The Spatial Distribution in infinite dimensional spaces and related Quantiles and Depths

Anirvan Chakraborty and Probal Chaudhuri

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

Abstract

The spatial distribution has been widely used to develop various nonparametric procedures for finite dimensional multivariate data. In this paper, we investigate the concept of spatial distribution for data in infinite dimensional Banach spaces. Many technical difficulties are encountered in such spaces that are primarily due to the non-compactness of the closed unit ball. In this work, we prove some Glivenko-Cantelli and Donsker type results for the empirical spatial distribution process in infinite dimensional spaces. The spatial quantiles in such spaces can be obtained by inverting the spatial distribution function. A Bahadur type asymptotic linear representation and the associated weak convergence results for the sample spatial quantiles in infinite dimensional spaces are derived. A study of the asymptotic efficiency of the sample spatial median relative to the sample mean is carried out for some standard probability distributions in function spaces. The spatial distribution can be used to define the spatial depth in infinite dimensional Banach spaces, and we study the asymptotic properties of the empirical spatial depth in such spaces. We also demonstrate the spatial quantiles and the spatial depth using some real and simulated functional data.

Keywords: Asymptotic relative efficiency, Bahadur representation, DD-plot, Donsker property, Gâteaux derivative, Glivenko-Cantelli property, Karhunen-Loève expansion, smooth Banach space.

1 Introduction

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., [13], [22], [27] and [30]). 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., [4], [12], [20], [26] and [29]). Nowadays, we often come across data, which are curves or functions and can be modeled as random observations from probability distributions in infinite dimensional spaces. The ECG curves of patients observed over a period of time, the spectrometry curves recorded for a range of wavelengths, the annual temperature curves of different places etc. are examples of such data. Many of the function spaces, where such data lie, are infinite dimensional Banach spaces. However, many of the well-known multivariate medians like the simplicial depth median (see [22]), and the simplicial volume median (see [27]) do not have meaningful extensions into such spaces. On the other hand, the spatial median as well as the spatial quantiles extend easily into infinite dimensional Banach spaces (see [12], [18] and [33]). The author of [17] proposed functional principal components using the sample spatial median and used those to analyze a data involving the movements of the lips. The authors of [8] 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. The spatial median has also been used in [11] to calculate the median profile for the electricity load data in France. Recently, the authors of [15] 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 in [21]. The principle quantile directions derived from these quantiles were used in [15] to detect outliers in a dataset of annual age-specific mortality rates of French males between the years 1899 and 2005.

The purpose of this article is to investigate the spatial distribution in infinite dimensional Banach spaces, and study their properties along with the spatial quantiles and the spatial depth. There are several mathematical difficulties in dealing with the probability distributions in such spaces. These are primarily due to the non-compactness of the closed unit ball in such spaces. In section 2, we prove some Glivenko-Cantelli and Donsker type results for the empirical spatial distribution process arising from data lying in infinite dimensional spaces. In section 3, we investigate the spatial quantiles in infinite dimensional spaces. A Bahadur type linear representation of the sample spatial quantiles and their 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. In section 4, we investigate the spatial depth and its asymptotic properties in infinite dimensional spaces. We also demonstrate how exploratory data analytic tools like the depth-depth plot (DD-plot) (see [23]) can be developed for data in infinite dimensional spaces using the spatial depth. The proofs of the theorems are given in the Appendix.

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

The spatial distribution was studied in the finite dimensional setup, and Glivenko-Cantelli and Donsker type results were obtained for the empirical spatial distribution process (see [20]). 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 in [6]. But the authors of [6] 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, i.e., the norm function in $\mathcal{X}$ is Gâteaux differentiable at each non-zero $\mathbf{x} \in \mathcal{X}$ with derivative, say, $SGN_{\mathbf{x}} \in \mathcal{X}^*$ (see, e.g., [3]). Here, $\mathcal{X}^*$ is the dual space of $\mathcal{X}$, i.e., the Banach space of all continuous real-valued linear functions on $\mathcal{X}$. Thus, $SGN_{\mathbf{x}}(\mathbf{h}) = \lim_{t \to 0} t^{-1}(||\mathbf{x} + t\mathbf{h}|| - ||\mathbf{x}||)$ for $\mathbf{h} \in \mathcal{X}$. If this limit is uniform over the set $\{\mathbf{h} \in \mathcal{X} : ||\mathbf{h}|| = 1\}$, then the norm is said to be Fréchet differentiable. If $\mathcal{X}$ is a Hilbert space, $SGN_{\mathbf{x}} = \mathbf{x}/||\mathbf{x}||$. When $\mathcal{X} = L_p[a,b]$ for some $p \in (1,\infty)$, which is the Banach space of all functions $\mathbf{x} : [a,b] \to \mathbb{R}$ satisfying $\int_a^b |\mathbf{x}(s)|^p ds < \infty$, then $SGN_{\mathbf{x}}(\mathbf{h}) = \int_a^b sign\{\mathbf{x}(s)\}|\mathbf{x}(s)|^{p-1}\mathbf{h}(s)ds/||\mathbf{x}||^{p-1}$ for all $\mathbf{h} \in \mathcal{X}$. The norm in any Hilbert space as well as $L_p[a,b]$ for a $p \in (1,\infty)$ is actually Fréchet differentiable. As a convention, we define $SGN_{\mathbf{x}} = \mathbf{0}$ 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 $S_{\mathbf{x}} = E\{SGN_{\mathbf{x}-\mathbf{X}}\}$. Throughout this article, the expectations of Banach valued random variables will be defined in the Bochner sense (see, e.g., [1, p. 100]). The empirical spatial distribution can be defined as $\widehat{S}_{\mathbf{x}} = n^{-1}\sum_{i=1}^n SGN_{\mathbf{x}-\mathbf{X}_i}$, 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 has been used to develop Wilcoxon–Mann–Whitney type tests for two sample problems in infinite dimensional spaces spaces (see [10]).

Associated with the spatial distribution is the corresponding empirical spatial distribution process $\{\widehat{S}_{\mathbf{x}} - S_{\mathbf{x}} : \mathbf{x} \in I\}$ 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 [20]). The following theorem states a Glivenko-Cantelli and a Donsker type result for the empirical spatial distribution process in a separable Hilbert space.

Theorem 2.1. *Let $\mathcal{X}$ be a separable Hilbert space, and $\mathcal{Z}$ be a finite dimensional subspace of $\mathcal{X}$.*

Then, $\widehat{S}_{\mathbf{x}}$ converges to $S_{\mathbf{x}}$ uniformly in $\mathcal{Z}$ in the weak topology of $\mathcal{X}$ almost surely. Further, if μ is non-atomic, then for any $d \geq 1$ and any continuous linear map $\mathbf{g} : \mathcal{X} \to \mathbb{R}^d$, the process $\{\mathbf{g}(\sqrt{n}(\widehat{S}_{\mathbf{x}} - S_{\mathbf{x}})) : \mathbf{x} \in \mathcal{Z}\}$ converges weakly to a d-variate Gaussian process on $\mathcal{Z}$.

