distributions in such spaces. Some of these depth functions have been used to develop nonparametric statistical methods for functional data (see, e.g., Fraiman and Muniz (2001), López-Pintado and Romo (2009) and Sun and Genton (2011)).

We have investigated the behaviour of various depth functions for some standard probability models that are widely used for data in infinite dimensional spaces, and this forms a part of the contents of `Chapter 4`. We have proved that the band depth and the half-region depth have degenerate behaviour for many standard probability distributions in function spaces. Further, the half-space depth (see, e.g., Donoho and Gasko (1992)) and the projection depth (see, e.g., Zuo and Serfling (2000)), which can be defined in infinite dimensional spaces have similar degenerate behaviour for many probability distributions in such spaces. On the other hand, we have shown that the modified versions of the band depth and the half-region depth, the integrated data depth and the infinite dimensional extension of the finite dimensional version of the spatial depth (see, e.g., Vardi and Zhang (2000) and Serfling (2002)) do not suffer from any such degenerate behaviour for similar probability distributions. We have also investigated the asymptotic properties of the empirical spatial depth in infinite dimensional spaces.

Associated with any depth function is the deepest point, which is the point in the sample space, where the depth function achieves its maximum value. The deepest point associated with various depth functions have been studied in finite dimensions, and it is a popular choice of the center or the median of a multivariate probability distribution (see, e.g., Small (1990) and Donoho and Gasko (1992)). It turns out that many of these deepest points do not have any meaningful extensions into infinite dimensional spaces. In `Chapter 4`, we have investigated the deepest points associated with some of the depth functions mentioned in the previous paragraph. We have proved that the integrated data depth, the infinite dimensional extension of the spatial depth, and the modified versions of both the band depth and the half-region depth yield statistically meaningful notions of deepest points for a large class of probability distributions in infinite dimensional spaces. The breakdown points of the deepest points based on these depth functions have been studied. The asymptotic consistency of the corresponding empirical deepest points have also been investigated. The contents of `Chapter 4` are partially based on Chakraborty and Chaudhuri (2014a) and Chakraborty and Chaudhuri (2014b).

The theory developed in this thesis and the proofs of the results involve several concepts from functional analysis, convex analysis in infinite dimensional spaces, and probability distributions in Banach spaces. For the sake of better readability and ready reference, we have included the relevant definitions and concepts as part of the `Appendix`.

Chapter 2

A Wilcoxon–Mann–Whitney type test for infinite dimensional data

For testing the equality of means of two functional datasets, Horváth et al. (2013) proposed two test statistics based on orthogonal projections of the difference between the sample mean functions. One of those statistics is same as Hotelling's T^2 statistic based on a finite number of such projections. The two sample version of the test statistics studied by Cuevas et al. (2004) and Zhang and Chen (2007) reduce to the L_2-norm of the difference between the sample mean functions, and this statistic was also studied by Zhang et al. (2010) for testing the equality of two mean functions. However, all of the above-mentioned tests for functional data perform poorly when the observations have non-Gaussian distributions with heavy-tails or the samples are contaminated with outliers. In a different direction, Bai and Saranadasa (1996), Fan and Lin (1998), Chen and Qin (2010) and Srivastava et al. (2013) studied some tests for comparing the means of two finite dimensional datasets for which the data dimension is larger than the sample size, and it grows with the sample size. These authors, however, worked in a setup, which is different from the infinite dimensional setup considered in this chapter.

In this chapter, we develop and study a Wilcoxon–Mann–Whitney type test for data lying in infinite dimensional spaces. For univariate data, the Wilcoxon–Mann–Whitney test is known to have better power than the t-test for several non-Gaussian distributions (see, e.g., Hájek et al. (1999)). Various extensions of the Wilcoxon–Mann–Whitney test have been studied for multivariate data in finite dimensional spaces (e.g., Puri and Sen (1971), Randles and Peters (1990), Liu and Singh (1993), Choi and Marden (1997), Chakraborty and Chaudhuri (1999) and Oja (1999)), and these extensions too outperform Hotelling's T^2 test for several non-Gaussian multivariate distributions. However, some of the Wilcoxon–Mann–Whitney type tests for finite dimensional data in $\mathbb{R}^d$ like those defined using simplices (e.g., Liu and Singh (1993) and Oja (1999)) or those based on interdirections (e.g., Randles and Peters (1990)), cannot be extended into infinite dimensional spaces due to their dependence on the finite dimensional coordinate system in $\mathbb{R}^d$. Further, a test that involves standardization by some covariance matrix computed from the sample (e.g., Puri and Sen (1971) and Oja (1999)) is challenging to extend for data lying in infinite dimensional spaces. This is because such an empirical covariance operator usually converges to a compact operator, which does not have a bounded inverse. So, one has to properly regularize the inverse of that empirical covariance operator as the sample size grows.

2.1 The test and its asymptotic distribution in Hilbert spaces

Let $\mathbf{X}$ be a random element in a separable Hilbert space $\mathcal{X}$. For any nonzero $\mathbf{x} \in \mathcal{X}$, denote $S_{\mathbf{x}} = \mathbf{x}/||\mathbf{x}||$, where $||\cdot||$ is the norm in $\mathcal{X}$ associated with the inner product $\langle\cdot,\cdot\rangle$ in $\mathcal{X}$. We define $S_{\mathbf{x}} = \mathbf{0}$ if $\mathbf{x} = \mathbf{0}$. The spatial rank of $\mathbf{x} \in \mathcal{X}$ with respect to the distribution of $\mathbf{X}$ is defined as $\Psi_{\mathbf{x}} = E(S_{\mathbf{x}-\mathbf{X}})$. Henceforth, the expectation of any random element in any Hilbert space will be in the Bochner sense. The spatial rank defined in this way has been studied in $\mathbb{R}^d$ by Chaudhuri (1996), Choi and Marden (1997), Oja (2010) and Hettmansperger and McKean (2011).

Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_m$ and $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_n$ be independent observations from two probability measures P and Q on $\mathcal{X}$. Define $\mu = E(S_{\mathbf{Y}-\mathbf{X}})$. Then the hypothesis $H_0 : \mu = 0$ is equivalent to the hypothesis that the spatial median of $\mathbf{Y} - \mathbf{X}$ is zero. If Q differs from P by a shift Δ in the location, then H_0 becomes the hypothesis $\Delta = 0$. When P and Q are symmetric about $\mathbf{m}_1$ and $\mathbf{m}_2$, their spatial medians are $\mathbf{m}_1$ and $\mathbf{m}_2$, which are also their means if the means exist. In this case, H_0 is equivalent to the hypothesis

$\mathbf{m}_1 = \mathbf{m}_2$. Our Wilcoxon–Mann–Whitney type statistic for testing $H_0 : \mu = 0$ against $H_1 : \mu \neq 0$ is defined as

$$W = (mn)^{-1} \sum_{i=1}^{m} \sum_{j=1}^{n} S_{\mathbf{Y}_j - \mathbf{X}_i} = (mn)^{-1} \sum_{i=1}^{m} \sum_{j=1}^{n} (\mathbf{Y}_j - \mathbf{X}_i)/||\mathbf{Y}_j - \mathbf{X}_i||. \tag{2.1}$$

Note that W is a Hilbert space valued U-statistic (e.g., Borovskikh (1996)) and is an unbiased estimator of μ. We reject the null hypothesis for large values of $||W||$. It is straightforward to verify that for any $c \in \mathbb{R}$, $\mathbf{a} \in \mathcal{X}$ and a bijective linear isometry B on $\mathcal{X}$, the hypotheses H_0, H_1, and the test statistic W remain invariant under the transformation $\mathbf{X} \mapsto c\mathsf{B}(\mathbf{X}) + \mathbf{a}$ and $\mathbf{Y} \mapsto c\mathsf{B}(\mathbf{Y}) + a$. If $\mathcal{X} = \mathbb{R}$, then $S_{\mathbf{x}} = sign(\mathbf{x})$, and W reduces to the univariate Wilcoxon–Mann–Whitney statistic. The statistic W can be easily computed using (2.1). We shall see later in this chapter that this test is robust against outliers in the data, it does not require any moment assumption unlike mean based tests and has good performance for heavy-tailed distributions.

We now study the asymptotic distribution of W. Define $\Gamma_1, \Gamma_2 : \mathcal{X} \to \mathcal{X}$ as

$$\begin{aligned} \Gamma_1 &= E\{E(S_{\mathbf{Y}-\mathbf{X}} \mid \mathbf{X}) \otimes E(S_{\mathbf{Y}-\mathbf{X}} \mid \mathbf{X})\} - \mu \otimes \mu, \\ \Gamma_2 &= E\{E(S_{\mathbf{Y}-\mathbf{X}} \mid \mathbf{Y}) \otimes E(S_{\mathbf{Y}-\mathbf{X}} \mid \mathbf{Y})\} - \mu \otimes \mu. \end{aligned}$$

Here, $\otimes$ denotes the tensor product in $\mathcal{X}$ (see the `Appendix`). So, Γ_1 and Γ_2 are continuous linear symmetric positive operators. We denote by $G(\mathbf{m}, \mathsf{C})$ the distribution of a Gaussian random element in a separable Hilbert space $\mathcal{X}$ with mean $\mathbf{m}$ and covariance C (see the `Appendix`).

Theorem 2.1.1. *Let $N = m + n$ and $m/N \to \gamma \in (0, 1)$ as $m, n \to \infty$. Then, for any two probability measures P and Q on $\mathcal{X}$, $(mn/N)^{1/2}(W - \mu)$ converges weakly to $G\{\mathbf{0}, (1-\gamma)\Gamma_1 + \gamma\Gamma_2\}$ as $m, n \to \infty$.*

Note that in the above theorem, we do not assume the existence of moments. Let c_α denote the $100(1-\alpha)$th percentile of the distribution of $||G(\mathbf{0}, (1-\gamma)\Gamma_1 + \gamma\Gamma_2)||$. If the test rejects H_0 for $||(mn/N)^{1/2}W|| > c_\alpha$, its asymptotic size will be α. It follows from Theorem 2.1.1 that the asymptotic power of our test will be 1 whenever $\mu \neq 0$. Suppose that both P and Q are symmetric, then our test is consistent whenever the two spatial medians are different. In particular, our test is consistent for location shift alternatives.

We now describe how to compute the critical value of our test statistic from a given sample. Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_m$ and $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_n$ be two samples taking values in a separable Hilbert space $\mathcal{X}$ with norm $||\cdot||$ induced by the inner product $\langle \cdot, \cdot \rangle$ in $\mathcal{X}$. We have already mentioned earlier that the statistic W can be easily computed using (2.1). The operator Γ_1 defined before Theorem 2.1.1 is estimated by

$$\widehat{\Gamma}_1 = \frac{1}{m-1}\left\{\sum_{i=1}^{m}\left(\frac{1}{n}\sum_{j=1}^{n}\frac{\mathbf{Y}_j - \mathbf{X}_i}{||\mathbf{Y}_j - \mathbf{X}_i||} - \widehat{\mu}\right) \otimes \left(\frac{1}{n}\sum_{j=1}^{n}\frac{\mathbf{Y}_j - \mathbf{X}_i}{||\mathbf{Y}_j - \mathbf{X}_i||} - \widehat{\mu}\right)\right\},$$

where $\widehat{\mu} = (mn)^{-1}\sum_{i=1}^{m}\sum_{j=1}^{n}(\mathbf{Y}_j - \mathbf{X}_i)/||\mathbf{Y}_j - \mathbf{X}_i||$ is the empirical version of μ. We estimate Γ_2 by $\widehat{\Gamma}_2$, which is defined similarly. Note that the asymptotic distribution of $||(mn/N)^{1/2}W||^2$ is a weighted sum of independent chi-square random variables each with one degree of freedom and the weights are the eigenvalues of the $(1-\gamma)\Gamma_1 + \gamma\Gamma_2$. This representation follows from the spectral decomposition of the compact self-adjoint operator $(1-\gamma)\Gamma_1 + \gamma\Gamma_2$ and can be deduced from Theorem IV.2.4 in page 213 and Proposition 1.9 in page 161 in Vakhania et al. (1987). The eigenvalues of $(1-\gamma)\Gamma_1 + \gamma\Gamma_2$ can be estimated by the eigenvalues of $(1-\gamma)\widehat{\Gamma}_1 + \gamma\widehat{\Gamma}_2$. The critical value c_α can now be obtained by simulating from the estimated asymptotic distribution of W. In practice, the sample observations are often obtained as finite dimensional approximations of elements in $\mathcal{X}$, e.g., functions observed on a finite grid of points on an interval or linear combinations of finitely many fixed elements of an orthonormal basis of $\mathcal{X}$. Then, in order to compute W, $\widehat{\Gamma}_1$ and $\widehat{\Gamma}_2$, we use these sample observations and the appropriate finite dimensional inner products and norms.

2.2 Asymptotic powers of different tests under shrinking location shifts

In the previous section, we have established the consistency of our test for models with fixed location shifts. We shall now derive the asymptotic distribution of our test statistic under appropriate sequences of shrinking location shifts. Recall that $N = m + n$ is the total size of the two samples. We assume the following.

(A) $\mathbf{Y}$ is distributed as $\mathbf{X} + \Delta_N$, where $\Delta_N = \delta(mn/N)^{-1/2}$ for some fixed nonzero $\delta \in \mathcal{X}$ and $N \geq 1$.

For some of the Wilcoxon–Mann–Whitney type tests studied in the finite dimensional setting, such alternative hypotheses have been shown to be contiguous to the null, and this leads to nondegenerate limiting distributions of the test statistics under those alternatives (e.g., Choi and Marden (1997), Chakraborty and Chaudhuri (1999) and Oja (1999)). For $\mathbf{h} \in \mathcal{X}$, define $\mathsf{J}_{\mathbf{x}} : \mathcal{X} \to \mathcal{X}$ as

$$\mathsf{J}_{\mathbf{x}}(\mathbf{h}) = E(||\mathbf{Y} - \mathbf{X} + \mathbf{x}||^{-1})\mathbf{h} - E\{\langle \mathbf{x} + \mathbf{Y} - \mathbf{X}, \mathbf{h}\rangle(\mathbf{x} + \mathbf{Y} - \mathbf{X})/||\mathbf{x} + \mathbf{Y} - \mathbf{X}||^3\}. \quad (2.2)$$

Theorem 2.2.1. *As before, let $N = m + n$, $m/N \to \gamma \in (0, 1)$ as $m, n \to \infty$. Also, assume that the distribution of $\mathbf{X}$ is nonatomic and $\mathsf{J}_{\mathbf{0}}$ exists. Then, under Assumption (A) described above, $(mn/N)^{1/2}W$ converges weakly to $G\{\mathsf{J}_{\mathbf{0}}(\delta), \Gamma_1\}$ as $m, n \to \infty$. Here, the expectation in $\mathsf{J}_{\mathbf{0}}$ is with respect to the common distribution of $\mathbf{X}$ and $\mathbf{Y}$ under the null hypothesis.*

Let $\mathbf{Y} - \mathbf{X} = \sum_{k=1}^{\infty} V_k \phi_k$ for an orthonormal basis $\{\phi_k\}_{k \geq 1}$ of $\mathcal{X}$. Then, the expectation defining $\mathsf{J}_{\mathbf{0}}$ is finite if any two dimensional marginal of $(V_1, V_2, \ldots)$ has a density that is bounded on bounded subsets of $\mathbb{R}^2$.

In order to compare the asymptotic power of our test with those of the tests available in Cuevas et al. (2004) and Horváth et al. (2013), we shall now study the asymptotic distributions of those test statistics under the sequences of shrinking shifts described in Assumption (A) above. For the two sample problem in $L_2[a, b]$, the test statistic studied by Cuevas et al. (2004) reduces to $T_{\text{CFF}} = m||\bar{\mathbf{X}} - \bar{\mathbf{Y}}||^2$. Horváth et al. (2013) studied the test statistics $T_{\text{HKR1}} = \sum_{k=1}^{L}(\langle \bar{\mathbf{X}} - \bar{\mathbf{Y}}, \widehat{\psi}_k\rangle)^2$ and $T_{\text{HKR2}} = \sum_{k=1}^{L} \widehat{\lambda}_k^{-1}(\langle \bar{\mathbf{X}} - \bar{\mathbf{Y}}, \widehat{\psi}_k\rangle)^2$. Here, the $\widehat{\lambda}_k$'s denote the eigenvalues of the empirical pooled covariance of the $\mathbf{X}_i$'s and the $\mathbf{Y}_j$'s in descending order of magnitudes, and the $\widehat{\psi}_k$'s are the corresponding empirical eigenfunctions. If $\mathcal{X} = \mathbb{R}^d$ and $L = d$, T_{HKR2} reduces to Hotelling's T^2 statistic, and $T_{\text{HKR1}} = m^{-1}T_{\text{CFF}}$.

Theorem 2.2.2. *Let $N = m + n$ and $m/N \to \gamma \in (0, 1)$ as $m, n \to \infty$. Then, under Assumption (A) described above, we have the following.*

(a) If $E(||\mathbf{X}||^2) < \infty$, $nN^{-1}T_{\text{CFF}}$ *converges weakly to* $\sum_{k=1}^{\infty} \lambda_k \chi^2_{(1)}(\beta_k^2/\lambda_k)$ *as* $m, n \to \infty$. *Here, the* λ_k*'s are the eigenvalues of the covariance* Σ *of* $\mathbf{X}$ *in decreasing order of magnitudes and the* ψ_k*'s are the eigenfunctions corresponding to the* λ_k*'s,* $\beta_k = \langle \delta, \psi_k \rangle$, *and* $\chi^2_{(1)}(\beta_k^2/\lambda_k)$ *denotes the noncentral chi-square variable with* 1 *degree of freedom and noncentrality parameter* β_k^2/λ_k *for* $k \geq 1$.
(b) Assume that for some $L \geq 1$, $\lambda_1 > \ldots > \lambda_L > \lambda_{L+1} > 0$. *If* $E(||\mathbf{X}||^4) < \infty$, $mnN^{-1}T_{\text{HKR1}}$ *converges weakly to* $\sum_{k=1}^{L} \lambda_k \chi^2_{(1)}(\beta_k^2/\lambda_k)$, *and* $mnN^{-1}T_{\text{HKR2}}$ *converges weakly to* $\sum_{k=1}^{L} \chi^2_{(1)}(\beta_k^2/\lambda_k)$ *as* $m, n \to \infty$.

For evaluating the asymptotic powers of different tests under shrinking location shifts, we have considered the random element in $\mathcal{X}$ defined as

$$\mathbf{X} = \sum_{k=1}^{\infty} 2^{1/2}\{(k-0.5)\pi\}^{-1} Z_k \sin\{(k-0.5)\pi t\}, \tag{2.3}$$

where the Z_k's are independent random variables for $k \geq 1$. We have considered two cases, namely, Z_k's having standard normal distributions, and $Z_k = U_k(V/5)^{-1/2}$, where the U_k's are independent standard normal variables and V has a chi-square distribution with 5 degrees of freedom independent of the U_k's for each $k \geq 1$. Both of these distributions satisfy the assumptions made in Theorems 2.2.1 and 2.2.2. These two cases correspond to the Karhunen-Loève expansions of the standard Brownian motion and the centered t process (e.g., Yu et al. (2007)) on $[0, 1]$ with 5 degrees of freedom and covariance kernel $K(t, s) = \min(t, s)$, respectively. We call them the sBm and the t(5) distributions, respectively. Recall that $\mathbf{Y}$ is distributed as $\mathbf{X} + \delta(mn/N)^{-1/2}$, and we consider three choices of δ, namely, $\delta_1(t) = c$, $\delta_2(t) = ct$ and $\delta_3(t) = ct(1-t)$, where $t \in [0, 1]$ and $c > 0$. For evaluating the asymptotic powers of different tests using Theorems 2.2.1 and 2.2.2, Γ_1 is estimated in a similar way as described in Section 2.1 using 1000 sample functions from the underlying distribution. The operator Σ is estimated by the sample covariance operator of $\mathbf{X}$. The eigenvalues and the eigenfunctions of Γ_1 and Σ are estimated by the eigenvalues and the eigenfunctions of their empirical counterparts. Also, $J_0(\delta)$ is estimated by its sample analog. The asymptotic powers are then computed by simulating from the appropriate asymptotic Gaussian distribution with these estimated parameters. For the tests based on T_{HKR1} and T_{HKR2}, the number

L is chosen using the cumulative variance method described in Horváth et al. (2013). Figure (2.1) shows the plots of the asymptotic powers of the tests based on W, T_{CFF}, T_{HKR1} and T_{HKR2} for the sBm and the t(5) distributions under shrinking location shifts.

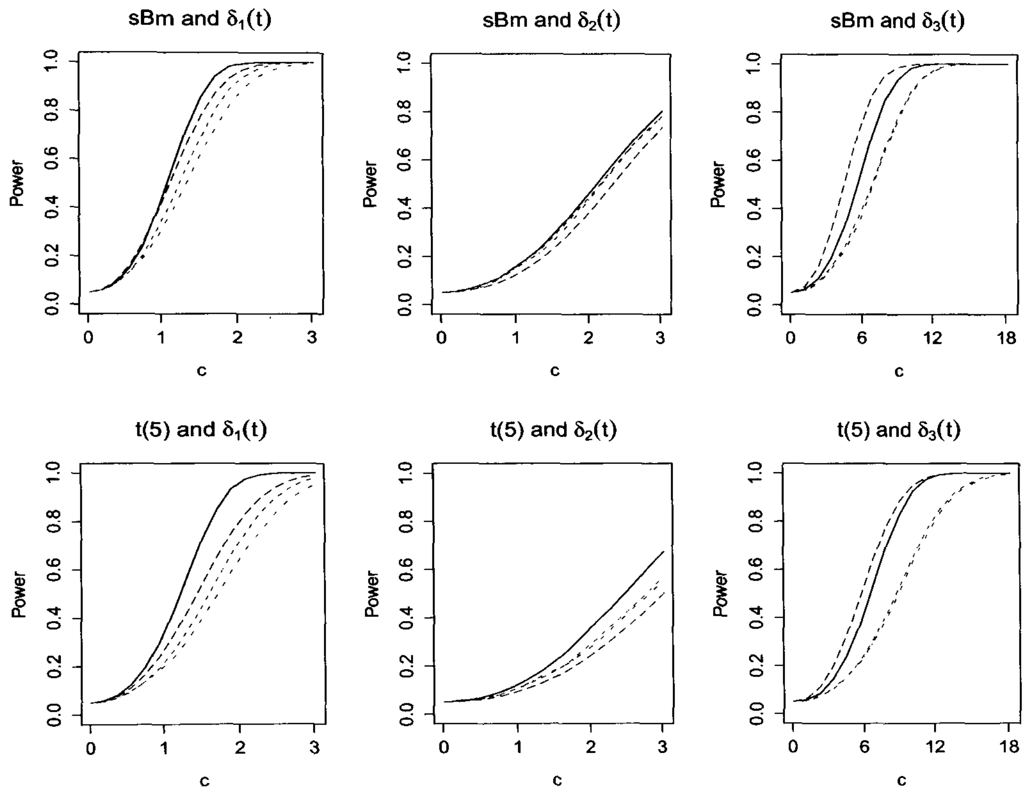

Figure 2.1: Plots of the asymptotic powers of the tests based on W (black curves), T_{CFF} (red - - curves), T_{HKR1} (violet - · - curves) and T_{HKR2} (green – – curves) with nominal level 5%.

It is seen from Figure 2.1 that our test based on W outperforms the tests based on T_{CFF} and T_{HKR1} for all the models considered except $\delta_2(t)$ under the sBm distribution, in which case, the powers of the three tests are similar. Our test outperforms the test based on T_{HKR2} for $\delta_2(t)$ under both the distributions, $\delta_1(t)$ under the t(5) distribution, and $\delta_1(t)$ under the sBm distribution for $c \geq 1$. Since the t(5) distribution is heavy-tailed, it has an adverse effect on the performances of the mean based tests, but our test based on spatial ranks remains unaffected. For the shift $\delta_3(t)$ under both the distributions, the test based on T_{HKR2} has significantly better performance than our test, but even in this case, its performance degrades significantly for the t(5) distribution as compared to the sBm distribution, but that does not happen for our test.

2.3 Empirical study of different tests

In this section, we will carry out a comparative study of the finite sample performances of various tests including the test based on W in terms of their robustness in presence of contamination in the samples, their empirical powers for simulated data, and their performances on real data.

2.3.1 Robustness study of different tests

It is known that the univariate Wilcoxon–Mann–Whitney test is robust against contamination of the samples by outliers while the tests based on the sample means like the t-test do not have such robustness. The influence function of the univariate Wilcoxon–Mann–Whitney test is bounded, which implies that the maximal asymptotic level and the minimal asymptotic power even for moderately high contamination proportions are not much different than those without contamination (see Chapter 7 in Hampel et al. (1986)). It can be shown that the influence function of our test based on W is $\mathsf{J}_{\mathbf{Q(0)}}\{E(S_{\mathbf{X}-\mathbf{y}} + S_{\mathbf{x}-\mathbf{Y}})\}$, where $\mathbf{x}, \mathbf{y} \in \mathcal{X}$, J is defined in Theorem 2.2.1, and $\mathbf{Q(0)}$ is the spatial median of the distribution of $\mathbf{Y} - \mathbf{X}$ for an independent copy $\mathbf{Y}$ of $\mathbf{X}$. This influence function is bounded in norm under the conditions of Proposition 2.1 in Cardot et al. (2013). Thus, it is expected that our test will be robust like the univariate Wilcoxon–Mann–Whitney test when the samples are contaminated by outliers. We now compare our test in terms of its size and power under contaminated data with the tests based on T_{CFF}, T_{HKR1}, T_{HKR2} as well as the two sample version of the test for linear models studied by Shen and Faraway (2004). We denote by T_{SF} the test statistic of the test studied by Shen and Faraway (2004).

The distribution of the uncontaminated data is taken to be the sBm distribution described in Section 2.2. We have considered contamination in the mean as well as in the covariance of the sBm distribution. For the mean contamination, we used the Gaussian

Table 2.1: Sizes of some tests with nominal level 5%

	Contamination proportion	W	T_{CFF}	T_{HKR1}	T_{HKR2}	T_{SF}
	0	5.0	4.9	4.9	6.1	3.1
Mean contamination in both the samples	1/15	5.6	35.2	34.6	7.0	3.4
	3/15	3.4	55.4	54.6	6.2	4.8
	5/15	5.6	56.0	55.4	6.2	5.6
Covariance contamination in both the samples	1/15	5.0	16.1	15.1	5.0	2.3
	3/15	3.7	33.9	30.7	4.8	2.2
	5/15	3.6	48.9	43.8	4.2	3.0
Covariance contamination in only one sample	1/15	4.8	11.2	10.4	5.3	3.1
	3/15	5.1	22.3	21.0	4.9	3.0
	5/15	3.7	31.8	29.0	4.5	2.1

distribution with mean function $\Delta(t) = 5$ and covariance kernel $K(t,s) = \min(t,s)$, $t, s \in [0,1]$. The covariance contamination is done using the zero mean Gaussian distribution with covariance kernel $K(t,s) = 16\min(t,s)$, $t,s \in [0,1]$. We have chosen $m = n = 15$, and each sample curve is observed at 250 equispaced points in $[0,1]$. The contamination proportions chosen are 1/15, 3/15 and 5/15. To keep the null hypothesis unchanged under mean contamination, we consider mean contamination in both the samples. The null hypothesis remains unaltered if one or both the samples have covariance contamination. For the power study under contamination, we consider the location shift $\Delta(t) = 0.6$, $t \in [0,1]$. The sizes and the powers of the tests are averaged over 1000 Monte Carlo simulations, and they are provided in Tables (2.1) and (2.2), respectively.

The size of our test based on spatial ranks is not significantly different from the nominal 5% level for all the contamination models considered. The sizes of the mean based tests using T_{CFF} and T_{HKR1} are larger than 34% in all cases of mean contamination. Their sizes are larger than 10% when one of the samples have covariance contamination, and these are comparatively higher when both the samples have covariance contamination. The size of the test using T_{SF} is always significantly smaller than the nominal

Table 2.2: Powers of some tests with nominal level 5%

	Contamination proportion	W	T_{HKR2}	T_{SF}
	0	90.9	86.2	70.4
Mean contamination in both the samples	1/15	71.2	32.0	27.6
	3/15	44.4	12.2	11.6
	5/15	26.6	8.8	9.0
Covariance contamination in both the samples	1/15	82.3	62.6	49.0
	3/15	63.9	39.1	25.2
	5/15	43.2	26.6	16.5
Covariance contamination only in the sample from the null	1/15	85.5	70.7	57.9
	3/15	77.3	53.8	40.9
	5/15	62.5	41.4	29.1
Covariance contamination only in the sample from the alternative	1/15	85.9	74.7	62.1
	3/15	76.7	54.4	41.5
	5/15	64.4	40.1	27.9

level for the contaminated models as well as the uncontaminated models considered later. The numerators and the denominators in the statistics T_{HKR2} and T_{SF} are possibly affected in a similar way in the presence of outliers, while the ratios as well as the sizes of the resulting tests remain relatively unaffected.

Since the tests based on T_{CFF} and T_{HKR1} have very high sizes under all the contamination models considered, we did not include them in the power study. Although there is a reduction in the powers of all the tests considered as the contamination proportion increases, this reduction is the least for our test based on spatial ranks among all the competing tests for all the models considered. The reduction in power is due to the increase in the variability of the sample as the contamination proportion increases. The effect of contamination on the mean based tests is more pronounced under mean contamination, when the powers of these tests become close to their sizes if the proportion of mean contamination is 1/3 or more in both the samples. In this situation, the mean shift is masked by the mean contamination, which is in the same direction as the mean shift but has a larger magnitude.

2.3.2 Empirical power study of different tests using simulated data

We shall now carry out a comparative study of the finite sample empirical powers of the tests considered in the previous subsection as well as some more tests available in the literature. A pointwise t test with an appropriate p-value correction for multiple testing was studied by Cox and Lee (2008) for testing the equality of means of two Gaussian functional datasets. Cuesta-Albertos and Febrero-Bande (2010) studied an analysis of variance test for functional data based on multiple testing using random univariate linear projections of the data. The two sample version of this test reduces to performing the univariate Wilcoxon–Mann–Whitney test for each projection followed by a p-value adjustment, and we have used 30 random projection directions for this test. Gretton et al. (2012) studied tests for comparing two probability distributions on metric spaces. We have used their test based on the asymptotic distribution of the unbiased statistic MMD_u^2 studied in Section 5 in that paper, and have chosen the radial basis kernel for this test. For comparing two finite dimensional probability distributions, Hall and Tajvidi (2002) studied a permutation test based on the ranks of the distances between the sample observations, while Rosenbaum (2005) studied a test based on a notion of adjacency. We have used the L_2-distance between the pointwise ranks as the distance function for implementing the latter test. Since the test statistic has a discrete distribution, we randomized the test to improve its size and power.

The form of the distribution of $\mathbf{X}$ is given by (2.3) in Section 2.2. Here, we consider the sBm and the t(5) distributions as in Section 2.2 as well as the t(1) distribution, for which $Z_k = U_k/V^{1/2}$, where the U_k's are independent standard normal variables and V has a chi-square distribution with 1 degree of freedom independent of the U_k's for each $k \geq 1$. The t(1) distribution is included to investigate the performance of our test and its competitors when the moment conditions on the probability distribution required by some of the competing tests fail to hold. We have chosen $m = n = 15$, and each sample curve is observed at 250 equispaced points in $[0, 1]$. The distribution of $\mathbf{Y}$ is same as that of $\mathbf{X} + \Delta_k$, $k = 1, 2, 3$, where $\Delta_1(t) = c$, $\Delta_2(t) = ct$ and $\Delta_3(t) = ct(1 - t)$ for $t \in [0, 1]$ and $c > 0$. All the sizes and the powers are evaluated by averaging the results of 1000 Monte-Carlo simulations. The powers of the tests based on W, T_{CFF},

T_{HKR1} and T_{HKR2} are plotted in Figure 2.2. We denote the test statistics of the test studied by Cox and Lee (2008), Cuesta-Albertos and Febrero-Bande (2010), Gretton et al. (2012), Hall and Tajvidi (2002) and Rosenbaum (2005) by T_{CL}, T_{CAFB}, T_{G}, T_{HT} and T_{R} respectively. The powers of these tests and that of our test based on W are plotted in Figure 2.3.

There is no significant difference between the size of our test and the nominal 5% level in all the models considered. The sizes of the competing tests except those based on T_{CAFB} and T_{SF} are also not significantly different from the nominal 5% level for the sBm and the t(5) distributions. The sizes of the tests based on T_{CFF}, T_{HKR1} and T_{HKR2} are approximately 1.5% for the heavy-tailed t(1) distribution, while those of the remaining tests are not significantly different from the nominal 5% level. The size of the test using T_{CAFB} is around 2.6% in all our simulations. The size of the test based on T_{SF} is significantly smaller than the nominal 5% level for all the distributions considered, and it is zero for the t(1) distribution. The standard errors of the estimated sizes of these tests range from 0.0037 to 0.0078 for all the distributions.

The tests based on T_{CFF}, T_{HKR1} and T_{HKR2} have extremely low powers and are significantly outperformed by our test based on W in the case of the t(1) distribution. The nonexistence of moments severely affects the performance of these mean based tests, but our test based on spatial ranks is less affected. As in the asymptotic power study in Section 2.2, our test has significantly higher power than the tests based on T_{CFF} and T_{HKR1} for $\Delta_1(t)$ and $\Delta_3(t)$ under the t(5) distributions. Under the t(5) distribution, our test is significantly more powerful than the test using T_{HKR2} for $c \geq 0.4$ in $\Delta_1(t) = c$ and $c \geq 0.6$ in $\Delta_2(t) = ct$. Heavy tails of the t(5) distribution adversely affect the powers of the mean based tests associated with T_{CFF}, T_{HKR1} and T_{HKR2}. The powers of our test and the tests based on T_{CFF} and T_{HKR1} are comparable for $\Delta_2(t)$ under the t(5) distribution, and the test based on T_{HKR2} significantly outperforms our test for $\Delta_3(t)$ under both the sBm and the t(5) distributions. However, the performances of the mean based tests degrade significantly under the t(5) distribution, while our test based on spatial ranks is not affected. Even under the light-tailed sBm distribution, our test outperforms the tests based on T_{CFF} and T_{HKR1} for $c \geq 0.3$ in $\Delta_1(t) = c$ and $c \geq 1.8$ in $\Delta_3(t) = ct(1-t)$, and it is as powerful as the test based on T_{HKR2} for $\Delta_1(t)$ and $\Delta_2(t)$.

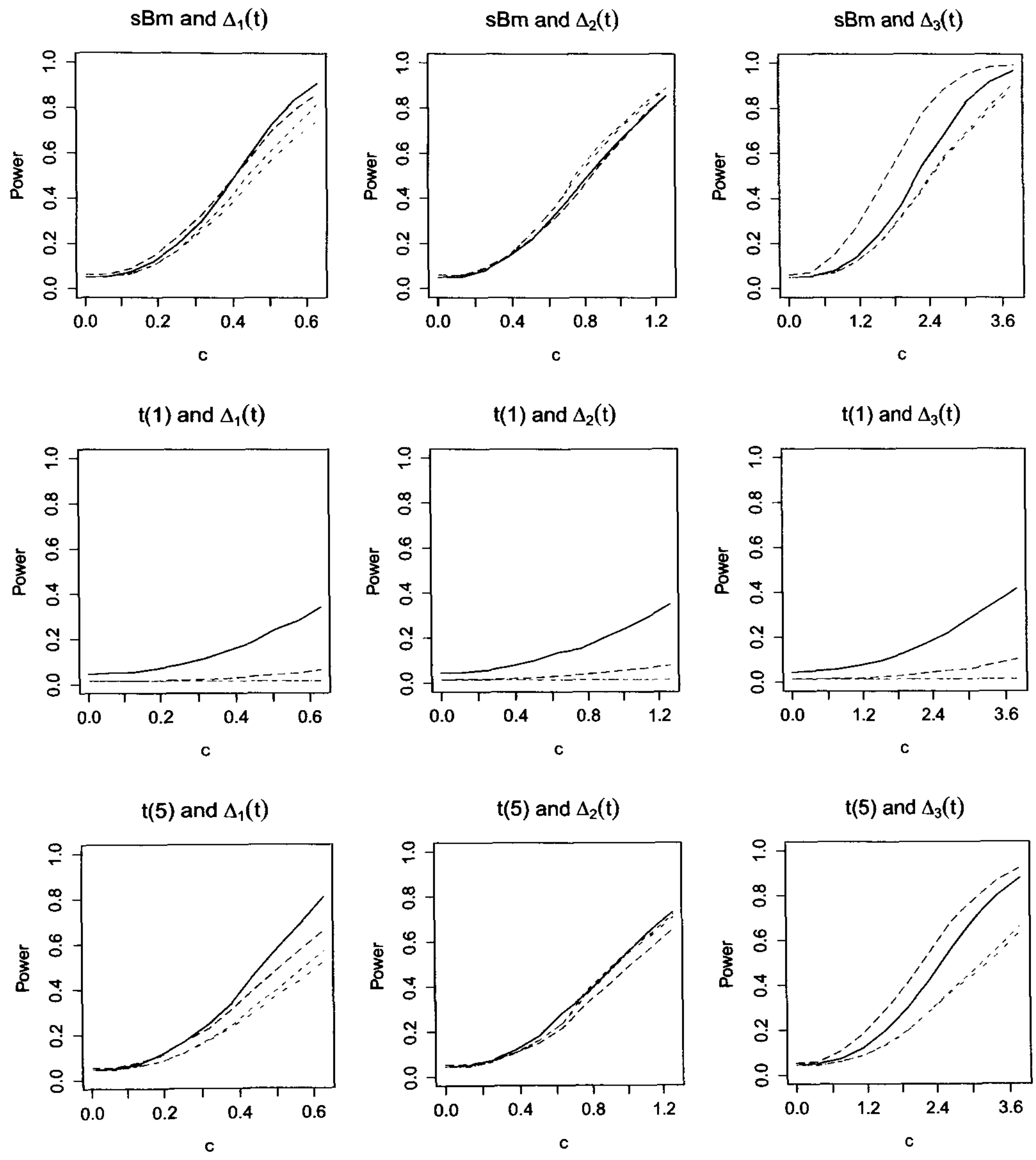

Figure 2.2: Plots of the finite sample powers of the tests based on W (black curves), T_{CFF} (red - - curves), T_{HKR1} (violet - · - curves) and T_{HKR2} (green – – curves) with nominal level of 5%.

But our test is significantly less powerful than the tests based on T_{CFF} and T_{HKR1} for $c \geq 0.7$ in $\Delta_2(t) = ct$.