The Glivenko-Cantelli and the Donsker type results in [20] for the empirical spatial distribution process in $\mathbb{R}^d$ follow from the above theorem as a straightforward corollary. The result stated in Theorem 2.1 is also true in Banach spaces like L_p spaces for some even integer $p > 2$ (see the remark after the proof of Theorem 2.1 given in the Appendix).

A probability measure in an infinite dimensional separable Banach space $\mathcal{X}$ (e.g., a non-degenerate 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., [1]). 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 $\mathcal{X}$, and for all $\mathbf{x} \neq \mathbf{0}, -\mathbf{h}$, we have $||SGN_{\mathbf{x}+\mathbf{h}} - SGN_{\mathbf{x}}|| \leq T(\mathbf{x})||\mathbf{h}||$.*

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 2.2. *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 non-atomic part of μ. Then, $\widehat{S}_{\mathbf{x}}$ converges to $S_{\mathbf{x}}$ uniformly over $\mathbf{x} \in K$ in the norm topology of $\mathcal{X}^*$ almost surely.
(b) Let μ be a non-atomic probability measure, Assumption (A) hold, and $\sup_{||\mathbf{x}|| \leq C} E_{\mu}\{T^2(\mathbf{x} - \mathbf{X})\} < \infty$ for each $C > 0$. If $\int_0^1 \sqrt{ln\ N(\varepsilon, K)}d\varepsilon < \infty$ for each $\varepsilon > 0$, then the empirical process $\widehat{\mathbf{S}}_{\mathbf{g}} = \{\mathbf{g}(\sqrt{n}(\widehat{S}_{\mathbf{x}} - S_{\mathbf{x}})) : \mathbf{x} \in K\}$ 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 $\mathbf{g} : \mathcal{X} \to \mathbb{R}^d$, $\widehat{\mathbf{S}}_{\mathbf{g}}$ 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 5.1 in Appendix). It follows from part (a) of the above theorem and the tightness of any probability measure in any complete separable metric space that $\int_{\mathcal{X}} ||\widehat{S}_{\mathbf{x}} - S_{\mathbf{x}}||^2 \mu(d\mathbf{x}) \to 0$ 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 $\sup_{\mathbf{x} \in K} ||\widehat{S}_{\mathbf{x}} - S_{\mathbf{x}}|| = O_P(1/\sqrt{n})$

and $\int_{\mathcal{X}} ||\widehat{S}_{\mathbf{x}} - S_{\mathbf{x}}||^2 \mu(d\mathbf{x}) = O_P(1/n)$ 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 $J = \int_0^1 \sqrt{ln\ N(\varepsilon, K)} d\varepsilon$. It is easy to verify that $J < \infty$ for every compact set K in any finite dimensional Banach space. The finiteness of J is also true for many compact sets in various infinite dimensional function spaces like the compact sets in L_p spaces for $p \in (1, \infty)$ whose elements have continuous partial derivatives up to order $r-1$ for some $r \geq 1$ and the r-th order partial derivatives are Holder continuous with a positive exponent (see, e.g., [19]).

3 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 non-atomic 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 [20]). 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. 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., [3]). Hilbert spaces and L_p spaces for $p \in (1, \infty)$ are strictly convex.

Theorem 3.1. *Let $\mathcal{X}$ be a smooth, strictly convex Banach space, and suppose that μ is non-atomic probability measure in $\mathcal{X}$. If μ is not entirely supported on a line in $\mathcal{X}$, the spatial distribution map $\mathbf{x} \mapsto S_{\mathbf{x}}$ is strictly monotone, i.e., $(S_{\mathbf{x}} - S_{\mathbf{y}})(\mathbf{x} - \mathbf{y}) > 0$ 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 3.1, 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 $E\{SGN_{\mathbf{Q}-\mathbf{X}}\} = \mathbf{u}$. 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 $\mathcal{X} \backslash A_\mu$ to $\mathcal{B}^*(\mathbf{0}, 1) \backslash \bigcup_{\mathbf{x} \in A_\mu} \overline{\mathcal{B}}^*(S_{\mathbf{x}}, p(\mathbf{x}))$ if the other conditions in Theorem 3.1 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 $\mathbf{Q} \in \mathcal{X}$, which is the usual definition of spatial quantiles (see [12] and [20]).

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, i.e., $||A(\mathbf{x})|| = ||\mathbf{x}||$ 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, e.g., [14]). 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 intractable in general. The author of [7] 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. The author of [17] 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 $\mathbf{x} = \sum_{k=1}^{\infty} x_k \phi_k$ (see, e.g., [14]). Note that if $\mathcal{X}$ is a Hilbert space and $\{\phi_1, \phi_2, \ldots\}$ is an orthonormal basis of $\mathcal{X}$, then it is a Schauder basis of $\mathcal{X}$. 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., [14, 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 $\mathbf{Q} \in \mathcal{Z}_n$.

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. We have first plotted the population spatial quantiles of the standard Brownian motion for $\mathbf{u} = \pm c\phi_k$, where $k = 1, 2, 3$ and $c = 0.25, 0.5, 0.75$ (see Figure 1). 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 motion, and each sample curve is observed at 250 equispaced points in $[0, 1]$. The real dataset considered here is available at `http://www.math.univ-toulouse.fr/∼staph/npfda/`, and it contains the spectrometric curves of $n = 215$ meat units measured at 100 wavelengths in the range 850 nm to 1050 nm along with the fat content of each unit categorized into two classes, namely, below and above 20%. The sample curves of the real data may be viewed as elements in $L_2[850, 1050]$ equipped with its usual norm. For each of the simulated and the real dataset, we have chosen $d(n) = [\sqrt{n}]$, 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 2 (respectively, Figure 3) shows the plots of the simulated dataset (respectively, real dataset) along with the sample spatial median and the sample spatial quantiles corresponding to $\mathbf{u} = \pm c\widehat{\phi}_k$ for $k = 1, 2, 3$ ($k = 1, 2$), 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 $k \geq 1$. 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 earlier. 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.

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 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 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). Moreover, the sample spatial quantiles of the two classes in the real dataset are different in their shapes.