Our test significantly outperforms the test based on T_{G} for $c \geq 0.2$ in $\Delta_1(t) = c$ and $c \geq 1.5$ in $\Delta_3(t) = ct(1-t)$ under the sBm and the t(5) distributions. Our test also outperforms this test for $c \geq 0.5$ in $\Delta_2(t) = ct$ under the sBm distribution. Our test is significantly more powerful than the test based on T_{CL} for $c \geq 2$ in $\Delta_3(t) = ct(1-t)$ under the t(1) distribution and for $c \geq 0.5$ in $\Delta_2(t) = ct$ under all the distributions. Further, our test is significantly more powerful than the tests using T_{SF}, T_{CAFB}, T_{HT} and T_{R} for almost all the shifts under the sBm and the t(5) distributions as well as for all large c values in the alternatives $\Delta_1(t) = c$, $\Delta_2(t) = ct$, and $\Delta_3(t) = ct(1-t)$ under the t(1) distribution. The test using T_{CAFB} is significantly less powerful than our test because the random projections failed to properly capture the directions of the location shifts. On the other hand, while the tests based on T_{HT}, T_{R} and T_{G} are consistent against general alternatives, which do not necessarily differ in their locations, such tests are often less powerful than tests for location like our test based on W, when the difference is only in the locations. Our test and the test based on T_{G} have comparable powers for $\Delta_1(t)$ and $\Delta_3(t)$ under the t(1) distribution and for $\Delta_2(t)$ under the t(5) distribution. The latter test is significantly more powerful than our test for $c \geq 0.7$ in $\Delta_2(t) = ct$ under the t(1) distribution. Our test is as powerful as the test based on T_{CL} for $\Delta_3(t)$ under the sBm and the t(5) distributions. However for $\Delta_1(t)$, the latter test is much more powerful than all the other tests for all of the distributions considered. This is because the coordinate random variable at $t = 0.0001$, which is closest to zero in our computations, has scale parameter equal to 0.0001. Consequently, for this coordinate and $\Delta_1(t)$, the adjusted p-values of the t test used in this test procedure are less than 0.05 for many of the simulations. The test based on T_{CL} rejects H_0 for such simulations resulting in the high power of this test. The standard errors of the estimated powers of all of these tests are of the order of 0.01 or less for all the distributions.

2.3.3 Analysis of the Spectrometry data

We have applied our test based on W as well as the other tests considered in the previous subsection to the Spectrometry data. This data is available at `http://www.math.univ-toulouse.fr/staph/npfda` and contains the spectrometric curves for 215 meat units measured at 100 wavelengths between 850 nm and 1050 nm. The data also contains the fat content of each meat unit, which is categorized into two classes, namely, "$\leq 20\%$" and "$> 20\%$". Each observation in this dataset can be viewed as an element in the

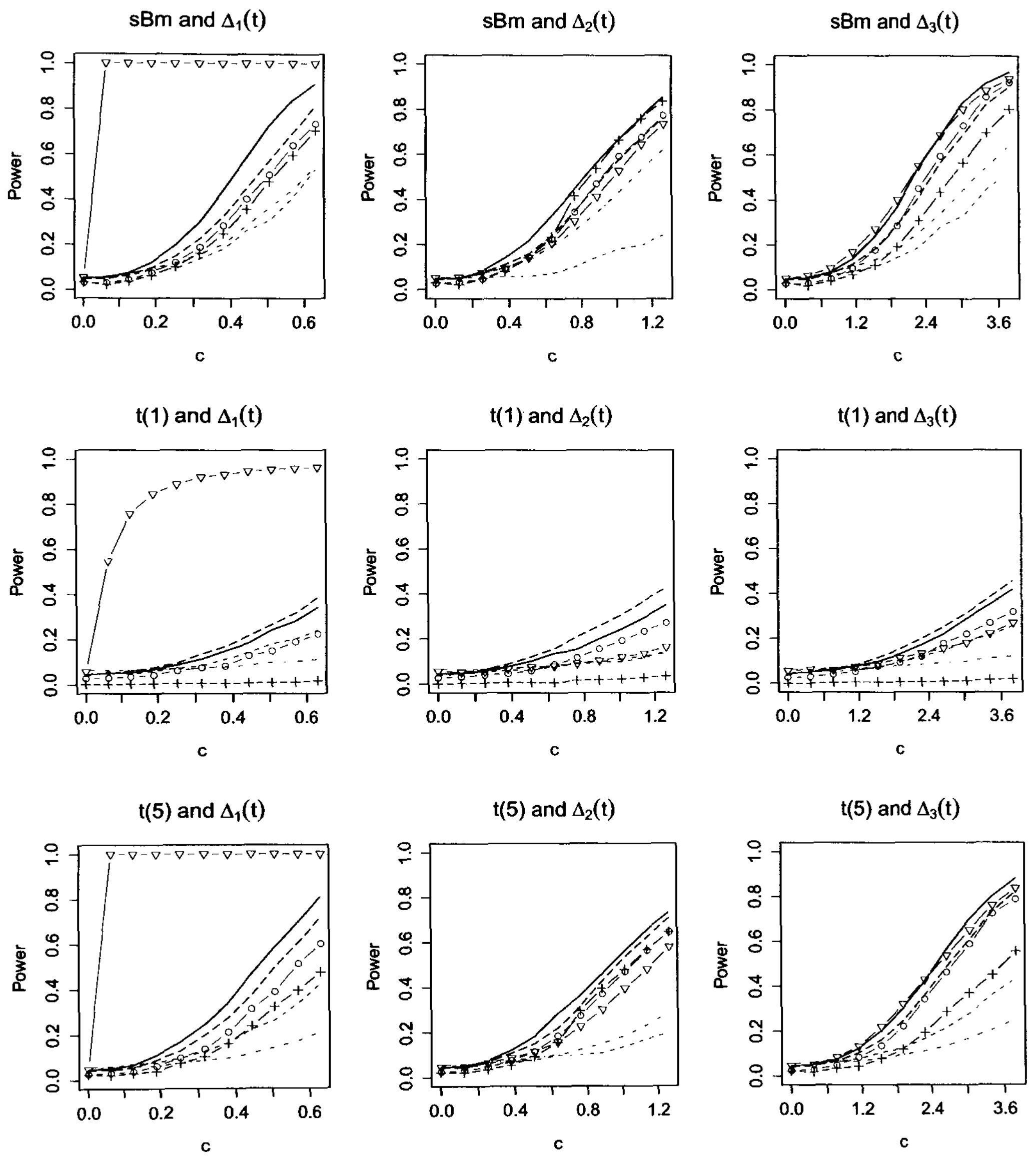

Figure 2.3: Plots of the finite sample powers of the tests based on W (black curves), T_{SF} (green - + - curves), T_{CL} (red - ∇ - curves), T_{CAFB} (purple - o - curves), T_{G} (blue – – curves), T_{HT} (brown - · - curves) and T_{R} (violet - - curves) with nominal level 5%.

separable Hilbert space $L_2[850, 1050]$.

For all the tests, the p-values are 0 up to two decimal places indicating significant difference in the two distributions. We have also applied these tests at nominal level 5%

to 1000 randomly chosen 20% subsamples of the dataset in order to evaluate their powers when the sample sizes are smaller. Our test significantly outperforms the competing tests, and its power is 0.832. The tests using T_{R} and T_{CL} have significantly less power than our test as they are based on coordinatewise ranks and coordinatewise t tests, respectively. They fail to properly capture the strong dependence structure in the data and have powers 0.452 and 0.462, respectively. The power of the test using T_{CAFB} is 0.406, which is also significantly less because the random projection directions failed to adequately capture the direction of the location shift. The powers of the tests using T_{G} and T_{HT} are 0.715 and 0.695, respectively, which are significantly smaller than that of our test. As discussed in subsection 2.3.2, these two tests are consistent for general alternatives, and they may perform worse than our test, when there is difference mainly in the locations of the two distributions. The powers of the tests based on T_{CFF}, T_{HKR1}, T_{HKR2} and T_{SF} are 0.712, 0.744, 0.778 and 0.709, respectively, and they are all significantly less powerful than our test. One of the advantages of our test based on spatial ranks is its robustness against outliers in the data unlike the mean based tests that use T_{CFF}, T_{HKR1}, T_{HKR2} and T_{SFF}. We now assess the impact of outliers in the data on these tests. We used the functional boxplot proposed in Sun and Genton (2011) for outlier detection. For comparing the effect of using different notions of functional depth on outlier detection, and hence on the performance of the tests, we considered two depth functions, namely, the modified band depth (López-Pintado and Romo (2009)) and the spatial depth (Chakraborty and Chaudhuri (2014b)) for this procedure. These two depth functions will be studied in detail in `Chapter 4`.

The functional boxplot using the modified band depth identified five observations from the class with fat content $\leq 20\%$ and three observations from the class with fat content $> 20\%$ as outliers. On the other hand, the functional boxplot using the spatial depth identified one more observation in the class with fat content $> 20\%$ as an outlier. So, the performance of the tests should not differ much for the two datasets. The plots of the original data along with the outliers (red curves) detected using both the procedures are given in Figure 2.4. We have marked the observation, which has been detected as an outlier only by the spatial depth as the red dotted curve.

The tests based on W, T_{CFF}, T_{HKR1}, T_{HKR2} and T_{SFF} are again applied to randomly

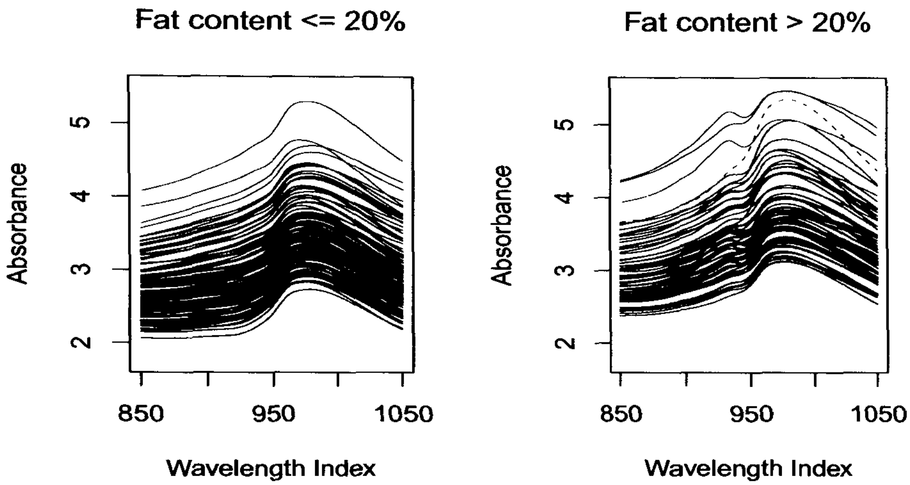

Figure 2.4: Plots of the samples curves of the Spectrometry data indicating the outlying curves.

chosen 20% subsamples of the two datasets obtained after removing the outliers detected by the two depth functions. The powers of the tests averaged over 1000 Monte-Carlo iterations are given in the following table, where FBP-MBD and FBP-SD denote the functional boxplot using the modified band depth and the spatial depth, respectively.

Table 2.3: Powers of some nominal 5% tests for the Spectrometry data.

Procedure	W	T_{CFF}	T_{HKR1}	T_{HKR2}	T_{SF}
FBP-MBD	0.870	0.851	0.862	0.868	0.809
FBP-SD	0.862	0.843	0.849	0.859	0.796

As expected and remarked earlier, there is no significant impact of the outlier detection procedure on the performance of the tests. Due to the robustness of our test based on spatial ranks, its power increases only marginally after removal of outliers. On the other hand, the lack of robustness of the mean based tests using T_{CFF}, T_{HKR1}, T_{HKR2} and T_{SF} is evident in the significant increase in their powers.

2.4 The test and its asymptotic distribution in general Banach spaces

Functional data are sometimes modelled as samples from probability distributions in an infinite dimensional Banach space. The Wilcoxon–Mann–Whitney type test discussed so far in this chapter can also be extended to the setup when data take values in a general Banach space. For this we will need some of the concepts introduced in the `Appendix`. Let $\mathcal{X}$ be a Banach space with norm $||\cdot||$, and denote its dual space by $\mathcal{X}^*$. Suppose that $\mathcal{X}$ is smooth, and denote the Gâteaux derivative of the norm $||\cdot||$ in $\mathcal{X}$ at a nonzero $\mathbf{x} \in \mathcal{X}$ by $S_{\mathbf{x}}$. The spatial rank of $\mathbf{x} \in \mathcal{X}$ with respect to the distribution of a random element $\mathbf{X} \in \mathcal{X}$ is defined as $\Psi_x = E(S_{\mathbf{x}-\mathbf{X}})$, where the expectation is in the Bochner sense. Hilbert spaces are smooth with $S_{\mathbf{x}} = \mathbf{x}/||\mathbf{x}||$. So, in a Hilbert space, $\Psi_{\mathbf{x}} = E\{(\mathbf{x}-\mathbf{X})/||\mathbf{x}-\mathbf{X}||\}$, and the spatial rank defined in this way coincides with the definition given in the first paragraph in Section 2.1 of this chapter. If $\mathcal{X} = L_p(I)$, where $I \subseteq \mathbb{R}^d$ for a $d \geq 1$, and $p \in (1,\infty)$, which is the Banach space of all functions $\mathbf{x} : I \to \mathbb{R}$ satisfying $\int_I |\mathbf{x}(\mathbf{s})|^p d\mathbf{s} < \infty$, then $S_{\mathbf{x}}(\mathbf{h}) = \int_I sign\{\mathbf{x}(\mathbf{s})\}|\mathbf{x}(\mathbf{s})|^{p-1}\mathbf{h}(\mathbf{s})ds/||\mathbf{x}||^{p-1}$ for all $\mathbf{x}, \mathbf{h} \in L_p(I)$.

Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_m$ and $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_n$ be independent observations from two probability measures P and Q on a smooth Banach space $\mathcal{X}$. Define $\mu = E(S_{\mathbf{Y}-\mathbf{X}})$. Our Wilcoxon–Mann–Whitney type statistic for testing the hypothesis $H_0 : \mu = \mathbf{0}$ against $H_1 : \mu \neq \mathbf{0}$ is defined as $W = (mn)^{-1}\sum_{i=1}^{m}\sum_{j=1}^{n} S_{\mathbf{Y}_j-\mathbf{X}_i}$.

Define $\Gamma_1, \Gamma_2 : \mathcal{X}^{**} \to \mathcal{X}^*$ to be the continuous linear symmetric positive operators given by

$$\Gamma_1(\mathbf{f}) = E[\mathbf{f}\{E(S_{\mathbf{Y}-\mathbf{X}} \mid \mathbf{X})\}E(S_{\mathbf{Y}-\mathbf{X}} \mid \mathbf{X})] - \mathbf{f}(\mu)\mu,$$

$$\Gamma_2(\mathbf{f}) = E[\mathbf{f}\{E(S_{\mathbf{Y}-\mathbf{X}} \mid \mathbf{Y})\}E(S_{\mathbf{Y}-\mathbf{X}} \mid \mathbf{Y})] - \mathbf{f}(\mu)\mu,$$

where $\mathbf{f} \in \mathcal{X}^{**}$. We denote by $G(\mathbf{m}, \mathsf{C})$ the distribution of a Gaussian random element in a separable Banach space $\mathcal{X}$ with mean $\mathbf{m}$ and covariance C. The next theorem gives the analog of Theorem 2.1.1 in general Banach spaces.

Theorem 2.4.1. *Let $N = m + n$ and $m/N \to \gamma \in (0,1)$ as $m, n \to \infty$. Also, assume* *that the dual space $\mathcal{X}^*$ is a separable and type 2 Banach space. Then, for any two probability measures P and Q on $\mathcal{X}$, $(mn/N)^{1/2}(W-\mu)$ converges weakly to $G\{\mathbf{0},(1-\gamma)\Gamma_1+\gamma\Gamma_2\}$ as $m,n\to\infty$.*

We next state the analog of Theorem 2.2.1 for data in general Banach spaces. Suppose that $\mathbf{Y}$ is distributed as $\mathbf{X}+\Delta_N$, where $\Delta_N=\delta(mn/N)^{-1/2}$ for some fixed nonzero $\delta\in\mathcal{X}$ and $N\geq 1$. This choice of shrinking alternatives have already been used earlier in this chapter. We assume that the norm in $\mathcal{X}$ is twice Gâteaux differentiable at every $\mathbf{x}\neq\mathbf{0}$. Let us denote the Hessian of the function $\mathbf{x}\mapsto E(||\mathbf{Y}-\mathbf{X}+\mathbf{x}||-||\mathbf{Y}-\mathbf{X}||)$ at $\mathbf{x}$ by $\mathsf{J}_{\mathbf{x}}:\mathcal{X}\to\mathcal{X}^*$ when it exists (see the `Appendix`). If $\mathcal{X}$ is a Hilbert space and $E(||\mathbf{Y}-\mathbf{X}+\mathbf{x}||^{-1})<\infty$, then $\mathsf{J}_{\mathbf{x}}$ is given by 2.2 in Section 2.2. Further, if $\mathcal{X}$ is a L_p space for some $2\leq p<\infty$, and $E(||\mathbf{Y}-\mathbf{X}+\mathbf{x}||^{-1})<\infty$, it can be shown that $maths f J_{\mathbf{x}}$ exists and is given by

$$\{\mathsf{J}_{\mathbf{x}}(\mathbf{z})\}(\mathbf{w})=(p-1)E\left[\frac{\int|\mathbf{Y}(s)-\mathbf{X}(s)+\mathbf{x}(s)|^{p-2}\mathbf{z}(s)\mathbf{w}(s)ds}{||\mathbf{Y}-\mathbf{X}+\mathbf{x}||^{p-1}}\right.$$
$$\left.-\frac{\left\{\int|\mathbf{Y}(s)-\mathbf{X}(s)+\mathbf{x}(s)|^{p-1}\mathbf{z}(s)ds\right\}\left\{\int|\mathbf{Y}(s)-\mathbf{X}(s)+\mathbf{x}(s)|^{p-1}\mathbf{w}(s)ds\right\}}{||\mathbf{Y}-\mathbf{X}+\mathbf{x}||^{2p-1}}\right],$$

where $\mathbf{z},\mathbf{w}$ and $\mathbf{x}\in\mathcal{X}$.

Theorem 2.4.2. *Let $N=m+n$, $m/N\to\gamma\in(0,1)$ as $m,n\to\infty$, and $\mathcal{X}^*$ is a separable and type 2 Banach space. Also, assume that the distribution of $\mathbf{X}$ is nonatomic and J_0 exists. Then, under the sequence of shrinking location shifts described at the beginning of this section, $(mn/N)^{1/2}W$ converges weakly to $G\{\mathsf{J}_0(\delta),\Gamma_1\}$ as $m,n\to\infty$. Here, the expectation in the definition of J_0 is taken with respect to the common distribution of $\mathbf{X}$ and $\mathbf{Y}$ under the null hypothesis.*

Finally, we discuss the implementation of our test when data lie in general Banach spaces. Note that for general Banach spaces, we no longer have the weighted chi-square representation of the asymptotic distribution as mentioned in Section 2.1 for Hilbert spaces. Note that in any smooth Banach space, we can estimate Γ_1 and Γ_2 by their empirical counterparts, which are defined in a similar way as discussed in Section 2 of this chapter. We can simulate from the asymptotic Gaussian distribution with these estimated covariance operators, and this leads to an estimate of the critical value of the test in smooth Banach spaces.

2.5 Paired sample Wilcoxon signed-rank type test in general Banach spaces

One can define a paired sample Wilcoxon signed-rank type test in smooth Banach spaces using spatial ranks in a similar way as the two sample Wilcoxon–Mann–Whitney test described in Section 2.4. Consider an independent and identically distributed sample $(\mathbf{X}_1, \mathbf{Y}_1), (\mathbf{X}_2, \mathbf{Y}_2) \ldots, (\mathbf{X}_n, \mathbf{Y}_n)$, and define $\mathbf{W}_i = \mathbf{Y}_i - \mathbf{X}_i$ for all $1 \leq i \leq n$. Define $\theta = E(S_{\mathbf{W}+\mathbf{W}'})$, where $\mathbf{W}'$ is an independet copy of $\mathbf{W}$. Then, the hypothesis $H_0 : \theta = \mathbf{0}$ is equivalent to the hypothesis that the spatial median of $\mathbf{W} + \mathbf{W}'$ is zero. Suppose that $\mathbf{Y} - \mathbf{X}$ has a symmetric distribution about some $\eta \in \mathcal{X}$, i.e., the distribution of $\mathbf{Y} - \mathbf{X} - \eta$ and $\eta - \mathbf{Y} + \mathbf{X}$ are the same. Then, it follows that H_0 becomes the hypothesis $\eta = \mathbf{0}$. This holds, in particular, if $\mathbf{X}$ and $\mathbf{Y} - \eta$ are exchangeable, i.e., the distributions of $(\mathbf{X}, \mathbf{Y} - \eta)$ and $(\mathbf{Y} - \eta, \mathbf{X})$ are the same. Analogous to the discussion in the second paragraph in Section 2.1, it follows that if the distribution of $\mathbf{Y} - \mathbf{X}$ is symmetric and its mean exists, then H_0 is equivalent to the hypothesis $E(\mathbf{Y} - \mathbf{X}) = \mathbf{0}$. Our Wilcoxon signed-rank type statistic for testing the hypothesis $H_0 : \theta = \mathbf{0}$ against $H_0 : \theta \neq \mathbf{0}$ is defined as $T_{SR} = 2\{n(n-1)\}^{-1} \sum_{1 \leq i < j \leq n} S_{\mathbf{W}_i + \mathbf{W}_j}$. We reject H_0 for large values of $||T_{SR}||$. Define $\Pi : \mathcal{X}^{**} \to \mathcal{X}^*$ as $\Pi(\mathbf{f}) = 4(E[\mathbf{f}\{E(S_{\mathbf{W}+\mathbf{W}'}|\mathbf{W})\}E(S_{\mathbf{W}+\mathbf{W}'}|\mathbf{W})] - \mathbf{f}(\theta)\theta)$, where $\mathbf{f} \in \mathcal{X}^{**}$. The next theorem gives the asymptotic distribution of T_{SR} for any fixed distribution of W.

Theorem 2.5.1. *Suppose that the dual space $\mathcal{X}^*$ is a separable and type 2 Banach spacc. Also assume that $\mathcal{X}^*$ is p-uniformly smooth for some $p \in (4/3, 2]$. Then, for any probability measure P on $\mathcal{X}$, $n^{1/2}(T_{SR} - \theta)$ converges weakly to $G(\mathbf{0}, \Pi)$ as $n \to \infty$.*

Thus, like our two sample Wilcoxon–Mann–Whitney type test, the paired sample Wilcoxon signed-rank type test will also have asymptotic size α. Further, it will be consistent against those alternatives, which satisfy $\theta \neq \mathbf{0}$. This holds if the distribution of $\mathbf{W}$ is symmetric and its spatial median is different from zero. In particular, this test is consistent for location shift alternatives.

We next describe how to compute the critical value of the paired sample Wilcoxon signed rank type test using its asymptotic distribution under the null hypothesis given in Theorem 2.5.1. Let $(\mathbf{X}_i, \mathbf{Y}_i)$, $1 \leq i \leq n$, be a sample from a probability distribution in $\mathcal{X}$. If $\mathcal{X}$ is a separable Hilbert space, then as in the case of the two sample Wilcoxon–Mann–Whitney type test, the asymptotic distribution of $||n^{1/2}T_{SR}||^2$ is also a weighted sum of independent chi-square random variables each with one degree of freedom, where the weights are the eigenvalues of the Π. The eigenvalues of Π can be estimated by the eigenvalues of $\widehat{\Pi}$, which is defined as

$$\widehat{\Pi} \;=\; \frac{4}{n-1}\left\{\sum_{i=1}^{n}\left(\frac{1}{n-1}\sum_{\substack{j=1\\ j\neq i}}^{n}\frac{\mathbf{W}_i+\mathbf{W}_j}{||\mathbf{W}_i+\mathbf{W}_j||}-\widehat{\theta}\right)\otimes\left(\frac{1}{n-1}\sum_{\substack{j=1\\ j\neq i}}^{n}\frac{\mathbf{W}_i+\mathbf{W}_j}{||\mathbf{W}_i+\mathbf{W}_j||}-\widehat{\theta}\right)\right\}.$$

Here, $\widehat{\theta} = 2\{n(n-1)\}^{-1}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(\mathbf{W}_i+\mathbf{W}_j)/||\mathbf{W}_i+\mathbf{W}_j||$. The critical value can now be obtained by simulating from the estimated asymptotic distribution of T_{SR}. On the other hand, if $\mathcal{X}$ is a general Banach space satisfying the assumptions of Theorem 2.5.1, we no longer have the weighted chi-square representation of the asymptotic distribution of T_{SR} under the null hypothesis. However, we can estimate Π by its empirical counterpart, which is defined in a similar way as the earlier definition in Hilbert spaces. We can then simulate from the asymptotic Gaussian distribution with this estimated covariance operator, and then compute the critical value of the test.

2.6 Mathematical details

Proof of Theorem 2.1.1. Observe that $W - \mu$ is a two-sample Hilbert space valued U-statistic with kernel $\mathbf{h}(\mathbf{x},\mathbf{y}) = S_{\mathbf{y}-\mathbf{x}} - \mu$ satisfying $E\{\mathbf{h}(\mathbf{X},\mathbf{Y})\} = \mathbf{0}$. By the Hoeffding decomposition for Hilbert space valued U-statistics (see Section 1.2 of Borovskikh (1996)), we have

$$W-\mu = \frac{1}{m}\sum_{i=1}^{m}E\{\mathbf{h}(\mathbf{X}_i,\mathbf{Y}) \mid \mathbf{X}_i\} + \frac{1}{n}\sum_{j=1}^{n}E\{\mathbf{h}(\mathbf{X},\mathbf{Y}_j) \mid \mathbf{Y}_j\} + \mathbf{R}_{m,n}.$$

So, $\mathbf{R}_{m,n} = (mn)^{-1}\sum_{i=1}^{m}\sum_{j=1}^{n}\widetilde{\mathbf{h}}(\mathbf{X}_i, \mathbf{Y}_j)$, where $\widetilde{\mathbf{h}}(\mathbf{x}, \mathbf{y}) = \mathbf{h}(\mathbf{x}, \mathbf{y}) - E\{\mathbf{h}(\mathbf{X}, \mathbf{Y}) \mid \mathbf{X} = \mathbf{x}\} - E\{\mathbf{h}(\mathbf{X}, \mathbf{Y}) \mid \mathbf{Y} = \mathbf{y}\}$. Since $E\{\widetilde{\mathbf{h}}(\mathbf{X}, \mathbf{Y}) \mid \mathbf{Y} = \mathbf{y}\} = E\{\widetilde{\mathbf{h}}(\mathbf{X}, \mathbf{Y}) \mid \mathbf{X} = \mathbf{x}\} = 0$, and $||\widetilde{\mathbf{h}}(\mathbf{x}, \mathbf{y})|| \leq 4$ for all $\mathbf{x}, \mathbf{y} \in \mathbf{X}$, we get

$$E(||\mathbf{R}_{m,n}||^2) = (mn)^{-1}E(||\widetilde{\mathbf{h}}(\mathbf{X}_1, \mathbf{Y}_1)||^2) \leq 16(mn)^{-1}. \tag{2.4}$$

Since $m/N \to \gamma \in (0,1)$, (2.4) implies that $(mn/N)^{1/2}\mathbf{R}_{m,n}$ converges to $\mathbf{0}$ in probability as $m, n \to \infty$. Now, $m^{-1/2}\sum_{i=1}^{m} E\{\mathbf{h}(\mathbf{X}_i, \mathbf{Y}) \mid \mathbf{X}_i\}$ and $n^{-1/2}\sum_{j=1}^{n} E\{\mathbf{h}(\mathbf{X}, \mathbf{Y}_j) \mid \mathbf{Y}_j\}$ converge weakly to $G(\mathbf{0}, \Gamma_1)$ and $G(\mathbf{0}, \Gamma_2)$, respectively, as $m, n \to \infty$ by the central limit theorem for independent and identically distributed random variables in a separable Hilbert space (see Theorem 7.5(i) of Araujo and Giné (1980)). This together with the assumption that $m/N \to \gamma$, and the fact that $(mn/N)^{1/2}R_{m,n}$ converges to $\mathbf{0}$ in probability complete the proof. □

Proof of Theorem 2.2.1. Define $\mu(\Delta_N) = E(S_{\mathbf{Y}-\mathbf{X}})$. Applying the Hoeffding decomposition for Hilbert space valued U-statistics as in the proof of Theorem 1, it follows that

$$\begin{aligned} W - \mu(\Delta_N) &= \frac{1}{m}\sum_{i=1}^{m}\{E(S_{\mathbf{Y}-\mathbf{X}_i} \mid \mathbf{X}_i) - \mu(\Delta_N)\} \\ &+ \frac{1}{n}\sum_{j=1}^{n}\{E(S_{\mathbf{Y}_j-\mathbf{X}} \mid \mathbf{Y}_j) - \mu(\Delta_N)\} + \mathbf{S}_{m,n}. \end{aligned} \tag{2.5}$$

Arguing as in the proof of Theorem 1, it can be shown that $E(||\mathbf{S}_{m,n}||^2) \leq 16(mn)^{-1}$ for each $m, n \geq 1$. Thus, $(mn/N)^{1/2}\mathbf{S}_{m,n} \to 0$ in probability as $m, n \to \infty$ under the sequence of shrinking shifts.

Note that $\mu(\Delta_N) = E(S_{\mathbf{Z}-\mathbf{X}+\Delta_N})$, where $\mathbf{Z}$ is an independent copy of $\mathbf{X}$. Also, from the definition of Fréchet derivative (see the `Appendix`) it follows that $\mathsf{J}_{\mathbf{x}}(\mathbf{h})$ is the Fréchet derivative of $E(S_{\mathbf{Z}-\mathbf{X}+\mathbf{x}})$ evaluated at $\mathbf{h} \in \mathcal{X}$. So, $(mn/N)^{1/2}\mu(\Delta_N)$ converges to $\mathsf{J}_0(\delta)$ as $m, n \to \infty$.

Let us write $\Phi_N(\mathbf{X}_i) = m^{-1/2}\{E(S_{\mathbf{Y}-\mathbf{X}_i} \mid \mathbf{X}_i) - \mu(\Delta_N)\}$. Note that $E\{\Phi_N(\mathbf{X}_i)\} = 0$. To prove the asymptotic Gaussianity of $\sum_{i=1}^{m}\Phi_N(\mathbf{X}_i)$, it is enough to show that the triangular array $\{\Phi_N(\mathbf{X}_1), \ldots, \Phi_N(\mathbf{X}_m)\}_{m=1}^{\infty}$ of rowwise independent and identically distributed random elements satisfy the conditions of Corollary 7.8 in Araujo and Giné (1980). Observe that for any $\epsilon > 0$,

$$\sum_{i=1}^{m} P\{||\Phi_N(\mathbf{X}_i)|| > \epsilon\} \leq \sum_{i=1}^{m} E\{||E(S_{\mathbf{Y}-\mathbf{X}_i} \mid \mathbf{X}_i) - \mu(\Delta_N)||^3\}/m^{3/2} \leq 8m^{-1/2}.$$

Thus, $\lim_{m\to\infty}\sum_{i=1}^{m} P\{||\Phi_N(\mathbf{X}_i)|| > \epsilon\} = 0$ for every $\epsilon > 0$, which ensures that condition (1) of Corollary 7.8 in Araujo and Giné (1980) holds.

We next verify condition (2) of Corollary 7.8 in Araujo and Giné (1980). Let us fix $\mathbf{f} \in \mathcal{X}$. Since $||S_{\mathbf{x}}|| = 1$ for all $\mathbf{x} \neq 0$, we can choose $\delta = 1$ in that condition (2). Then,

$$\sum_{i=1}^{m} E\{\langle \mathbf{f}, \Phi_N(\mathbf{X}_i)\rangle^2\} = m^{-1}\sum_{i=1}^{m} E[\{W_{N,i} - E(W_{N,i})\}^2], \tag{2.6}$$

where $W_{N,i} = \langle \mathbf{f}, E(S_{\mathbf{Y}-\mathbf{X}_i} \mid \mathbf{X}_i)\rangle$. Since the $\mathbf{X}_i$'s are identically distributed, the right hand side in (2.6) simplifies to $E[\{W_{N,1} - E(W_{N,1})\}^2]$. Note that $W_{N,1} = \langle \mathbf{f}, E(S_{\mathbf{Z}-\mathbf{X}_1+\Delta_N} \mid \mathbf{X}_1)\rangle$, where $\mathbf{Z}$ is an independent copy of $\mathbf{X}_1$. The dominated convergence theorem implies that

$$E(S_{\mathbf{Z}-\mathbf{X}_1+\Delta_N} \mid \mathbf{X}_1) \longrightarrow E(S_{\mathbf{Z}-\mathbf{X}_1} \mid \mathbf{X}_1) \tag{2.7}$$

as $m, n \to \infty$ for almost all values of $\mathbf{X}_1$. Thus, we get the convergence of $E(W_{N,1})$ to $E\{\langle \mathbf{f}, E(S_{\mathbf{Z}-\mathbf{X}_1} \mid \mathbf{X}_1)\rangle\}$ as $m, n \to \infty$. Similarly, it follows that $E(W_{N,1}^2)$ converges to $E\{\langle \mathbf{f}, E(S_{\mathbf{Z}-\mathbf{X}_1} \mid \mathbf{X}_1)\rangle^2\}$ as $m, n \to \infty$. So, $\sum_{i=1}^{m} E\{\langle \mathbf{f}, \Phi_N(\mathbf{X}_i)\rangle^2\} \to \Gamma_1(\mathbf{f}, \mathbf{f})$ as $m, n \to \infty$, where Γ_1 is as defined before Theorem 1 in Section 2 in the paper. This completes the verification of condition (2) of Corollary 7.8 in Araujo and Giné (1980).

Finally, for the verification of condition (3) of Corollary 7.8 in Araujo and Giné (1980), suppose that $\{\mathcal{F}_k\}_{k\geq 1}$ is a sequence of finite dimensional subspaces of $\mathcal{X}$ such that $\mathcal{F}_k \subseteq \mathcal{F}_{k+1}$ for all $k \geq 1$, and the closure of $\bigcup_{k=1}^{\infty} \mathcal{F}_k$ is $\mathcal{X}$. Such a sequence of subspaces exists because of the separability of $\mathcal{X}$. For any $\mathbf{x} \in \mathcal{X}$ and any $k \geq 1$, we define $d(\mathbf{x}, \mathcal{F}_k) = \inf\{||\mathbf{x} - \mathbf{y}|| : \mathbf{y} \in \mathcal{F}_k\}$. It is straightforward to verify that for every $k \geq 1$, the map $\mathbf{x} \mapsto d(\mathbf{x}, \mathcal{F}_k)$ is continuous and bounded on any closed ball in $\mathcal{X}$. It follows from (2.7) that $\mu(\Delta_N) \to 0$ as $m, n \to \infty$, and

$$\begin{aligned}\sum_{i=1}^{m} E[d^2\{\Phi_N(\mathbf{X}_i), \mathcal{F}_k\}] &= m^{-1}\sum_{i=1}^{m} E[d^2\{E(S_{\mathbf{Z}-\mathbf{X}_i+\Delta_N} \mid \mathbf{X}_i) - \mu(\Delta_N), \mathcal{F}_k\}] \\ &= E[d^2\{E(S_{\mathbf{Z}-\mathbf{X}_1+\Delta_N} \mid \mathbf{X}_1) - \mu(\Delta_N), \mathcal{F}_k\}] \\ &\longrightarrow E[d^2\{E(S_{\mathbf{Z}-\mathbf{X}_1} \mid \mathbf{X}_1), \mathcal{F}_k\}]\end{aligned}$$

as $m, n \to \infty$. From the choice of the $\mathcal{F}_k$'s, it can be shown that $d(\mathbf{x}, \mathcal{F}_k) \to 0$ as $k \to \infty$ for all $\mathbf{x} \in \mathcal{X}$. So, $E[d^2\{E(S_{\mathbf{Z}-\mathbf{X}_1} \mid \mathbf{X}_1), \mathcal{F}_k\}]$ converges to zero as $k \to \infty$, which completes the verification of condition (3) of Corollary 7.8 in Araujo and Giné (1980).

Thus, $\sum_{i=1}^{m} \Phi_N(\mathbf{X}_i)$ converges weakly to a centered Gaussian random element in $\mathcal{X}$ as $m, n \to \infty$. Further, its asymptotic covariance is Γ_1, which was obtained while checking condition (2) of Corollary 7.8 in Araujo and Giné (1980). It follows from similar arguments that $n^{-1/2}\sum_{j=1}^{n} E\{\mathbf{h}(\mathbf{X}, \mathbf{Y}_j) \mid \mathbf{Y}_j\}$ also converges weakly to a Gaussian random element in $\mathcal{X}$ with the same distribution as $m, n \to \infty$. Thus, it follows from (2.5) that

$$(mn/N)^{1/2}\{W - \mu(\Delta_N)\} \longrightarrow G(\mathbf{0}, \Gamma_1)$$

weakly as $m, n \to \infty$ under the sequence of shrinking shifts. This, together with the fact that $(mn/N)^{1/2}\mu(\Delta_N)$ converges to $\mathrm{J}_\mathbf{0}(\delta)$ as $m, n \to \infty$ completes the proof of the theorem. □

Proof of Theorem 2.2.2. (a) Let us observe that $nN^{-1}T_{\mathrm{CFF}} = mnN^{-1}||\bar{\mathbf{X}} - \bar{\mathbf{Y}}||^2$. For each $N \geq 1$, $\mathbf{Y}$ has the same distribution as that of $\mathbf{Z} + \Delta_N$, where $\mathbf{Z}$ is an independent copy of $\mathbf{X}$. Now, by the central limit theorem for independent and identically distributed random elements in a separable Hilbert space (see Theorem 7.5(i) in Araujo and Giné (1980)), it follows that $(mn/N)^{1/2}(\bar{\mathbf{Z}} - \bar{\mathbf{X}})$ converges weakly to $G(\mathbf{0}, \Sigma)$ as $m, n \to \infty$. Thus, $(mn/N)^{1/2}(\bar{\mathbf{Y}} - \bar{\mathbf{X}})$, which has the same distribution as that of $(mn/N)^{1/2}(\bar{\mathbf{Z}} - \bar{\mathbf{X}} + \Delta_N)$, converges weakly to $G(\delta, \Sigma)$ as $m, n \to \infty$. Now, the distribution of $||G(\delta, \Sigma)||^2$ is the same as that of $\sum_{k=1}^{\infty} \lambda_k \chi^2_{(1)}(\beta_k^2/\lambda_k)$ using the spectral decomposition of the compact self-adjoint operator Σ. This proves part (a) of the proposition.