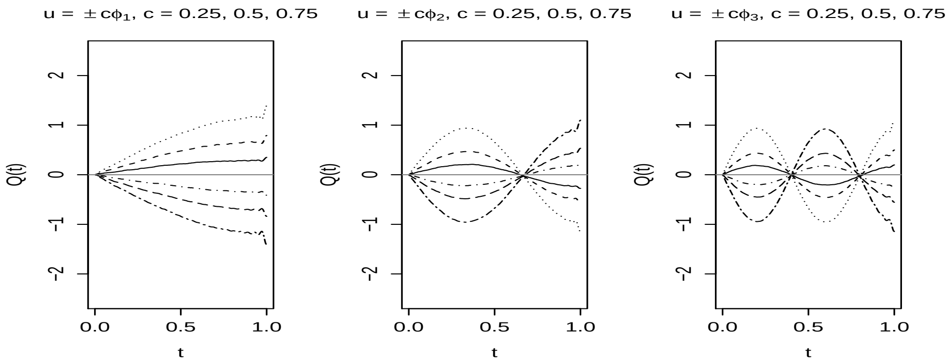

Figure 1: *The plots of the spatial quantiles of the standard Brownian motion including the spatial median (horizontal line through zero in all the plots). For each* $k = 1, 2, 3$*, the spatial quantiles corresponding to* $\mathbf{u} = c\phi_k$ *for* $c = 0.25$*,* 0.5 *and* 0.75 *are given by the solid (—), the dashed (- - -) and the dotted (· · ·) curves, respectively, while those corresponding to* $\mathbf{u} = -c\phi_k$ *for these c values are given by the dot-dashed (– · –), the long-dashed (– –) and the two-dashed (– - –) curves, respectively.*

3.1 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. 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., [3]). The norm in any Hilbert space or any L_p space for $p \in (1, \infty)$ is locally uniformly rotund.

Theorem 3.1.1. *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 non-atomic 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., [12] and [20]), 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 $J_{\mathbf{Q}}$, which is a symmetric bounded bilinear function from $\mathcal{X} \times \mathcal{X}$ into $\mathbb{R}$ satisfying

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

for any $\mathbf{h} \in \mathcal{X}$. We define the continuous linear operator $\widetilde{J}_{\mathbf{Q}} : \mathcal{X} \to \mathcal{X}$ associated with $J_{\mathbf{Q}}$ by the equation $\langle \widetilde{J}_{\mathbf{Q}}(\mathbf{h}), \mathbf{v} \rangle = J_{\mathbf{Q}}(\mathbf{h}, \mathbf{v})$ for every $\mathbf{h}, \mathbf{v} \in \mathcal{X}$. We define the Hessian $J_{n,\mathbf{Q}}$ of the

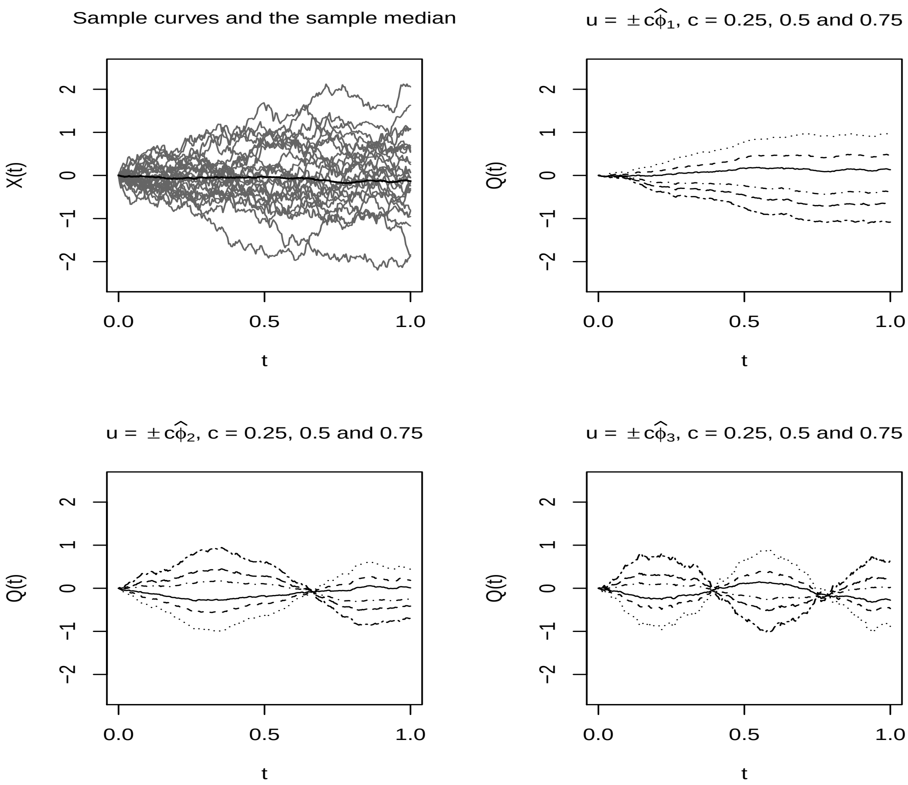

Figure 2: *The plots of the simulated data along with the sample spatial median (bold curve in the top left plot), and other sample spatial quantiles (in the remaining plots). For each* $k = 1, 2, 3$*, the sample spatial quantiles corresponding to* $\mathbf{u} = c\phi_k$ *for* $c = 0.25,\ 0.5$ *and* 0.75 *are given by the solid (—), the dashed (- - -) and the dotted (· · ·) curves, respectively, while those corresponding to* $\mathbf{u} = -c\phi_k$ *for these* c *values are given by the dot-dashed (– · –), the long-dashed (– –) and the two-dashed (– - –) curves, respectively.*

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 continuous linear operator associated with $J_{n,\mathbf{Q}}$ is denoted by $\widetilde{J}_{n,\mathbf{Q}}$. 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. 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 3.1.2 below.

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

As discussed after Assumption (A) in Section 2, 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 2.2. Also, it holds under the same

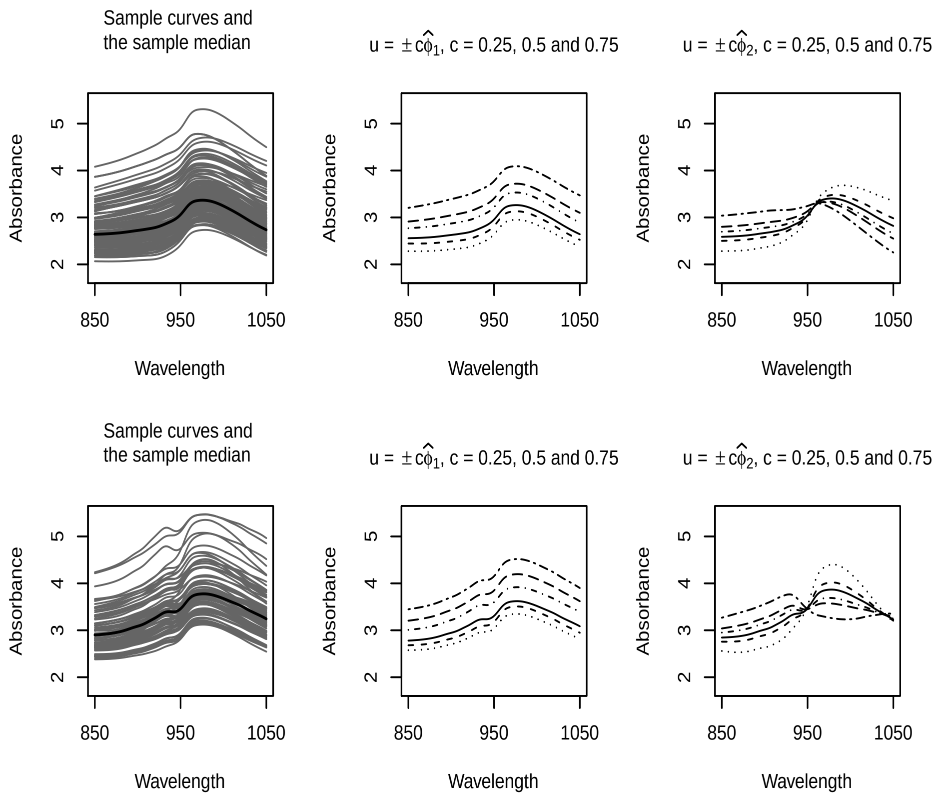

Figure 3: *The plots of the spectrometric data and the sample spatial quantiles. The plots in the first column show the observations for fat content* $\leq 20\%$ *and* $> 20\%$ *along with the sample spatial medians (bold curves). For each* $k = 1, 2$*, the sample spatial quantiles corresponding to* $\mathbf{u} = c\phi_k$ *for* $c = 0.25,\ 0.5$ *and* 0.75 *are given by the solid (—), the dashed (- - -) and the dotted (· · ·) curves, respectively, while those corresponding to* $\mathbf{u} = -c\phi_k$ *for these c values are given by the dot-dashed (– · –), the long-dashed (– –) and the two-dashed (– - –) curves, respectively, in the plots in the second and the third columns.*

situation as discussed after Theorem 2.2.

Theorem 3.1.2. *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* $n \to \infty$.

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

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

The Bahadur-type representation of the sample spatial **u**-quantile in finite dimensional Euclidean spaces (see, e.g., [12] and [20]) 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 $n \to \infty$.

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 3.1.2, 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 3.1.2, 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 [36] 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 [32] and Theorem 2.1 in [37], 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 [5] 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., [28]). 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 3.1.3. *Suppose that the assumptions of Theorem 3.1.2 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 covariance of* $\mathbf{Z_u}$ *is given by* $V_{\mathbf{u}} = [\widetilde{J}_{\mathbf{Q}(\mathbf{u})}]^{-1} \Lambda_{\mathbf{u}} [\widetilde{J}_{\mathbf{Q}(\mathbf{u})}]^{-1}$, *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* $\langle .,. \rangle$ *denotes the inner product in* $\mathcal{X}$*.*

A random element $\mathbf{Z}$ in the separable Hilbert space $\mathcal{X}$ is said to have a Gaussian distribution with mean $\mathbf{m} \in \mathcal{X}$ and covariance $\mathbf{C} : \mathcal{X} \to \mathcal{X}$ if for any $\mathbf{l} \in \mathcal{X}$, $\langle \mathbf{l}, \mathbf{Z} \rangle$ has a Gaussian distribution on $\mathbb{R}$ with mean $\langle \mathbf{l}, \mathbf{m} \rangle$ and variance $\langle \mathbf{C}(\mathbf{l}), \mathbf{l} \rangle = E\{(\langle \mathbf{l}, \mathbf{Z} - \mathbf{m} \rangle)^2\}$ (see, e.g., [1]).

3.2 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 3.1.2, 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 3.1.3. 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 $K(t,s) = 0.5(t^{2\mathrm{H}} + s^{2\mathrm{H}} - |t-s|^{2\mathrm{H}})$ 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., [38]) 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 [28]). 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 3.1.3. The asymptotic efficiency of $\widehat{\mathbf{Q}}(\mathbf{0})$ relative to $\overline{\mathbf{X}}$ can be defined as $trace(\Sigma)/trace(V_{\mathbf{0}})$. The traces of Σ and $V_{\mathbf{0}}$ are defined as $\sum_{k=1}^{\infty} \langle \Sigma\psi_k, \psi_k \rangle$ and $\sum_{k=1}^{\infty} \langle V_{\mathbf{0}}\psi_k, \psi_k \rangle$, 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.

4 Spatial depth and the DD-plot in Banach spaces

In the finite dimensional setup, the spatial distribution has been used to define the spatial depth (see [29] and [35]). Likewise, the spatial depth at $\mathbf{x}$ in a smooth Banach space $\mathcal{X}$ with respect to the probability distribution of a random element $\mathbf{X} \in \mathcal{X}$ can be defined as $SD(\mathbf{x}) = 1 - ||S_{\mathbf{x}}||$, and its empirical version is given by $\widehat{SD}(\mathbf{x}) = 1 - ||\widehat{S}_{\mathbf{x}}||$. Here, $S_{\mathbf{x}}$ and $\widehat{S}_{\mathbf{x}}$ are as defined in Section 2. There are a few other notions of depth function for data in infinite dimensional function spaces (see, e.g., [16], [24], [25] and [31]). However, as shown in [9], some of these depth functions exhibit degeneracy for certain types of functional data, and hence are not very useful.

We will now discuss some properties of the spatial depth function in Banach spaces. The spatial distribution function $S_{\mathbf{x}}$ 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 $SGN_{L(\mathbf{x})-L(\mathbf{X})}(\mathbf{h}) = SGN_{A(\mathbf{x})-A(\mathbf{X})}(A(\mathbf{h}')) = SGN_{\mathbf{x}-\mathbf{X}}(\mathbf{h}') = SGN_{\mathbf{x}-\mathbf{X}}(A^{-1}(\mathbf{h})) = (A^{-1})^*(SGN_{\mathbf{x}-\mathbf{X}}(\mathbf{h}))$ 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 $S_{L(\mathbf{x})}$ is the spatial distribution at $L(\mathbf{x})$ with respect to the probability distribution of $L(\mathbf{X})$, we have $S_{L(\mathbf{x})} = (A^{-1})^*(S_{\mathbf{x}})$, where $S_{\mathbf{x}}$ 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 [18] that if $\mathcal{X}$ is a strictly convex Banach space, and the distribution of $\mathbf{X}$ is non-atomic 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 $SD(\mathbf{m}) = 1$. It follows from the last assertion in Theorem 3.1 that if the norm in $\mathcal{X}$ is Fréchet differentiable and the distribution of $\mathbf{X}$ is non-atomic, 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 $\mathbf{y} \neq \mathbf{0}$. 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 [22] and [40] for the finite dimensional setting.