(b) Let $\mathbf{v} = (\langle \bar{\mathbf{X}} - \bar{\mathbf{Y}}, \psi_1\rangle, \langle \bar{\mathbf{X}} - \bar{\mathbf{Y}}, \psi_2\rangle, \ldots, \langle \bar{\mathbf{X}} - \bar{\mathbf{Y}}, \psi_L\rangle)$ and $\widetilde{\beta} = (\beta_1, \beta_2, \ldots, \beta_L)$. It follows from the central limit theorem in $\mathbb{R}^L$ that $(mn/N)^{1/2}\{\mathbf{v} - (mn/N)^{-1/2}\widetilde{\beta}\}$ converges weakly to $N_L(\mathbf{0}, \Lambda_L)$ as $m, n \to \infty$ under the given sequence of shrinking shifts, where Λ_L is the diagonal matrix $Diag(\lambda_1, \ldots, \lambda_L)$. Thus, under the given sequence of shifts, $(mn/N)^{1/2}v$ converges weakly to a $N_L(\widetilde{\beta}, \Lambda_L)$ distribution as $m, n \to \infty$.

From arguments similar to those in the proof of Theorem 5.3 in Horváth and Kokoszka (2012), and using the assumptions in the present theorem, we get

$$\max_{1\le k\le L} (mn/N)^{1/2} |\langle \bar{\mathbf{X}} - \bar{\mathbf{Y}}, \widehat{\psi}_k - \widehat{c}_k\psi_k\rangle| = o_P(1) \tag{2.8}$$

as $m, n \to \infty$ under this sequence of shifts. Here $\widehat{\psi}_k$ is the empirical version of ψ_k and $\widehat{c}_k = sign(\langle \widehat{\psi}_k, \psi_k\rangle)$. In view of (2.8), the limiting distribution of $mnN^{-1}\sum_{k=1}^{L}(\langle \bar{\mathbf{X}} - \bar{\mathbf{Y}}, \widehat{\psi}_k\rangle)^2$ is the same as that of $mnN^{-1}\sum_{k=1}^{L}(\langle \bar{\mathbf{X}} - \bar{\mathbf{Y}}, \widehat{c}_k\psi_k\rangle)^2 = mnN^{-1}||\mathbf{v}||^2$, and the latter converges weakly to $||N_L(\widetilde{\beta}, \Lambda_L)||^2$ as $m, n \to \infty$. Thus, $mnN^{-1}T_{\text{HKR1}}$ converges weakly to $\sum_{k=1}^{L} \lambda_k \chi^2_{(1)}(\beta_k^2/\lambda_k)$ under the given sequence of shrinking shifts as $m, n \to \infty$.

It also follows using similar arguments as in the proof of Theorem 5.3 in Horváth and Kokoszka (2012) that under the assumptions of the present theorem, we have

$$\max_{1\le k\le L} (mn/N)^{1/2} \widehat{\lambda}_k^{-1/2} |\langle \bar{\mathbf{X}} - \bar{\mathbf{Y}}, \widehat{\psi}_k - \widehat{c}_k\psi_k\rangle| = o_P(1)$$

as $m, n \to \infty$ under the given sequence of shifts. Similar arguments as in the case of T_{HKR1} now yield the asymptotic distribution of $mnN^{-1}T_{\text{HKR2}}$, and this completes the proof. □

Proof of Theorem 2.4.1. Similar to the proof of Theorem 2.1.1, we use the Hoeffding decomposition for $W - \mu$ to get

$$W - \mu = \frac{1}{m}\sum_{i=1}^{m} E\{\mathbf{h}(\mathbf{X}_i, \mathbf{Y}) \mid \mathbf{X}_i\} + \frac{1}{n}\sum_{j=1}^{n} E\{\mathbf{h}(\mathbf{X}, \mathbf{Y}_j) \mid \mathbf{Y}_j\} + \mathbf{R}_{m,n},$$

where $\mathbf{R}_{m,n} = (mn)^{-1}\sum_{i=1}^{m}\sum_{j=1}^{n} \widetilde{\mathbf{h}}(\mathbf{X}_i, \mathbf{Y}_j)$ as in Theorem 2.1.1. Let $\Phi(\mathbf{X}_i) =$

$\sum_{j=1}^{n} \tilde{\mathbf{h}}(\mathbf{X}_i, \mathbf{Y}_j)$. Using the definition of type 2 Banach spaces, we get

$$\begin{aligned} & E(||\mathbf{R}_{m,n}||^2 \mid \mathbf{Y}_j; j = 1, 2, \dots, n) \\ = \ & \frac{1}{m^2 n^2} E\left\{ \left\| \sum_{i=1}^{m} \Phi(\mathbf{X}_i) \right\|^2 \mid \mathbf{Y}_j; j = 1, \dots, n \right\} \\ \leq \ & \frac{b}{m^2 n^2} \sum_{i=1}^{m} E\left\{ ||\Phi(\mathbf{X}_i)||^2 \mid \mathbf{Y}_j; j = 1, \dots, n \right\}. \end{aligned} \tag{2.9}$$

Taking expectations of both sides of 2.9 with respect to $\mathbf{Y}_j$ for $1 \leq j \leq n$, we get

$$E(||\mathbf{R}_{m,n}||^2) \leq \frac{b}{mn^2} E\left\{ \left\| \sum_{j=1}^{n} \tilde{\mathbf{h}}(\mathbf{X}_1, \mathbf{Y}_j) \right\|^2 \right\}. \tag{2.10}$$

Since $E\{\tilde{\mathbf{h}}(\mathbf{X}, \mathbf{Y}) \mid \mathbf{X} = \mathbf{x}\} = \mathbf{0}$ for all $\mathbf{x} \in \mathcal{X}$, once again from the definition of type 2 Banach spaces, we get

$$\begin{aligned} E\left\{ \left\| \sum_{j=1}^{n} \tilde{\mathbf{h}}(\mathbf{X}_1, \mathbf{Y}_j) \right\|^2 \right\} & \leq bE\left[\sum_{j=1}^{n} E\left\{ \left\| \tilde{\mathbf{h}}(\mathbf{X}_1, \mathbf{Y}_j) \right\|^2 \mid \mathbf{X}_1 \right\} \right] \\ & = bnE\left\{ \left\| \tilde{\mathbf{h}}(\mathbf{X}_1, \mathbf{Y}_1) \right\|^2 \right\}. \end{aligned} \tag{2.11}$$

Using the boundedness of the kernel and (2.10) and (2.11), we have

$$E(||\mathbf{R}_{m,n}||^2) \leq \frac{b^2}{mn} E\left\{ \left\| \tilde{\mathbf{h}}(\mathbf{X}_1, \mathbf{Y}_1) \right\|^2 \right\} \leq \frac{16b^2}{mn}.$$

Hence, $(mn/N)^{1/2}\mathbf{R}_{m,n}$ converges to $\mathbf{0}$ in probability as $m, n \to \infty$. The weak convergence of the first two terms in the Hoeffding decomposition after proper scaling follows from similar arguments as in the proof of Theorem 2.1.1, and this completes the proof.

□

Proof of Theorem 2.4.2. Define $\mu = \mu(\Delta_N) = E(S_{\mathbf{Y}-\mathbf{X}})$. Applying the Hoeffding decomposition as in the proof of the previous theorem, it follows that

$$W - \mu(\Delta_N) = \frac{1}{m}\sum_{i=1}^{m}\{E(S_{\mathbf{Y}-\mathbf{X}_i} \mid \mathbf{X}_i) - \mu(\Delta_N)\}$$
$$+\frac{1}{n}\sum_{j=1}^{n}\{E(S_{\mathbf{Y}_j-\mathbf{X}} \mid \mathbf{Y}_j) - \mu(\Delta_N)\} + \widetilde{\mathbf{R}}_{m,n}. \tag{2.12}$$

Arguing as in the proof of Theorem 2.4.1, it can be shown that $E(||\widetilde{\mathbf{R}}_{m,n}||^2) \leq 16b^2/mn$ for each $m, n \geq 1$. Thus, $(mn/N)^{1/2}\widetilde{\mathbf{R}}_{m,n} \to \mathbf{0}$ in probability as $m, n \to \infty$ under the sequence of shrinking shifts.

Note that $\mu(\Delta_N) = E(S_{\mathbf{Z}-\mathbf{X}+\Delta_N})$, where $\mathbf{Z}$ is an independent copy of $\mathbf{X}$. So, it follows from the equivalent definition of Hessian given in the `Appendix` that

$$(mn/N)^{1/2}\mu(\Delta_N) \longrightarrow \mathsf{J}_{\mathbf{0}}(\delta) \tag{2.13}$$

as $m, n \to \infty$.

The asymptotic Gaussianity of the first two terms on the right hand side of (2.12) after proper scaling will follow from arguments similar to those in the proof of Theorem 2.2.1. Let us write $\Phi_N(\mathbf{X}_i) = m^{-1/2}\{E(S_{\mathbf{Y}-\mathbf{X}_i} \mid \mathbf{X}_i) - \mu(\Delta_N)\}$. Condition (1) of Corollary 7.8 in Araujo and Giné (1980) holds as in the proof of Theorem 2.2.1.

For verifying condition (2) of Corollary 7.8 in Araujo and Giné (1980), let us fix $\mathbf{f} \in \mathcal{X}^{**}$. We have

$$\sum_{i=1}^{m} E[\mathbf{f}^2\{\Phi_N(\mathbf{X}_i)\}] = m^{-1}\sum_{i=1}^{m} E[\{U_{N,i} - E(U_{N,i})\}^2] = E[\{U_{N,1} - E(U_{N,1})\}^2], \tag{2.14}$$

where $U_{N,i} = \mathbf{f}\{E(S_{\mathbf{Y}-\mathbf{X}_i} \mid \mathbf{X}_i)\}$. Note that $U_{N,1} = \mathbf{f}\{E(S_{\mathbf{Z}-\mathbf{X}_1+\Delta_N} \mid \mathbf{X}_1)\}$, where $\mathbf{Z}$ is an independent copy of $\mathbf{X}_1$. Since the norm in $\mathcal{X}$ is assumed to be twice Gâteaux differentiable, it follows from Theorem 4.6.15(a) and Proposition 4.6.16 in Borwein and Vanderwerff (2010) that the norm in $\mathcal{X}$ is Fréchet differentiable. This in turn implies that the map $\mathbf{x} \mapsto S_{\mathbf{x}}$ is continuous on $\mathcal{X}\backslash\{\mathbf{0}\}$ (see, e.g., Corollary 4.2.12 in Borwein and Vanderwerff (2010)). The rest of the proof involving verifying conditions (2) and (3) of Corollary 7.8 in Araujo and Giné (1980) is exactly same as that of the proof of Theorem 2.2.1. □

Proof of Theorem 2.5.1. Similar to the proof of Theorem 2.4.1, we use the Hoeffding decomposition for $T_{SR} - \theta$ to get

$$T_{SR} - \theta = \frac{2}{n} \sum_{i=1}^{n} [E\{S_{\mathbf{W}_i + \mathbf{W}'} \mid \mathbf{W}_i\} - \theta] + \mathbf{s}_n,$$

where $\mathbf{s}_n = 2\{n(n-1)\}^{-1} \sum_{1 \leq i < j \leq n} S_{\mathbf{W}_i + \mathbf{W}'} - \frac{2}{n} \sum_{i=1}^{n} E\{S_{\mathbf{W}_i + \mathbf{W}'} \mid \mathbf{W}_i\} + \theta$. Arguing as in the proof of Theorem 2.4.1, it can be shown that $n^{1/2}\mathbf{s}_n$ converges to $\mathbf{0}$ in probability as $n \to \infty$. The weak convergence of the first term in the Hoeffding decomposition after proper scaling follows from similar arguments as in the proof of Theorem 2.4.1, and this completes the proof. □

Chapter 3

The spatial distribution and quantiles in infinite dimensional spaces

For a univariate probability distribution, median is a well-known and popular choice of its center. It has several desirable statistical properties, which include equivariance under monotone transformations and asymptotic consistency under very general conditions. The univariate median and other quantiles have been extended in a number of ways for multivariate data and distributions in finite dimensional spaces (see, e.g., Oja (1983), Liu (1990), Small (1990) and Donoho and Gasko (1992)). In particular, the spatial median, the spatial quantiles and the associated spatial distribution function in finite dimensional Euclidean spaces have been extensively studied (see, e.g., Brown (1983), Chaudhuri (1996), Koltchinskii (1997), Möttönen et al. (1997) and Serfling (2002)). However, many of the well-known multivariate medians like the simplicial depth median (see Liu (1990)), and the simplicial volume median (see Oja (1983)) do not have meaningful extensions into infinite dimensional spaces. On the other hand, the spatial median as well as the spatial quantiles extend easily into infinite dimensional Banach spaces (see Valadier (1984), Kemperman (1987) and Chaudhuri (1996)). Gervini (2008) proposed functional principal components using the sample spatial median and used those to analyze a data involving the movements of the lips. The spatial median has also been used by Chaouch and Goga (2012) to calculate the median profile for the electricity load data in France. Fraiman and Pateiro-López (2012) studied some direction-based quantiles for probability distributions in infinite dimensional Hilbert spaces. These quantiles are defined for unit direction vectors in such spaces, and they extend the finite dimensional quantiles considered by Kong and Mizera (2012). The principle quantile directions derived from these quantiles were used by Fraiman and Pateiro-López (2012) to detect outliers in a dataset of annual age-specific mortality rates of French males between the years 1899 and 2005. Recently, Cardot et al. (2013) considered an updation based estimator of the spatial median and used it to compute the profile of a typical television audience in France throughout a single day.

In this chapter, we investigate the spatial distribution in infinite dimensional Banach spaces and study their properties along with the spatial quantiles. There are several mathematical difficulties in dealing with the probability distributions in such spaces. These are primarily due to the noncompactness of the closed unit ball in such spaces. We prove some Glivenko-Cantelli and Donsker type results for empirical spatial distribution processes arising from data lying in infinite dimensional spaces. A Bahadur type linear representation of the sample spatial quantile and its asymptotic Gaussianity are derived. We also study the asymptotic efficiency of the sample spatial median relative to the sample mean for some well-known probability distributions in function spaces.

3.1 The spatial distribution and the associated empirical processes in Banach spaces

For probability distributions in $\mathbb{R}^d$, the spatial distribution is a special case of the M-distribution function, which was studied in detail in Koltchinskii (1997). Consider the map $f : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ such that for every $\mathbf{x} \in \mathbb{R}^d$, $f(\cdot, \mathbf{x})$ is a convex function. Then, for any random vector $\mathbf{X} \in \mathbb{R}^d$, a subgradient of the map $\mathbf{x} \mapsto E\{f(\mathbf{x}, \mathbf{X})\}$ is called the M-distribution function of $\mathbf{X}$ with respect to f. If $f(\mathbf{x}, \mathbf{y}) = ||\mathbf{x} - \mathbf{y}|| - ||\mathbf{y}||$, where $|| \cdot ||$ is the usual Euclidean norm, the M-distribution function is the spatial distribution function, whose value at $\mathbf{x}$ with respect to the probability distribution of $\mathbf{X}$ is $E\{(\mathbf{x} - \mathbf{X})/||\mathbf{x} - \mathbf{X}||\}$. If $d = 1$, the spatial distribution simplifies to $2F(x) - 1$, where F

is the cumulative distribution function of X. Koltchinskii (1997) showed that under certain conditions, the M-distribution function characterizes the probability distribution of a random vector like the cumulative distribution function. In that paper, Glivenko-Cantelli and Donsker type results were also proved for the empirical M-distribution process. These results are similar to those obtained for the empirical process associated with the cumulative distribution function in the finite dimensional multivariate setting. For probability distributions in the space of real-valued functions on an interval, a notion of distribution functional was studied by Bugni et al. (2009). But those authors did not study any Glivenko-Cantelli or Donsker type result for the empirical processes associated with their distribution functionals. Further, there is no natural extension of the cumulative distribution function for probability distributions in general infinite dimensional Banach spaces.

In this section, we study the spatial distribution in infinite dimensional Banach spaces and obtain Glivenko-Cantelli and Donsker type results for the associated empirical processes. Let $\mathcal{X}$ be a smooth Banach space, and denote the Gâteaux derivative (see the `Appendix`) of the norm $||\cdot||$ in $\mathcal{X}$ at a nonzero $\mathbf{x} \in \mathcal{X}$ by $S_{\mathbf{x}}$. As a convention, we define $S_{\mathbf{x}} = \mathbf{0}$ if $\mathbf{x} = \mathbf{0}$. Let $\mathbf{X}$ be a random element in $\mathcal{X}$, and denote its probability distribution by μ. The spatial distribution at $\mathbf{x} \in \mathcal{X}$ with respect to μ is defined as $\Psi_{\mathbf{x}} = E\{S_{\mathbf{x}-\mathbf{X}}\}$. The empirical spatial distribution can be defined as $\widehat{\Psi}_{\mathbf{x}} = n^{-1}\sum_{i=1}^{n} S_{\mathbf{x}-\mathbf{X}_i}$, where $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ are i.i.d. observations from a probability distribution μ in $\mathcal{X}$. The empirical spatial distribution was used in the `Chapter 2` to develop Wilcoxon–Mann–Whitney type tests for two sample problems in infinite dimensional spaces. Later in this chapter, the spatial quantile will be defined in infinite dimensional spaces as the inverse of the spatial distribution.

Associated with the spatial distribution is the corresponding empirical spatial distribution process $\{\widehat{\Psi}_{\mathbf{x}} - \Psi_{\mathbf{x}} : \mathbf{x} \in I\}$ indexed by $I \subseteq \mathcal{X}$. This is a Banach space valued stochastic process indexed by the elements in a Banach space. When $\mathcal{X} = \mathbb{R}^d$ equipped with the Euclidean norm, the Glivenko-Cantelli and the Donsker type results hold for the empirical spatial distribution process with $I = \mathbb{R}^d$ (see Koltchinskii (1997)). The following theorem states a Glivenko-Cantelli and a Donsker type result for the empirical spatial distribution process in a separable Hilbert space.

Theorem 3.1.1. *Let $\mathcal{X}$ be a separable Hilbert space, and $\mathcal{Z}$ be a finite dimensional subspace of $\mathcal{X}$. Then, $\widehat{\Psi}_{\mathbf{x}}$ converges to $\Psi_{\mathbf{x}}$ uniformly in $\mathcal{Z}$ in the weak topology of $\mathcal{X}$ almost surely. Further, if μ is nonatomic, then for any $d \geq 1$ and any continuous linear map $\mathbf{g} : \mathcal{X} \to \mathbb{R}^d$, the process $\{\mathbf{g}(\sqrt{n}(\widehat{\Psi}_{\mathbf{x}} - \Psi_{\mathbf{x}})) : \mathbf{x} \in \mathcal{Z}\}$ converges weakly to a d-variate Gaussian process on $\mathcal{Z}$.*

The Glivenko-Cantelli and the Donsker type results in Koltchinskii (1997) for the empirical spatial distribution process in $\mathbb{R}^d$ follow from the above theorem as a straightforward corollary. The result stated in Theorem 3.1.1 is also true in Banach spaces like L_p spaces for some even integer $p > 2$ (see the remark after the proof of Theorem 3.1.1 given in the section on mathematical details).

A probability measure in an infinite dimensional separable Banach space $\mathcal{X}$ (e.g., a nondegenerate Gaussian measure) may assign zero probability to all finite dimensional subspaces. However, since $\mathcal{X}$ is separable, for any $\varepsilon > 0$, we can find a compact set $K \subseteq \mathcal{X}$ such that $\mu(K) > 1 - \varepsilon$ (see, e.g., Araujo and Giné (1980)). Thus, given any measurable set $V \subseteq \mathcal{X}$, there exists a compact set such that the probability content of V outside this compact set is as small as we want. The next theorem gives the asymptotic properties of the empirical spatial distribution process uniformly over any compact subset of $\mathcal{X}$. We state an assumption that is required for the next theorem.

Assumption (A). *There exists a map $T(\mathbf{x}) : \mathcal{X}\backslash\{\mathbf{0}\} \to (0, \infty)$, which is measurable with respect to the usual Borel σ-field of $\mathcal{X}$ such that for all $\mathbf{x} \neq \mathbf{0}, -\mathbf{h}$, we have $||S_{\mathbf{x}+\mathbf{h}} - S_{\mathbf{x}}|| \leq T(\mathbf{x})||\mathbf{h}||$.*

Assumption (A) holds if $\mathcal{X}$ is a a Hilbert space or a L_p space for some $p \in [2, \infty)$, and in the former case we can choose $T(\mathbf{x}) = 2/||\mathbf{x}||$. For any set $A \subset \mathcal{X}$, we denote by $N(\varepsilon, A)$ the minimum number of open balls of radii ε and centers in A that are needed to cover A.

Theorem 3.1.2. *Let $\mathcal{X}^*$ be a separable Banach space, and $K \subseteq \mathcal{X}$ be a compact set.*
(a) Suppose that Assumption (A) holds, and $\sup_{||\mathbf{x}|| \leq C} E_{\mu_1}\{T(\mathbf{x} - \mathbf{X})\} < \infty$ for each $C > 0$, where μ_1 is the nonatomic part of μ. Then, $\widehat{\Psi}_{\mathbf{x}}$ converges to $\Psi_{\mathbf{x}}$ uniformly over $\mathbf{x} \in K$ in the norm topology of $\mathcal{X}^*$ almost surely.
(b) Let μ be a nonatomic probability measure, $\sup_{||\mathbf{x}|| \leq C} E_{\mu}\{T^2(\mathbf{x} - \mathbf{X})\} < \infty$ for each

$C > 0$, and Assumption (A) hold. If $\int_0^1 \sqrt{ln\ N(\varepsilon, K)}d\varepsilon < \infty$ for each $\varepsilon > 0$, then the empirical process $\widehat{\Upsilon}_{\mathbf{g}} = \{\mathbf{g}(\sqrt{n}(\widehat{\Psi}_{\mathbf{x}} - \Psi_{\mathbf{x}})) : \mathbf{x} \in K\}$ converges weakly to a d-variate Gaussian process on K for any $d \geq 1$ and any continuous linear function $\mathbf{g} : \mathcal{X}^* \to \mathbb{R}^d$. Further, if $\mathcal{X}$ is a separable Hilbert space, then for any Lipschitz continuous function $\mathbf{g} : \mathcal{X} \to \mathbb{R}^d$, $\widehat{\Upsilon}_{\mathbf{g}}$ converges weakly to a $\mathbb{R}^d$-valued stochastic process on K.

If μ is a purely atomic measure, the Glivenko-Cantelli type result in part (a) of the above theorem holds over the entire space $\mathcal{X}$ (see Lemma 3.3.1 in the section on mathematical details). It follows from part (a) of the above theorem and the tightness of any probability measure in any complete separable metric space that $\int_{\mathcal{X}} ||\widehat{\Psi}_{\mathbf{x}} - \Psi_{\mathbf{x}}||^2 \mu(d\mathbf{x}) \to 0$ as $n \to \infty$ almost surely. If we choose $d = 1$ and $\mathbf{g}(\mathbf{x}) = ||\mathbf{x}||$ in the second statement in part (b) of the above theorem, it follows that $\sup_{\mathbf{x} \in K} ||\widehat{\Psi}_{\mathbf{x}} - \Psi_{\mathbf{x}}|| = O_P(1/\sqrt{n})$ and $\int_{\mathcal{X}} ||\widehat{\Psi}_{\mathbf{x}} - \Psi_{\mathbf{x}}||^2 \mu(d\mathbf{x}) = O_P(1/n)$ as $n \to \infty$.

Let $\mathcal{X}$ be a separable Hilbert space and $\mathbf{X} = \sum_{k=1}^{\infty} X_k \psi_k$ for an orthonormal basis $\{\psi_k\}_{k \geq 1}$ of $\mathcal{X}$. Then, the moment condition assumed in part (a) (respectively, part (b)) of the above theorem holds if some bivariate (respectively, trivariate) marginal of $(X_1, X_2, \ldots)$ has a density under μ_1 (respectively, μ) that is bounded on bounded subsets of $\mathbb{R}^2$ (respectively, $\mathbb{R}^3$).

Let $\mathcal{I} = \int_0^1 \sqrt{ln\ N(\varepsilon, K)}d\varepsilon$. We will now explore some situations under which the entropy condition in part (b) of Theorem 3.1.2 holds, i.e., $I < \infty$. It is easy to verify that $\mathcal{I} < \infty$ for every compact set K in any finite dimensional Banach space. The finiteness of $\mathcal{I}$ is also true for many compact sets in various infinite dimensional function spaces (see, e.g., Kolmogorov and Tihomirov (1961)), and we discuss some examples below.

1. Consider the space $\mathbb{C}(B)$ of continuous real-valued functions defined on a closed and bounded box B in $\mathbb{R}^d$ and equipped with the supremum norm. For some $C > 0$ and $r \geq 1$, consider the subset $F_q^B(C)$ of $\mathbb{C}(B)$ whose elements have partial derivatives upto order r, the r-th order partial derivatives are Holder continuous with exponent $\alpha \in (0, 1]$, and the remainder terms in their Taylor expansions at $\mathbf{x} + \mathbf{h}$ about $\mathbf{x}$ are uniformly bounded by $C||\mathbf{h}||^q$ for all $\mathbf{x}, \mathbf{x} + \mathbf{h} \in B$. Here, $q = r + \alpha$. This will be the situation, for example, when we have functional data in the space $\mathbb{C}(B)$, and for modelling purposes the sample space is assumed to

have certain smoothness properties. Then, for any compact subset K of $F_q^B(C)$, we have $ln\ N(\varepsilon, K) \sim (1/\varepsilon)^{(d/q)}$. Therefore, $I < \infty$ if $d < 2q$. It is important to note that K is totally bounded in $L_p(B,\mu)$ for any $p \in [1,\infty)$ with its ε-entropy of the same order for any finite measure μ because $\mathbb{C}(B) \subset L_p(B,\mu)$ when B is compact. Thus if we take $\mathcal{X} = L_p(B,\mu)$ for some $p \in [2,\infty)$ and the compact subset to be K as described above, we have the Glivenko-Cantelli and Donsker properties as in part (b) of Theorem 3.1.2.

2. Let (Z,ρ) be any metric space and A be a compact subset of Z. Let us consider the subset of $\mathbb{C}(A)$, which satisfy Holder's condition with a positive, nonconvex function ϕ, i.e., any $f : A \to \mathbb{R}$ in this set satisfies $|f(\mathbf{x}_1) - f(\mathbf{x}_2)| \leq \phi(\rho(\mathbf{x}_1, \mathbf{x}_2))$ for every $\mathbf{x}_1, \mathbf{x}_2 \in A$. Consider a subset K of such functions that are bounded on A by a constant C. Suppose that $N(\varepsilon, A) \sim o(ln\ (1/\varepsilon))$ and $N(\varepsilon, A)$ is of the same order as $N(k\varepsilon, A)$ for an arbitrary $k > 0$. Then, $ln\ N(\varepsilon, K) \sim cN(\phi^{-1}(\varepsilon), A)$ for a universal constant $c > 0$. Thus, K is totally bounded. Hence, $I < \infty$ if $\int_0^1 \sqrt{N(\phi^{-1}(\varepsilon), A)} < \infty$, which is true if $\int_0^1 \sqrt{-ln\ \phi^{-1}(\varepsilon)} < \infty$ by the assumption on $N(\varepsilon, A)$. Examples of such positive, nonconvex function ϕ include $\phi(u) = u^{1/r}$ for some $r > 1$. Similar to the previous example, if we take $\mathcal{X} = L_p(A,\mu)$ for a σ-finite measure μ on A and $p \in [2,\infty)$, then for the compact sets described above, the Glivenko-Cantelli and the Donsker properties as in part (b) of Theorem 3.1.2 hold.

3.2 Spatial quantiles in Banach spaces

An important property of the spatial distribution in finite dimensional Euclidean spaces is its strict monotonicity for a class of nonatomic probability distributions. This along with its continuity and the surjective property was used to define the spatial quantile as the inverse of the spatial distribution in these spaces (see Koltchinskii (1997)). The following result shows that even in a class of infinite dimensional Banach spaces, we have the strict monotonicity, the surjectivity and the continuity of the spatial distribution map.

Theorem 3.2.1. *Let $\mathcal{X}$ be a smooth, strictly convex Banach space, and suppose that μ is nonatomic probability measure in $\mathcal{X}$. If μ is not entirely supported on a line in $\mathcal{X}$, the spatial distribution map $\mathbf{x} \mapsto \Psi_{\mathbf{x}}$ is strictly monotone, i.e., $(\Psi_{\mathbf{x}} - \Psi_{\mathbf{y}})(\mathbf{x} - \mathbf{y}) > 0$ for all $\mathbf{x}, \mathbf{y} \in \mathcal{X}$ with $\mathbf{x} \neq \mathbf{y}$. The range of the spatial distribution map is the entire open unit ball in $\mathcal{X}^*$ if $\mathcal{X}$ is reflexive (i.e., $\mathcal{X} = \mathcal{X}^{**}$). If the norm in $\mathcal{X}$ is Fréchet differentiable, the spatial distribution map is continuous.*

Under the conditions of Theorem 3.2.1, for any $\mathbf{u}$ in the open unit ball $\mathcal{B}^*(\mathbf{0}, 1)$ in $\mathcal{X}^*$, the spatial $\mathbf{u}$-quantile $\mathbf{Q}(\mathbf{u})$ can be defined as the inverse, evaluated at $\mathbf{u}$, of the spatial distribution map. Thus, $\mathbf{Q}(\mathbf{u})$ is the solution of the equation $E\{S_{\mathbf{Q}-\mathbf{X}}\} = \mathbf{u}$. When μ has atoms, we can define $\mathbf{Q}(\mathbf{u})$ by *appropriately inverting* the spatial distribution map, which is now a continuous bijection from $\mathcal{X} \setminus A_\mu$ to $\mathcal{B}^*(\mathbf{0}, 1) \setminus \bigcup_{\mathbf{x} \in A_\mu} \overline{\mathcal{B}}^*(S_{\mathbf{x}}, p(\mathbf{x}))$ if the other conditions in Theorem 3.2.1 hold but it is discontinuous at each $\mathbf{x} \in A_\mu$. Here, A_μ denotes the set of atoms of μ, $p(\mathbf{x}) = P(\mathbf{X} = \mathbf{x})$ for $\mathbf{x} \in A_\mu$, and $\mathcal{B}^*(\mathbf{z}, r)$ and $\overline{\mathcal{B}}^*(\mathbf{z}, r)$ denote the open and the closed balls in $\mathcal{X}^*$, respectively, with radius r and center $\mathbf{z} \in \mathcal{X}^*$. Even if μ has atoms, it can be shown that $\mathbf{Q}(\mathbf{u})$ is the minimizer of $E\{||\mathbf{Q} - \mathbf{X}|| - ||\mathbf{X}||\} - \mathbf{u}(\mathbf{Q})$ with respect to $\mathbf{Q} \in \mathcal{X}$. Spatial quantiles have been defined in $\mathbb{R}^d$ through such a minimization problem in Chaudhuri (1996) and Koltchinskii (1997). The former paper also mentioned about the extension of spatial quantiles into general Banach spaces. The properties of spatial quantiles for probability distributions in $\mathbb{R}^d$ equipped with the l_p-norm for some $p \in [1, \infty)$ was studied by Chakraborty (2001). Suppose that we have a unimodal probability density function in $\mathbb{R}^d$. If the density function is a strictly decreasing function of $||\mathbf{x}||_p$, where $|| \cdot ||_p$ is the l_p-norm, then it can be easily shown that the density contours and the contours of the spatial quantiles computed using the l_p-norm coincide.

Note that the central quantiles correspond to small values of $||\mathbf{u}||$, while the extreme quantiles correspond to larger values of $||\mathbf{u}||$. Further, $\mathbf{u}/||\mathbf{u}||$ gives the direction of the proximity/remoteness of $\mathbf{Q}(\mathbf{u})$ relative the center of the probability distribution. For example, let $\mathbf{X} = (X_1, X_2, \ldots)$ be a nondegenerate random element symmetric about zero in l_p for some $p \in (1, \infty)$. So, the spatial median of $\mathbf{X}$ is zero. For any $\mathbf{u}$ in the open unit ball of $l_q = l_p^*$, where $1/p + 1/q = 1$, the spatial $\mathbf{u}$-quantile $\mathbf{Q}(\mathbf{u}) = (q_1, q_2, \ldots)$

of $\mathbf{X}$ will satisfy the equation $E\{sign(q_k - X_k)|q_k - X_k|^{p-1}/||\mathbf{Q}(\mathbf{u}) - \mathbf{X}||^{p-1}\} = u_k$ for all $k \geq 1$. If $||\mathbf{u}||$ is close to zero, then it follows from the symmetry of the distribution of X_k that q_k should also be close to zero for all $k \geq 1$. Further, if the q_k's are large, the corresponding $\mathbf{Q}(\mathbf{u})$ is an extreme quantile of the distribution of $\mathbf{X}$.

The spatial quantile possesses an equivariance property under the class of affine transformations $L : \mathcal{X} \to \mathcal{X}$ of the form $L(\mathbf{x}) = cA(\mathbf{x}) + \mathbf{a}$, where $c > 0$, $\mathbf{a} \in \mathcal{X}$ and $A : \mathcal{X} \to \mathcal{X}$ is a linear surjective isometry for all $\mathbf{x} \in \mathcal{X}$. Using the surjective property of A it follows that minimizing $E\{||\mathbf{Q} - L(\mathbf{X})|| - ||L(\mathbf{X})||\} - \mathbf{u}(\mathbf{Q})$ over $\mathbf{Q} \in \mathcal{X}$ is equivalent to minimizing $E\{||A(\mathbf{Q}') - A(\mathbf{X})|| - ||A(\mathbf{X})||\} - \mathbf{u}(A(\mathbf{Q}'))$ over $\mathbf{Q}' \in \mathcal{X}$, where $\mathbf{Q} = L(\mathbf{Q}')$. The last minimization problem is the same as minimizing $E\{||\mathbf{Q}' - \mathbf{X}|| - ||\mathbf{X}||\} - (A^*(\mathbf{u}))(\mathbf{Q}')$ over $\mathbf{Q}' \in \mathcal{X}$ by virtue of the isometry of A. Here, $A^* : \mathcal{X}^* \to \mathcal{X}^*$ denotes the adjoint of A (see the Appendix). Thus, the spatial $\mathbf{u}$-quantile of the distribution of $L(\mathbf{X})$ equals $L(\mathbf{Q}(A^*(\mathbf{u})))$, where $\mathbf{Q}(A^*(\mathbf{u}))$ is the $A^*(\mathbf{u})$-quantile of the distribution of $\mathbf{X}$.

The sample spatial $\mathbf{u}$-quantile can be defined as the minimizer over $\mathbf{Q} \in \mathcal{X}$ of $n^{-1}\sum_{i=1}^{n}\{||\mathbf{Q} - \mathbf{X}_i|| - ||\mathbf{X}_i||\} - \mathbf{u}(\mathbf{Q})$. Note that this minimization problem is an infinite dimensional one and is thus difficult to solve in general. Cadre (2001) proposed an alternative estimator of the spatial median (i.e., when $\mathbf{u} = \mathbf{0}$) by considering the above empirical minimization problem only over the data points. However, as mentioned by that author, this estimator will be inconsistent when the population spatial median lies outside the support of the distribution. Gervini (2008) proposed an algorithm for computing the sample spatial median in Hilbert spaces. However, the idea does not extend to spatial quantiles or into general Banach spaces.

We shall now discuss a computational procedure for sample spatial quantiles in a Banach space. We assume that $\mathcal{X}$ is a Banach space having a Schauder basis $\{\phi_1, \phi_2, \ldots\}$, say, so that for any $\mathbf{x} \in \mathcal{X}$, there exists a unique sequence of real numbers $\{x_k\}_{k\geq 1}$ such that $\mathbf{x} = \sum_{k=1}^{\infty} x_k\phi_k$. Let $\mathcal{Z}_n = span\{\phi_1, \phi_2, \ldots, \phi_{d(n)}\}$, where $d(n)$ is a positive integer depending on the sample size n. Define $\mathbf{z}^{(n)} = \sum_{k=1}^{d(n)} a_k\phi_k$, where $\mathbf{z} = \sum_{k=1}^{\infty} a_k\phi_k$. We will assume that $||\mathbf{z}^{(n)}|| \leq ||\mathbf{z}||$ for all $n \geq 1$ and $\mathbf{z} \in \mathcal{X}$. Note that if $\mathcal{X}$ is a Hilbert space, and $\{\phi_1, \phi_2, \ldots\}$ is an orthonormal basis of $\mathcal{X}$, then $\mathbf{z}^{(n)}$ is the orthogonal projection of $\mathbf{z}$ onto $\mathcal{Z}_n$. For each $k \geq 1$, define $\widetilde{\phi}_k$ to be the continuous linear functional on $\mathcal{X}$ given by $\widetilde{\phi}_k(\mathbf{z}) = a_k$. Let us assume that $\{\widetilde{\phi}_1, \widetilde{\phi}_2, \ldots\}$ is a Schauder basis of $\mathcal{X}^*$. Define $\mathbf{u}^{(n)} = \sum_{k=1}^{d(n)} b_k \widetilde{\phi}_k$, where $\mathbf{u} \in \mathcal{B}^*(\mathbf{0}, 1)$ and $\mathbf{u} = \sum_{k=1}^{\infty} b_k \widetilde{\phi}_k$. We also assume that $||\mathbf{u}^{(n)}|| \leq ||\mathbf{u}||$ for all $n \geq 1$ and $\mathbf{u} \in \mathcal{B}^*(\mathbf{0}, 1)$. The above assumptions concerning the Schauder bases of a Banach space and its dual space hold for any separable Hilbert space and any L_p space with $p \in (1, \infty)$ (see, e.g., (Fabian et al., 2001, pp. 166-169)). We compute the sample spatial $\mathbf{u}$-quantile $\widehat{\mathbf{Q}}(\mathbf{u})$ as the minimizer of $n^{-1} \sum_{i=1}^{n} \{||\mathbf{Q} - \mathbf{X}_i^{(n)}|| - ||\mathbf{X}_i^{(n)}||\} - \mathbf{u}^{(n)}(\mathbf{Q})$ over $\mathbf{Q} \in \mathcal{Z}_n$.

For all the numerical studies in this chapter, we have chosen $d(n) = [\sqrt{n}]$. In our simulated data examples, sample quantiles computed with this choice of $d(n)$ approximate the true quantiles quite well. We will later show that this choice of $d(n)$ ensures the consistency of sample quantiles in a class of Banach spaces, and is sufficient to prove their asymptotic Gaussianity in separable Hilbert spaces (cf. Theorems 3.2.1.1 and 3.2.1.3).

We now demonstrate the spatial quantiles using some simulated and real data. We have considered the random element $\mathbf{X} = \sum_{k=1}^{\infty} \lambda_k Y_k \phi_k$ in $L_2[0,1]$. Here, the Y_k's are independent $N(0,1)$ random variables, $\lambda_k = \{(k - 0.5)\pi\}^{-1}$ and $\phi_k(t) = \sqrt{2} sin\{(k - 0.5)\pi t\}$ for $k \geq 1$. Note that $\mathbf{X}$ has the distribution of the standard Brownian motion on $[0,1]$ with ϕ_k being the eigenfunction associated with the eigenvalue λ_k^2 of the covariance kernel of the standard Brownian motion. Figure 3.1 shows the plots of the population spatial quantiles of the standard Brownian motion for $\mathbf{u} = \pm c\phi_k$, where $k = 1, 2, 3$ and $c = 0.25, 0.5, 0.75$. We have also plotted the spatial median, which is presented as the horizontal line through zero in all the plots. For each $k = 1, 2, 3$, the spatial quantiles corresponding to $\mathbf{u} = c\phi_k$ for $c = 0.25$, 0.5 and 0.75 are given by the black, the red and the violet solid curves, respectively, while those corresponding to $\mathbf{u} = -c\phi_k$ for these c values are given by the black, the red and the violet dashed curves, respectively. Note that $\lambda_1 Y_1$, $\lambda_2 Y_2$ and $\lambda_3 Y_3$ account for 81.1%, 9% and 3.24%, respectively, of the total variation $E(||\mathbf{X}||^2) = \sum_{k=1}^{\infty} Var(\lambda_k Y_k) = \sum_{k=1}^{\infty} \lambda_k^2$ in the Brownian motion process. For computing the population spatial quantiles, we generated a large sample of size $n = 2500$ from the standard Brownian motion and computed the sample spatial quantiles with $d(n) = [\sqrt{n}]$ and $\mathcal{Z}_n = span\{\phi_1, \phi_2, \ldots, \phi_{d(n)}\}$.

Our simulated data consists of $n = 50$ sample curves from the standard Brownian

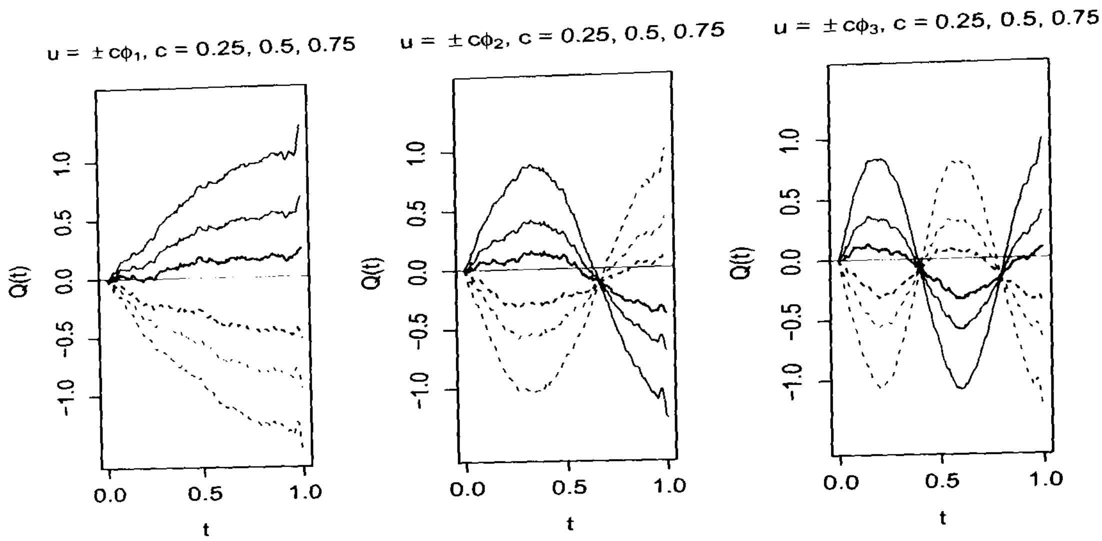

Figure 3.1: Plots of the spatial quantiles of the standard Brownian motion including the spatial median.

motion, and each sample curve is observed at 250 equispaced points in $[0,1]$. The real dataset considered here is the Spectrometry data used earlier in Chapter 2. For each of the simulated and the real dataset, we have chosen $d(n) = \lceil\sqrt{n}\rceil$, and $\mathcal{Z}_n$ is constructed using the eigenvectors associated with the $d(n)$ largest eigenvalues of the sample covariance matrix. For computing the sample spatial quantiles for both the simulated and the real data, we have first computed the sample spatial quantiles for the centered data obtained by subtracting the sample mean from each observation, and then added back the sample mean to the computed sample spatial quantiles. Figure 3.2 shows the plots of the simulated dataset along with the sample spatial median (the purple curve in the top left plot) and those of the sample spatial quantiles corresponding to $\mathbf{u} = \pm c\widehat{\phi}_k$ for $k = 1,2,3$, where $c = 0.25, 0.5, 0.75$ and $\widehat{\phi}_k$ is the eigenvector associated with the kth largest eigenvalue of the sample covariance matrix for $k \geq 1$. Figure 3.3 shows the plots of the real dataset along with the sample spatial medians (the purple curves in the plots in the first column) and those of the sample spatial quantiles corresponding to $\mathbf{u} = \pm c\widehat{\phi}_k$ for $k = 1,2$, where $c = 0.25, 0.5, 0.75$ and $\widehat{\phi}_k$ is the eigenvector associated with the kth largest eigenvalue of the sample covariance matrix for $k \geq 1$. In both the figures, for each k, the sample spatial quantiles corresponding to $\mathbf{u} = c\widehat{\phi}_k$ for $c = 0.25$, 0.5 and 0.75 are given by the black, the red and the violet solid curves, respectively, while those corresponding to $\mathbf{u} = -c\widehat{\phi}_k$ for these c values are given by the black, the red and the violet dashed curves, respectively. The percentage of the total variation in the simulated data explained by the first three sample eigenvectors is almost same as the population values mentioned in the previous paragraph. For each of the two classes in the real dataset, the first two sample eigenvectors account for about 99.5% of the total variation in that class.

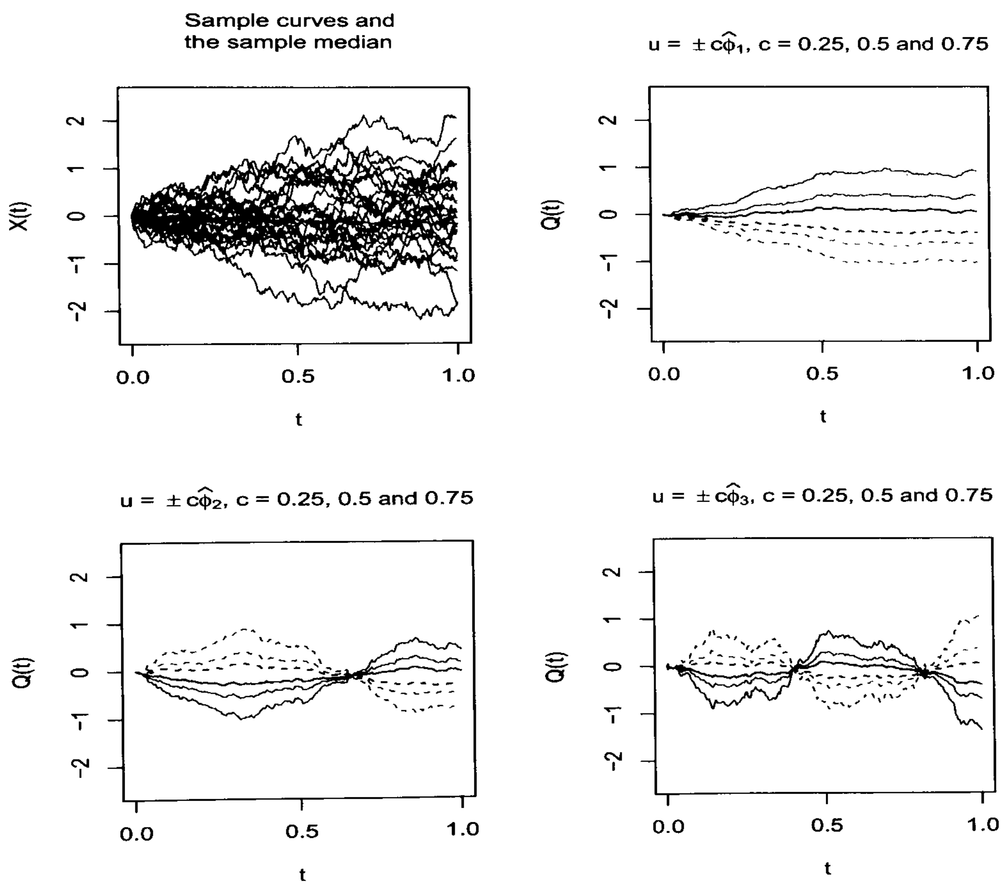

Figure 3.2: Plots of the simulated data from the standard Brownian motion along with the sample spatial quantiles.

For each k, the spatial $\mathbf{u}$-quantiles of the standard Brownian motion corresponding to $\mathbf{u} = c\phi_k$ and $-c\phi_k$ exhibit an ordering, where the spatial $\mathbf{u}$-quantile associated with a smaller c value is relatively closer to the spatial median than the spatial $\mathbf{u}$-quantile associated with a larger c value (see Figure 3.1). A similar ordering is also seen for the sample spatial quantiles of both the simulated and the two classes in the real dataset. The sample spatial median for the simulated data is close to the zero function (see Figure 3.2), which is the spatial median of the standard Brownian motion. There is a noticeable difference in the locations of the sample spatial median and the sample spatial quantiles corresponding to $\mathbf{u} = \pm c\widehat{\phi}_1$ between the two classes in the real dataset (see Figure 3.3). Moreover, the sample spatial quantiles of the two classes in the real dataset are different in their shapes.

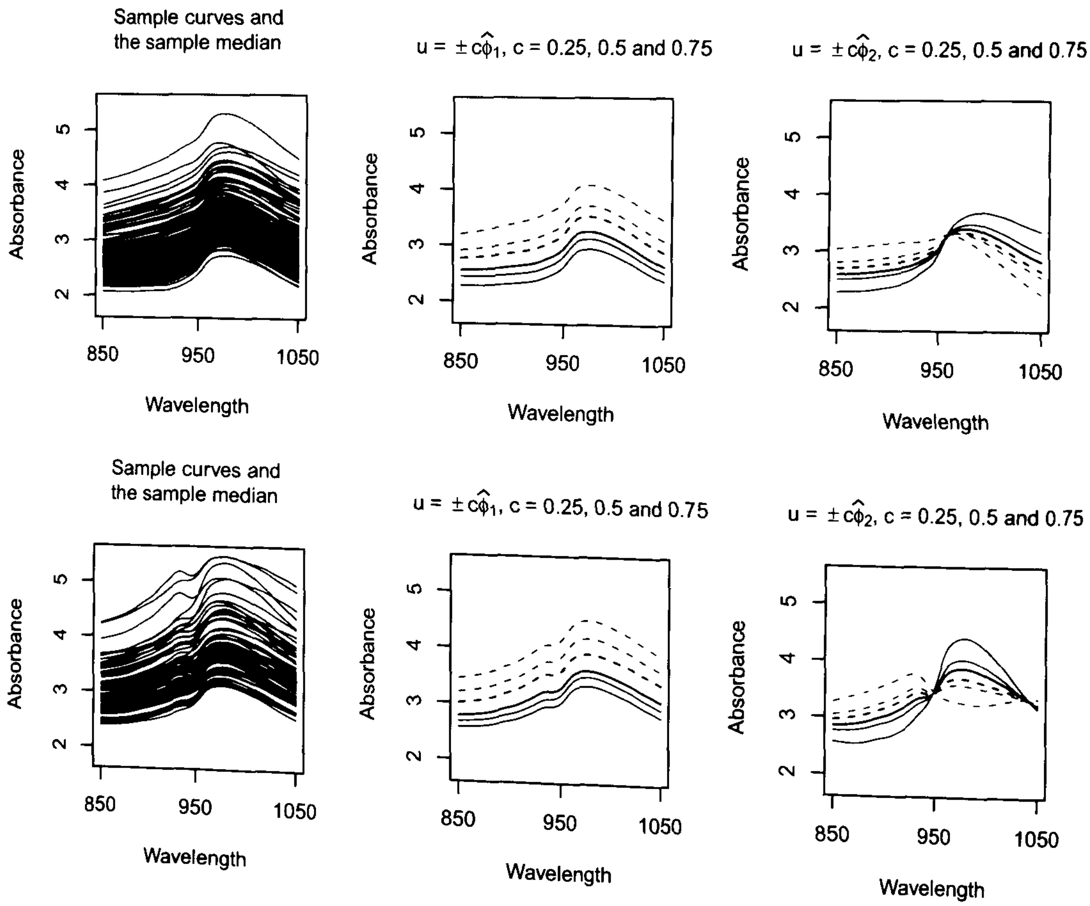

Figure 3.3: Plots of the spectrometric data along with the sample spatial quantiles.

3.2.1 Asymptotic properties of sample spatial quantiles

The following theorem gives the strong consistency of $\widehat{\mathbf{Q}}(\mathbf{u})$ in the norm topology for a class of Banach spaces.

Theorem 3.2.1.1. *Suppose that $\mathcal{X}$ is a separable, reflexive Banach space such that the norm in $\mathcal{X}$ is locally uniformly rotund, and assume that μ is nonatomic and not entirely supported on a line in $\mathcal{X}$. Then, $||\widehat{\mathbf{Q}}(\mathbf{u}) - \mathbf{Q}(\mathbf{u})|| \to 0$ as $n \to \infty$ almost surely if $d(n) \to \infty$ as $n \to \infty$.*

Since $\widehat{\mathbf{Q}}(\mathbf{u})$ is a nonlinear function of the data, in order to study its asymptotic distribution, we need to approximate it by a suitable linear function of the data. In finite dimensions, this is achieved through a Bahadur type asymptotic linear representation (see, e.g., Chaudhuri (1996) and Koltchinskii (1997)), and our next theorem gives a similar representation in infinite dimensional Hilbert spaces. Consider the real-valued function $g(\mathbf{Q}) = E\{||\mathbf{Q} - \mathbf{X}|| - ||\mathbf{X}||\} - \mathbf{u}(\mathbf{Q})$ defined on a Hilbert space $\mathcal{X}$, and denote its Hessian at $\mathbf{Q} \in \mathcal{X}$ by $\mathrm{J}_{\mathbf{Q}}$, which is a continuous linear operator on $\mathcal{X}$ (see the Appendix). We denote by $\widetilde{\mathrm{J}}_{\mathbf{Q}}$ the symmetric bounded bilinear functional from $\mathcal{X} \times \mathcal{X}$ into $\mathbb{R}$ associated with $\mathrm{J}_{\mathbf{Q}}$, which satisfies

$$\lim_{t \to 0} \left| g(\mathbf{Q} + t\mathbf{h}) - g(\mathbf{Q}) - tE\left\{ \frac{\mathbf{Q} - \mathbf{X}}{||\mathbf{Q} - \mathbf{X}||} - \mathbf{u} \right\}(\mathbf{h}) - \frac{t^2}{2}\widetilde{\mathrm{J}}_{\mathbf{Q}}(\mathbf{h}, \mathbf{h}) \right| / t^2 = 0$$

for any $\mathbf{h} \in \mathcal{X}$. Here, $\widetilde{\mathrm{J}}_{\mathbf{Q}}(\mathbf{h}, \mathbf{v}) = \langle \mathrm{J}_{\mathbf{Q}}(\mathbf{h}), \mathbf{v} \rangle$ for every $\mathbf{h}, \mathbf{v} \in \mathcal{X}$. We define the Hessian $\mathrm{J}_{n,\mathbf{Q}}$ of the function $g_n(\mathbf{Q}) = E\{||\mathbf{Q} - \mathbf{X}^{(n)}|| - ||\mathbf{X}^{(n)}||\} - \mathbf{u}^{(n)}(\mathbf{Q})$, which is defined on $\mathcal{Z}_n$, in a similar way. The symmetric bounded bilinear functional associated with $\mathrm{J}_{n,\mathbf{Q}}$ is denoted by $\widetilde{\mathrm{J}}_{n,\mathbf{Q}}$. Here, we consider an orthonormal basis of $\mathcal{X}$ (which is a Schauder basis), and $\mathcal{Z}_n$ is as chosen as in Section 3.2. Let $\mathbf{Q}_n(\mathbf{u}) = \arg\min_{\mathbf{Q} \in \mathcal{Z}_n} g_n(\mathbf{Q})$ and define $\mathrm{B}_n(\mathbf{u}) = ||\mathbf{Q}_n(\mathbf{u}) - \mathbf{Q}(\mathbf{u})||$. It can be shown that $\mathrm{B}_n(\mathbf{u}) \to 0$ as $n \to \infty$. We make the following assumption, which will be required for Theorem 3.2.1.2 below.

ASSUMPTION (B). *Suppose that μ is nonatomic and not entirely supported on a line in $\mathcal{X}$, and $\sup_{\mathbf{Q} \in \mathcal{Z}_n, ||\mathbf{Q}|| \le C} E\{||\mathbf{Q} - \mathbf{X}^{(n)}||^{-2}\} < \infty$ for each $C > 0$ and all appropriately large n.*

As discussed after Assumption (A) in Section 3.1, if $\mathcal{X}$ is a Hilbert space, we can choose $T(\mathbf{x}) = 2/||\mathbf{x}||$ in that assumption. Thus, Assumption (B) can be viewed as a $d(n)$-dimensional analog of the moment condition assumed in part (b) of Theorem 3.1.2. Also, it holds under the same situation as discussed after Theorem 3.1.2.

Theorem 3.2.1.2. *Let $\mathcal{X}$ be a separable Hilbert space, and Assumption (B) hold. Then, the following Bahadur type asymptotic linear representation holds if for some $\alpha \in (0, 1/2]$, $d(n)/n^{1-2\alpha}$ tends to a positive constant as $n \to \infty$.*

$$\widehat{\mathbf{Q}}(\mathbf{u}) - \mathbf{Q}_n(\mathbf{u}) \;=\; -\frac{1}{n}\sum_{i=1}^{n}[\mathrm{J}_{n,\mathbf{Q}_n(\mathbf{u})}]^{-1}\left(\frac{\mathbf{Q}_n(\mathbf{u}) - \mathbf{X}_i^{(n)}}{||\mathbf{Q}_n(\mathbf{u}) - \mathbf{X}_i^{(n)}||} - \mathbf{u}^{(n)}\right) + \mathbf{R}_n,$$

where $\mathbf{R}_n = O((ln\ n)/n^{2\alpha})$ as $n \to \infty$ almost surely.

The Bahadur-type representation of the sample spatial $\mathbf{u}$-quantile in finite dimensional Euclidean spaces (see, e.g., Chaudhuri (1996) and Koltchinskii (1997)) can be obtained as a straightforward corollary of the above theorem by choosing $\alpha = 1/2$. Under the assumptions of the preceding theorem, if $\alpha \in (1/4, 1/2]$, we have the asymptotic Gaussianity of $\sqrt{n}(\widehat{\mathbf{Q}}(\mathbf{u}) - \mathbf{Q}_n(\mathbf{u}))$ as $n \to \infty$.

The extension of the above Bahadur-type representation into general Banach spaces is a challenging task mainly due to two reasons. First, although some version of Bernstein type exponential bounds as in Fact 3.3.1 are available in general Banach spaces (see, e.g., Theorem 2.1 in Yurinskiĭ (1976)), those bounds are not adequate for extending the proof from Hilbert spaces into general Banach spaces. Next, the lower bound of $\widetilde{\mathrm{J}}_{n,\mathbf{Q}}(\mathbf{h},\mathbf{h})/||\mathbf{h}||^2$ in Fact 3.3.3 is not always true in general Banach spaces. For instance, let $\mathcal{X} = l_4$ and $\mathbf{X} = (X_1, X_2, \ldots)$ be a zero mean Gaussian random element in $\mathcal{X}$. Let $\mathcal{Z}_n = span\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_{d(n)}\}$, where $\mathbf{e}_k = (I(j = k) : j \geq 1)$, $k \geq 1$, which form the canonical Schauder basis for l_4. Let $\mathbf{h}_n = \mathbf{e}_{d(n)} \in \mathcal{Z}_n$. Then, for any $\mathbf{Q} = (q_1, q_2, \ldots) \in \mathcal{Z}_n$, it can be shown that $\widetilde{\mathrm{J}}_{n,\mathbf{Q}}(\mathbf{h}_n, \mathbf{h}_n)/||\mathbf{h}_n||^2 \leq 3E[(q_{d(n)} - X_{d(n)})^2/||\mathbf{Q} - \mathbf{X}^{(n)}||^3]$. It can also be shown that the last term converges to zero as $n \to \infty$ by observing that $|q_{d(n)} - X_{d(n)}| \to 0$ almost surely and $d(n) \to \infty$ as $n \to \infty$. This clearly implies that the lower bound in Fact 3.3.3 does not hold in this case.

We shall now discuss some situations when $\mathrm{B}_n(\mathbf{u}) = ||\mathbf{Q}_n(\mathbf{u}) - \mathbf{Q}(\mathbf{u})||$ satisfies $\lim_{n\to\infty} \sqrt{n}\mathrm{B}_n(\mathbf{u}) = 0$. This along with the weak convergence of $\sqrt{n}(\widehat{\mathbf{Q}}(\mathbf{u}) - \mathbf{Q}_n(\mathbf{u}))$

stated above will give the asymptotic Gaussianity of $\sqrt{n}(\widehat{\mathbf{Q}}(\mathbf{u})-\mathbf{Q}(\mathbf{u}))$ as $n \to \infty$. Under the assumptions of Theorem 3.2.1.2, it can be shown that for some constants $b_1, b_2 > 0$, we have $\mathrm{B}_n(\mathbf{u}) \leq b_1 r_n + b_2 s_n$ for all large n, where $r_n = E\{||\mathbf{X}-\mathbf{X}^{(n)}||/||\mathbf{Q}(\mathbf{u})-\mathbf{X}||\}$ and $s_n = ||\mathbf{u}-\mathbf{u}^{(n)}||$. Let us take $\mathcal{X} = L_2([a,b],\nu)$, which is the space of all real-valued functions $\mathbf{x}$ on $[a,b] \subseteq \mathbb{R}$ with ν a probability measure on $[a,b]$ such that $\int \mathbf{x}^2(t)\nu(dt) < \infty$. Suppose $\mathbf{X}$ has the Karhunen-Loève expansion $\mathbf{X} = \mathbf{m} + \sum_{k=1}^{\infty} \lambda_k Y_k \phi_k$, where the Y_k's are uncorrelated random variables with zero means and unit variances, the λ_k^2's and the ϕ_k's are the eigenvalues and the eigenfunctions, respectively, of the covariance of $\mathbf{X}$. Let $\mathcal{Z}_n = span\{\phi_1, \phi_2, \ldots, \phi_{d(n)}\}$. Under the assumptions of Theorem 3.2.1.2, it can be shown that $\lim_{n\to\infty} \sqrt{n} r_n = 0$ if $\lim_{n\to\infty} \sqrt{n}||\mathbf{m} - \mathbf{m}^{(n)}|| = 0$ and $\lim_{n\to\infty} n \sum_{k>d(n)} \lambda_k^2 = 0$. The latter is true for some $\alpha > 1/4$ if $\lim_{k\to\infty} k^2 \lambda_k = 0$ (e.g., if the λ_k's decay geometrically as $k \to \infty$). We now discuss some conditions that are sufficient to ensure $\lim_{n\to\infty} \sqrt{n}||\mathbf{m} - \mathbf{m}^{(n)}|| = 0$ as well as $\lim_{n\to\infty} \sqrt{n} s_n = 0$ (implying that $\lim_{n\to\infty} \sqrt{n}\mathrm{B}_n(\mathbf{u}) = 0$) in separable Hilbert spaces. If $\mathcal{X} = L_2([0,1],\nu)$, where ν is the uniform distribution, and $\{\phi_k\}_{k\geq 1}$ is the set of standard Fourier basis functions, then Theorem 4.4 in Vretblad (2003) describes those $\mathbf{x} \in \mathcal{X}$ for which $\lim_{n\to\infty} \sqrt{n}||\mathbf{x} - \mathbf{x}^{(n)}|| = 0$ holds. It follows from that theorem that a sufficient condition for $\lim_{n\to\infty} \sqrt{n}||\mathbf{x} - \mathbf{x}^{(n)}|| = 0$ to hold is that $\mathbf{x}$ is thrice differentiable on $[0,1]$, $\mathbf{x}(0) = \mathbf{x}(1)$, and its right hand derivative at 0 equals its left hand derivative at 1 for each of the three derivatives. On the other hand, if $\{\phi_k\}_{k\geq 1}$ is either the set of normalized Chebyshev or Legendre polynomials, which form orthonormal bases of $\mathcal{X}$ when ν is the uniform and the $Beta(1/2, 1/2)$ distributions, respectively, then $\mathbf{x} \in \mathcal{X}$ satisfying $\lim_{n\to\infty} \sqrt{n}||\mathbf{x} - \mathbf{x}^{(n)}|| = 0$ can be obtained using Theorem 4.2 in Trefethen (2008) and Theorem 2.1 in Wang and Xiang (2012), respectively. Next, let $\mathcal{X} = L_2(\mathbb{R},\nu)$, where ν is the normal distribution with zero mean and variance 1/2, and $\phi_k(t) \propto exp\{-At^2\}h_k(A't), t \in \mathbb{R}, k \geq 1$ for an appropriate $A \geq 0$ and $A' > 0$, where $\{h_k\}_{k\geq 1}$ is the set of Hermite polynomials. Then, $\mathbf{x} \in \mathcal{X}$ satisfying $\lim_{n\to\infty} \sqrt{n}||\mathbf{x} - \mathbf{x}^{(n)}|| = 0$ can be obtained from the conditions of the theorem in p. 385 in Boyd (1984) for $j \geq 5$. An important special case in this setup is the Gaussian process with the Gaussian covariance kernel, which is used in classification and regression problems (see, e.g., Rasmussen and Williams (2006)). The eigenvalues of this kernel decay geometrically, which implies that $\lim_{n\to\infty} n \sum_{k>d(n)} \lambda_k^2 = 0$ for some $\alpha > 1/4$. Summarizing this discussion, we have the following theorem.

Theorem 3.2.1.3. *Suppose that the assumptions of Theorem 3.2.1.2 hold. Also, assume that for some* $\alpha \in (1/4, 1/2]$, $\sqrt{n}s_n \to 0$, $\sqrt{n}||\mathbf{m} - \mathbf{m}^{(n)}|| \to 0$ *and* $n\sum_{k>d(n)} \lambda_k^2 \to 0$ *as* $n \to \infty$. *Then, there exists a zero mean Gaussian random element* $\mathbf{Z_u}$ *such that* $\sqrt{n}(\widehat{\mathbf{Q}}(\mathbf{u}) - \mathbf{Q}(\mathbf{u}))$ *converges weakly to* $\mathbf{Z_u}$ *as* $n \to \infty$. *The co-variance of* $\mathbf{Z_u}$ *is given by* $V_\mathbf{u} = [\mathrm{J}_{\mathbf{Q}(\mathbf{u})}]^{-1}\Lambda_\mathbf{u}[\mathrm{J}_{\mathbf{Q}(\mathbf{u})}]^{-1}$, *where* $\Lambda_\mathbf{u} : \mathcal{X} \to \mathcal{X}$ *satisfies* $\langle \Lambda_\mathbf{u}(\mathbf{z}), \mathbf{w}\rangle = E\left\{\left\langle \frac{\mathbf{Q}(\mathbf{u})-\mathbf{X}}{||\mathbf{Q}(\mathbf{u})-\mathbf{X}||} - \mathbf{u}, \mathbf{z}\right\rangle \left\langle \frac{\mathbf{Q}(\mathbf{u})-\mathbf{X}}{||\mathbf{Q}(\mathbf{u})-\mathbf{X}||} - \mathbf{u}, \mathbf{w}\right\rangle\right\}$ *for* $\mathbf{z}, \mathbf{w} \in \mathcal{X}$, *and* $\langle\cdot,\cdot\rangle$ *denotes the inner product in* $\mathcal{X}$.

3.2.2 Asymptotic efficiency of the sample spatial median

We will now study the asymptotic efficiency of the sample spatial median $\widehat{\mathbf{Q}}(\mathbf{0})$ relative to the sample mean $\overline{\mathbf{X}}$ when $\mathbf{X}$ has a symmetric distribution in a Hilbert space $\mathcal{X}$ about some $\mathbf{m} \in \mathcal{X}$. In this case, $\mathbf{Q}(\mathbf{0}) = E(\mathbf{X}) = \mathbf{m}$. We assume that $E(||\mathbf{X}||^2) < \infty$, and let Σ be the covariance of $\mathbf{X}$. Note that $\mathbf{Q}_n(\mathbf{0}) = \mathbf{m}^{(n)}$, and following the discussion after Theorem 3.2.1.2, it can be shown that under the conditions of that theorem and if $\sqrt{n}||\mathbf{m} - \mathbf{m}^{(n)}|| \to 0$ as $n \to \infty$, we have the weak convergence of $\sqrt{n}(\widehat{\mathbf{Q}}(\mathbf{0}) - \mathbf{m})$ to $\mathbf{Z_0}$ as $n \to \infty$. Here, $\mathbf{Z_0}$ is a Gaussian random element with zero mean and covariance $V_\mathbf{0}$ as in Theorem 3.2.1.3. On the other hand, using the central limit theorem in Hilbert spaces, we have the weak convergence of $\sqrt{n}(\overline{\mathbf{X}} - \mathbf{m})$ to a Gaussian random element with zero mean and covariance Σ.

For our asymptotic efficiency study, we have first considered $\mathbf{X} = \mathbf{m} + \sum_{k=1}^{\infty} \lambda_k Y_k \phi_k$ in $L_2[0,1]$ with Y_k's having independent standard normal distributions, and the λ_k^2's and the ϕ_k's being the eigenvalues and the eigenfunctions of the covariance kernel $K(t,s) = 0.5(t^{2H} + s^{2H} - |t-s|^{2H})$ for H ranging from 0.1 to 0.9. In this case, $\mathbf{X}$ has the distribution of a fractional Brownian motion on $[0,1]$ with mean $\mathbf{m}$ and Hurst index H. We have also considered t processes (see, e.g., Yu et al. (2007)) on $[0,1]$ with mean $\mathbf{m}$, degrees of freedom $r \geq 3$ and covariance kernel $K(t,s) = \min(t,s)$. In this case, $\mathbf{X} = \mathbf{m} + \sum_{k=1}^{\infty} \lambda_k Y_k \phi_k$ with $Y_k = Z_k/\sqrt{W/r}$ for $r \geq 3$, where the Z_k's are independent standard normal variables, and W is an independent chi-square variable with r degrees of freedom.

Here, the λ_k^2's and the ϕ_k's are the eigenvalues and the eigenfunctions, respectively, of the covariance kernel $K(t,s) = \min(t,s)$. We have also included in our study the distributions of $\mathbf{X} = \mathbf{m} + \sum_{k=1}^{\infty} \lambda_k Y_k \phi_k$ in $L_2(\mathbb{R}, \nu)$ corresponding to all the choices of the Y_k's mentioned above. Here, ν is the normal distribution with zero mean and variance 1/2, the λ_k^2's and the ϕ_k's are the eigenvalues and the eigenfunctions, respectively, of the Gaussian covariance kernel $K(t,s) = exp\{-(t-s)^2\}$ (see Section 4.3 in Rasmussen and Williams (2006)). These processes on $\mathbb{R}$ are the Gaussian and the t processes with r degrees of freedom for $r \geq 3$, respectively, having mean $\mathbf{m}$ and the Gaussian covariance kernel. The mean function $\mathbf{m}$ of each of the processes considered above is assumed to satisfy $\sqrt{n}||\mathbf{m} - \mathbf{m}^{(n)}|| \to 0$ as $n \to \infty$ so that we can apply Theorem 3.2.1.3. The asymptotic efficiency of $\widehat{\mathbf{Q}}(\mathbf{0})$ relative to $\overline{\mathbf{X}}$ can be defined as $trace(\Sigma)/trace(V_0)$. The traces of Σ and V_0 are defined as $\sum_{k=1}^{\infty} \langle \Sigma\psi_k, \psi_k \rangle$ and $\sum_{k=1}^{\infty} \langle V_0\psi_k, \psi_k \rangle$, respectively, where $\{\psi_k\}_{k\geq 1}$ is an orthonormal basis of the Hilbert space $\mathcal{X}$. It can be shown that both the infinite sums are convergent, and their values are independent of the choice of $\{\psi_k\}_{k\geq 1}$. For numerically computing the efficiency, each of the two infinite dimensional covariances are replaced by the D-dimensional covariance matrix of the distribution of $(\mathbf{X}(t_1), \mathbf{X}(t_2), \ldots, \mathbf{X}(t_D))$, where D is appropriately large. For the processes in $L_2[0,1]$, $t_1, t_2, \ldots, t_D$ are chosen to be equispaced points in $[0,1]$, while for the processes in $L_2(\mathbb{R}, \nu)$, these points are chosen randomly from the distribution ν. These choices ensures that for any $\mathbf{x} \in L_2[0,1]$ or $L_2(\mathbb{R}, \nu)$, $||\mathbf{x}||^2$ can be approximated by the average of $\mathbf{x}^2(t)$ over these D points. For our numerical evaluation of the asymptotic efficiencies, we have chosen $D = 200$.

The efficiency of $\widehat{\mathbf{Q}}(\mathbf{0})$ relative to $\overline{\mathbf{X}}$ for the fractional Brownian motion decreases from 0.923 to 0.718 as the value of H increases from 0.1 to 0.9. For the Brownian motion (i.e., when $H = 0.5$) this efficiency is 0.83. For the t-processes in $[0,1]$, this efficiency is 2.135 for 3 degrees of freedom, and it decreases with the increase in the degrees of freedom. The efficiency remains more than 1 up to 9 degrees of freedom, when its value is 1.006. This efficiency for the Gaussian process in $L_2(\mathbb{R}, \nu)$ is 0.834. The efficiency for the t-processes in $L_2(\mathbb{R}, \nu)$ is 2.247 for 3 degrees of freedom, and it decreases with the increase in the degrees of freedom. As before, this efficiency remains more than 1 up to 9 degrees of freedom, when its value is 1.013.

3.3 Mathematical details

Lemma 3.3.1. *Suppose that $\mathcal{X}^*$ is a separable Banach space. If μ is atomic, then $\sup_{\mathbf{x}\in\mathcal{X}} ||\widehat{\Psi}_{\mathbf{x}} - \Psi_{\mathbf{x}}|| \to 0$ as $n \to \infty$ almost surely.*

Proof. Define $\widehat{p}(\mathbf{y}) = n^{-1}\sum_{i=1}^{n} I(\mathbf{X}_i = \mathbf{y})$ and $p(\mathbf{y}) = P(\mathbf{X} = \mathbf{y})$ for $\mathbf{y} \in A_\mu$, where A_μ denotes the set of atoms of μ. By the strong law of large numbers, $\lim_{n\to\infty}\widehat{p}(\mathbf{y}) = p(\mathbf{y})$ almost surely for each $\mathbf{y} \in A_\mu$. Observe that $\sup_{\mathbf{x}\in\mathcal{X}} ||\widehat{\Psi}_{\mathbf{x}} - \Psi_{\mathbf{x}}|| \leq \sum_{\mathbf{y}\in A_\mu} |\widehat{p}(\mathbf{y}) - p(\mathbf{y})| = 2 - 2\sum_{\mathbf{y}\in A_\mu}\min\{\widehat{p}(\mathbf{y}), p(\mathbf{y})\}$. Since $\min\{\widehat{p}(\mathbf{y}), p(\mathbf{y})\} \leq p(\mathbf{y})$, the proof is complete by the dominated convergence theorem. □

Proof of Theorem 3.1.1. Let us write $\mu = \rho\mu_1 + (1-\rho)\mu_2$, where μ_1 and μ_2 are the nonatomic and the atomic parts of μ, respectively. Let $N_n = \sum_{i=1}^{n} I(\mathbf{X}_i \notin A_\mu)$, where A_μ is the set of atoms of μ. Denote by $\widehat{\mu}_1$ and $\widehat{\mu}_2$ the empirical probability distributions corresponding to μ_1 and μ_2, respectively. Here, as well as in other proofs in this section, we will denote the inner product in a Hilbert space by $\langle\cdot,\cdot\rangle$. Observe that for any $\mathbf{x} \in \mathcal{Z}$ and $\mathbf{l} \in \mathcal{X}$,

$$
\begin{aligned}
\left|\left\langle \mathbf{l}, \widehat{\Psi}_{\mathbf{x}} - \Psi_{\mathbf{x}}\right\rangle\right| \leq\ & \frac{N_n}{n}\left|E_{\widehat{\mu}_1}\left(\left\langle \mathbf{l}, \frac{\mathbf{x}-\mathbf{X}}{||\mathbf{x}-\mathbf{X}||}\right\rangle\right) - E_{\mu_1}\left(\left\langle \mathbf{l}, \frac{\mathbf{x}-\mathbf{X}}{||\mathbf{x}-\mathbf{X}||}\right\rangle\right)\right| \\
& + \left|\frac{N_n}{n}E_{\mu_1}\left(\left\langle \mathbf{l}, \frac{\mathbf{x}-\mathbf{X}}{||\mathbf{x}-\mathbf{X}||}\right\rangle\right) - \rho E_{\mu_1}\left(\left\langle \mathbf{l}, \frac{\mathbf{x}-\mathbf{X}}{||\mathbf{x}-\mathbf{X}||}\right\rangle\right)\right| \\
& + \frac{(n-N_n)}{n}\left|E_{\widehat{\mu}_2}\left(\left\langle \mathbf{l}, \frac{\mathbf{x}-\mathbf{X}}{||\mathbf{x}-\mathbf{X}||}\right\rangle\right) - E_{\mu_2}\left(\left\langle \mathbf{l}, \frac{\mathbf{x}-\mathbf{X}}{||\mathbf{x}-\mathbf{X}||}\right\rangle\right)\right| \\
& + \left|\frac{n-N_n}{n}E_{\mu_2}\left(\left\langle \mathbf{l}, \frac{\mathbf{x}-\mathbf{X}}{||\mathbf{x}-\mathbf{X}||}\right\rangle\right) - (1-\rho)E_{\mu_2}\left(\left\langle \mathbf{l}, \frac{\mathbf{x}-\mathbf{X}}{||\mathbf{x}-\mathbf{X}||}\right\rangle\right)\right| \\
\leq\ & \left|\left\langle \mathbf{l}, E_{\widehat{\mu}_1}\left(\frac{\mathbf{x}-\mathbf{X}}{||\mathbf{x}-\mathbf{X}||}\right) - E_{\mu_1}\left(\frac{\mathbf{x}-\mathbf{X}}{||\mathbf{x}-\mathbf{X}||}\right)\right\rangle\right| \\
& + \left|\left\langle \mathbf{l}, E_{\widehat{\mu}_2}\left(\frac{\mathbf{x}-\mathbf{X}}{||\mathbf{x}-\mathbf{X}||}\right) - E_{\mu_2}\left(\frac{\mathbf{x}-\mathbf{X}}{||\mathbf{x}-\mathbf{X}||}\right)\right\rangle\right| + 2\left|N_n/n - \rho\right|
\end{aligned}
$$

In other words,

$$
\begin{aligned}
\left|\left\langle \mathbf{l}, \widehat{\Psi}_{\mathbf{x}} - \Psi_{\mathbf{x}}\right\rangle\right| \leq & \left|\left\langle \mathbf{l}, E_{\widehat{\mu}_1}\left(\frac{\mathbf{x}-\mathbf{X}}{||\mathbf{x}-\mathbf{X}||}\right) - E_{\mu_1}\left(\frac{\mathbf{x}-\mathbf{X}}{||\mathbf{x}-\mathbf{X}||}\right)\right\rangle\right| \\
& + ||\mathbf{l}||\left\|E_{\widehat{\mu}_2}\left(\frac{\mathbf{x}-\mathbf{X}}{||\mathbf{x}-\mathbf{X}||}\right) - E_{\mu_2}\left(\frac{\mathbf{x}-\mathbf{X}}{||\mathbf{x}-\mathbf{X}||}\right)\right\| + 2\left|N_n/n - \rho\right|. \quad (3.1)
\end{aligned}
$$

The third term in the right hand side of (3.1) converges to zero as $n \to \infty$ almost surely by the strong law of large numbers. By Lemma 3.3.1, the second term in the right hand side of (3.1) converges to zero uniformly over $\mathbf{x} \in \mathcal{X}$ as $n \to \infty$ almost surely.