It follows from Theorem 3.1 that if $\mathcal{X}$ is a reflexive Banach space and the distribution of $\mathbf{X}$ is non-atomic, then $SD(\mathbf{x})$ takes all values in $(0, 1]$ as $\mathbf{x}$ varies over $\mathcal{X}$. Also, 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. On the other hand, many of the well-known depths for infinite dimensional data like the half-space depth, the band depth and the half-region depth do not possess such regular behavior and exhibit degeneracy for many Gaussian distributions (see [9]).

We will next study the properties of the empirical spatial depth in smooth Banach spaces. A Banach space $\mathcal{X}$ is said to be of type 2 (see, e.g., [1]) if there exists a constant $\gamma > 0$ such that for any $n \geq 1$ and independent zero mean random elements $\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_n$ in $\mathcal{X}$ with $E\{||\mathbf{U}_i||^2\} < \infty$ for all $i = 1, 2, \ldots, n$, we have $E\{|| \sum_{i=1}^n \mathbf{U}_i||^2\} \leq \gamma \sum_{i=1}^n E\{||\mathbf{U}_i||^2\}$. Examples of type 2 spaces include Hilbert spaces and L_p spaces with $p \geq 2$. Type 2 Banach spaces are the only Banach spaces, where the central limit theorem will hold for every sequence of i.i.d. random elements, whose squared norms have finite expectations. Let $\mathbf{C} : \mathcal{X}^* \to \mathcal{X}^{**}$ be a symmetric nonnegative definite continuous linear operator. A random element $\mathbf{X}$ in a separable Banach space $\mathcal{X}$ is said to have a Gaussian distribution with mean $\mathbf{m} \in \mathcal{X}$ and covariance $\mathbf{C}$ if for any $\mathbf{l} \in \mathcal{X}^*$, $\mathbf{l}(\mathbf{X})$ has a Gaussian distribution on $\mathbb{R}$ with mean $\mathbf{l}(\mathbf{m})$ and variance $(\mathbf{C}(\mathbf{l}))(\mathbf{l})$ (see, e.g., [1]). If $\mathcal{X}$ is a Hilbert space, this definition coincides with the one given after Theorem 3.1.3.

Theorem 4.1. *Suppose that the assumptions of part (a) of Theorem 2.2 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 $SGN_{S_{\mathbf{x}}}(\mathbf{W})$ if $S_{\mathbf{x}} \neq \mathbf{0}$. If $S_{\mathbf{x}} = \mathbf{0}$, $\sqrt{n}(\widehat{SD}(\mathbf{x}) - SD(\mathbf{x}))$ converges weakly to $||\mathbf{V}||$. Here, $\mathbf{W}$ and $\mathbf{V}$ are zero mean Gaussian random elements in $\mathcal{X}^*$.*

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 [23]). A DD-plot is a scatterplot of the depth values of the 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. Here, we consider the problem of constructing DD-plots for data in infinite dimensional spaces. It follows from [9] that the half-space depth and the simplicial depth, which have been used by the authors of [23] 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 the spatial depth (see Figure 4). 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 $L_2[0, 1]$ (see Section 3). 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 Section 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, $S_{\mathbf{x}}$ 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 first (respectively, second) plot in Figure 4 is the DD-plot for the two samples from the standard Brownian motion (respectively, the fractional Brownian motion). The third plot is the DD-plot for the two samples from the standard Brownian motion and the fractional Brownian motion. The axes of the first and the second DD-plots 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 $\circ$'s and the $\times$'s represent the sample observations of the two samples. The vertical and the horizontal axis of the third DD-plot correspond to the depth values with respect to the empirical distributions of the standard Brownian motion and the fractional Brownian motion, respectively, and the $\circ$'s and the $\times$'s represent the samples from these two distributions, respectively. In the first two DD-plots, the $\circ$'s and the $\times$'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 third DD-plot, all the $\circ$'s and the $\times$'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 fourth plot in Figure 4) 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 $\circ$'s and the $\times$'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.

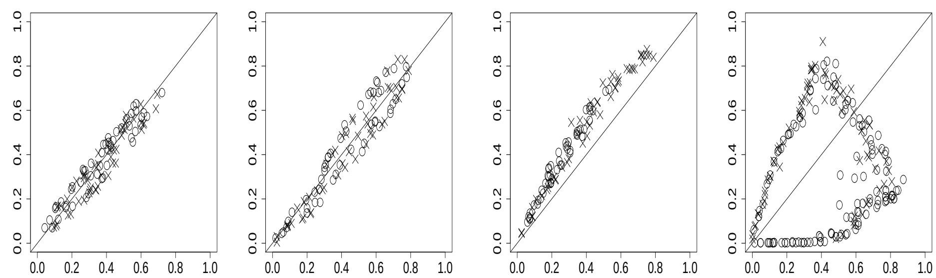

Figure 4: *The DD-plots for the simulated and the real data. The first (respectively, second) figure from the left is the DD-plot for the two samples from the standard Brownian motion (respectively, the fractional Brownian motion). The third figure from the left is the DD-plot for the two samples from the standard Brownian motion and the fractional Brownian motion. The fourth figure is the DD-plot for the two samples in the spectrometric data.*

5 Appendix: The proofs

The proofs involve several concepts and techniques from probability theory in Banach spaces and convex analysis. Readers are referred to [1] for an exposition on probability theory in Banach spaces. We refer to [14] for an exposition on the theory of Banach spaces, and [3] for the relevant details on convex analysis in Banach spaces.