Let us next consider the class of functions

$$\mathcal{G} = \{\psi_{\mathbf{x}} : \mathcal{X} \to \mathbb{R},\ \psi_{\mathbf{x}}(\mathbf{s}) = \langle \mathbf{l}, \mathbf{x} - \mathbf{s} \rangle I(\mathbf{x} \neq \mathbf{s}) / ||\mathbf{x} - \mathbf{s}||;\ \mathbf{x} \in \mathcal{Z}\}.$$

Similar arguments as those in the proofs of Theorems 5.5 and 5.6 in pp. 471 – 474 in Koltchinskii (1997) show that $\mathcal{G}$ is a VC-subgraph class. Since μ_1 is nonatomic, the functions in $\mathcal{G}$ are almost surely μ_1-continuous. Thus, using the separability of $\mathcal{X}$, we get that $\mathcal{G}$ is a pointwise separable class (see p. 116 in van der Vaart and Wellner (1996)) with an envelope function that is unity everywhere. Thus, it follows from Theorem 2.6.8 in van der Vaart and Wellner (1996) that $\mathcal{G}$ is a Glivenko-Cantelli class with respect to the measure μ_1, which implies that the first term in the right hand side of (3.1) converges uniformly over $\mathbf{x} \in \mathcal{Z}$ as $n \to \infty$ almost surely.

Since $\mathcal{X}$ is separable, it has a countable dense subset $\mathcal{L}$. So,

$$\lim_{n\to\infty} \sup_{\mathbf{x}\in\mathcal{Z}} \left| \left\langle \mathbf{l}, E_{\widehat{\mu}_1}\left(\frac{\mathbf{x}-\mathbf{X}}{||\mathbf{x}-\mathbf{X}||}\right) - E_{\mu_1}\left(\frac{\mathbf{x}-\mathbf{X}}{||\mathbf{x}-\mathbf{X}||}\right)\right\rangle \right| = 0 \quad \forall\, \mathbf{l} \in \mathcal{L} \tag{3.2}$$

as $n \to \infty$ almost surely. Note that both the expectations in (3.2) above are bounded in norm by 1. Using this fact, equation (3.2) and the fact that $\mathcal{L}$ is dense in $\mathcal{X}$, we get the proof.

For the second part of the theorem, note that it is enough to prove the result for $d = 1$. By the Riesz representation theorem, for any continuous linear map $\mathbf{g} : \mathcal{X} \to \mathbb{R}$, there exists $\mathbf{l} \in \mathcal{X}$ satisfying $\mathbf{g}(\mathbf{x}) = \langle \mathbf{l}, \mathbf{x} \rangle$ for every $\mathbf{x} \in \mathcal{X}$. Let us consider the class of functions $\mathcal{G}$ defined above in the proof of the first part of this theorem. If μ itself is nonatomic, it follows from the arguments in that proof by replacing μ_1 with μ that $\mathcal{G}$ is a VC-subgraph class. This along with Theorem 2.6.8 in van der Vaart and Wellner (1996) implies that $\mathcal{G}$ is a Donsker class with respect to μ. This completes the proof of the theorem. □

REMARK: Suppose that $\mathcal{X}$ is a L_p space for an even integer $p > 2$. Using arguments similar to those used in deriving (3.1), we get an analogous bound for $\mathbf{l}(S_{\mathbf{x}-\mathbf{X}})$ for any $\mathbf{x} \in \mathcal{Z}$ and $\mathbf{l} \in \mathcal{X}$. In this case, $\mathcal{G}$ in the proof of Theorem 3.1.1 is to be defined as $\mathcal{G} = \{\psi_{\mathbf{x}} : \mathcal{X} \to \mathbb{R},\ \psi_{\mathbf{x}}(\mathbf{s}) = \mathbf{l}(S_{\mathbf{x}-\mathbf{s}});\ \mathbf{x} \in \mathcal{Z}\}$, and $\mathbf{g}$ in that theorem is to be chosen a function from $\mathcal{X}^*$ into $\mathbb{R}^d$. Using arguments similar to those in the proof of Theorem 3.1.1, it can be shown that $\mathcal{G}$ is a VC-subgraph and a pointwise separable class, and hence a Glivenko-Cantelli and a Donsker class. So, the assertions of Theorem 3.1.1 hold in this case as well.

The following fact is a generalization of the Bernstein inequality for probability distributions in separable Hilbert spaces, and it will be used in the proof of Theorem 3.1.2(b).

Fact 3.3.1. (Yurinskiĭ, 1976, p. 491) *Let $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_n$ be independent random elements in a separable Hilbert space $\mathcal{X}$ satisfying $E(\mathbf{Y}_i) = \mathbf{0}$ for $1 \leq i \leq n$. Suppose that for some $h > 0$ and $u_i > 0$, we have $E(||\mathbf{Y}_i||^m) \leq (m!/2)u_i^2 h^{m-2}$ for $1 \leq i \leq n$ and all $m \geq 2$. Let $U_n^2 = \sum_{i=1}^n u_i^2$. Then, for any $v > 0$, $P(||\sum_{i=1}^n \mathbf{Y}_i|| \geq vU_n) \leq 2exp\{-(v^2/2)(1 + 1.62(vh/U_n))^{-1}\}$.*

Proof of Theorem 3.1.2. (a) As in the proof of Theorem 3.1.1, we get

$$\begin{aligned} ||\widehat{\Psi}_{\mathbf{x}} - \Psi_{\mathbf{x}}|| \ \leq \ & ||E_{\widehat{\mu}_1}\{S_{\mathbf{x}-\mathbf{X}}\} - E_{\mu_1}\{S_{\mathbf{x}-\mathbf{X}}\}|| + \\ & ||E_{\widehat{\mu}_2}\{S_{\mathbf{x}-\mathbf{X}}\} - E_{\mu_2}\{S_{\mathbf{x}-\mathbf{X}}\}|| + 2\,|N_n/n - \rho|\,. \end{aligned} \tag{3.3}$$

Further, the second and the third terms in the right hand side of the inequality in (3.3) converge to zero as $n \to \infty$ almost surely by the same arguments as in the proof of Theorem 3.1.1. Note that the convergence of the second term is uniform in $\mathcal{X}$ as before.

Now, for an $\varepsilon > 0$, consider an ε-net $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{N(\varepsilon)}$ of K. The first term in the right hand side of the inequality in (3.3) is bounded above by

$$\begin{aligned} & ||E_{\widehat{\mu}_1}\{S_{\mathbf{x}-\mathbf{X}}\} - E_{\widehat{\mu}_1}\{S_{\mathbf{v}_j-\mathbf{X}}\}|| + ||E_{\mu_1}\{S_{\mathbf{x}-\mathbf{X}}\} - E_{\mu_1}\{S_{\mathbf{v}_j-\mathbf{X}}\}|| \\ & \quad + \max_{1 \leq l \leq N(\varepsilon)} ||E_{\widehat{\mu}_1}\{S_{\mathbf{v}_l-\mathbf{X}}\} - E_{\mu_1}\{S_{\mathbf{v}_l-\mathbf{X}}\}||, \end{aligned}$$

where $||\mathbf{x} - \mathbf{v}_j|| < \varepsilon$. Using Assumption (A) in Section 3.1, it follows that

$$\begin{aligned} ||E_{\widehat{\mu}_1}\{S_{\mathbf{x}-\mathbf{x}}\} - E_{\widehat{\mu}_1}\{S_{\mathbf{v}_j-\mathbf{x}}\}|| &\leq E_{\widehat{\mu}_1}\{T(\mathbf{v}_j - \mathbf{X})\}||\mathbf{x} - \mathbf{v}_j|| \\ &\leq 2\varepsilon E_{\mu_1}\{T(\mathbf{v}_j - \mathbf{X})\}, \end{aligned} \tag{3.4}$$

for all n sufficiently large almost surely. Further,

$$||E_{\mu_1}\{S_{\mathbf{x}-\mathbf{x}}\} - E_{\mu_1}\{S_{\mathbf{v}_j-\mathbf{x}}\}|| \leq \varepsilon E_{\mu_1}\{T(\mathbf{v}_j - \mathbf{X})\}. \tag{3.5}$$

Using (3.4) and (3.5), the moment condition in the theorem and the fact that $\max_{1 \leq l \leq N(\varepsilon)} ||E_{\widehat{\mu}_1}\{ S_{\mathbf{v}_l-\mathbf{x}}\} - E_{\mu_1}\{S_{\mathbf{v}_l-\mathbf{x}}\}||$ converges to zero as $n \to \infty$ almost surely, we get the proof of part (a) of the theorem.

(b) As argued in the proof of Theorem 3.1.1, it is enough to consider the case $d = 1$. Using Theorems 1.5.4 and 1.5.7 in van der Vaart and Wellner (1996), it follows that we only need to prove the asymptotic equicontinuity in probability of $\widehat{\Upsilon}_{\mathbf{g}}$ with respect to the norm in $\mathcal{X}$. Further, since μ is assumed to be nonatomic, the map $\mathbf{x} \mapsto \mathbf{g}(\sqrt{n}(\widehat{\Psi}_{\mathbf{x}} - \Psi_{\mathbf{x}}))$ is almost surely μ-continuous. Since K is compact, it follows that the process $\widehat{\Upsilon}_{\mathbf{g}}$ is separable (see p. 115 in van der Vaart and Wellner (1996)). Thus, in view of Corollary 2.2.8 in van der Vaart and Wellner (1996) and the assumption of the finiteness of the integral $\int_0^1 \sqrt{ln\ N(\varepsilon, K)}$ for each $\varepsilon > 0$, we will have the asymptotic equicontinuity in probability of $\widehat{\Upsilon}_{\mathbf{g}}$ if we can show the sub-Gaussianity of the process (see p. 101 in van der Vaart and Wellner (1996)) with respect to the metric induced by the norm in $\mathcal{X}$. Since $\mathbf{g} \in \mathcal{X}^{**}$, the empirical process $\widehat{\Upsilon}_{\mathbf{g}} = \{\sqrt{n}[n^{-1}\sum_{i=1}^{n} \mathbf{g}(S_{\mathbf{x}-\mathbf{x}_i}) - E\{\mathbf{g}(S_{\mathbf{x}-\mathbf{x}_i})\}] : \mathbf{x} \in K\}$. Using the Bernstein inequality for real-valued random variables and the assumptions in the theorem, we have

$$P(|\widehat{\Upsilon}_{\mathbf{g}}(\mathbf{x}) - \widehat{\Upsilon}_{\mathbf{g}}(\mathbf{y})| > t) \leq 2exp\left\{-t^2/a_1||\mathbf{x} - \mathbf{y}||^2\right\} \quad \forall\, n$$

for a suitable constant $a_1 > 0$. This proves the sub-Gaussianity of the process and completes the proof of the first statement in part (b) of the theorem.

For proving the second statement in part (b) of the theorem, we will need Fact 3.3.1

stated earlier. Using this, we have

$$\begin{aligned} P(|\widehat{\Upsilon}_{\mathbf{g}}(\mathbf{x}) - \widehat{\Upsilon}_{\mathbf{g}}(\mathbf{y})| > t) &\leq P(\sqrt{n}||(\widehat{\Psi}_{\mathbf{x}} - \widehat{\Psi}_{\mathbf{y}}) - (\Psi_{\mathbf{x}} - \Psi_{\mathbf{y}})|| > t) \\ &\leq 2exp\{-t^2/a_2||\mathbf{x} - \mathbf{y}||^2\} \quad \forall\, n \end{aligned}$$

for an appropriate constant $a_2 > 0$. This proves the sub-Gaussianity of the process, and hence its weak convergence to a tight stochastic process. □

Proof of Theorem 3.2.1. Since $\mathcal{X}$ is strictly convex, and μ is not completely supported on a straight line in $\mathcal{X}$, the map $\mathbf{x} \mapsto E\{||\mathbf{x}-\mathbf{X}|| - ||\mathbf{X}||\}$ is strictly convex. Thus, using exercise 4.2.12 in Borwein and Vanderwerff (2010), we have the strict monotonicity of the spatial distribution map. Let $\widetilde{g}(\mathbf{y}, \mathbf{v}) = E\{||\mathbf{y} - \mathbf{X}|| - ||\mathbf{X}||\} - \mathbf{v}(\mathbf{y})$, where $\mathbf{y} \in \mathcal{X}$ and $\mathbf{v} \in \mathcal{B}^*(\mathbf{0}, 1)$. Since $\mathcal{X}$ is reflexive, it follows from Remark 3.5 in Kemperman (1987) that there exists a minimizer of $\widetilde{g}$ in $\mathcal{X}$. Let us denote it by $\mathbf{x}(\mathbf{v})$. So, $\widetilde{g}(\mathbf{x}(\mathbf{v}), \mathbf{v}) \leq \widetilde{g}(\mathbf{y}, \mathbf{v})$ for all $\mathbf{y} \in \mathcal{X}$. Equivalently, $\mathbf{v}\{\mathbf{y} - \mathbf{x}(\mathbf{v})\} \leq E\{||\mathbf{y} - \mathbf{X}|| - ||\mathbf{x}(\mathbf{v}) - \mathbf{X}||\}$ for all $\mathbf{y} \in \mathcal{X}$. Since μ is nonatomic, it follows that the map $\mathbf{x} \mapsto E\{||\mathbf{x} - \mathbf{X}|| - ||\mathbf{X}||\}$ is Gâteaux differentiable everywhere. So, using the previous inequality and Corollary 4.2.5 in Borwein and Vanderwerff (2010), we have $\Psi_{\mathbf{x}(\mathbf{v})} = E\{S_{\mathbf{x}(\mathbf{v})-\mathbf{X}}\} = \mathbf{v}$. This proves that the range of the spatial distribution map is the whole of $\mathcal{B}^*(\mathbf{0}, 1)$. Since the norm in $\mathcal{X}$ is Fréchet differentiable on $\mathcal{X}\backslash\{\mathbf{0}\}$ and μ is nonatomic, the map $\mathbf{x} \mapsto E\{||\mathbf{x}-\mathbf{X}|| - ||\mathbf{X}||\}$ is Fréchet differentiable everywhere. The continuity property of the spatial distribution map is now a consequence of Corollary 4.2.12 in Borwein and Vanderwerff (2010). □

The next result can be obtained by suitably modifying the arguments in the second paragraph in the proof of Theorem 3.1.1 in Chaudhuri (1996).

Fact 3.3.2. *If $\mathcal{X}$ is a Banach space, there exists $C_1 > 0$ (depending on $\mathbf{u}$) such that $||\widehat{\mathbf{Q}}(\mathbf{u}) - \mathbf{Q}(\mathbf{u})|| \leq C_1$ for all sufficiently large n almost surely.*

Proof of Theorem 3.2.1.1. From the assumptions in the theorem and Theorem 2.17 and Remark 3.5 in Kemperman (1987), it follows that $\mathbf{Q}(\mathbf{u})$ exists and is unique. Let $\widehat{g}_n(\mathbf{Q}) = n^{-1}\sum_{i=1}^{n}\{||\mathbf{Q} - \mathbf{X}_i^{(n)}|| - ||\mathbf{X}_i^{(n)}||\} - \mathbf{u}^{(n)}(\mathbf{Q})$ for $\mathbf{Q} \in \mathcal{X}$. We will first prove the result when $\mathbf{X}$ is assumed to be bounded almost surely, i.e., for some $M > 0$,

$P(||\mathbf{X}|| \leq M) = 1$. Now, it follows from arguments similar to those in the proof of Lemma 2(i) in Cadre (2001) that $\sup_{||\mathbf{Q}|| \leq C} |\widehat{g}_n(\mathbf{Q}) - g_n(\mathbf{Q})| \to 0$ as $n \to \infty$ almost surely for any $C > 0$. We next show that $g(\widehat{\mathbf{Q}}(\mathbf{u})) \to g(\mathbf{Q}(\mathbf{u}))$ as $n \to \infty$ almost surely. Note that

$$0 \leq g(\widehat{\mathbf{Q}}(\mathbf{u})) - g(\mathbf{Q}(\mathbf{u})) = [g(\widehat{\mathbf{Q}}(\mathbf{u})) - g_n(\widehat{\mathbf{Q}}(\mathbf{u}))] - [g(\mathbf{Q}(\mathbf{u})) - g_n(\mathbf{Q}(\mathbf{u}))] + [g_n(\widehat{\mathbf{Q}}(\mathbf{u})) - g_n(\mathbf{Q}(\mathbf{u}))]. \tag{3.6}$$

Observe that for any $\mathbf{Q}$, $|g(\mathbf{Q}) - g_n(\mathbf{Q})| \leq 2E\{||\mathbf{X} - \mathbf{X}^{(n)}||\} + ||\mathbf{Q}||\ ||\mathbf{u} - \mathbf{u}^{(n)}||$, which implies that

$$\sup_{||\mathbf{Q}|| \leq C} |g(\mathbf{Q}) - g_n(\mathbf{Q})| \to 0, \tag{3.7}$$

as $n \to \infty$ almost surely for any $C > 0$. Further,

$$\begin{aligned} & g_n(\widehat{\mathbf{Q}}(\mathbf{u})) - g_n(\mathbf{Q}(\mathbf{u})) \\ = \ & [g_n(\widehat{\mathbf{Q}}(\mathbf{u})) - \widehat{g}_n(\widehat{\mathbf{Q}}(\mathbf{u}))] + [\widehat{g}_n(\widehat{\mathbf{Q}}(\mathbf{u})) - \widehat{g}_n(\mathbf{Q}^{(n)}(\mathbf{u}))] \\ & + [\widehat{g}_n(\mathbf{Q}^{(n)}(\mathbf{u})) - g_n(\mathbf{Q}^{(n)}(\mathbf{u}))] + [g_n(\mathbf{Q}^{(n)}(\mathbf{u})) - g_n(\mathbf{Q}(\mathbf{u}))]. \end{aligned} \tag{3.8}$$

In the notation of Section 3.2, $\mathbf{Q}^{(n)}(\mathbf{u}) = \sum_{k=1}^{d(n)} q_k\phi_k$, where $\mathbf{Q} = \sum_{k=1}^{\infty} q_k\phi_k$ for a Schauder basis $\{\phi_1, \phi_2, \ldots\}$ of $\mathcal{X}$. The first and the third terms in the right hand side of (3.8) are bounded above by $\sup_{||\mathbf{Q}|| \leq C_2} |\widehat{g}_n(\mathbf{Q}) - g_n(\mathbf{Q})|$ for all sufficiently large n almost surely. Here, $C_2 = C_1 + 2||\mathbf{Q}(\mathbf{u})||$, and C_1 is as in Fact 3.3.2. The second term in the right hand side of (3.8) is negative because $\widehat{\mathbf{Q}}(\mathbf{u})$ is a minimizer of $\widehat{g}_n$. The fourth term in the right hand side of (3.8) is bounded above by $2||\mathbf{Q}^{(n)}(\mathbf{u}) - \mathbf{Q}(\mathbf{u})||$. So,

$$g_n(\widehat{\mathbf{Q}}(\mathbf{u})) - g_n(\mathbf{Q}(\mathbf{u})) \leq 2 \sup_{||\mathbf{Q}|| \leq C_2} |\widehat{g}_n(\mathbf{Q}) - g_n(\mathbf{Q})| + 2||\mathbf{Q}^{(n)}(\mathbf{u}) - \mathbf{Q}(\mathbf{u})||$$

for all sufficiently large n almost surely. Combining (3.6), (3.7) and the previous inequality, we get $g(\widehat{\mathbf{Q}}(\mathbf{u})) \to g(\mathbf{Q}(\mathbf{u}))$ as $n \to \infty$ almost surely.

Let us now observe that for any random element $\mathbf{X}$ in the separable Banach space $\mathcal{X}$ and any fixed $\varepsilon > 0$, there exists $M > 0$ such that $P(||\mathbf{X}|| > M) < \varepsilon/C_1$. So, we have

$|g(\widehat{\mathbf{Q}}(\mathbf{u})) - g(\mathbf{Q}(\mathbf{u}))| \leq \varepsilon + |\overline{g}(\widehat{\mathbf{Q}}(\mathbf{u})) - \overline{g}(\mathbf{Q}(\mathbf{u}))|$ for all sufficiently large n almost surely. Here, $\overline{g}(\mathbf{Q}) = E\{(||\mathbf{Q} - \mathbf{X}|| - ||\mathbf{X}||)I(||\mathbf{X}|| \leq M)\} - \mathbf{u}(\mathbf{Q})$. Thus, letting $\varepsilon \to 0$, we have $g(\widehat{\mathbf{Q}}(\mathbf{u})) \to g(\mathbf{Q}(\mathbf{u}))$ as $n \to \infty$ almost surely for those random elements in $\mathcal{X}$ that are not necessarily almost surely bounded. Now, using Theorems 1 and 3 in Asplund (1968), it follows that $||\widehat{\mathbf{Q}}(\mathbf{u}) - \mathbf{Q}(\mathbf{u})|| \to 0$ as $n \to \infty$ almost surely. □

The Hessian of the function $g_n(\mathbf{Q})$ is

$$\mathrm{J}_{n,\mathbf{Q}}(\mathbf{h}) = E\left\{\frac{\mathbf{h}}{||\mathbf{Q} - \mathbf{X}^{(n)}||} - \frac{\langle \mathbf{h}, \mathbf{Q} - \mathbf{X}^{(n)}\rangle(\mathbf{Q} - \mathbf{X}^{(n)})}{||\mathbf{Q} - \mathbf{X}^{(n)}||^3}\right\}.$$

The next result is the $d(n)$-dimensional analog of Proposition 2.1 in Cardot et al. (2013), and can be obtained by suitably modifying the proof of that proposition.

Fact 3.3.3. *Suppose that the assumptions of Theorem 3.2.1.2 hold. Then, for each $C > 0$, there exists $b, B \in (0, \infty)$ with $b < B$ such that for all appropriately large n we have $b||\mathbf{h}||^2 \leq \widetilde{\mathrm{J}}_{n,\mathbf{Q}}(\mathbf{h}, \mathbf{h}) \leq B||\mathbf{h}||^2$ for any $\mathbf{Q}$, $\mathbf{h} \in \mathcal{Z}_n$ with $||\mathbf{Q}|| \leq C$.*

Lemma 3.3.4. *Suppose that the assumptions of Theorem 3.2.1.2 hold and $C > 0$ is arbitrary. Then, there exist $b', B' \in (0, \infty)$ such that for all appropriately large n and any $\mathbf{Q}$, $\mathbf{h}$, $\mathbf{z} \in \mathcal{Z}_n$ with $||\mathbf{Q} - \mathbf{Q}_n(\mathbf{u})|| \leq C$, we have*

$$\begin{aligned}
\left\|E\left\{\frac{\mathbf{Q} - \mathbf{X}^{(n)}}{||\mathbf{Q} - \mathbf{X}^{(n)}||} - \mathbf{u}^{(n)}\right\}\right\| &\geq b'||\mathbf{Q} - \mathbf{Q}_n(\mathbf{u})||, \\
\sup_{||\mathbf{h}||=||\mathbf{v}||=1} |\widetilde{\mathrm{J}}_{n,\mathbf{Q}}(\mathbf{h}, \mathbf{v}) - \widetilde{\mathrm{J}}_{n,\mathbf{Q}_n(\mathbf{u})}(\mathbf{h}, \mathbf{v})| &\leq B'||\mathbf{Q} - \mathbf{Q}_n(\mathbf{u})||, \qquad \textit{and} \\
\left\|E\left\{\frac{\mathbf{Q} - \mathbf{X}^{(n)}}{||\mathbf{Q} - \mathbf{X}^{(n)}||} - \mathbf{u}^{(n)}\right\} - \mathrm{J}_{n,\mathbf{Q}_n(\mathbf{u})}(\mathbf{Q} - \mathbf{Q}_n(\mathbf{u}))\right\| &\leq B'||\mathbf{Q} - \mathbf{Q}_n(\mathbf{u})||^2.
\end{aligned}$$

Proof. Consider any $\mathbf{h} \in \mathcal{Z}_n$ such that $||\mathbf{h}|| = 1$. A first order Taylor expansion of the function $E\left\{\frac{\mathbf{Q}-\mathbf{X}^{(n)}}{||\mathbf{Q}-\mathbf{X}^{(n)}||} - \mathbf{u}^{(n)}\right\}(\mathbf{h})$ about $\mathbf{Q}_n(\mathbf{u})$ yields

$$E\left\{\frac{\mathbf{Q} - \mathbf{X}^{(n)}}{||\mathbf{Q} - \mathbf{X}^{(n)}||} - \mathbf{u}^{(n)}\right\}(\mathbf{h}) = \widetilde{\mathrm{J}}_{n,\widetilde{\mathbf{Q}}}(\mathbf{Q} - \mathbf{Q}_n(\mathbf{u}), \mathbf{h}), \tag{3.9}$$

where $||\widetilde{\mathbf{Q}} - \mathbf{Q}_n(\mathbf{u})|| < ||\mathbf{Q} - \mathbf{Q}_n(\mathbf{u})||$. Choosing $\mathbf{h} = (\mathbf{Q} - \mathbf{Q}_n(\mathbf{u}))/||\mathbf{Q} - \mathbf{Q}_n(\mathbf{u})||$ and using Fact 3.3.3, we have the first inequality.

The second inequality follows from the definition of $\widetilde{\mathrm{J}}_{n,Q}$, the upper bound in Fact 3.3.3 and some straight-forward algebra.

From (3.9), we get

$$\begin{aligned} & \left| E\left\{ \frac{\mathbf{Q}-\mathbf{X}^{(n)}}{||\mathbf{Q}-\mathbf{X}^{(n)}||} - \mathbf{u}^{(n)} \right\}(\mathbf{h}) - \widetilde{\mathrm{J}}_{n,\mathbf{Q}_n(\mathbf{u})}(\mathbf{Q}-\mathbf{Q}_n(\mathbf{u}),\mathbf{h}) \right| \\ = & \ |\widetilde{\mathrm{J}}_{n,\widetilde{\mathbf{Q}}}(\mathbf{Q}-\mathbf{Q}_n(\mathbf{u}),\mathbf{h}) - \widetilde{\mathrm{J}}_{n,\mathbf{Q}_n(\mathbf{u})}(\mathbf{Q}-\mathbf{Q}_n(\mathbf{u}),\mathbf{h})| \\ \leq & \ B'||\mathbf{Q}-\mathbf{Q}_n(\mathbf{u})||^2, \ \text{ since } ||\widetilde{\mathbf{Q}}-\mathbf{Q}_n(\mathbf{u})|| < ||\mathbf{Q}-\mathbf{Q}_n(\mathbf{u})||. \end{aligned}$$

Taking supremum over $||\mathbf{h}|| = 1$ and using the definition of $\mathrm{J}_{n,\mathbf{Q}}$, we have the proof of the third inequality. □

Proposition 3.3.5. *Suppose that the assumptions of Theorem 3.2.1.2 hold. Then, $||\widehat{\mathbf{Q}}(\mathbf{u}) - \mathbf{Q}_n(\mathbf{u})|| = O(\delta_n)$ as $n \to \infty$ almost surely, where $\delta_n \sim \sqrt{ln\ n}/n^{\alpha}$ and α is as in Theorem 3.2.1.2.*

Proof. From Fact 3.3.2 and the behavior of $\mathbf{Q}_n(\mathbf{u})$ discussed before Assumption (B) in Section 3.2.1, we get the existence of $C_3 > 0$ satisfying $||\widehat{\mathbf{Q}}(\mathbf{u}) - \mathbf{Q}_n(\mathbf{u})|| \leq C_3$ for all sufficiently large n almost surely. Define $\mathrm{G}_n = \{\mathbf{Q}_n(\mathbf{u}) + \sum_{j\leq d(n)} \beta_j\varphi_j : n^4\beta_j$ is an integer in $[-C_3, C_3]$ and $||\sum_{j\leq d(n)} \beta_j\varphi_j|| \leq C_3\}$, and $\mathcal{Z}_n = span\{\varphi_1, \varphi_2, \ldots, \varphi_{d(n)}\}$, where $\{\varphi_j\}_{j\geq 1}$ is an orthonormal basis of $\mathcal{X}$. Let us define the event

$$\mathrm{E}_n = \left\{ \max_{\mathbf{Q}\in\mathrm{G}_n} \left\| \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\mathbf{Q}-\mathbf{X}_i^{(n)}}{||\mathbf{Q}-\mathbf{X}_i^{(n)}||} - \mathbf{u}^{(n)} \right) - E\left(\frac{\mathbf{Q}-\mathbf{X}^{(n)}}{||\mathbf{Q}-\mathbf{X}^{(n)}||} - \mathbf{u}^{(n)} \right) \right\| \leq C_4\delta_n \right\}.$$

Note that $\left\|\frac{\mathbf{Q}-\mathbf{X}^{(n)}}{||\mathbf{Q}-\mathbf{X}^{(n)}||} - \mathbf{u}^{(n)}\right\| \leq 2$ for all $\mathbf{Q} \in \mathcal{Z}_n$ and $n \geq 1$. So, using Fact 3.3.1, there exists $C_5 > 0$ such that $P(\mathrm{E}_n^c) \leq 2(3C_3n^4)^{d(n)}exp\{-nC_5^2\delta_n^2\}$ for all appropriately large n. Using the definition of δ_n given in the statement of the proposition, C_5 in the previous inequality can be chosen in such a way that $\sum_{n=1}^{\infty} P(\mathrm{E}_n^c) < \infty$. Thus,

$$P(\mathrm{E}_n \text{ occurs for all sufficiently large } n) = 1. \tag{3.10}$$

We next define the event $\mathrm{F}_n = \left\{\max_{\mathbf{Q}\in\mathrm{G}_n} \sum_{i=1}^{n} I_{\{||\mathbf{Q}-\mathbf{X}_i^{(n)}||\leq n^{-2}\}} \leq C_6 n\delta_n^2\right\}$. Note that $M_n' = \max_{\mathbf{Q}\in\mathrm{G}_n} E\{||\mathbf{Q}-\mathbf{X}^{(n)}||^{-1}\} < \infty$ for all appropriately large n in view of Assumption (B) in Section 3.2.1. Further, $M'_n \geq M'_{n+k}$ for all $k \geq 1$ and $n \geq 1$. Then, $P(||\mathbf{Q} - \mathbf{X}^{(n)}|| \leq n^{-2}) \leq M'_n n^{-2} \leq C_6\delta_n^2/2$ for any $\mathbf{Q} \in \mathrm{G}_n$ and all appropriately large n (the first inequality follows from the Markov inequality). Therefore, $Var\{I(||\mathbf{Q} - \mathbf{X}^{(n)}|| \leq n^{-2})\} \leq C_6\delta_n^2/2$ for any $\mathbf{Q} \in \mathrm{G}_n$ and all appropriately large n. The Bernstein inequality for real-valued random variables implies that there exists $C_7 > 0$ such that $P(\mathrm{F}_n^c) \leq (3C_3n^4)^{d(n)}exp\{-nC_7\delta_n^2\}$ for all appropriately large n. As before, C_7 in the previous inequality can be chosen in such a way that $\sum_{n=1}^{\infty} P(\mathrm{F}_n^c) < \infty$, which implies that

$$P(\mathrm{F}_n \text{ occurs for all sufficiently large } n) = 1. \tag{3.11}$$

Now consider a point in G_n nearest to $\widehat{\mathbf{Q}}(\mathbf{u})$, say, $\overline{\mathbf{Q}}_n(\mathbf{u})$. Then, $||\widehat{\mathbf{Q}}(\mathbf{u}) - \overline{\mathbf{Q}}_n(\mathbf{u})|| \leq C_8 d(n)/n^4$ for a constant $C_8 > 0$. Note that

$$\left\| \frac{\widehat{\mathbf{Q}}(\mathbf{u}) - \mathbf{X}_i^{(n)}}{||\widehat{\mathbf{Q}}(\mathbf{u}) - \mathbf{X}_i^{(n)}||} - \frac{\overline{\mathbf{Q}}_n(\mathbf{u}) - \mathbf{X}_i^{(n)}}{||\overline{\mathbf{Q}}_n(\mathbf{u}) - \mathbf{X}_i^{(n)}||} \right\| \leq \frac{2||\widehat{\mathbf{Q}}(\mathbf{u}) - \overline{\mathbf{Q}}_n(\mathbf{u}))||}{||\overline{\mathbf{Q}}_n(\mathbf{u}) - \mathbf{X}_i^{(n)}||}. \tag{3.12}$$

Then, for a constant $C_9 > 0$, we have

$$\begin{aligned}
\left\| \frac{1}{n}\sum_{i=1}^{n} \frac{\overline{\mathbf{Q}}_n(\mathbf{u}) - \mathbf{X}_i^{(n)}}{||\overline{\mathbf{Q}}_n(\mathbf{u}) - \mathbf{X}_i^{(n)}||} - \mathbf{u}^{(n)} \right\| &\leq \left\| \frac{1}{n}\sum_{i=1}^{n} \frac{\widehat{\mathbf{Q}}(\mathbf{u}) - \mathbf{X}_i^{(n)}}{||\widehat{\mathbf{Q}}(\mathbf{u}) - \mathbf{X}_i^{(n)}||} - \mathbf{u}^{(n)} \right\| \qquad (3.13)\\
&+ \left\| \frac{1}{n}\sum_{i=1}^{n} \left\{ \frac{\overline{\mathbf{Q}}_n(\mathbf{u}) - \mathbf{X}_i^{(n)}}{||\overline{\mathbf{Q}}_n(\mathbf{u}) - \mathbf{X}_i^{(n)}||} - \frac{\widehat{\mathbf{Q}}(\mathbf{u}) - \mathbf{X}_i^{(n)}}{||\widehat{\mathbf{Q}}(\mathbf{u}) - \mathbf{X}_i^{(n)}||} \right\} \right\| \\
\leq &\left\| \frac{1}{n}\sum_{i=1}^{n} \frac{\widehat{\mathbf{Q}}(\mathbf{u}) - \mathbf{X}_i^{(n)}}{||\widehat{\mathbf{Q}}(\mathbf{u}) - \mathbf{X}_i^{(n)}||} - \mathbf{u}^{(n)} \right\| + 2C_8 d(n) n^{-2} \\
&+ \frac{2}{n}\sum_{i=1}^{n} I\{||\overline{\mathbf{Q}}_n(\mathbf{u}) - \mathbf{X}_i^{(n)}|| \leq n^{-2}\} \quad \text{(using (3.12))} \\
\leq &\left\| \frac{1}{n}\sum_{i=1}^{n} \frac{\widehat{\mathbf{Q}}(\mathbf{u}) - \mathbf{X}_i^{(n)}}{||\widehat{\mathbf{Q}}(\mathbf{u}) - \mathbf{X}_i^{(n)}||} - \mathbf{u}^{(n)} \right\| + C_9\delta_n^2 \quad \text{(using (3.11))}.
\end{aligned}$$

It follows from arguments similar to those used in the proof of Theorem 4.11 in Kemperman (1987) that $\left\| \sum_{i=1}^{n} \frac{\widehat{\mathbf{Q}}(\mathbf{u}) - \mathbf{X}_i^{(n)}}{||\widehat{\mathbf{Q}}(\mathbf{u}) - \mathbf{X}_i^{(n)}||} - n\mathbf{u}^{(n)} \right\| \leq 1$. Combining this with (3.13), we get

$$\left\| \sum_{i=1}^{n} \frac{\overline{\mathbf{Q}}_n(\mathbf{u}) - \mathbf{X}_i^{(n)}}{||\overline{\mathbf{Q}}_n(\mathbf{u}) - \mathbf{X}_i^{(n)}||} - n\mathbf{u}^{(n)} \right\| \le 3C_7 n\delta_n \tag{3.14}$$

for all sufficiently large n almost surely. Suppose that $\mathbf{Q} \in \mathrm{G}_n$ and $||\mathbf{Q}-\mathbf{Q}_n(\mathbf{u})|| > C_{10}\delta_n$ for some $C_{10} > 0$. Then, it follows from (3.10) and the first inequality in Lemma 3.3.4 that $\left\| \sum_{i=1}^{n} \frac{\mathbf{Q}-\mathbf{X}_i^{(n)}}{||\mathbf{Q}-\mathbf{X}_i^{(n)}||} - n\mathbf{u}^{(n)} \right\| \ge (C_{10}b' - C_4)n\delta_n$ for all sufficiently large n almost surely. If we choose C_{10} such that $C_{10}b' - C_4 > 4C_7$, then in view of (3.14), we must have $||\overline{\mathbf{Q}}_n(\mathbf{u}) - \mathbf{Q}_n(\mathbf{u})|| \le C_{10}\delta_n$ for all sufficiently large n almost surely. This implies that for a constant $C_{11} > 0$, $||\widehat{\mathbf{Q}}(\mathbf{u}) - \mathbf{Q}_n(\mathbf{u})|| \le C_{11}\delta_n$ for all sufficiently large n almost surely. This completes the proof. □

Proof of Theorem 3.2.1.2. Let H_n denote the collection of points from G_n, which satisfy $||\mathbf{Q} - \mathbf{Q}_n(\mathbf{u})|| \le C_{11}\delta_n$. Let us define for $\mathbf{Q} \in \mathcal{Z}_n$,

$$\Gamma_n(\mathbf{Q}, \mathbf{X}_i) = \frac{\mathbf{Q}_n(\mathbf{u}) - \mathbf{X}_i^{(n)}}{||\mathbf{Q}_n(\mathbf{u}) - \mathbf{X}_i^{(n)}||} - \frac{\mathbf{Q} - \mathbf{X}_i^{(n)}}{||\mathbf{Q} - \mathbf{X}_i^{(n)}||} + E\left\{ \frac{\mathbf{Q} - \mathbf{X}^{(n)}}{||\mathbf{Q} - \mathbf{X}^{(n)}||} - \mathbf{u}^{(n)} \right\},$$

$$\text{and } \Delta_n(\mathbf{Q}) = E\left\{ \frac{\mathbf{Q} - \mathbf{X}^{(n)}}{||\mathbf{Q} - \mathbf{X}^{(n)}||} - \frac{\mathbf{Q}_n(\mathbf{u}) - \mathbf{X}^{(n)}}{||\mathbf{Q}_n(\mathbf{u}) - \mathbf{X}^{(n)}||} \right\} - \mathbf{J}_{n,\mathbf{Q}_n(\mathbf{u})}(\mathbf{Q} - \mathbf{Q}_n(\mathbf{u})).$$

Using Assumption (B) in Section 3.2.1, it follows that for a constant $C_{12} > 0$,

$$\begin{aligned} E\,||\Gamma_n(\mathbf{Q}, \mathbf{X})||^2 &\le 2E\left\| \frac{\mathbf{Q}_n(\mathbf{u}) - \mathbf{X}_i^{(n)}}{||\mathbf{Q}_n(\mathbf{u}) - \mathbf{X}_i^{(n)}||} - \frac{\mathbf{Q} - \mathbf{X}_i^{(n)}}{||\mathbf{Q} - \mathbf{X}_i^{(n)}||} \right\|^2 \\ &\quad + 2\left\| E\left\{ \frac{\mathbf{Q}_n(\mathbf{u}) - \mathbf{X}^{(n)}}{||\mathbf{Q}_n(\mathbf{u}) - \mathbf{X}^{(n)}||} \right\} - E\left\{ \frac{\mathbf{Q} - \mathbf{X}^{(n)}}{||\mathbf{Q} - \mathbf{X}^{(n)}||} \right\} \right\|^2 \\ &\le C_{12}||\mathbf{Q} - \mathbf{Q}_n(\mathbf{u})||^2. \end{aligned}$$

So, in view of Fact 3.3.1, there exists a constant $C_{13} > 0$ such that

$$\max_{\mathbf{Q} \in \mathrm{H}_n} \left\| \frac{1}{n} \sum_{i=1}^{n} \Gamma_n(\mathbf{Q}, \mathbf{X}_i) \right\| \le C_{13}\delta_n^2, \tag{3.15}$$

for all sufficiently large n almost surely. Using the third inequality in Lemma 3.3.4, there exists a constant $C_{14} > 0$ such that $||\Delta_n(\mathbf{Q})|| \le C_{14}||\mathbf{Q} - \mathbf{Q}_n(\mathbf{u})||^2$ for all appropriately large n. This along with (3.15) and the definitions of Γ_n and $\Delta_n(\mathbf{Q})$ yield

$$\mathsf{J}_{n,\mathbf{Q}_n(\mathbf{u})}(\mathbf{Q} - \mathbf{Q}_n(\mathbf{u})) = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\mathbf{Q}_n(\mathbf{u}) - \mathbf{X}_i^{(n)}}{||\mathbf{Q}_n(\mathbf{u}) - \mathbf{X}_i^{(n)}||} - \frac{\mathbf{Q} - \mathbf{X}_i^{(n)}}{||\mathbf{Q} - \mathbf{X}_i^{(n)}||}\right\} + \widetilde{\mathbf{R}}_n(\mathbf{Q}),$$

where $\max_{\mathbf{Q}\in \mathrm{H}_n} ||\widetilde{\mathbf{R}}_n(\mathbf{Q})|| = O(\delta_n^2)$ as $n \to \infty$ almost surely. From Fact 3.3.3, it follows that the operator norm of $\mathsf{J}_{n,\mathbf{Q}_n(\mathbf{u})}$ is uniformly bounded away from zero, and $[\mathsf{J}_{n,\mathbf{Q}_n(\mathbf{u})}]^{-1}$ is defined on the whole of $\mathcal{Z}_n$ for all appropriately large n. It follows that for a constant $C_{15} > 0$, $\max_{\mathbf{Q}\in \mathrm{H}_n} ||[\mathsf{J}_{n,\mathbf{Q}_n(\mathbf{u})}]^{-1}(\widetilde{\mathbf{R}}_n(\mathbf{Q}))|| \leq C_{15}\delta_n^2$ for all sufficiently large n almost surely.

Hence, choosing $\mathbf{Q} = \overline{\mathbf{Q}}_n(\mathbf{u})$, and utilizing inequality (3.13) in the proof of Proposition 3.3.5, we get

$$\widehat{\mathbf{Q}}(\mathbf{u}) - \mathbf{Q}_n(\mathbf{u}) = \frac{1}{n}\sum_{i=1}^{n}[\mathsf{J}_{n,\mathbf{Q}_n(\mathbf{u})}]^{-1}\left\{\frac{\mathbf{Q}_n(\mathbf{u}) - \mathbf{X}_i^{(n)}}{||\mathbf{Q}_n(\mathbf{u}) - \mathbf{X}_i^{(n)}||} - \mathbf{u}^{(n)}\right\} + \mathbf{R}_n,$$

where $||\mathbf{R}_n|| = O(\delta_n^2)$ as $n \to \infty$ almost surely. □

Proof of Theorem 3.2.1.3. Since $\mathbf{U}_n = n^{-1}\sum_{i=1}^{n}\left(\frac{\mathbf{Q}_n(\mathbf{u})-\mathbf{X}_i^{(n)}}{||\mathbf{Q}_n(\mathbf{u})-\mathbf{X}_i^{(n)}||} - \mathbf{u}^{(n)}\right)$ is a sum of uniformly bounded, independent, zero mean random elements in the separable Hilbert space $\mathcal{X}$, we get that $||\sqrt{n}\mathbf{U}_n||$ is bounded in probability as $n \to \infty$ in view of Fact 3.3.1. We will show that $\sqrt{n}\{[\mathsf{J}_{n,\mathbf{Q}_n(\mathbf{u})}]^{-1}(\mathbf{U}_n) - [\mathsf{J}_{\mathbf{Q}(\mathbf{u})}]^{-1}(\mathbf{U}_n)\} \to \mathbf{0}$ in probability as $n \to \infty$. Note that for each $C > 0$, every $\mathbf{Q} \in \mathcal{X}$ satisfying $||\mathbf{Q}|| \leq C$ and all appropriately large n, $\widetilde{\mathsf{J}}_{n,\mathbf{Q}}$ and $\mathsf{J}_{n,\mathbf{Q}}$ can be defined from $\mathcal{X} \times \mathcal{X} \to \mathbb{R}$ and $\mathcal{X} \to \mathcal{X}$, respectively, by virtue of Assumption (B) in Section 3.2.1. Further, the bound obtained in the second inequality in Lemma 3.3.4 actually holds (up to a constant multiple) for all appropriately large n, any $C > 0$ and any $\mathbf{Q}$, $\mathbf{h}$, $\mathbf{v} \in \mathcal{X}$, which satisfy $||\mathbf{Q}|| \leq C$. Thus, $||\mathsf{J}_{n,\mathbf{Q}_n(\mathbf{u})} - \mathsf{J}_{n,\mathbf{Q}(\mathbf{u})}|| \leq B''||\mathbf{Q}_n(\mathbf{u}) - \mathbf{Q}(\mathbf{u})||$ for a constant $B'' > 0$ and all appropriately large n. Since $||\mathbf{X}^{(n)} - \mathbf{X}|| \to 0$ as $n \to \infty$ almost surely, it follows from Assumption (B) in Section 3.2.1 that $||\mathsf{J}_{n,\mathbf{Q}(\mathbf{u})} - \mathsf{J}_{\mathbf{Q}(\mathbf{u})}|| \to 0$ as $n \to \infty$. Since $\mathbf{Q}_n(\mathbf{u}) \to \mathbf{Q}(\mathbf{u})$, we now have $||\mathsf{J}_{n,\mathbf{Q}_n(\mathbf{u})} - \mathsf{J}_{\mathbf{Q}(\mathbf{u})}|| \to 0$ as $n \to \infty$. It follows from Proposition 2.1 in Cardot et al. (2013) that the linear operator $\mathsf{J}_{\mathbf{Q}(\mathbf{u})}$ has a bounded inverse, which is defined on the whole of $\mathcal{X}$. Using the fact that $||\sqrt{n}\mathbf{U}_n||$ is bounded in probability as $n \to \infty$ we get that

$$\begin{aligned} & \sqrt{n}\ \|\{J_{n,\mathbf{Q}_n(\mathbf{u})}\}^{-1}(\mathbf{U}_n) - \{J_{\mathbf{Q}(\mathbf{u})}\}^{-1}(\mathbf{U}_n)\| \\ \leq\ & \sqrt{n}\ \|\{J_{n,\mathbf{Q}_n(\mathbf{u})}\}^{-1} - \{J_{\mathbf{Q}(\mathbf{u})}\}^{-1}\|\ \|(\mathbf{U}_n)\| \\ \leq\ & \|\{J_{\mathbf{Q}(\mathbf{u})}\}^{-1}\|\ \|J_{n,\mathbf{Q}_n(\mathbf{u})} - J_{\mathbf{Q}(\mathbf{u})}\|\ \|\{J_{n,\mathbf{Q}_n(\mathbf{u})}\}^{-1}\|\ \|\sqrt{n}\mathbf{U}_n\| \\ \stackrel{P}{\rightarrow}\ & 0 \text{ as } n \to \infty. \end{aligned}$$

The convergence in probability asserted above holds because the operator norm of $J_{n,\mathbf{Q}_n(\mathbf{u})}$ is uniformly bounded away from zero by Fact 3.3.3. The asymptotic Gaussianity of $\{J_{\mathbf{Q}(\mathbf{u})}\}^{-1}(\sqrt{n}\mathbf{U}_n)$ follows from the central limit theorem for a triangular array of rowwise independent Hilbert space valued random elements (see, e.g., Corollary 7.8 in Araujo and Giné (1980)). □

Chapter 4

Data depth in infinite dimensional spaces

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. 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 introduced in Liu (1990) and discussed subsequently in Zuo and Serfling (2000). These properties have been utilized in developing various statistical procedures. Depth-weighted L-type location estimators like trimmed means have been considered in Liu (1990), Donoho and Gasko (1992), Liu et al. (1999), 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 et al. (2012)). Another useful application of depths is in constructing depth contours (see, e.g., Donoho and Gasko (1992), He and Wang (1997) and Mosler (2002)), which determine central and outlying regions of a probability distribution. These contours and regions are useful in outlier detection.

We mentioned in `Chapter 2` that the concept of median has been extended in several ways for probability distributions in finite dimensional Euclidean spaces. In particular, the median has been defined as the point in the sample space with the highest depth value with respect to an appropriate depth function, i.e., the deepest point with respect to that depth function. (see, e.g., Donoho and Gasko (1992); Liu et al. (1999); Small (1990); Zuo and Serfling (2000)).

Some of the depth functions like the half-space depth (HD) (see, e.g., Donoho and Gasko (1992)), the projection depth (PD) (see, e.g., Zuo and Serfling (2000)) and the spatial depth (SD) (see, e.g., Vardi and Zhang (2000) and Serfling (2002)), which were originally defined for finite dimensional data, can have natural extensions into infinite dimensional spaces as we shall see in subsequent sections. As discussed in the Introduction, there have been some work on developing notions of depth for functional data (see, e.g., Fraiman and Muniz (2001), López-Pintado and Romo (2009), López-Pintado and Romo (2011)), and these have been used for developing various statistical methods for such data by López-Pintado and Romo (2006), López-Pintado and Romo (2009) and Sun and Genton (2011). In this chapter, we shall investigate the properties of some of the depth functions for probability distributions in infinite dimensional spaces and the associated deepest points.

4.1 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 by 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 $\theta(\cdot)$ and $\sigma(\cdot)$ 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 $\langle \cdot, \cdot \rangle$ 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 4.1.1. *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 $\theta(\cdot)$ and $\sigma(\cdot)$ 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 by 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 4.5). This, for instance, holds whenever $\mathbf{Y}$ is a sequence of independent nondegenerate 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 4.1.2. *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* $\theta(\cdot)$ *as the mean and* $\sigma(\cdot)$ *as the standard deviation in the definition of* $PD(\mathbf{x})$.

The degeneracy of HD stated in Theorems 4.1.1 and 4.1.2 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)).

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 4.1.1 or 4.1.2, it is easy to verify using the arguments in the proofs of those theorems (see Section 4.5) 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.

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 center 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 $\mathcal{X}$.

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 nondegenerate interval, then for Gaussian and many other distributions of $\mathbf{X}$ in $\mathcal{X}$, $IDD(\mathbf{X})$ will have a nondegenerate 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.

4.2 Depths based on coordinate random variables

In this section, we shall discuss depths that use the underlying coordinate system of the sample space. 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). 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

$$BD(\mathbf{x}) = \sum_{j=2}^{J} P\left(\min_{i=1,\ldots,j} X_{i,t} \leq x_t \leq \max_{i=1,\ldots,j} X_{i,t},\ \forall\, t \in [0,1]\right) \text{ and } \quad (4.1)$$

$$HRD(\mathbf{x}) = \min\{P(X_t \leq x_t,\ \forall\, t \in [0,1]),\ P(X_t \geq x_t,\ \forall\, t \in [0,1])\}, \quad (4.2)$$

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) 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. Recently, Sun and Genton (2011) used BD and a modified version of BD to construct boxplots for functional data.

López-Pintado and Romo (2009, 2011) defined finite dimensional versions of BD and HRD 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 $\mathbf{x} = (x_1, x_2, \ldots, x_d)$,

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

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 4.2.1. *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 nonatomic 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 4.2.2. *Let $\mathbf{X} = \{X_t\}_{t\in[0,1]}$ 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 nonatomic 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.*

A Feller process is a strong Markov process whose transition probability function satisfies certain continuity properties. Important examples of Feller processes include Brownian motions, Brownian bridges etc. Feller processes have been used for modelling data in physical and biological sciences (see, e.g., Böttcher (2010) for a review and related references). We refer to Revuz and Yor (1991) for an exposition on Feller processes. 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 4.2.2, using the arguments in the proofs of Lemma 4.5.1 and 4.5.2 (see Section 4.5), 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 $\mathbf{Y}$.

It was observed by López-Pintado and Romo (2009, 2011) that both of BD and HRD tend to take small values if the sample consists of irregular (nonsmooth) curves that cross one another often. To overcome this problem, these authors 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. 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

$\mathbf{X} = \{X_t\}_{t\in[0,1]} \in C[0,1]$,

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

where $\lambda(\cdot)$ is the Lebesgue measure on $[0,1]$. Fraiman and Muniz (2001) defined the integrated data depth (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. Fraiman and Muniz (2001) used this depth function to construct trimmed means. These authors showed that the empirical ID is a strongly and uniformly consistent estimator of its population counterpart. They used ID to categorize extremal and central curves in the data consisting of 100 curves used to build the NASDAQ 100 index. 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 $\mathbf{x} = \{x_t\}_{t\in[0,1]} \in C[0,1]$,

$$\begin{aligned} MBD(\mathbf{x}) &= \sum_{j=2}^{J} E\left[\int_0^1 I\left(\min_{i=1,\dots,j} X_{it} \le x_t \le \max_{i=1,\dots,j} X_{it}\right) dt\right] \\ &= \sum_{j=2}^{J} \int_0^1 \left[1 - F_t^j(x_t-) - (1-F_t(x_t))^j\right] dt \ \text{ and} \qquad (4.3) \\ MHRD(\mathbf{x}) &= \min\left\{E\left[\int_0^1 I(X_t \le x_t)dt\right],\ E\left[\int_0^1 I(X_t \ge x_t)dt\right]\right\} \\ &= \min\left\{\int_0^1 P(X_t \le x_t)dt,\ \int_0^1 P(X_t \ge x_t)dt\right\}. \qquad (4.4) \end{aligned}$$

We now discuss some useful properties of MBD, MHRD and ID. It is easy to see from (4.3) 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 (4.3) and (4.4) 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 nondegenerate 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 in the previous paragraph 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 center of symmetry of any probability distribution, both of these depth functions will be able to discriminate between two distributions with distinct centers of symmetry.

Note that in view of (4.3) and (4.4), both of MBD and MHRD are invariant under coordinatewise strictly monotone transformations. This property also holds for ID if all the univariate depths D_t's are invariant under such transformations as well. For the next theorem, in the definition of ID, we shall assume $D_t(\cdot) = \psi(F_t(\cdot))$ 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 4.2.3. *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 conclusions in (a) and (b) above also 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 nonconstant function, the support of the distribution of $ID(\widetilde{\mathbf{X}})$ is actually a closed nondegenerate 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 4.5.4 in Section 4.5).

4.2.1 The deepest point

In this subsection, we study the behaviour of the deepest point based on the depths discussed in the previous subsection. We mentioned earlier that in spite of the degeneracy of HD and PD, the corresponding medians are well-defined. However, for many commonly used stochastic models for functional data including those in Theorem 4.2.2, there is no meaningful notion of the deepest point associated with BD and HRD.

Theorem 4.2.4. *Let* $\mathbf{X}$ *be a Feller process in* $C[0,1]$ *starting at* $a_0 \in \mathbb{R}$, *which is symmetric about* $\mathbf{a} = \{a_t\}_{t\in[0,1]} \in C[0,1]$, *i.e., the distribution of* $\mathbf{X}-\mathbf{a}$ *is same as that of* $\mathbf{a}-\mathbf{X}$. *Also, assume that* X_t *has a continuous distribution for all* $t \in (0,1]$ *and any finite dimensional marginal* $(X_{t_1}, X_{t_2}, \ldots, X_{t_d})$ *of* $\mathbf{X}$ *has a positive density in a neighbourhood of its center of symmetry* $(a_{t_1}, a_{t_2}, \ldots, a_{t_d})$, *where* $0 < t_1 < t_2 < \ldots < t_d \leq 1$. *Then,* $BD(\mathbf{x}) = 0$ *for all* $\mathbf{x} \in C[0,1]$. *Moreover,* $HRD(\mathbf{a}) = 0$.

Since any reasonable depth function should assign maximum value to the point of symmetry of a distribution, and it should preferably be the unique maximizer, it follows from the above theorem that neither of these two depth functions yields any useful notion of deepest point for such stochastic processes. It will be appropriate to note here that Kuelbs and Zinn (2012) recently proved that the half-region depth vanishes at every point in the space of sample functions for a large class of stochastic processes, which include processes in $C[0,1]$.

Henceforth, we shall assume that for each $t \in [0,1]$, the univariate depths D_t's in the definition of ID is maximized at the median of X_t if X_t has a continuous distribution. The next result gives a description of points, which maximize the depth functions MBD, MHRD and ID. For this, let us write $\mathbf{m} = \{m_t\}_{t\in[0,1]}$, where m_t denotes the median of X_t for $t \in [0,1]$.

Theorem 4.2.5. *Let* $\mathbf{X}$ *be a random element in* $C[0,1]$ *such that* X_t *has a continuous distribution, which is strictly increasing in a neighbourhood of* m_t *for each* $t \in [0,1]$. *Then,* $\mathbf{m}$ *is a maximizer of MBD, MHRD and ID. Any* $\mathbf{m}^*$, *which equals* $\mathbf{m}$ *for all* $t \in [0,1]$ *except on a set of Lebesgue measure zero, is also a maximizer of MBD and ID. Moreover, any* $\mathbf{m}^{**} = \{m_t^{**}\}_{t\in[0,1]}$ *satisfying* $\int_0^1 F_t(m_t^{**})dt = 1/2$ *is a maximizer of MHRD.*

It is clear that $\mathbf{m}^{**}$, which satisfies $\int_0^1 F_t(m_t^{**})dt = 1/2$, may differ from $\mathbf{m}$ on a set of positive Lebesgue measure unlike $\mathbf{m}^*$. Moreover, although $\mathbf{m}^{**}$ is a maximizer of MHRD, its components m_t^{**} may be far from being univariate medians. For instance, let us consider a standard Brownian motion $\mathbf{X} = \{X_t\}_{t\in[0,1]}$. Define the function $\mathbf{f} = \{f_t\}_{t\in[0,1]}$, where f_t is the α^{th} percentile of X_t for $t \in [0,1/2)$, and f_t is the $(1-\alpha)^{th}$ percentile of X_t for $t \in [1/2,1]$. Here $\alpha \in (0,1)$ can be as small or as large as we like.

Then, any such $\mathbf{f}$ is a maximizer of MHRD.

Let $\widehat{F}_t$ denote the empirical distribution of X_t for each $t \in [0,1]$. Then, the empirical MBD of $\mathbf{x} = \{x_t\}_{t\in[0,1]}$ is given by

$$\sum_{j=2}^{J} \int_0^1 \left[1 - \widehat{F}_t^j(x_t-) - (1 - \widehat{F}_t(x_t))^j\right] dt.$$

The empirical MHRD of $\mathbf{x}$ is given by $\min\left\{\int_0^1 \widehat{F}_t(x_t)dt,\ 1 - \int_0^1 \widehat{F}_t(x_t-)dt\right\}$. Also, the empirical ID of $\mathbf{x}$ is given by $\int_0^1 \widehat{D}_t(x_t)dt$, where $\widehat{D}_t$'s denote the empirical versions of the coordinatewise depths D_t's. Let us denote the empirical coordinatewise median by $\widehat{\mathbf{m}} = \{\widehat{m}_t\}_{t\in[0,1]}$, where $\widehat{m}_t$ is the median of the empirical distribution of X_t for all $t \in [0,1]$. Here, we use the conventional definition of median, i.e., if n is even, $\widehat{m}_t = (X_{(n/2),t} + X_{(n/2+1),t})/2$, and if n is odd, $\widehat{m}_t = X_{((n+1)/2),t}$. We can assume that $\widehat{D}_t(x_t)$ is maximized at $\widehat{m}_t$ for all $t \in [0,1]$, which is true for almost any depth function for univariate data. It can be verified that the empirical coordinatewise median $\widehat{\mathbf{m}} = \{\widehat{m}_t\}_{t\in[0,1]}$ is a maximizer of the empirical versions of all of the three depth functions mentioned above if the distribution of X_t is continuous for all $t \in [0,1]$. Further, any $\widehat{\mathbf{m}}^*$, which differ from $\widehat{\mathbf{m}}$ only on a Lebesgue null set is also a maximizer of the empirical MBD and the empirical ID. Also, there is no unique maximizer of the empirical MHRD, and any $\widehat{\mathbf{m}}^{**} = \{\widehat{m}_t^{**}\}_{t\in[0,1]}$ satisfying $\int_0^1 \widehat{F}_t(\widehat{m}_t^{**})dt = 1/2$ will be a maximizer of the empirical MHRD like its population counterpart.

It is easy to see that the empirical coordinatewise median is equivariant under any coordinatewise monotone transformation given by $\mathbf{x} \mapsto \Psi(\mathbf{x})$, where $\mathbf{x} = \{x_t\}_{t\in[0,1]}$, $\Psi(\mathbf{x}) = \{\psi_t(x_t)\}_{t\in[0,1]}$, and ψ_t is a monotone function for each $t \in [0,1]$. These include location shifts $\mathbf{x} \mapsto \mathbf{x} + \mathbf{c}$, where $\mathbf{c} \in C[0,1]$, as well as coordinatewise scale transformations $\mathbf{x} \mapsto \mathbf{x}_1$, where $\mathbf{x}_1 = \{a_t x_t\}_{t\in[0,1]}$ and $a_t > 0$ for each $t \in [0,1]$.

Assume that $\int_0^1 |X_t|dt < \infty$ with probability 1. This assumption holds if with probability one, the process $\mathbf{X}$ has continuous paths (e.g., the standard Brownian motion on $[0,1]$). Then, it can be shown that $\widehat{\mathbf{m}}$ is a minimizer of the function $\int_0^1 \sum_{i=1}^n |X_{i,t} - x_t|dt$ over $\mathbf{x} \in C[0,1]$, which satisfies $\int_0^1 |x_t|dt < \infty$. Here $\mathbf{X}_i = \{X_{i,t}\}_{t\in[0,1]}$ are the sample observations. In other words, $\widehat{\mathbf{m}}$ is an empirical spatial median of the distribution of

$\mathbf{X}$ in the Banach space of real-valued absolutely integrable functions defined on $[0, 1]$. Hence, from Theorem 2.10 in Kemperman (1987), we get that $\widehat{\mathbf{m}}$ has 50% breakdown point.

The following result asserts the uniform strong consistency of the empirical coordinatewise median. Here the uniformity is over a subset of I the size of which grows with the sample size at an appropriate rate.

Fact 4.2.1.1. *Suppose that X_t has a density f_t in a neighbourhood of m_t for each $t \in [0, 1]$. Assume that for some $c_0, \eta_0 > 0$, we have $\inf_{|x - m_{t_k}| < \eta_0} f_{t_k}(x) > c_0$ for all $1 \le k \le d_n$ and $d_n \ge 1$. Then, if $\log(d_n) = o(n)$ as $n \to \infty$, we have $\sup_{1 \le k \le d_n} |\widehat{m}_{t_k} - m_{t_k}| \to 0$ as $n \to \infty$ almost surely.*

The proof of the above result can be obtained using the arguments in the proof of Corollary 6 in Kosorok and Ma (2007). These authors considered the coordinatewise median for high dimensional data when the dimension increases with the sample size. Using sharp uniform bounds on the marginal empirical processes corresponding to the coordinate variables, they obtained the rate of convergence of $\widehat{\mathbf{m}}$ under the same set of assumptions used in Fact 4.2.1.1 (see Corollary 6 in Kosorok and Ma (2007)).

In practice, the process $\mathbf{X}$ is observed at d_n grid points in $[0, 1]$. We can then construct a functional estimator of $\mathbf{m}$ from $\widehat{\mathbf{m}}$ as follows. For any point $t \in [0, 1]$, which is not a grid point, we can define $\widehat{m}_t$ by the average of the empirical medians corresponding to its k nearest grid points. Here $k \ge 1$ is a fixed integer. In that case, the uniform consistency of this functional estimator over the whole of $[0, 1]$ can be derived from Fact 4.2.1.1. For this derivation, let us assume that the grid points become dense in $[0, 1]$ as $n \to \infty$. In other words, for each $t \in [0, 1]$, any fixed neighbourhood of t will contain infinitely many grid points as $n \to \infty$. Let us also assume that the population deepest point $\mathbf{m} \in C[0, 1]$. Then, under the conditions assumed in Fact 4.2.1.1, it is easy to show that $\sup_{t \in [0,1]} |\widehat{m}_t - m_t| \to 0$ as $n \to \infty$ almost surely.

4.3 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)). We can define the spatial depth at $\mathbf{x}$ in any smooth Banach space $\mathcal{X}$ with respect to the probability distribution of a random element $\mathbf{X} \in \mathcal{X}$ as $SD(\mathbf{x}) = 1 - ||\Psi_{\mathbf{x}}||$, and its empirical version is given by $\widehat{SD}(\mathbf{x}) = 1 - ||\widehat{\Psi}_{\mathbf{x}}||$. Here, $\Psi_{\mathbf{x}}$ and $\widehat{\Psi}_{\mathbf{x}}$ are the spatial distribution at $\mathbf{x}$ and its empirical version as defined in `Chapter 3`.

Let us also mention here that an alternative definition of spatial depth in $\mathbb{R}^d$ was considered by the authors of Vardi and Zhang (2000). These authors defined the spatial depth of $\mathbf{x} \in \mathbb{R}^d$ relative to the distribution of a random element $\mathbf{X} \in \mathbb{R}^d$ as $1 - \inf\{w \geq 0 : \text{spatial median of } (w\delta_{\mathbf{x}} + F)/(1+w) = \mathbf{x}\}$. Here $\delta_{\mathbf{x}}$ denotes the point mass at $\mathbf{x}$, and F denotes the distribution of $\mathbf{X}$. It can be shown using the characterization of spatial median given in Theorem 4.14 in Kemperman (1987) that one gets $1 - \max\{0, [1 - SD(\mathbf{x}) - P(\mathbf{X} = \mathbf{x})]\}$ as the depth of $\mathbf{x}$ according to this definition. So, this definition of spatial depth coincides with the definition in the previous paragraph if F is nonatomic.

Spatial depth function inherits many of its interesting properties from finite dimensions. The spatial distribution function $\Psi_{\mathbf{x}}$ possesses an invariance property under the class of affine transformations $L : \mathcal{X} \to \mathcal{X}$ of the form $L(\mathbf{x}) = cA(\mathbf{x}) + \mathbf{a}$, where $c > 0$, $\mathbf{a} \in \mathcal{X}$ and $A : \mathcal{X} \to \mathcal{X}$ is a linear surjective isometry. By the definition of Gâteaux derivative and using the isometry of A, we have $S_{L(\mathbf{x})-L(\mathbf{X})}(\mathbf{h}) = S_{A(\mathbf{x})-A(\mathbf{X})}(A(\mathbf{h}')) = S_{\mathbf{x}-\mathbf{X}}(\mathbf{h}') = S_{\mathbf{x}-\mathbf{X}}(A^{-1}(\mathbf{h})) = (A^{-1})^*(S_{\mathbf{x}-\mathbf{X}}(\mathbf{h}))$ for any $\mathbf{x}, \mathbf{h} \in \mathcal{X}$. Here, $\mathbf{h} = A(\mathbf{h}')$, and $(A^{-1})^* : \mathcal{X}^* \to \mathcal{X}^*$ denotes the adjoint of A^{-1}. Thus, if $\Psi_{L(\mathbf{x})}$ is the spatial distribution at $L(\mathbf{x})$ with respect to the probability distribution of $L(\mathbf{X})$, we have $\Psi_{L(\mathbf{x})} = (A^{-1})^*(\Psi_{\mathbf{x}})$, where $\Psi_{\mathbf{x}}$ is the spatial distribution at $\mathbf{x}$ with respect to the probability distribution of $\mathbf{X}$. This implies that the spatial depth is invariant under such affine transformations in the sense that the spatial depth at $L(\mathbf{x})$ with respect to the distribution of $L(\mathbf{X})$ is same as the spatial depth at $\mathbf{x}$ with respect to the distribution of $\mathbf{X}$.

It follows from Remark 3.5 and Theorems 2.17 and 4.14 in Kemperman (1987) that if $\mathcal{X}$ is a strictly convex Banach space, and the distribution of $\mathbf{X}$ is nonatomic and not entirely contained on a line in $\mathcal{X}$, then $SD(\mathbf{x})$ has a unique maximizer at the spatial median (say, $\mathbf{m}$) of $\mathbf{X}$, and $SD(\mathbf{m}) = 1$ (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)). It follows from Theorem 3.2.1 in Chapter 3 that if the norm in $\mathcal{X}$ is Fréchet differentiable and the distribution of $\mathbf{X}$ is nonatomic, then $SD(\mathbf{x})$ is a continuous function in $\mathbf{x}$. Moreover, in such cases, $SD(\mathbf{x} + n\mathbf{y}) \to 0$ as $n \to \infty$ for any $\mathbf{x}, \mathbf{y} \in \mathcal{X}$ with $\mathbf{y} \neq \mathbf{0}$ (cf. the "FD3" property in Mosler and Polyakova (2012), Theorem 1 in Liu (1990) and property "P4" in Zuo and Serfling (2000, p. 464)). This implies that the spatial depth function vanishes at infinity along any ray through any point in $\mathcal{X}$. The above properties of $SD(\mathbf{x})$ are among the desirable properties of any statistical depth function listed in Liu (1990) and Zuo and Serfling (2000) for the finite dimensional setting.

A natural question that now arises is whether SD suffers from degeneracy similar to what was observed in the case of some of the depth functions discussed earlier or whether the distribution of SD is well spread out. It follows from Theorem 3.2.1 in Chapter 3 that if $\mathcal{X}$ is a reflexive Banach space and the distribution of $\mathbf{X}$ is nonatomic, then $SD(\mathbf{x})$ takes all values in $(0, 1]$ as $\mathbf{x}$ varies over $\mathcal{X}$. 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 separable Hilbert spaces including Gaussian probabilities.

Theorem 4.3.1. *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 $[0, 1]$.*

It is easy to show that if $SD(\mathbf{x})$ is continuous in $\mathbf{x}$, then $SD(\mathbf{x})$ takes all values in $(0, w] \subseteq (0, 1]$ as $\mathbf{x}$ varies over a closed subspace $\mathcal{W}$ of $\mathcal{X}$, where $w = \sup_{\mathbf{x} \in \mathcal{W}} SD(\mathbf{x})$.

In particular, $w = 1$ if $\mathcal{W}$ contains the spatial median of $\mathbf{X}$. It can be shown that the support of a Gaussian distribution in a separable Banach space is the closure of the translation of a subspace of $\mathcal{X}$ by the mean (which is also the spatial median) of that distribution. So, if the norm in that space is Fréchet differentiable, then $SD(\mathbf{x})$ is continuous in $\mathbf{x}$ and it takes all values in $(0, 1]$ as $\mathbf{x}$ varies over the support of that distribution. The properties of the spatial depth discussed above imply that it induces a meaningful center-outward ordering of the points in these spaces, and can be used to develop depth-based statistical procedures for data from such distributions.

Let us observe that the definition of SD using Gateaux derivatives is not applicable for probability distributions in the space $C[0, 1]$, where most of the other functional depths like BD, MBD, HRD, MHRD and ID are defined. This is because the norm in $C[0, 1]$ is not Gateaux differentiable everywhere (see, e.g., (Fabian et al., 2001, Ex. 8.28, p. 267)). However, since $C[0, 1] \subseteq L_2[0, 1]$, for any probability distribution on $C[0, 1]$, SD can be 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 4.3.1, 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-Loève expansion of $\mathbf{X}$, which will then be a sequence of independent Gaussian random variables, and consequently, the conditions of Theorem 4.3.1 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 4.2.3. Indeed, even if $\mathbf{X}$ admits a Karhunen-Loève type expansion in such a case, the sequence of coefficients need not satisfy the conditions of Theorem 4.3.1. 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 4.3.2. *Consider the process* $\mathbf{X} = \{X_t\}_{t\in[0,1]} = \{g(t, Y_t)\}_{t\in[0,1]}$ *as in Theorem 4.2.3. 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* $\tilde{\mathbf{X}}$ *is an independent copy of* $\mathbf{X}$.

It follows from arguments that are very similar to those in Remark 4.5.4 in Section 4.5 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 4.3.1 and 4.3.2 imply that like ID, MBD and MHRD, SD can also be used for various depth-based statistical procedures for data in infinite dimensional spaces. 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 nonatomic 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.

So far in this section, we showed that like MBD, MHRD and ID, SD also takes all values in a nondegenerate interval and its distribution is well spread out for a large class of distributions. We next study the asymptotic properties of the empirical spatial depth in smooth Banach spaces. The uniform consistency of the empirical ID, MBD and MHRD have already been studied by Fraiman and Muniz (2001), López-Pintado and Romo (2009) and López-Pintado and Romo (2011). Further, under the conditions of Theorem 4.2.2, when BD and HRD are degenerate, the uniform consistency of their empirical versions is trivially true. This follows from Theorem 4 in López-Pintado and Romo (2009) and Theorem 3 in López-Pintado and Romo (2011).

Theorem 4.3.3. *Suppose that the assumptions of part (a) of Theorem 3.1.2 in Chapter 3 hold. Then,* $\sup_{\mathbf{x}\in K} |\widehat{SD}(\mathbf{x}) - SD(\mathbf{x})| \to 0$ *as* $n \to \infty$ *almost surely for every compact set* $K \subseteq \mathcal{X}$. *Suppose that the norm function in* $\mathcal{X}^*$ *is Fréchet differentiable, and* $\mathcal{X}^*$ *is a separable and type 2 Banach space. Then,* $\sqrt{n}(\widehat{SD}(\mathbf{x}) - SD(\mathbf{x}))$ *converges weakly to* $S_{\Psi_{\mathbf{x}}}(\mathbf{W})$ *if* $\Psi_{\mathbf{x}} \neq \mathbf{0}$. *If* $\Psi_{\mathbf{x}} = \mathbf{0}$, $\sqrt{n}(\widehat{SD}(\mathbf{x}) - SD(\mathbf{x}))$ *converges weakly to* $-||\mathbf{V}||$. *Here,* $\mathbf{W}$ *and* $\mathbf{V}$ *are zero mean Gaussian random elements in* $\mathcal{X}^*$.

4.3.1 The deepest point

In `Chapter 3`, we studied the spatial median and its properties for arbitrary probability measures (see Theorem 3.2.1 and the discussion following it). In particular, the spatial median is unique if the underlying Banach space $\mathcal{X}$ is strictly convex, and the distribution of $\mathbf{X}$ is not entirely supported on a line in $\mathcal{X}$. Further, if the distribution of $\mathbf{X}$ is nonatomic, the unique spatial median is the only point in $\mathcal{X}$, which satisfies $E(S_{\mathbf{x}-\mathbf{X}}) = \mathbf{0}$. This implies that the spatial depth of the spatial median is 1. Equivalently, the spatial median is the deepest point associated with the spatial depth. When we have a sample from a probability distribution in $\mathcal{X}$, it can be verified using Theorem 4.14 in Kemperman (1987) that the sample spatial depth of the empirical spatial median $\widehat{\mathbf{m}}$ is 1 whenever it is not one of the data points. In other words, in such situations, the empirical spatial median $\widehat{\mathbf{m}}$ is the unique maximizer of the sample spatial depth. It is easy to see that the empirical spatial median is equivariant under location shifts and homogeneous scale transformations $\mathbf{x} \mapsto a\mathbf{x}$, where $a > 0$. Moreover, it is also equivariant under any linear isometry. Further, it is known that the empirical spatial median has 50% breakdown point (see Theorem 2.10 in Kemperman (1987)). It has been shown in Chakraborty and Chaudhuri (2014a) that under certain conditions, the empirical spatial median is consistent in any separable Hilbert space $\mathcal{X}$. Note that in `Chapter 3`, we studied an estimator of the spatial median based on finite dimensional approximations, and it was proved that this estimator is consistent in a class of Banach spaces including Hilbert spaces (see Theorem 3.2.1.1).

4.4 Demonstration using real and simulated data

In the preceding sections, we 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 from a zero mean Gaussian random vector $\mathbf{X} = (X_1, X_2, \ldots, X_d)$ with $Cov(X_k, X_l) = r^{-|k-l|}/(kl)^2$, 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 4.1). 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 4.1 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 4.1.

For the colon data, we have prepared another dotplot (see Fig. 4.2), 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 tissue 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 et al. (2012)) constructed using SD will yield good results for this dataset.