Lemma 5.1. *Suppose that $\mathcal{X}^*$ is a separable Banach space. If μ is atomic, then $\sup_{\mathbf{x}\in\mathcal{X}} ||\widehat{S}_{\mathbf{x}} - S_{\mathbf{x}}|| \to 0$ 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 $\sup_{\mathbf{x}\in\mathcal{X}} ||\widehat{S}_{\mathbf{x}} - S_{\mathbf{x}}|| \leq \sum_{\mathbf{y}\in A_\mu} |\widehat{p}(\mathbf{y}) - p(\mathbf{y})| = 2 - 2\sum_{\mathbf{y}\in A_\mu} \min\{\widehat{p}(\mathbf{y}), p(\mathbf{y})\}$. Since $\min\{\widehat{p}(\mathbf{y}), p(\mathbf{y})\} \leq p(\mathbf{y})$, the proof is complete by the dominated convergence theorem. □

Proof of Thm. 2.1. Let us write $\mu = \rho\mu_1 + (1-\rho)\mu_2$, where μ_1 and μ_2 are the non-atomic 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 $\langle .,. \rangle$. Observe that for any $\mathbf{x} \in \mathcal{Z}$ and $\mathbf{l} \in \mathcal{X}$,

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

In other words,

$$\begin{aligned}
\left|\left\langle \mathbf{l}, \widehat{S}_{\mathbf{x}} - S_{\mathbf{x}} \right\rangle\right| \leq \left|\left\langle \mathbf{l}, E_{\widehat{\mu}_1}\left(\frac{\mathbf{x}-\mathbf{X}}{||\mathbf{x}-\mathbf{X}||}\right) - E_{\mu_1}\left(\frac{\mathbf{x}-\mathbf{X}}{||\mathbf{x}-\mathbf{X}||}\right)\right\rangle\right| + \\
||\mathbf{l}|| \left\| E_{\widehat{\mu}_2}\left(\frac{\mathbf{x}-\mathbf{X}}{||\mathbf{x}-\mathbf{X}||}\right) - E_{\mu_2}\left(\frac{\mathbf{x}-\mathbf{X}}{||\mathbf{x}-\mathbf{X}||}\right)\right\| + 2\left|N_n/n - \rho\right|.
\end{aligned} \tag{5.1}$$

The third term in the right hand side of (5.1) converges to zero as $n \to \infty$ *almost surely* by the strong law of large numbers. By Lemma 5.1, the second term in the right hand side of (5.1) 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 [20] show that $\mathcal{G}$ is a VC-subgraph class. Since μ_1 is non-atomic, 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 point-wise separable class (see p. 116 in [34]) with an envelope function that is unity everywhere. Thus, it follows from Theorem 2.6.8 in [34] 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 (5.1) 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,

$$\lim_{n\to\infty} \sup_{\mathbf{x}\in\mathcal{Z}} \left| \left\langle \mathbf{l}, E_{\widehat{\mu}_1}\left(\frac{\mathbf{x}-\mathbf{X}}{||\mathbf{x}-\mathbf{X}||}\right) - E_{\mu_1}\left(\frac{\mathbf{x}-\mathbf{X}}{||\mathbf{x}-\mathbf{X}||}\right)\right\rangle \right| = 0 \quad \forall\, \mathbf{l} \in \mathcal{L} \tag{5.2}$$

as $n \to \infty$ *almost surely*. Note that both the expectations in (5.2) above are bounded in norm by 1. Using this fact, equation (5.2) 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 non-atomic, 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 [34] implies that $\mathcal{G}$ is a Donsker class with respect to μ. This completes the proof of the theorem. □

Remark: Suppose that $\mathcal{X} = L_p$ for an even integer $p > 2$. Using arguments similar to those used in deriving (5.1), we get an analogous bound for $\mathbf{l}(SGN_{\mathbf{x}-\mathbf{X}})$ for any $\mathbf{x} \in \mathcal{Z}$ and $\mathbf{l} \in \mathcal{X}$. In this case, $\mathcal{G}$ in the proof of Theorem 2.1 is to be defined as $\mathcal{G} = \{\psi_{\mathbf{x}} : \mathcal{X} \to \mathbb{R},\ \psi_{\mathbf{x}}(\mathbf{s}) = \mathbf{l}(SGN_{\mathbf{x}-\mathbf{s}});\ \mathbf{x} \in \mathcal{Z}\}$, 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 2.1, it can be shown that $\mathcal{G}$ is a VC-subgraph and a point-wise separable class, and hence a Glivenko-Cantelli and a Donsker class. So, the assertions of Theorem 2.1 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 2.2(b).

Fact 5.2. [39, 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 $1 \le i \le n$. Suppose that for some $h > 0$ and $u_i > 0$, we have $E(||\mathbf{Y}_i||^m) \le (m!/2)u_i^2 h^{m-2}$ for $1 \le i \le n$ and all $m \ge 2$. Let $U_n^2 = \sum_{i=1}^n u_i^2$. Then, for any $v > 0$, $P(||\sum_{i=1}^n \mathbf{Y}_i|| \ge vU_n) \le 2exp\{-(v^2/2)(1 + 1.62(vh/U_n))^{-1}\}$.*

Proof of Thm. 2.2. (a) As in the proof of Theorem 2.1, we get

$$\begin{aligned} ||\widehat{S}_{\mathbf{x}} - S_{\mathbf{x}}|| \quad &\le \quad ||E_{\widehat{\mu}_1}\{SGN_{\mathbf{x}-\mathbf{X}}\} - E_{\mu_1}\{SGN_{\mathbf{x}-\mathbf{X}}\}|| + \\ &\quad\quad ||E_{\widehat{\mu}_2}\{SGN_{\mathbf{x}-\mathbf{X}}\} - E_{\mu_2}\{SGN_{\mathbf{x}-\mathbf{X}}\}|| + 2\,|N_n/n - \rho|\,. \end{aligned} \tag{5.3}$$

Further, the second and the third terms in the right hand side of the inequality in (5.3) converge to zero as $n \to \infty$ *almost surely* by the same arguments as in the proof of Theorem 2.1. 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 (5.3) is bounded above by

$$||E_{\widehat{\mu}_1}\{SGN_{\mathbf{x}-\mathbf{X}}\} - E_{\widehat{\mu}_1}\{SGN_{\mathbf{v}_j-\mathbf{X}}\}|| + ||E_{\mu_1}\{SGN_{\mathbf{x}-\mathbf{X}}\} - E_{\mu_1}\{SGN_{\mathbf{v}_j-\mathbf{X}}\}|| \\ + \max_{1\le l\le N(\varepsilon)} ||E_{\widehat{\mu}_1}\{SGN_{\mathbf{v}_l-\mathbf{X}}\} - E_{\mu_1}\{SGN_{\mathbf{v}_l-\mathbf{X}}\}||,$$

where $||\mathbf{x} - \mathbf{v}_j|| < \varepsilon$. Using Assumption (A) in Section 2, it follows that

$$\begin{aligned} ||E_{\widehat{\mu}_1}\{SGN_{\mathbf{x}-\mathbf{X}}\} - E_{\widehat{\mu}_1}\{SGN_{\mathbf{v}_j-\mathbf{X}}\}|| &\le E_{\widehat{\mu}_1}\{T(\mathbf{v}_j - \mathbf{X})\}||\mathbf{x} - \mathbf{v}_j|| \\ &\le 2\varepsilon E_{\mu_1}\{T(\mathbf{v}_j - \mathbf{X})\}, \end{aligned} \tag{5.4}$$

for all n sufficiently large *almost surely*. Further,

$$||E_{\mu_1}\{SGN_{\mathbf{x}-\mathbf{X}}\} - E_{\mu_1}\{SGN_{\mathbf{v}_j-\mathbf{X}}\}|| \le \varepsilon E_{\mu_1}\{T(\mathbf{v}_j - \mathbf{X})\}. \tag{5.5}$$

Using (5.4) and (5.5), the moment condition in the theorem and the fact that $\max_{1\le l\le N(\varepsilon)} ||E_{\widehat{\mu}_1}\{ SGN_{\mathbf{v}_l-\mathbf{X}}\} - E_{\mu_1}\{SGN_{\mathbf{v}_l-\mathbf{X}}\}||$ 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 2.1, it is enough to consider the case $d = 1$. Using Theorems 1.5.4 and 1.5.7 in [34], it follows that we only need to prove the asymptotic equicontinuity *in probability* of $\widehat{\mathbf{S}}_{\mathbf{g}}$ with respect to the norm in $\mathcal{X}$. Further, since μ is assumed to be non-atomic, the map $\mathbf{x} \mapsto \mathbf{g}(\sqrt{n}(\widehat{S}_{\mathbf{x}} - S_{\mathbf{x}}))$ is *almost surely* μ-continuous. Since K is compact, it follows that the process $\widehat{\mathbf{S}}_{\mathbf{g}}$ is separable (see p. 115 in [34]). Thus, in view of Corollary 2.2.8 in [34] 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 $\widehat{\mathbf{S}}_{\mathbf{g}}$ if we can show the sub-Gaussianity of the process (see p. 101 in [34]) with respect to the metric induced by the norm in $\mathcal{X}$. Since $\mathbf{g} \in \mathcal{X}^{**}$, the empirical process $\widehat{\mathbf{S}}_{\mathbf{g}} = \{\sqrt{n}[n^{-1}\sum_{i=1}^n \mathbf{g}(SGN_{\mathbf{x}-\mathbf{X}_i}) - E\{\mathbf{g}(SGN_{\mathbf{x}-\mathbf{X}_i})\}] : \mathbf{x} \in K\}$. Using the Bernstein inequality for real-valued random variables and the assumptions in the theorem, we have

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

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 5.2 stated earlier. Using this, we have

$$\begin{aligned} P(|\widehat{\mathbf{S}}_{\mathbf{g}}(\mathbf{x}) - \widehat{\mathbf{S}}_{\mathbf{g}}(\mathbf{y})| > t) &\le P(\sqrt{n}||(\widehat{S}_{\mathbf{x}} - \widehat{S}_{\mathbf{y}}) - (S_{\mathbf{x}} - S_{\mathbf{y}})|| > t) \\ &\le 2exp\left\{-t^2/a_2||\mathbf{x} - \mathbf{y}||^2\right\} \quad \forall\ n \end{aligned}$$

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 Thm. 3.1. 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 [3], 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 [18] 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 non-atomic, 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 [3], we have $S_{\mathbf{x}(\mathbf{v})} = E\{SGN_{\mathbf{x}(\mathbf{v})-\mathbf{X}}\} = \mathbf{v}$. 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 non-atomic, 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 [3]. □

The next result can be obtained by suitably modifying the arguments in the second paragraph in the proof of Theorem 3.1.1 in [12].

Fact 5.3. *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 Thm. 3.1.1. From the assumptions in the theorem and Theorem 2.17 and Remark 3.5 in [18], 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 [7] 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

$$0 \leq g(\widehat{\mathbf{Q}}(\mathbf{u})) - g(\mathbf{Q}(\mathbf{u})) = [g(\widehat{\mathbf{Q}}(\mathbf{u})) - g_n(\widehat{\mathbf{Q}}(\mathbf{u}))] - [g(\mathbf{Q}(\mathbf{u})) - g_n(\mathbf{Q}(\mathbf{u}))] + [g_n(\widehat{\mathbf{Q}}(\mathbf{u})) - g_n(\mathbf{Q}(\mathbf{u}))]. \tag{5.6}$$

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

$$\sup_{||\mathbf{Q}|| \leq C} |g(\mathbf{Q}) - g_n(\mathbf{Q})| \to 0, \tag{5.7}$$

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

$$\begin{aligned} & g_n(\widehat{\mathbf{Q}}(\mathbf{u})) - g_n(\mathbf{Q}(\mathbf{u})) \\ = \;& [g_n(\widehat{\mathbf{Q}}(\mathbf{u})) - \widehat{g}_n(\widehat{\mathbf{Q}}(\mathbf{u}))] + [\widehat{g}_n(\widehat{\mathbf{Q}}(\mathbf{u})) - \widehat{g}_n(\mathbf{Q}^{(n)}(\mathbf{u}))] \\ & + [\widehat{g}_n(\mathbf{Q}^{(n)}(\mathbf{u})) - g_n(\mathbf{Q}^{(n)}(\mathbf{u}))] + [g_n(\mathbf{Q}^{(n)}(\mathbf{u})) - g_n(\mathbf{Q}(\mathbf{u}))]. \end{aligned} \tag{5.8}$$

In the notation of Section 3, $\mathbf{Q}^{(n)}(\mathbf{u}) = \sum_{k=1}^{d(n)} q_k\phi_k$, 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 (5.8) 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 5.3. The second term in the right hand side of (5.8) is negative because $\widehat{\mathbf{Q}}(\mathbf{u})$ is a minimizer of $\widehat{g}_n$. The fourth term in the right hand side of (5.8) 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 (5.6), (5.7) 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 [2], 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

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

The next result is the $d(n)$-dimensional analog of Proposition 2.1 in [8], and can be obtained by suitably modifying the proof of that proposition.

Fact 5.4. *Suppose that the assumptions of Theorem 3.1.2 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 $b||\mathbf{h}||^2 \leq J_{n,\mathbf{Q}}(\mathbf{h}, \mathbf{h}) \leq B||\mathbf{h}||^2$ for any $\mathbf{Q}$, $\mathbf{h} \in \mathcal{Z}_n$ with $||\mathbf{Q}|| \leq C$.*

Lemma 5.5. *Suppose that the assumptions of Theorem 3.1.2 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*

$$\left\| E\left\{ \frac{\mathbf{Q} - \mathbf{X}^{(n)}}{||\mathbf{Q} - \mathbf{X}^{(n)}||} - \mathbf{u}^{(n)} \right\} \right\| \geq b'||\mathbf{Q} - \mathbf{Q}_n(\mathbf{u})||,$$

$$\sup_{||\mathbf{h}||=||\mathbf{v}||=1} |J_{n,\mathbf{Q}}(\mathbf{h}, \mathbf{v}) - J_{n,\mathbf{Q}_n(\mathbf{u})}(\mathbf{h}, \mathbf{v})| \leq B'||\mathbf{Q} - \mathbf{Q}_n(\mathbf{u})||, \qquad \text{and}$$

$$\left\| E\left\{ \frac{\mathbf{Q} - \mathbf{X}^{(n)}}{||\mathbf{Q} - \mathbf{X}^{(n)}||} - \mathbf{u}^{(n)} \right\} - \widetilde{J}_{n,\mathbf{Q}_n(\mathbf{u})}(\mathbf{Q} - \mathbf{Q}_n(\mathbf{u})) \right\| \leq B'||\mathbf{Q} - \mathbf{Q}_n(\mathbf{u})||^2.$$