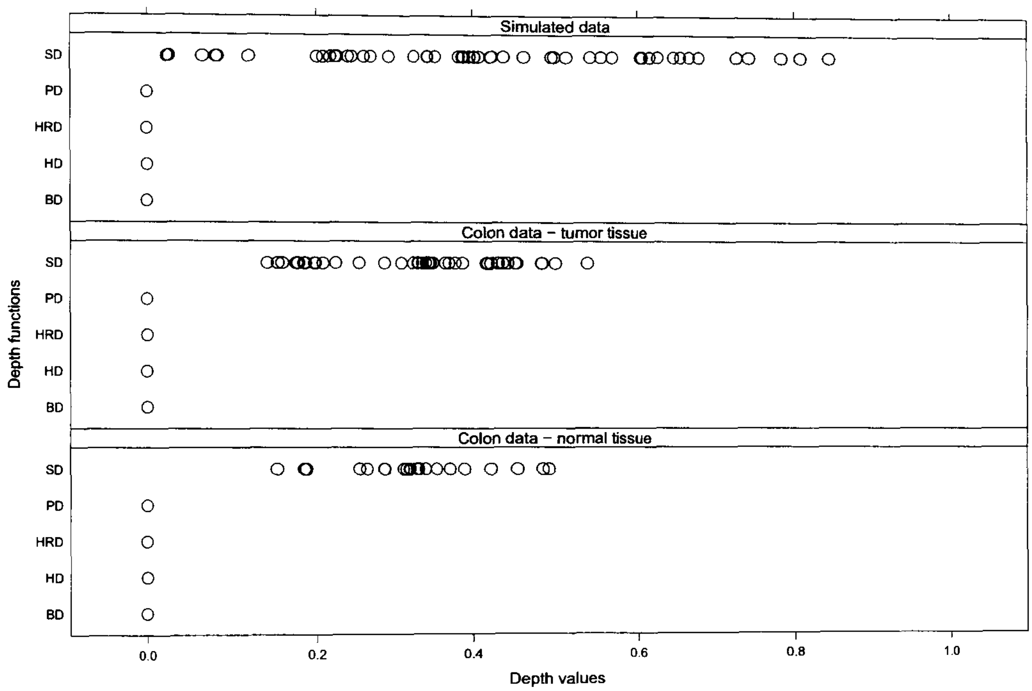

Figure 4.1: Dotplots of the empirical depth values for the simulated datasets and the colon data.

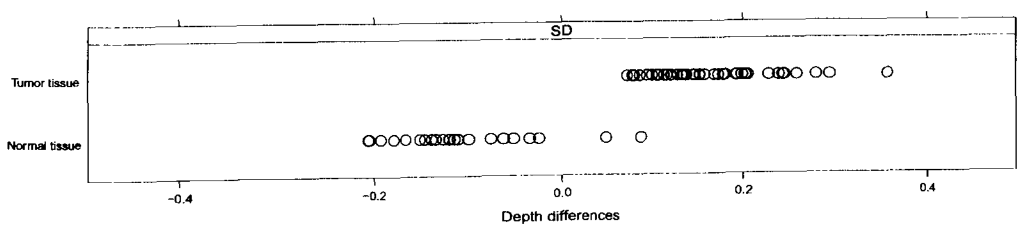

Figure 4.2: Dotplots of the empirical depth differences based on the spatial depth for the colon data.

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

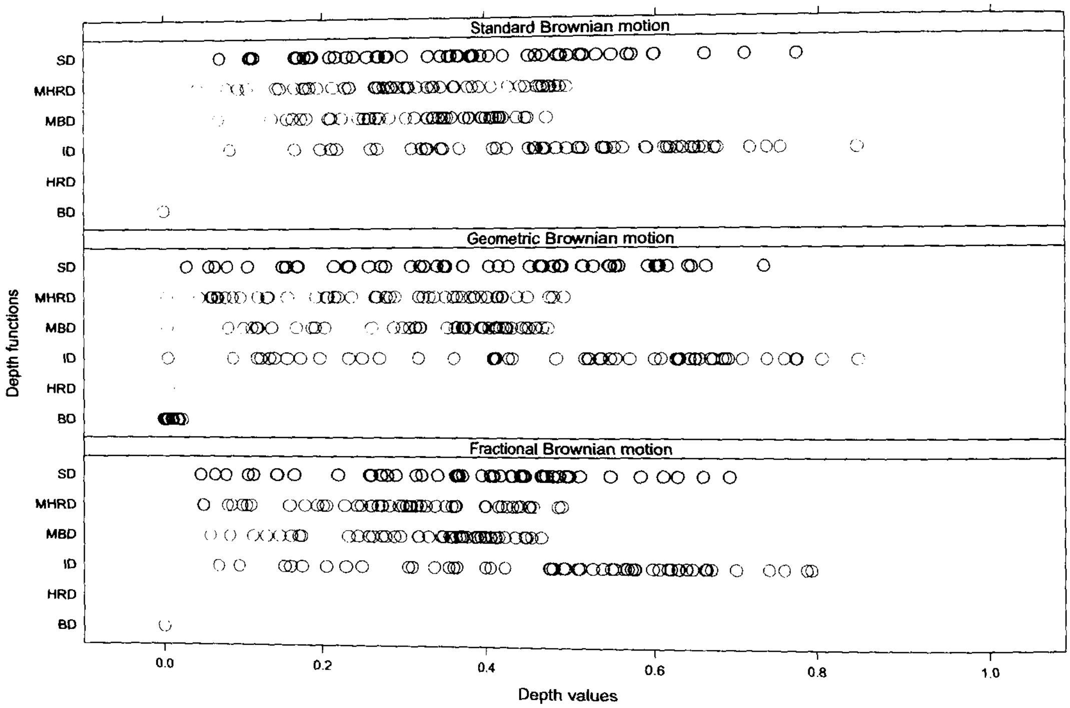

Figure 4.3: Dotplots of the empirical depth values for some simulated datasets.

$X_t = exp((r - \sigma^2/2)t + \sigma B_t)$, 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 curves are obtained through monotone spline smoothing available in the R package "fda", and these are recorded at $d = 101$ equispaced ages in the interval $[1,18]$. We derived the acceleration curves from the smoothed growth curves. For these functional datasets, we calculated MBD by taking $J = 2$ as suggested by López-Pintado and Romo (2009), and D_t in the definition of ID was taken to be SD for each t, which is same as the depth function used by Fraiman and Muniz (2001).

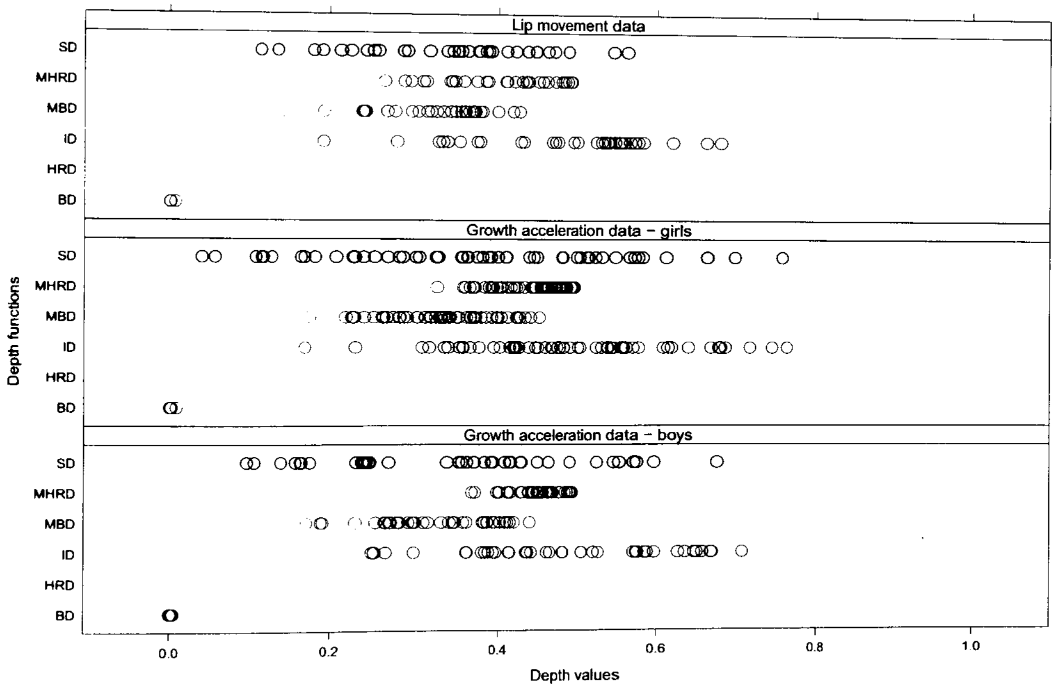

Figure 4.4: Dotplots of the empirical depth values for the lip movement data and the growth acceleration data.

As shown in the dotplots in Figures 4.3 and 4.4, 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 fractional Brownian motion (see Figure 4.3). For the geometric 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 4.3). 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 4.4). 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 4.4).

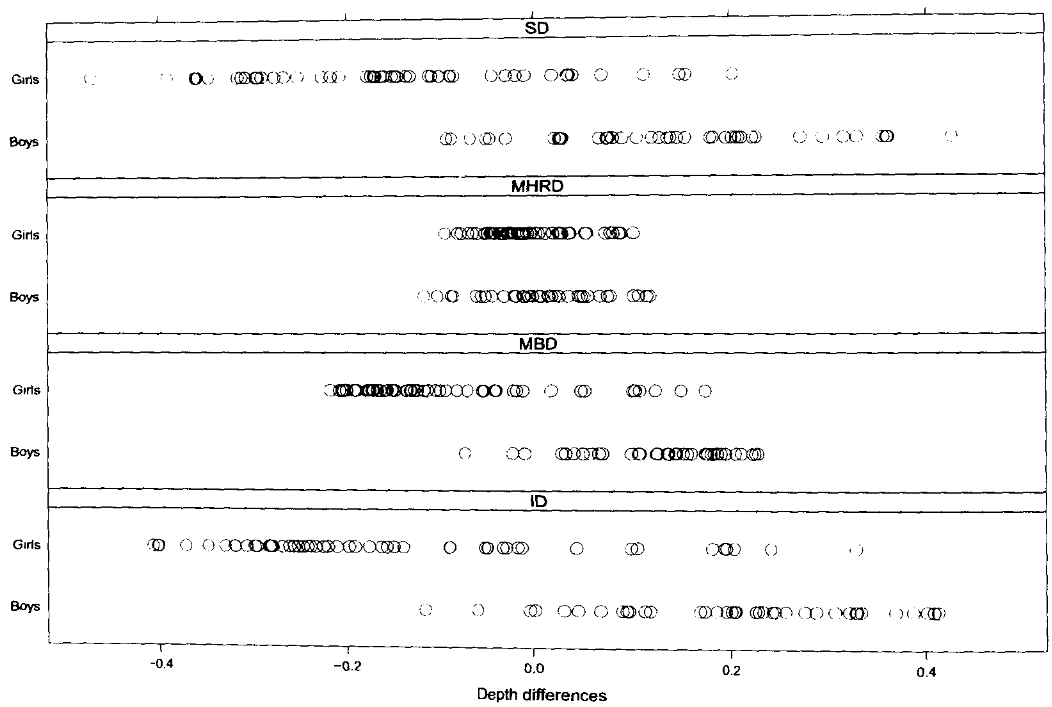

Figure 4.5: Dotplots of the empirical depth differences for the growth accceleration data.

For the growth acceleration data, Figure 4.5 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 et al. (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.

4.4.1 Demonstration of the empirical deepest point

We now demonstrate the empirical deepest points using some simulated and real datasets. For both the simulated and the real datasets, the empirical spatial median is computed using the finite dimensional approximation method studied in Chapter 3 by considering the data as a random sample in an appropriate separable Hilbert space. Each of the simulated datasets that we consider has 25 observations. One dataset is generated from the standard Brownian motion in $[0, 1]$, and the other two datasets are generated from fractional Brownian motions in $[0, 1]$ with Hurst indices $H = 0.3$ and $H = 0.7$. The sample curves are observed at 1000 equispaced points in $[0, 1]$. Figure 4.6 shows the plots of the sample curves along with the empirical coordinatewise medians (the blue curves) and the empirical spatial medians (the red curves) for these three distributions. It is observed from the plots that both of the empirical medians are close to the zero function, which is the spatial median as well as the coordinatewise median for all three distributions.

The real data considered here is the growth acceleration dataset, which was also used earlier in this section. Figure 4.7 shows the plots of the growth acceleration curves of the boys and the girls along with the empirical coordinatewise medians (the blue curves) and the empirical spatial medians (the red curves). It is seen from Figure 4.7 that both of the empirical medians are close to the central curves in each dataset.

4.4.2 DD-plot in infinite dimensional spaces

In the finite dimensional setup, an exploratory data analytic tool for checking whether two given samples arise from the same distribution or not is the depth-depth plot (DD-plot) (see Liu et al. (1999)). A DD-plot is a scatterplot of the depth values of the

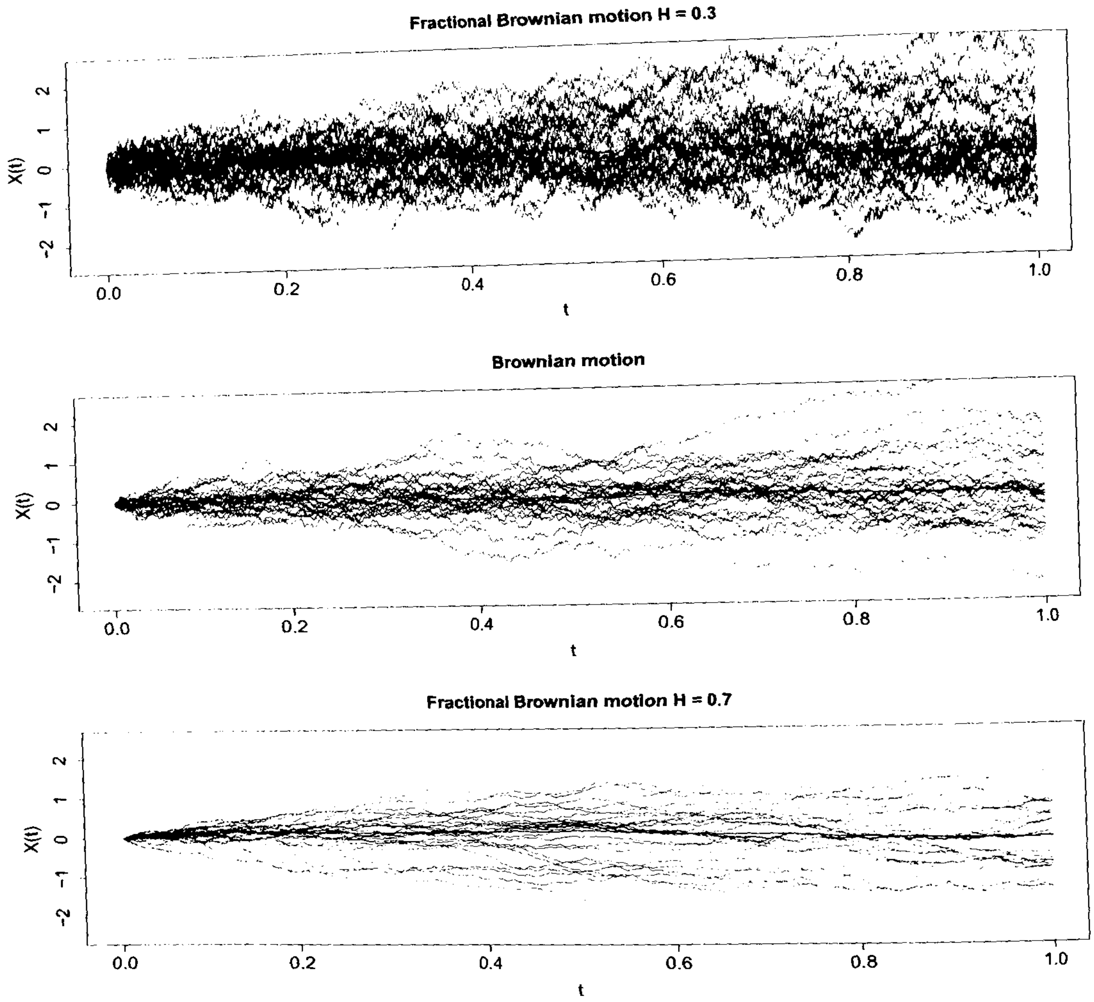

Figure 4.6: Plots of some simulated functional data along with the empirical coordinatewise medians and the empirical spatial medians.

data points in the pooled sample with respect to the empirical distributions of the two samples. It can be used to detect differences in location, scale etc. let us now consider the problem of constructing DD-plots for data in infinite dimensional spaces. It follows from some of the earlier results that the half-space depth and the simplicial depth, which have been used by the authors of Liu et al. (1999) for constructing DD-plots for data in finite dimensional spaces, cannot be used for constructing DD-plots in infinite dimensional spaces.

We have prepared DD-plots for some real and simulated functional data using SD,

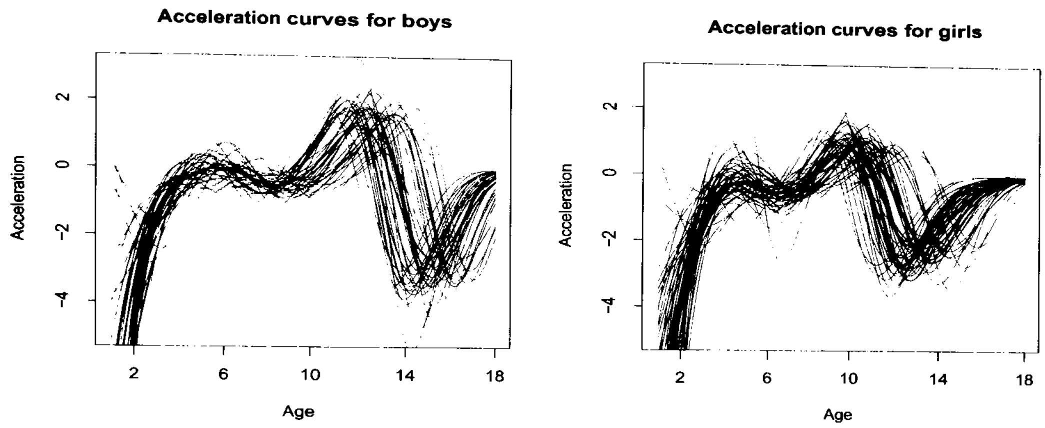

Figure 4.7: Plots of the acceleration curves for boys and girls along with the empirical coordinatewise medians and the empirical spatial medians.

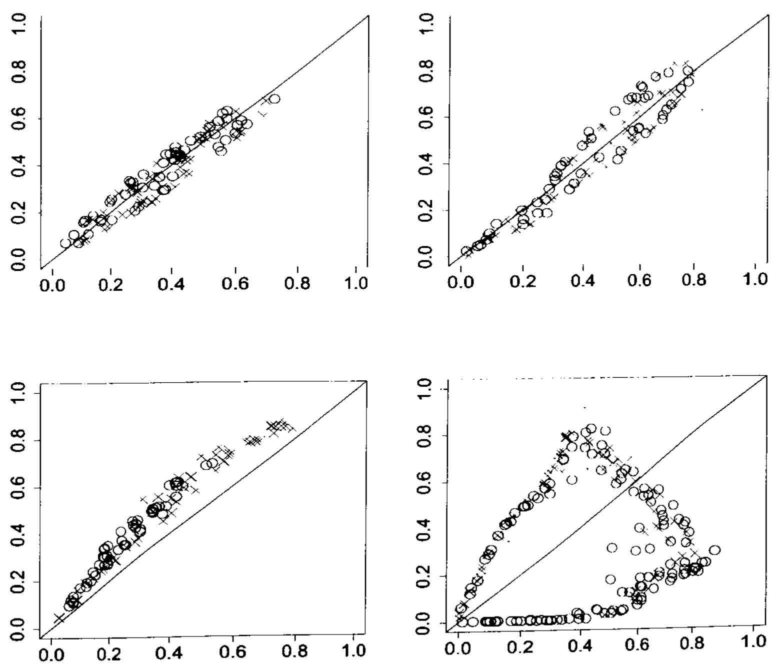

Figure 4.8: DD-plots for the simulated datasets and the spectrometric data using SD.

ID, MBD and MHRD (see Figures 4.8, 4.9, 4.10 and 4.11, respectively). The simulated datasets are samples from the standard Brownian motion and the fractional Brownian motion with H = 0.9. Both of these processes have Karhunen-Loève expansions in

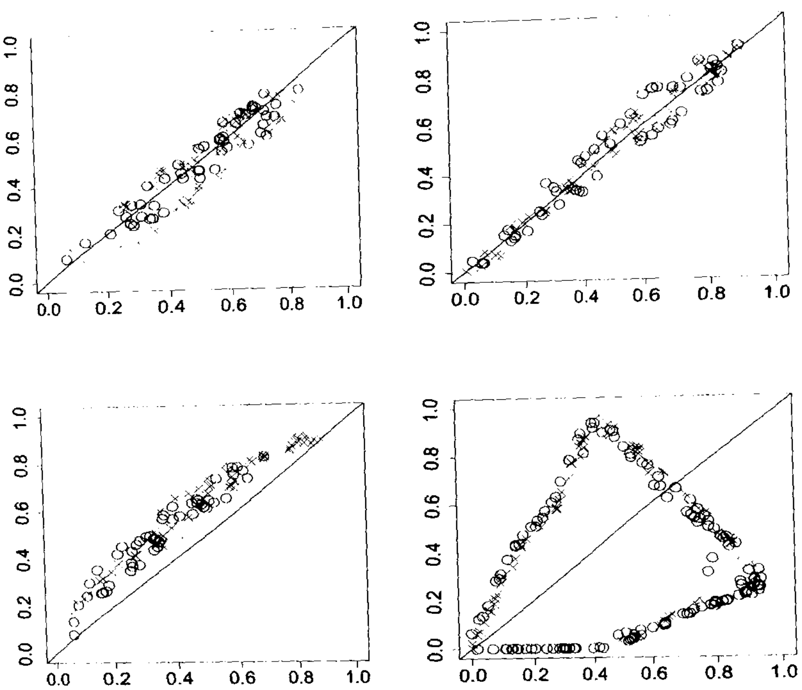

Figure 4.9: DD-plots for the simulated datasets and the spectrometric data using ID.

$L_2[0,1]$. Each simulated data consists of $n = 50$ samples, and the sample curves are observed at 250 equispaced points on $[0,1]$. The real data is the Spectrometry data used in `Chapter` 3, which can be viewed as a random sample from a probability distribution in $L_2[850, 1050]$. Since the sample spaces for the simulated and the real datasets considered here are Hilbert spaces, $\Psi_{\mathbf{x}}$ in the definition of SD simplifies to $E\{(\mathbf{x} - \mathbf{X})/||\mathbf{x} - \mathbf{X}||\}$. The norm in this expression is computed as the norm of the Euclidean space whose dimension is the number of values of the argument over which the sample functions in the dataset are observed.

The plots in the first row in Figure 4.8 are the DD-plots using SD for the two samples from the standard Brownian motion and the fractional Brownian motion. The first plot in the second row is the DD-plot for the two samples from the standard Brownian motion and the fractional Brownian motion. The axes of the DD-plots in the first row correspond to the depth values with respect to the empirical distributions of the standard Brownian motion and the fractional Brownian motion, respectively. In each of those plots, the black o's and the red $\times$'s represent the sample observations of the

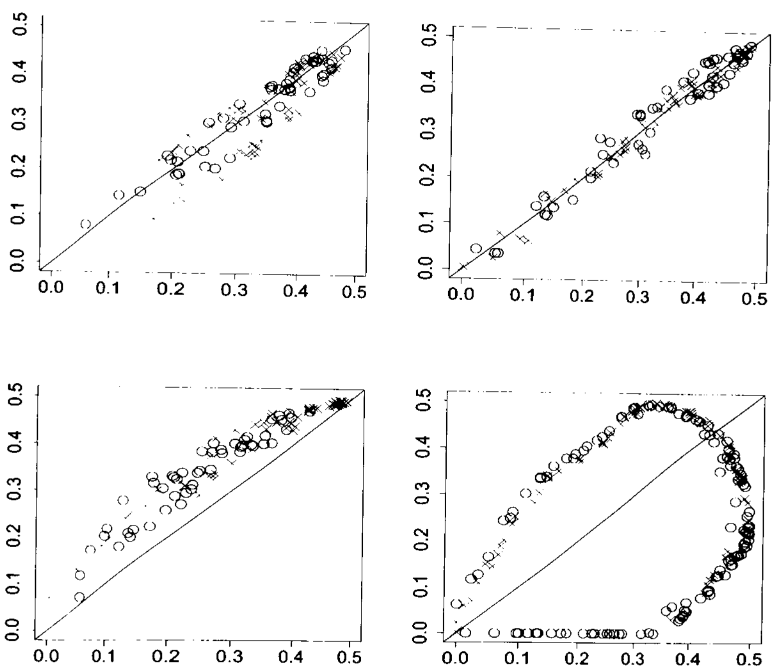

Figure 4.10: DD-plots for the simulated datasets and the spectrometric data using MBD.

two samples. The vertical and the horizontal axis of the first DD-plot in the second row correspond to the depth values with respect to the empirical distributions of the standard Brownian motion and the fractional Brownian motion, respectively, and the black o's and the red $\times$'s represent the samples from these two distributions, respectively. In the DD-plots in the first row, the black o's and the red $\times$'s are clustered around the 45° line through the origin. So, the observations from each of the two samples have similar depth values with respect to both the samples. This indicates that there is not much difference between the two underlying populations in each case. In the first DD-plot in the second row of Figure 4.8, all the black o's and the red $\times$'s lie above the 45° line through the origin in the shape of an arch. So, all the observations in the sample from the fractional Brownian motion have higher depth values with respect to the empirical distribution of the sample from the standard Brownian motion. This indicates that the former population has less spread than the latter one. The horizontal and the vertical axes of the DD-plot for the spectrometric data (see the second plot in the second

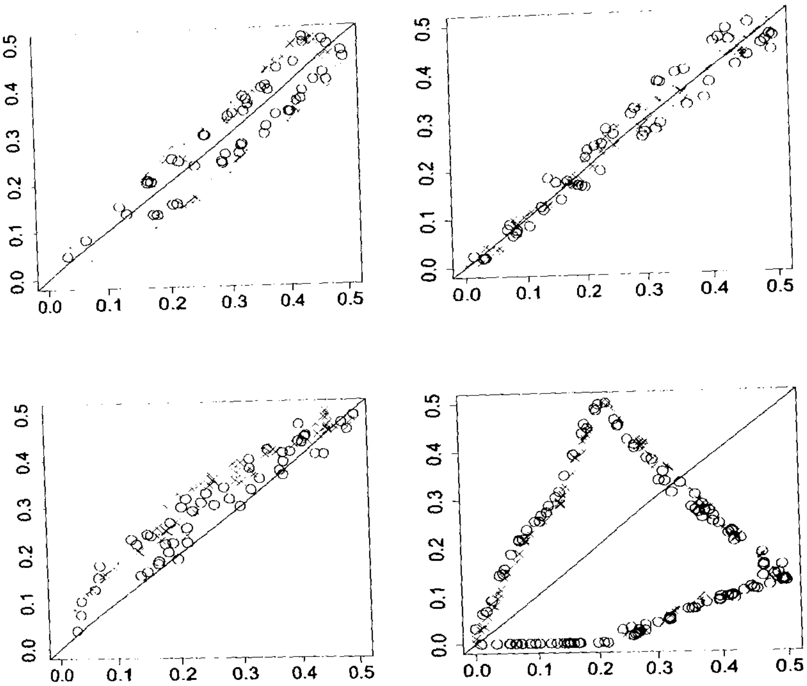

Figure 4.11: DD-plots for the simulated datasets and the spectrometric data using MHRD.

row in Figure 4.8) correspond to the spatial depth values with respect to the empirical distribution of the classes with fat content $\leq 20\%$ and $> 20\%$, respectively, and the black o's and the red $\times$'s represent the samples from these two classes, respectively. It is seen that the observations from both the samples are almost evenly spread out below and above the 45° line through the origin in the shape of a triangle. One side of the triangle is formed by the line joining the points with approximate coordinates $(0.4, 0.8)$ and $(0.8, 0.2)$, and the vertex opposite to that side is the origin. This type of DD-plot indicates a difference in location between the two samples. The points around the aforementioned side of the triangle lie in the overlapping region of the two samples, and have moderate to high depth values with respect to the empirical distributions of both the samples.

Analogous descriptions as in the previous paragraph applies to the plots in Figures 4.9, 4.10 and 4.11. The plots in each of these three figures are quite similar to the corresponding DD-plots using SD in Figure 4.8 with the exceptions being the DD-plot for the spectrometric data using MBD in Figure 4.10 and the first DD-plot in the second row in Figure 4.11. The shape of the DD-plot for the spectrometric data using MBD in Figure 4.10 is slightly different from the triangular shape seen, when it is constructed using SD, ID and MHRD. However, like those DD-plots, it also indicates a difference in the location between the two samples of the spectrometric data. The points around the top right corner of the DD-plot lie in the overlapping region of the two samples, and have moderate to high depth values with respect to the empirical distributions of both the samples. The first DD- plot in the second row in Figure 4.11 is more inclined towards the 45° line than the corresponding arch shaped DD-plots seen in the other three figures. This indicates that MHRD is less accurate in distinguishing the difference in scale between the Brownian motion and the fractional Brownian motion considered than SD, MBD and ID. In view of the generalizability of SD and its desirable properties discussed earlier, it is recommended to use SD to construct DD-plots for data in infinite dimensional spaces.

4.5 Mathematical details

Proof of Theorem 4.1.1. 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

$$\begin{aligned} HD(\mathbf{x}) &= \inf_{\mathbf{u} \in l_2} P(\langle \mathbf{u}, \mathbf{X} - \mathbf{x} \rangle \geq 0) \leq \inf_{d \geq 1} \inf_{\mathbf{v} \in \mathbb{R}^d} P(\mathbf{v}'\mathbf{X}(d) \geq \mathbf{v}'\mathbf{x}(d)) \\ &= \inf_{d \geq 1} \inf_{\mathbf{v} \in \mathbb{R}^d} P(\mathbf{v}'\mathbf{Y}(d) \geq \mathbf{v}'\mathbf{y}(d)) \\ &\leq \inf_{d \geq 1} \inf_{\mathbf{v} \in \mathbb{R}^d : \mathbf{v}'\mathbf{y}(d) > 0} P(\mathbf{v}'\mathbf{Y}(d) \geq \mathbf{v}'\mathbf{y}(d)), \end{aligned} \tag{4.5}$$

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 (4.5) and Chebyshev inequality that

$$HD(\mathbf{x}) \leq \inf_{d\geq 1} \inf_{\mathbf{v}:\mathbf{v}'\mathbf{y}(d)>0} \frac{Var(\mathbf{v}'\mathbf{Y}(d))}{(\mathbf{v}'\mathbf{y}(d))^2} = \inf_{d\geq 1} \inf_{\mathbf{v}:\mathbf{v}'\mathbf{y}(d)>0} \frac{\sum_{k=1}^d v_k^2\tau_k^2}{\left[\sum_{k=1}^d v_k y_k\right]^2}. \quad (4.6)$$

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

$$HD(\mathbf{x}) \leq \inf_{d\geq 1} \left[\sum_{k=1}^d y_k^2/\tau_k^2\right]^{-1}. \quad (4.7)$$

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

$$d^{-1}\sum_{k=1}^d Y_k^2/\tau_k^2 \to 1 \text{ a.s.} \Rightarrow \inf_{d\geq 1} \left[\sum_{k=1}^d Y_k^2/\tau_k^2\right]^{-1} = 0 \text{ a.s.} \quad (4.8)$$

(4.7) and (4.8) 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 $\mathbf{x} \in l_2$,

$$\begin{aligned}
\frac{1-PD(\mathbf{x})}{PD(\mathbf{x})} &= \sup_{\mathbf{u}\in l_2} \frac{|\langle \mathbf{u},\mathbf{x}\rangle - E(\langle \mathbf{u},\mathbf{X}\rangle)|}{\sqrt{Var(\langle \mathbf{u},\mathbf{X}\rangle)}} \geq \sup_{d\geq 1}\sup_{\mathbf{v}\in\mathbb{R}^d} \frac{|\mathbf{v}'\mathbf{x}(d) - E(\mathbf{v}'\mathbf{X}(d))|}{\sqrt{Var(\mathbf{v}'\mathbf{X}(d))}} \\
&\geq \sup_{d\geq 1}\sup_{\mathbf{v}\in\mathbb{R}^d} \frac{|\mathbf{v}'\mathbf{y}(d)|}{\sqrt{Var(\mathbf{v}'\mathbf{Y}(d))}} \geq \sup_{d\geq 1}\sup_{\mathbf{v}\in\mathbb{R}^d} \frac{\left|\sum_{k=1}^d v_k y_k\right|}{\sqrt{\sum_{k=1}^d v_k^2\tau_k^2}} \\
&= \sup_{d\geq 1}\sum_{k=1}^d \frac{y_k^2}{\tau_k^2}. \quad (4.9)
\end{aligned}$$

As in the case of HD, in view of the moment and the mixing conditions on the Y_k's assumed in the theorem, (4.8) and (4.9) now imply that $PD(\mathbf{x}) = 0$ for all $\mathbf{x}$ in a subset of l_2 with μ-measure one.

□

Proof of Theorem 4.1.2. 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 4.1.1 and using $\mathbf{Y}_d$ in place of $\mathbf{Y}(d)$. □

Proof of Theorem 4.2.1. 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 $E\{BD(\widetilde{\mathbf{X}})\} = \sum_{j=2}^{J} P(\min_{1\le i\le j} X_{i,k} \le \widetilde{X}_k \le \max_{1\le i\le j} X_{i,k},\ \forall\ k \ge 1)$. So, $E\{BD(\widetilde{\mathbf{X}})\} = 0$ iff $P(\min_{1\le i\le j} X_{i,k} \le \widetilde{X}_k \le \max_{1\le i\le j} X_{i,k},\ \forall\ k \ge 1) = 0$ for all $2 \le j \le J$. Consequently, it is enough to show that for any $2 \le j \le J$, the event $\{\min_{1\le i\le j} X_{i,k} > \widetilde{X}_k\} \cup \{\max_{1\le i\le j} X_{i,k} < \widetilde{X}_k\}$ occurs for some $k \ge 1$ with probability one. Now, the sequence $(\min_{1\le i\le j} X_{i,1} - \widetilde{X}_1, \min_{1\le i\le j} X_{i,2} - \widetilde{X}_2, \ldots)$ is α-mixing for any $1 \le j \le J$, and its mixing coefficients satisfy the conditions assumed in the theorem. On the other hand, $P(\min_{1\le i\le j} X_{i,k} > \widetilde{X}_k) = 2^{-j}$ for all $k \ge 1$, by the continuity of the distributions of the X_k's. So, using Corollary 4 in Hansen (1991), we have $d^{-1} \sum_{k=1}^{d} I(\min_{1\le i\le j} X_{i,k} > \widetilde{X}_k) \to 2^{-j}$ as $d \to \infty$ with probability one for all $1 \le j \le J$. So, the event $\{\min_{1\le i\le j} X_{i,k} > \widetilde{X}_k\}$ actually occurs for infinitely many $k \ge 1$ 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 4.5.1. *Let $\{X_t\}_{t\in[0,1]}$ be a Feller processes in $C[0,1]$ satisfying the conditions of Theorem 4.2.2. Let $\mathbf{X}_i = \{X_{i,t}\}_{t\in[0,1]}$, $i = 1, 2, \ldots, J$, denote independent copies of $\mathbf{X}$, and define $T_j = \inf\{t > 0 : \min_{1\le i\le j} X_{i,t} > x_0\}$ and $S_j = \inf\{t > 0 : \max_{1\le i\le j} X_{i,t} < x_0\}$ for $1 \le j \le J$. Then, $P(T_j = 0) = P(S_j = 0) = 1$ for all $1 \le j \le J$.*

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

$$P(T_j = 0) = \lim_{s\downarrow 0} P(T_j \le s) \ge 2^{-j} \quad \text{and} \quad P(S_j = 0) = \lim_{s\downarrow 0} P(S_j \le s) \ge 2^{-j}. \tag{4.10}$$

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 (4.10). □

Lemma 4.5.2. *Let $\{X_t\}_{t\in[0,1]}$ be a Feller process on $C[0,1]$ satisfying the conditions of Theorem 4.2.2. 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 4.5.1, 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 4.2.2. We first prove the result for BD using similar ideas as in the proof of Theorem 4.2.1. From the definition of BD in (4.1), we have

$$\begin{aligned} E\{BD(\widetilde{\mathbf{X}})\} &= \sum_{j=2}^{J} P\left(\min_{1\leq i\leq j} X_{i,t} \leq \widetilde{X}_t \leq \max_{1\leq i\leq j} X_{i,t},\ \forall\, t \in [0,1]\right) \\ &\leq \sum_{j=2}^{J} P\left(\min_{1\leq i\leq j} X_{i,t} \leq \widetilde{X}_t,\ \forall\, t \in [0,1]\right) \\ &= \sum_{j=2}^{J} E\left\{P\left(\min_{1\leq i\leq j} X_{i,t} \leq \widetilde{X}_t,\ \forall\, t \in [0,1] \,\middle|\, \mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_J\right)\right\} \end{aligned} \tag{4.11}$$

For any fixed j, let $\mathbf{z} = \{z_t\}_{t\in[0,1]}$ be a realization of the process $\{\min_{1\leq i\leq j} X_{i,t}\}_{t\in[0,1]}$. Then, from Lemma 4.5.1, it follows that $\mathbf{z}$ satisfies, with probability one, the assumptions made on the function $\mathbf{f}$ in Lemma 4.5.2. So, using Lemma 4.5.2, we have $P(z_t \leq \widetilde{X}_t,\ \forall\, t \in [0,1]) = 0$ for all $\mathbf{z}$ in a set of probability one. Hence, the expectation in (4.11) 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 4.5.1 and similar arguments as above. □

Lemma 4.5.3. *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 4.2.3. (a) Since the process $\mathbf{Y} = \{Y_t\}_{t\in[0,1]}$ has almost surely continuous sample paths, Lemma 4.5.3 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 4.5.3, we have $\mathbf{x}_p = \mathbf{G}(\mathbf{y}_p) \in C[0,1]$.

Note that by strict monotonicity of $g(t,\cdot)$ 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

4.5.3, 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 $g(t,\cdot)$ 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 4.5.4. *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 4.2.4. Since $X_0 = a_0$ with probability one, if $\mathbf{y} = \{y_t\}_{t\in[0,1]} \in C[0,1]$ is such that $y_0 \neq a_0$, then $\mathbf{y}$ cannot be contained in any band formed by the $\mathbf{X}_i$'s. Consequently, $BD(\mathbf{y}) = 0$. So, it is enough to prove the result for any $\mathbf{y} = \{y_t\}_{t\in[0,1]} \in C[0,1]$ satisfying $y_0 = x_0$. Let $\mathbf{y}_m = (y_{1/2^m}, y_{2/2^m}, \ldots, y_{2^m/2^m})$. It follows from part 2 of Theorem 1 in López-Pintado and Romo (2009) that $BD_m(\mathbf{y}_m) \leq BD_m(\mathbf{a}_m)$ for all $m \geq 1$, where the function BD_m is the m-dimensional band depth calculated using the distribution of $\mathbf{X}_m = (X_{1/2^m}, X_{2/2^m}, \ldots, X_{2^m/2^m})$. Since

$$\begin{aligned} &\{\min_{1\leq i\leq j} X_{i,k/2^{m+1}} \leq x_{k/2^{m+1}} \leq \max_{1\leq i\leq j} X_{i,k/2^{m+1}} \ \forall\ k = 1,2,\ldots,2^{m+1}\} \\ \subseteq\ &\{\min_{1\leq i\leq j} X_{i,k/2^m} \leq x_{k/2^m} \leq \max_{1\leq i\leq j} X_{i,k/2^m} \ \forall\ k = 1,2,\ldots,2^m\} \end{aligned}$$

for all $1 \leq j \leq J$, it follows that $BD_{m+1}(\mathbf{y}_{m+1}) \leq BD_m(\mathbf{y}_m)$ for all $m \geq 1$. So,

$$\lim_{m\to\infty} BD_m(\mathbf{y}_m)$$
$$= \sum_{j=2}^{J} P\left(\min_{1\leq i\leq j} X_{i,k/2^m} \leq x_{k/2^m} \leq \max_{1\leq i\leq j} X_{i,k/2^m}\ \forall\ k = 1,2,\ldots,2^m \text{ and } m \geq 1\right)$$

Then, using the almost sure uniform continuity of the sample paths, we have

$$BD(\mathbf{y}) = \lim_{m\to\infty} BD_m(\mathbf{y}_m) \leq \lim_{m\to\infty} BD_m(\mathbf{a}_m) = BD(\mathbf{a}).$$

The proof will be complete if we show that $BD(\mathbf{a}) = 0$. Let us now consider the multivariate Feller process $\{(X_{1,t}, X_{2,t}, \ldots, X_{j,t})\}_{t\in[0,1]}$ for $1 \leq j \leq J$. Define $\mathbf{Y} = \{Y_t\}_{t\in[0,1]} = \mathbf{X} - \mathbf{a}$. Since $\{X_t\}_{t\in[0,1]}$ is a Feller process starting at $x_0 = a_0$ and symmetric about $\mathbf{a}$, $\mathbf{Y}$ is a Feller process starting at 0 and symmetric about $\mathbf{0}$. Let $T_j = \inf\{t > 0 : \min_{1\leq i\leq j} X_{i,t} > a_t\} = \inf\{t > 0 : \min_{1\leq i\leq j} Y_{i,t} > 0\}$ and $S_j = \inf\{t > 0 : \max_{1\leq i\leq j} X_{i,t} < a_t\} = \inf\{t > 0 : \max_{1\leq i\leq j} Y_{i,t} < 0\}$. From the continuity of the sample paths and using Propositions 2.16 and 2.17 in Revuz and Yor (1991), we get that $P(T_j = 0) = 0$ or 1 and $P(S_j = 0) = 0$ or 1 for all $1 \leq j \leq J$. Since $P(T_j = 0) = \lim_{t\downarrow 0} P(T_j \leq t) \geq 2^{-j}$ and $P(S_j = 0) = \lim_{t\downarrow 0} P(S_j \leq t) \geq 2^{-j}$, we have $P(T_j = 0) = P(S_j = 0) = 1$ for all $1 \leq j \leq J$.

The proof of the fact that the half-region depth of $\mathbf{a}$ is 0 follows from the above arguments after taking $J = 1$. □

Proof of Theorem 4.2.5. From the definition of MBD we have

$$\begin{aligned} MBD(\mathbf{x}) &= \sum_{j=2}^{J} E\left[\int_0^1 I\left(\min_{i=1,\ldots,j} X_{i,t} \leq x_t \leq \max_{i=1,\ldots,j} X_{i,t}\right) dt\right] \\ &= \sum_{j=2}^{J} \left[\int_0^1 P\left(\min_{i=1,\ldots,j} X_{i,t} \leq x_t \leq \max_{i=1,\ldots,j} X_{i,t}\right) dt\right] \\ &= \sum_{j=2}^{J} \int_0^1 \left[1 - F_t^j(x_t) - (1 - F_t(x_t))^j\right] dt, \end{aligned} \qquad (4.12)$$

where the second equality in the above follows from Fubini's theorem. Here F_t denotes the distribution of X_t for $t \in [0,1]$. For each $t \in [0,1]$, the integrand in (4.12) is maximized iff $F_t(x_t) = 1/2$, which can be easily verified using standard calculus. This implies that the term in the right hand side of equation (4.12) is maximized iff $F_t(x_t) = 1/2$ for all $t \in [0,1]$, except perhaps on a subset of I with Lebesgue measure zero. Hence, MBD is maximized at $\mathbf{m}$, and also at any $\mathbf{m}^*$, which equals $\mathbf{m}$ outside a Lebesgue null set.

Since D_t is maximized at m_t for each $t \in [0,1]$, it follows that the depth function $ID(\mathbf{x}) = \int_0^1 D_t(x_t)dt$ is maximized at $\mathbf{m}$, and at any $\mathbf{m}^*$, which equals $\mathbf{m}$ except on a Lebesgue null set.

For MHRD, we have using Fubini's theorem

$$\begin{aligned} MHRD(\mathbf{x}) &= \min\left\{E\left[\int_0^1 I(X_t \le x_t)dt\right],\ E\left[\int_0^1 I(X_t \ge x_t)dt\right]\right\} \\ &= \min\left\{\left[\int_0^1 P(X_t \le x_t)dt\right],\ \left[\int_0^1 P(X_t \ge x_t)dt\right]\right\} \\ &= \min\left\{\int_0^1 F_t(x_t)dt,\ 1 - \int_0^1 F_t(x_t)dt\right\}. \end{aligned} \tag{4.13}$$

The maximum value of the right hand side of (4.13) is 1/2. Since $F_t(m_t) = 1/2$ for all $t \in [0,1]$, $\mathbf{m}$ is a maximizer of MHRD. Further, any $\mathbf{m}^{**} = \{m_t^{**}\}_{t\in[0,1]}$ satisfying $\int_0^1 F_t(m_t^{**})dt = 1/2$ will also maximize MHRD. □

Proof of Theorem 4.3.1. 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 $d \ge 1$ 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,

$$\begin{aligned} &P(||\widetilde{\mathbf{X}} - \mathbf{x}|| < 2\eta) > P(||\widetilde{\mathbf{X}} - \mathbf{x}[d]|| < \eta) \\ &> P\left(\sum_{k\le M}(\widetilde{X}_k - x_k)^2 < \frac{\eta^2}{2} \;\middle|\; \sum_{k>M}\widetilde{X}_k^2 < \frac{\eta^2}{2}\right) P\left(\sum_{k>M}\widetilde{X}_k^2 < \frac{\eta^2}{2}\right). \end{aligned} \tag{4.14}$$

Using Markov inequality, we get

$$P\left(\sum_{k>M}\widetilde{X}_k^2<\frac{\eta^2}{2}\right) > 1-\frac{\sum_{k>M}E(\widetilde{X}_k^2)}{\eta^2/2} > 1/2. \tag{4.15}$$

(4.14) and (4.15) now imply that

$$P(||\widetilde{\mathbf{X}}-\mathbf{x}||<2\eta) > \frac{1}{2}P\left(\sum_{k\le M}(\widetilde{X}_k-x_k)^2<\frac{\eta^2}{2}\;\Bigg|\;\sum_{k>M}\widetilde{X}_k^2<\frac{\eta^2}{2}\right). \tag{4.16}$$

From the conditional full support assumption on the X_k's, it follows that the expression on the right hand side of the inequality (4.16) 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 nonatomic, SD is a continuous function on l_2 as mentioned in Section 4.3. 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 4.3, 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 4.5.5. *The set $H_0=\mathbf{G}(C_{y_0}[0,1])$ is convex. Here, $\mathbf{G}$ is as in Lemma 4.5.1 and $C_{y_0}[0,1]$ is as in the proof of Theorem 4.2.3.*

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 $g(t,\cdot)$, 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 $g(t,\cdot)$ 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

$$g(t_n,q_{t_n})\to\lambda g(t,f_t)+(1-\lambda)g(t,h_t)=g(t,q_t) \tag{4.17}$$

as $n\to\infty$. Suppose now, if possible, $q_{t_n}\not\to q_t$ as $n\to\infty$. Then, there exists $\epsilon_0>0$ and a subsequence $\{t_{n_j}\}_{j\ge1}$ such that $|q_{t_{n_j}}-q_t|>\epsilon_0$ for all $j\ge1$. A further subsequence of $\{t_{n_j}\}_{j\ge1}$ 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 (4.17) 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 4.5.6. *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 4.3.2.*

Proof. Fix $\mathbf{f} \in C_{y_0}[0, 1]$ and $\eta > 0$. Let $||\cdot||$ denote the supremum norm on $C[0, 1]$ as before, and $||\cdot||_2$ denote the usual norm on $L_2[0, 1]$. Since $||\mathbf{y}||_2 \leq ||\mathbf{y}||$ 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 4.5.3, 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 4.3.2. 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 4.3, 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 4.5.5, 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 nonatomicity 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 $||\cdot||_2$ 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 $g(t, \cdot)$ is a strictly increasing continuous function for each $t \in [0, 1]$ with $g(t, s) \to \infty$ as $s \to \infty$, 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 4.5.6, 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 $[0,1]$. □

Proof of Theorem 4.3.3. The proof of the first statement follows directly from part (a) of Theorem 3.1.2 after using the inequality $|||\mathbf{x}|| - ||\mathbf{y}||| \leq ||\mathbf{x}-\mathbf{y}||$, which holds for any $\mathbf{x}, \mathbf{y} \in \mathcal{X}$.

Let us next consider the case $\Psi_{\mathbf{x}} \neq \mathbf{0}$. From the Fréchet differentiability of the norm in $\mathcal{X}^*$, we have $\widehat{SD}(\mathbf{x}) - SD(\mathbf{x}) = S_{\Psi_{\mathbf{x}}}(\widehat{\Psi}_{\mathbf{x}} - \Psi_{\mathbf{x}}) + o(||\widehat{\Psi}_{\mathbf{x}} - \Psi_{\mathbf{x}}||)$. The central limit theorem for i.i.d. random elements in $\mathcal{X}^*$ (see, e.g., Araujo and Giné (1980)) implies that $\sqrt{n}(\widehat{\Psi}_{\mathbf{x}} - \Psi_{\mathbf{x}})$ converges weakly to a zero mean Gaussian random element $\mathbf{W} \in \mathcal{X}^*$ as $n \to \infty$. In particular, $\sqrt{n}||\widehat{\Psi}_{\mathbf{x}} - \Psi_{\mathbf{x}}||$ is bounded in probability as $n \to \infty$. Since the map $S_{\Psi_{\mathbf{x}}} : \mathcal{X}^* \to \mathbb{R}$ is continuous, we now have the result for $\Psi_{\mathbf{x}} \neq \mathbf{0}$ using the continuous mapping theorem.

Now, we consider the case $\Psi_{\mathbf{x}} = \mathbf{0}$. In this case, $\widehat{SD}(\mathbf{x}) - SD(\mathbf{x}) = -||\widehat{\Psi}_{\mathbf{x}}||$. The central limit theorem for i.i.d. random elements in $\mathcal{X}^*$ yields that $\sqrt{n}\widehat{\Psi}_{\mathbf{x}}$ converges weakly to a zero mean Gaussian random element $\mathbf{V} \in \mathcal{X}^*$ as $n \to \infty$. Finally, the continuous mapping theorem completes the proof in view of the continuity of the norm function in any Banach space. □

Appendix A

Appendix: Some definitions and concepts in Banach spaces

Let $\mathcal{X}$ be a Banach space with norm $||\cdot||$. The dual space of $\mathcal{X}$, denoted by $\mathcal{X}^*$, is the Banach space of all real-valued continuous linear functions on $\mathcal{X}$. We denote by $\mathcal{X}^{**}$ the dual of $\mathcal{X}^*$. A Banach space is called reflexive is there exists a surjective isomorphism from $\mathcal{X}$ to $\mathcal{X}^{**}$, and in such cases $\mathcal{X}$ can be identified with $\mathcal{X}^{**}$. Hilbert spaces and L_p spaces for $p \in (1, \infty)$ are examples of reflexive Banach spaces.

The norm in $\mathcal{X}$ is said to be Gâteaux differentiable at a nonzero $\mathbf{x} \in \mathcal{X}$ with derivative, say, $S_{\mathbf{x}} \in \mathcal{X}^*$ if

$$\lim_{t \to 0} t^{-1}(||\mathbf{x} + t\mathbf{h}|| - ||\mathbf{x}||) = S_{\mathbf{x}}(\mathbf{h})$$

for all $\mathbf{h} \in \mathcal{X}$. If the above limit is uniform in $\mathbf{h}$ for $||\mathbf{h}|| < 1$, then the norm is said to be Fréchet differentiable at $\mathbf{x}$. Equivalently, the norm is Fréchet differentiable if

$$\lim_{\mathbf{h} \to \mathbf{0}} ||\mathbf{h}||^{-1}\{||\mathbf{x} + \mathbf{h}|| - ||\mathbf{x}|| - S_{\mathbf{x}}(\mathbf{h})\} = 0.$$

We define $S_{\mathbf{x}} = \mathbf{0}$ if $\mathbf{x} = \mathbf{0}$. The Banach space $\mathcal{X}$ is said to be smooth (Fréchet smooth) if the norm in $\mathcal{X}$ is Gâteaux (Fréchet) differentiable at every nonzero $\mathbf{x} \in \mathcal{X}$. Norms in Hilbert spaces and L_p spaces are Fréchet differentiable. For a Hilbert space, $S_{\mathbf{x}} = \mathbf{x}/||\mathbf{x}||$. When $\mathcal{X} = L_p(\mathbb{R}^d)$, we have $S_{\mathbf{x}}(\mathbf{h}) = \int_{\mathbb{R}^d} sign\{\mathbf{x}(\mathbf{s})\}|\mathbf{x}(\mathbf{s})|^{p-1}\mathbf{h}(\mathbf{s})d\mathbf{s}/||\mathbf{x}||^{p-1}$ for $\mathbf{x}, \mathbf{h} \in \mathcal{X}$.

The norm in $\mathcal{X}$ is said to be twice Gâteaux differentiable at $\mathbf{x} \neq \mathbf{0}$ with Hessian (or second order Gâteaux derivative) $\mathsf{H}_{\mathbf{x}}$, a continuous linear map from $\mathcal{X}$ to $\mathcal{X}^*$, if

$$\lim_{t \to 0} t^{-1}(S_{\mathbf{x}+t\mathbf{h}} - S_{\mathbf{x}}) = \mathsf{H}_{\mathbf{x}}(\mathbf{h})$$

for all $\mathbf{h} \in \mathcal{X}$. Here, the limit is assumed to exist in the norm topology of $\mathcal{X}^*$. In such cases, the following Taylor expansion holds for each $\mathbf{h} \in \mathcal{X}$.

$$||\mathbf{x} + t\mathbf{h}|| = ||\mathbf{x}|| + tS_{\mathbf{x}} + \frac{1}{2}t^2\{\mathsf{H}_{\mathbf{x}}(\mathbf{h})\}(\mathbf{h}) + \mathbf{r}(t),$$

where $\mathbf{r}(t)/t^2 \to 0$ as $t \to 0$. Norms in Hilbert spaces and L_p spaces for $p \in [2, \infty)$ are twice Gâteaux differentiable. For a Hilbert space $\mathcal{X}$ with inner product $\langle \cdot, \cdot \rangle$, $\mathsf{H}_{\mathbf{x}} = ||\mathbf{x}||^{-1}\{\mathsf{I} - (\mathbf{x} \otimes \mathbf{x})/||\mathbf{x}||^2\}$ for any $\mathbf{x} \in \mathcal{X}$. Here, I is the identity operator on $\mathcal{X}$, and $\otimes$ denotes the tensor product on $\mathcal{X}$. The latter is defined as $\mathbf{x} \otimes \mathbf{x} : \mathcal{X} \to \mathcal{X}$, where $\langle (\mathbf{x} \otimes \mathbf{x})(\mathbf{z}), \mathbf{w} \rangle = \langle \mathbf{x}, \mathbf{z} \rangle \langle \mathbf{x}, \mathbf{w} \rangle$. If $\mathcal{X} = L_p(\mathbb{R}^d)$, then

$$\begin{aligned} \{\mathsf{H}_{\mathbf{x}}(\mathbf{z})\}(\mathbf{w}) &= (p-1)E\left[\frac{\int_{\mathbb{R}^d} |\mathbf{x}(\mathbf{s})|^{p-2}\mathbf{z}(\mathbf{s})\mathbf{w}(\mathbf{s})ds}{||\mathbf{x}||^{p-1}}\right. \\ &\quad - \left.\frac{\left\{\int_{\mathbb{R}^d} |\mathbf{x}(\mathbf{s})|^{p-1}\mathbf{z}(\mathbf{s})ds\right\}\left\{\int_{\mathbb{R}^d} |\mathbf{x}(\mathbf{s})|^{p-1}\mathbf{w}(\mathbf{s})ds\right\}}{||\mathbf{x}||^{2p-1}}\right], \end{aligned}$$

for $\mathbf{x}, \mathbf{z}, \mathbf{w} \in \mathcal{X}$. We refer to Chapters 4 and 5 in Borwein and Vanderwerff (2010) for further details.

For any Banach space $\mathcal{X}$, a sequence $\mathbf{x}_n$ in $\mathcal{X}$ is said to converge in the weak topology of $\mathcal{X}$ (or converges weakly) to $\mathbf{x} \in \mathcal{X}$ iff $\mathbf{y}(\mathbf{x}_n) \to \mathbf{y}(\mathbf{x})$ for each $\mathbf{y} \in \mathcal{X}^*$. In other words, the weak topology on $\mathcal{X}$ is the smallest topology with respect to which all elements in $\mathcal{X}^*$ are continuous.

A Banach space $\mathcal{X}$ is said to be strictly convex if for any $\mathbf{x} \neq \mathbf{y} \in \mathcal{X}$ satisfying $||\mathbf{x}|| = ||\mathbf{y}|| = 1$, we have $||(\mathbf{x}+\mathbf{y})/2|| < 1$ (see, e.g., Borwein and Vanderwerff (2010)). Hilbert spaces and L_p spaces for $p \in (1, \infty)$ are strictly convex.

The norm in a Banach space $\mathcal{X}$ is said to be locally uniformly rotund if for any sequence $\{\mathbf{x}_n\}_{n \geq 1} \in \mathcal{X}$ and any $\mathbf{x} \in \mathcal{X}$ satisfying $||\mathbf{x}_n|| = ||\mathbf{x}|| = 1$ for all $n \geq 1$, $\lim_{n \to \infty} ||\mathbf{x}_n + \mathbf{x}|| = 2$ implies $\lim_{n \to \infty} ||\mathbf{x}_n - \mathbf{x}|| = 0$ (see, e.g., Borwein and Vanderwerff

(2010)). The norm in any Hilbert space or any L_p space for $p \in (1, \infty)$ is locally uniformly rotund. It is easy to verify that if the norm in a Banach space is locally uniformly rotund, then the Banach space is strictly convex (see, e.g., Borwein and Vanderwerff (2010)).

A separable Banach space $\mathcal{X}$ is said to have a Schauder basis $\{\phi_1, \phi_2, \ldots\}$ if for any $\mathbf{x} \in \mathcal{X}$, there exists a unique sequence of real numbers $\{x_k\}_{k\geq 1}$ such that $\mathbf{x} = \sum_{k=1}^{\infty} x_k \phi_k$ (see, e.g., Fabian et al. (2001)). If $\mathcal{X}$ is a separable Hilbert space and $\{\phi_1, \phi_2, \ldots\}$ is an orthonormal basis of $\mathcal{X}$, then it is a Schauder basis of $\mathcal{X}$. Further, any L_p space for $p \in (1, \infty)$ admits a Schauder basis.

A continuous functional $\mathsf{T} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is symmetric, nonnegative definite and bilinear if it satisfies

$$
\begin{aligned}
&(i) && \mathsf{T}(\mathbf{x}, \mathbf{y}) = \mathsf{T}(\mathbf{y}, \mathbf{x}),\\
&(ii) && \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \mathsf{T}(\mathbf{x}_i, \mathbf{x}_j) \geq 0,\\
&(iii) && \mathsf{T}(a\mathbf{x} + \mathbf{y}, \mathbf{z}) = a\mathsf{T}(\mathbf{x}, \mathbf{z}) + \mathsf{T}(\mathbf{y}, \mathbf{z}),\ \ \mathsf{T}(\mathbf{x}, a\mathbf{y} + \mathbf{z}) = a\mathsf{T}(\mathbf{x}, \mathbf{y}) + \mathsf{T}(\mathbf{x}, \mathbf{z}),
\end{aligned}
$$

for every $n \geq 1$, $a, a_1, \ldots, a_n \in \mathbb{R}$, and $\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathcal{X}$. Associated with such a functional T is a unique continuous linear symmetric nonnegative operator $\widetilde{\mathsf{T}} : \mathcal{X} \to \mathcal{X}^*$ defined as $\{\widetilde{\mathsf{T}}(\mathbf{x})\}(\mathbf{y}) = \mathsf{T}(\mathbf{x}, \mathbf{y})$, where $\mathbf{x}, \mathbf{y} \in \mathcal{X}$. Note that $\widetilde{\mathsf{T}}$ is symmetric because $\{\widetilde{\mathsf{T}}(\mathbf{x})\}(\mathbf{y}) = \{\widetilde{\mathsf{T}}(\mathbf{y})\}(\mathbf{x})$, and it is nonnegative since $\{\widetilde{\mathsf{T}}(\mathbf{x})\}(\mathbf{x}) \geq 0$ for all $\mathbf{x}, \mathbf{y} \in \mathcal{X}$.

Let $\mathcal{X}$ and $\mathcal{Y}$ be two Banach spaces, and $\mathsf{T} : \mathcal{X} \to \mathcal{Y}$ be a continuous linear operator. Then, the adjoint operator $\mathsf{T}^* : \mathcal{Y}^* \to \mathcal{X}^*$ of T is defined by the equation $\{\mathsf{T}^*(\mathbf{y}^*)\}(\mathbf{x}) = \mathbf{y}^*\{\mathsf{T}(\mathbf{x})\}$, where $\mathbf{x} \in \mathcal{X}$ and $\mathbf{y}^* \in \mathcal{Y}^*$.

A Banach space $\mathcal{X}$ is said to be of type 2 if there exists a constant $b > 0$ such that for any $n \geq 1$ and independent zero mean random elements $\mathbf{U}_1, \ldots, \mathbf{U}_n$ in $\mathcal{X}$ satisfying $E(||\mathbf{U}_i||^2) < \infty$ for all $i = 1, \ldots, n$, we have

$$E(||\sum_{i=1}^{n} \mathbf{U}_i||^2) \leq b \sum_{i=1}^{n} E(||\mathbf{U}_i||^2).$$

Type 2 Banach spaces are the only Banach spaces where any sum of i.i.d. random elements in that space, whose norm has finite second moment, will satisfy the central limit theorem. Hilbert spaces and L_p spaces with $p \in [2, \infty)$ are examples of type 2 Banach spaces. We refer to Section 7 of Chapter 3 in Araujo and Giné (1980) for further details.

A Banach space $\mathcal{X}$ is said to be p-uniformly smooth for some $p \in (1, 2]$ if for every $q \geq 1$ there exists a constant $\alpha_q > 0$ such that for any zero mean martingale sequence $(\mathbf{M}_m, \mathcal{G}_m)_{m \geq 1}$ in $\mathcal{X}$, we have $E(||\mathbf{M}_m||^q) \leq \alpha_q \sum_{i=1}^{m} E(||\mathbf{M}_i - \mathbf{M}_{i-1}||^p)^{q/p}$. Here, the sequence $(\mathbf{M}_m)_{m \geq 1}$ is adapted to the filtration $(\mathcal{G}_m)_{m \geq 1}$. Any 2-uniformly smooth Banach space is of type 2. Hilbert spaces are 2-uniformly smooth, and L_p spaces are $\widetilde{p}$-uniformly smooth, where $\widetilde{p} = \min(p, 2)$ for $p \in (1, \infty)$. We refer to Borovskikh (1996) for further details.

Let $\mathbf{X}$ be a random element in a separable Banach space $\mathcal{X}$. Suppose that there exists a sequence $\{\mathbf{X}_n\}_{n \geq 1}$ of measurable simple functions in $\mathcal{X}$ such that $\mathbf{X}_n \to \mathbf{X}$ almost surely and $E(||\mathbf{X}_n - \mathbf{X}||) \to 0$ as $n \to \infty$. Then, we say that $\mathbf{X}$ has an expectation in the Bochner sense, and the Bochner integral of $\mathbf{X}$ equals $\lim_{n \to \infty} E(\mathbf{X}_n)$. Here, $E(\mathbf{X}_n) = \sum_{k=1}^{M_n} \mathrm{x}_{nk} P(A_{nk})$ if $\mathbf{X}_n = \sum_{k=1}^{M_n} \mathrm{x}_{nk} I(A_{nk})$. It is known that the Bochner expectation of $\mathbf{X}$ exists if $E(||\mathbf{X}||) < \infty$. We refer to Section 2 in Chapter 3 of Araujo and Giné (1980) for more details. Further, for two random elements $\mathbf{X}$ and $\mathbf{Y}$ with finite Bochner expectations, the conditional expectation of $\mathbf{X}$ given $\mathbf{Y}$ exists and can be properly defined (see Section 4 in Chapter II of Vakhania et al. (1987)).

By a Gaussian random element $\mathbf{W}$ in a separable Banach space $\mathcal{X}$ with mean $\mathbf{m} \in \mathcal{X}$ and covariance C, we mean that for all $\mathbf{l} \in \mathcal{X}^*$, $\mathbf{l}(\mathbf{W})$ has a Gaussian distribution on $\mathbb{R}$ with mean $\mathbf{l}(\mathbf{m})$ and variance $\{\mathsf{C}(\mathbf{l})\}(\mathbf{l})$ (see Section 2.4 in Chapter IV of Vakhania et al. (1987)). Here, $\mathsf{C} : \mathcal{X}^* \to \mathcal{X}$ is a continuous linear symmetric positive operator.

Bibliography

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

Asplund, E. (1968). Fréchet differentiability of convex functions. *Acta Mathematica,* 121:31–47.

Bai, Z. and Saranadasa, H. (1996). Effect of high dimension: by an example of a two sample problem. *Statistica Sinica,* 6(2):311–329.

Borovskikh, Y. V. (1996). *U-statistics in Banach spaces.* VSP, Utrecht.

Borwein, J. M. and Vanderwerff, J. D. (2010). *Convex functions: constructions, characterizations and counterexamples,* volume 109 of *Encyclopedia of Mathematics and its Applications.* Cambridge University Press, Cambridge.

Böttcher, B. (2010). Feller processes: The next generation in modeling. Brownian motion, Lévy processes and beyond. *PLoS ONE,* 5(12):e15102.

Boyd, J. P. (1984). Asymptotic coefficients of Hermite function series. *Journal of Computational Physics,* 54(3):382–410.

Brown, B. M. (1983). Statistical uses of the spatial median. *Journal of the Royal Statistical Society. Series B. Methodological,* 45(1):25–30.

Bugni, F. A., Hall, P., Horowitz, J. L., and Neumann, G. R. (2009). Goodness-of-fit tests for functional data. *The Econometrics Journal,* 12(S1):S1–S18.

Cadre, B. (2001). Convergent estimators for the L_1-median of a Banach valued random variable. *Statistics. A Journal of Theoretical and Applied Statistics*, 35(4):509–521.

Cardot, H., Cénac, P., and Zitt, P.-A. (2013). Efficient and fast estimation of the geometric median in hilbert spaces with an averaged stochastic gradient algorithm. *Bernoulli*, 19(1):18–43.

Chakraborty, A. and Chaudhuri, P. (2014a). The deepest point for distributions in infinite dimensional spaces. *Statistical Methodology, (Special Issue: Re-sampling and Contemporary Inference – A tribute to Kesar Singh)*, 20:27–39.

Chakraborty, A. and Chaudhuri, P. (2014b). On data depth in infinite dimensional spaces. *Annals of the Institute of Statistical Mathematics*, 66(2):303–324.

Chakraborty, A. and Chaudhuri, P. (2014c). The spatial distribution in infinite dimensional spaces and related quantiles and depths. *The Annals of Statistics*, 42(3):1203–1231.

Chakraborty, A. and Chaudhuri, P. (2014d). A Wilcoxon-Mann-Whitney type test for infinite dimensional data. Accepted for publication in *Biometrika.*

Chakraborty, B. (2001). On affine equivariant multivariate quantiles. *Annals of the Institute of Statistical Mathematics*, 53(2):380–403.

Chakraborty, B. and Chaudhuri, P. (1999). On affine invariant sign and rank tests in one- and two-sample multivariate problems. In *Multivariate analysis, design of experiments, and survey sampling*, volume 159 of *Statist. Textbooks Monogr.*, pages 499–522. Dekker, New York.

Chaouch, M. and Goga, C. (2012). Using complex surveys to estimate the l_1-median of a functional variable: application to electricity load curves. *International Statistical Review*, 80(1):40–59.

Chaudhuri, P. (1996). On a geometric notion of quantiles for multivariate data. *Journal of the American Statistical Association*, 91(434):862–872.

Chen, S. X. and Qin, Y.-L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *The Annals of Statistics*, 38(2):808–835.

Choi, K. and Marden, J. (1997). An approach to multivariate rank tests in multivariate analysis of variance. *Journal of the American Statistical Association*, 92(440):1581–1590.

Cox, D. D. and Lee, J. S. (2008). Pointwise testing with functional data using the Westfall-Young randomization method. *Biometrika*, 95(3):621–634.

Cuesta-Albertos, J. A. and Febrero-Bande, M. (2010). A simple multiway ANOVA for functional data. *TEST*, 19(3):537–557.

Cuesta-Albertos, J. A. and Nieto-Reyes, A. (2008). The random Tukey depth. *Computational Statistics & Data Analysis*, 52(11):4979–4988.

Cuevas, A., Febrero, M., and Fraiman, R. (2004). An anova test for functional data. *Computational Statistics & Data Analysis*, 47(1):111–122.

Cuevas, A. and Fraiman, R. (2009). On depth measures and dual statistics. A methodology for dealing with general data. *Journal of Multivariate Analysis*, 100(4):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(4):1803–1827.

Dutta, S. and Ghosh, A. K. (2012). On robust classification using projection depth. *Annals of the Institute of Statistical Mathematics*, 64(3):657–676.

Dutta, S., Ghosh, A. K., and Chaudhuri, P. (2011). Some intriguing properties of Tukey's half-space depth. *Bernoulli*, 17(4):1420–1434.

Fabian, M., Habala, P., Hájek, P., Montesinos Santalucía, V., Pelant, J., and Zizler, V. (2001). *Functional analysis and infinite-dimensional geometry*. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC, 8. Springer-Verlag, New York.

Fan, J. and Lin, S.-K. (1998). Test of significance when data are curves. *Journal of the American Statistical Association*, 93(443):1007–1021.

Fraiman, R. and Muniz, G. (2001). Trimmed means for functional data. *Test*, 10(2):419–440.

Fraiman, R. and Pateiro-López, B. (2012). Quantiles for finite and infinite dimensional data. *Journal of Multivariate Analysis*, 108:1–14.

Gervini, D. (2008). Robust functional estimation using the median and spherical principal components. *Biometrika*, 95(3):587–600.

Ghosh, A. K. and Chaudhuri, P. (2005). On data depth and distribution-free discriminant analysis using separating surfaces. *Bernoulli*, 11(1):1–27.

Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B., and Smola, A. (2012). A kernel two-sample test. *Journal of Machine Learning Research (JMLR)*, 13:723–773.

Guasoni, P. (2006). No arbitrage under transaction costs, with fractional Brownian motion and beyond. *Mathematical Finance. An International Journal of Mathematics, Statistics and Financial Economics*, 16(3):569–582.

Hájek, J., Šidák, Z., and Sen, P. K. (1999). *Theory of rank tests*. Probability and Mathematical Statistics. Academic Press Inc., San Diego, CA, second edition.

Hall, P. and Tajvidi, N. (2002). Permutation tests for equality of distributions in high-dimensional settings. *Biometrika*, 89(2):359–374.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (1986). *Robust statistics*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York. The approach based on influence functions.

Hansen, B. E. (1991). Strong laws for dependent heterogeneous processes. *Econometric Theory*, 7(2):213–221.

He, X. and Wang, G. (1997). Convergence of depth contours for multivariate datasets. *The Annals of Statistics*, 25(2):495–504.

Hettmansperger, T. P. and McKean, J. W. (2011). *Robust nonparametric statistical methods*, volume 119 of *Monographs on Statistics and Applied Probability*. CRC Press, Boca Raton, FL, second edition.

Horváth, L. and Kokoszka, P. (2012). *Inference for functional data with applications*. Springer Series in Statistics. Springer, New York.

Horváth, L., Kokoszka, P., and Reeder, R. (2013). Estimation of the mean of functional time series and a two-sample problem. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, 75(1):103–122.

Jörnsten, R. (2004). Clustering and classification based on the L_1 data depth. *Journal of Multivariate Analysis*, 90(1):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)*, pages 217–230. North-Holland, Amsterdam.

Kolmogorov, A. N. and Tihomirov, V. M. (1961). ε-entropy and ε-capacity of sets in functional space. *American Mathematical Society Translations*, 17:277–364.

Koltchinskii, V. I. (1997). M-estimation, convexity and quantiles. *The Annals of Statistics*, 25(2):435–477.

Kong, L. and Mizera, I. (2012). Quantile tomography: using quantiles with multivariate data. *Statistica Sinica*, 22(4):1589–1610.

Kosorok, M. R. and Ma, S. (2007). Marginal asymptotics for the "large p, small n" paradigm: with applications to microarray data. *The Annals of Statistics*, 35(4):1456–1486.

Kuelbs, J. and Zinn, J. (2012). Half-region depth for stochastic processes. Manuscript submitted.

Lehmann, E. L. (1975). *Nonparametrics: statistical methods based on ranks*. Holden-Day Inc., San Francisco, Calif. With the special assistance of H. J. M. d'Abrera, Holden-Day Series in Probability and Statistics.

Li, J., Cuesta-Albertos, J. A., and Liu, R. Y. (2012). DD-classifier: Nonparametric classification procedure based on DD-plot. *Journal of the American Statistical Association*, 107(498):737–753.

Liu, R. Y. (1990). On a notion of data depth based on random simplices. *The Annals of Statistics*, 18(1):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(3):783–858. With discussion and a rejoinder by Liu and Singh.

Liu, R. Y. and Singh, K. (1993). A quality index based on data depth and multivariate rank tests. *Journal of the American Statistical Association*, 88(421):252–260.

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

López-Pintado, S. and Romo, J. (2009). On the concept of depth for functional data. *Journal of the American Statistical Association*, 104(486):718–734.

López-Pintado, S. and Romo, J. (2011). A half-region depth for functional data. *Computational Statistics & Data Analysis*, 55(4):1679–1695.

Mosler, K. (2002). *Multivariate dispersion, central regions and depth*, volume 165 of *Lecture Notes in Statistics*. Springer-Verlag, Berlin. The lift zonoid approach.

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

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

Möttönen, J., Oja, H., and Tienari, J. (1997). On the efficiency of multivariate spatial sign and rank tests. *The Annals of Statistics*, 25(2):542–552.

Oja, H. (1983). Descriptive statistics for multivariate distributions. *Statistics & Probability Letters*, 1(6):327–332.

Oja, H. (1999). Affine invariant multivariate sign and rank tests and corresponding estimates: a review. *Scandinavian Journal of Statistics. Theory and Applications*, 26(3):319–343.

Oja, H. (2010). *Multivariate nonparametric methods with R*, volume 199 of *Lecture Notes in Statistics*. Springer, New York. An approach based on spatial signs and ranks.

Puri, M. L. and Sen, P. K. (1971). *Nonparametric methods in multivariate analysis*. John Wiley & Sons, Inc., New York-London-Sydney.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional data analysis*. Springer Series in Statistics. Springer, New York, second edition.

Randles, R. H. and Peters, D. (1990). Multivariate rank tests for the two-sample location problem. *Communications in Statistics. Theory and Methods*, 19(11):4225–4238 (1991).

Rasmussen, C. E. and Williams, C. K. I. (2006). *Gaussian processes for machine learning*. Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA.

Revuz, D. and Yor, M. (1991). *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin.

Rosenbaum, P. R. (2005). An exact distribution-free test comparing two multivariate distributions based on adjacency. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, 67(4):515–530.

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., pages 25–38. Birkhäuser, Basel.

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

Shen, Q. and Faraway, J. (2004). An F test for linear models with functional responses. *Statistica Sinica*, 14(4):1239–1257.

Small, C. G. (1990). A survey of multidimensional medians. *International Statistical Review*, 58(3):263–277.

Srivastava, M. S., Katayama, S., and Kano, Y. (2013). A two sample test in high dimensional data. *Journal of Multivariate Analysis*, 114:349–358.

Sun, Y. and Genton, M. G. (2011). Functional boxplots. *Journal of Computational and Graphical Statistics*, 20(2):316–334.

Trefethen, L. N. (2008). Is Gauss quadrature better than Clenshaw-Curtis? *SIAM Review*, 50(1):67–87.

Vakhania, N. N., Tarieladze, V. I., and Chobanyan, S. A. (1987). *Probability distributions on Banach spaces*, volume 14 of *Mathematics and its Applications (Soviet Series)*. D. Reidel Publishing Co., Dordrecht. Translated from the Russian and with a preface by Wojbor A. Woyczynski.

Valadier, M. (1984). La multi-application médianes conditionnelles. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 67(3):279–282.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes*. Springer Series in Statistics. Springer-Verlag, New York. With applications to statistics.

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(4):1423–1426 (electronic).

Vretblad, A. (2003). *Fourier analysis and its applications*, volume 223 of *Graduate Texts in Mathematics*. Springer-Verlag, New York.

Wang, H. and Xiang, S. (2012). On the convergence rates of Legendre approximation. *Mathematics of Computation*, 81(278):861–877.

Yu, S., Tresp, V., and Yu, K. (2007). Robust multi-task learning with t-processes. In *Proceedings of the 24th International Conference on Machine Learning (Oregon, 2007)*, pages 1103–1110. ACM New York, NY, USA.

Yurinskiĭ, V. V. (1976). Exponential inequalities for sums of random vectors. *Journal of Multivariate Analysis*, 6(4):473–499.

Zhang, C., Peng, H., and Zhang, J.-T. (2010). Two samples tests for functional data. *Communications in Statistics. Theory and Methods*, 39(4):559–578.

Zhang, J.-T. and Chen, J. (2007). Statistical inferences for functional data. *The Annals of Statistics*, 35(3):1052–1079.

Zuo, Y. (2006). Multidimensional trimming based on projection depth. *The Annals of Statistics*, 34(5):2211–2251.

Zuo, Y. and Serfling, R. (2000). General notions of statistical depth function. *The Annals of Statistics*, 28(2):461–482.