Proof. For any $||\mathbf{h}|| = 1$, 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

$$E\left\{ \frac{\mathbf{Q} - \mathbf{X}^{(n)}}{||\mathbf{Q} - \mathbf{X}^{(n)}||} - \mathbf{u}^{(n)} \right\}(\mathbf{h}) = J_{n,\widetilde{\mathbf{Q}}}(\mathbf{Q} - \mathbf{Q}_n(\mathbf{u}), \mathbf{h}), \tag{5.9}$$

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 5.4, we have the first inequality.

The second inequality follows from the definition of $J_{n,Q}$, the upper bound in Fact 5.4 and some straight-forward algebra.

From (5.9), we get

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

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

Proposition 5.6. *Suppose that the assumptions of Theorem 3.1.2 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 3.1.2.*

Proof. From Fact 5.3 and the behavior of $\mathbf{Q}_n(\mathbf{u})$ discussed before Assumption (B) in Section 3.1, 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 5.2, there exists $C_5 > 0$ such that $P(\mathrm{E}_n^c) \leq 2(3C_3 n^4)^{d(n)} exp\{-nC_5^2 \delta_n^2\}$ 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,

$$P(\mathrm{E}_n \text{ occurs for all sufficiently large } n) = 1. \tag{5.10}$$

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 3.1. 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 $P(\mathrm{F}_n^c) \leq (3C_3 n^4)^{d(n)} exp\{-nC_7 \delta_n^2\}$ 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

$$P(\mathrm{F}_n \text{ occurs for all sufficiently large } n) = 1. \tag{5.11}$$

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

$$\left\| \frac{\widehat{\mathbf{Q}}(\mathbf{u}) - \mathbf{X}_i^{(n)}}{||\widehat{\mathbf{Q}}(\mathbf{u}) - \mathbf{X}_i^{(n)}||} - \frac{\overline{\mathbf{Q}}_n(\mathbf{u}) - \mathbf{X}_i^{(n)}}{||\overline{\mathbf{Q}}_n(\mathbf{u}) - \mathbf{X}_i^{(n)}||} \right\| \le \frac{2||\widehat{\mathbf{Q}}(\mathbf{u}) - \overline{\mathbf{Q}}_n(\mathbf{u}))||}{||\overline{\mathbf{Q}}_n(\mathbf{u}) - \mathbf{X}_i^{(n)}||}. \tag{5.12}$$

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

$$\begin{aligned} \left\| \frac{1}{n} \sum_{i=1}^{n} \frac{\overline{\mathbf{Q}}_n(\mathbf{u}) - \mathbf{X}_i^{(n)}}{||\overline{\mathbf{Q}}_n(\mathbf{u}) - \mathbf{X}_i^{(n)}||} - \mathbf{u}^{(n)} \right\| &\le \left\| \frac{1}{n} \sum_{i=1}^{n} \frac{\widehat{\mathbf{Q}}(\mathbf{u}) - \mathbf{X}_i^{(n)}}{||\widehat{\mathbf{Q}}(\mathbf{u}) - \mathbf{X}_i^{(n)}||} - \mathbf{u}^{(n)} \right\| \\ &+ \left\| \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{\overline{\mathbf{Q}}_n(\mathbf{u}) - \mathbf{X}_i^{(n)}}{||\overline{\mathbf{Q}}_n(\mathbf{u}) - \mathbf{X}_i^{(n)}||} - \frac{\widehat{\mathbf{Q}}(\mathbf{u}) - \mathbf{X}_i^{(n)}}{||\widehat{\mathbf{Q}}(\mathbf{u}) - \mathbf{X}_i^{(n)}||} \right\} \right\| \\ &\le \left\| \frac{1}{n} \sum_{i=1}^{n} \frac{\widehat{\mathbf{Q}}(\mathbf{u}) - \mathbf{X}_i^{(n)}}{||\widehat{\mathbf{Q}}(\mathbf{u}) - \mathbf{X}_i^{(n)}||} - \mathbf{u}^{(n)} \right\| + 2C_8 d(n) n^{-2} \\ &\quad + \frac{2}{n} \sum_{i=1}^{n} I\{||\overline{\mathbf{Q}}_n(\mathbf{u}) - \mathbf{X}_i^{(n)}|| \le n^{-2}\} \quad \text{(using (5.12))} \\ &\le \left\| \frac{1}{n} \sum_{i=1}^{n} \frac{\widehat{\mathbf{Q}}(\mathbf{u}) - \mathbf{X}_i^{(n)}}{||\widehat{\mathbf{Q}}(\mathbf{u}) - \mathbf{X}_i^{(n)}||} - \mathbf{u}^{(n)} \right\| + C_9 \delta_n^2 \quad \text{(using (5.11))}. \end{aligned} \tag{5.13}$$

It follows from arguments similar to those used in the proof of Theorem 4.11 in [18] that $\left\| \sum_{i=1}^{n} \frac{\widehat{\mathbf{Q}}(\mathbf{u}) - \mathbf{X}_i^{(n)}}{||\widehat{\mathbf{Q}}(\mathbf{u}) - \mathbf{X}_i^{(n)}||} - n\mathbf{u}^{(n)} \right\| \le 1$. Combining this with (5.13), we get

$$\left\| \sum_{i=1}^{n} \frac{\overline{\mathbf{Q}}_n(\mathbf{u}) - \mathbf{X}_i^{(n)}}{||\overline{\mathbf{Q}}_n(\mathbf{u}) - \mathbf{X}_i^{(n)}||} - n\mathbf{u}^{(n)} \right\| \le 3C_7 n \delta_n \tag{5.14}$$

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 (5.10) and the first inequality in Lemma 5.5 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 (5.14), 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 Thm. 3.1.2. 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 $\mathbf{Q} \in \mathcal{Z}_n$,

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

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

Using Assumption (B) in Section 3.1, it follows that for a constant $C_{12} > 0$,

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

So, in view of Fact 5.2, there exists a constant $C_{13} > 0$ such that

$$\max_{\mathbf{Q}\in \mathrm{H}_n} \left\|\frac{1}{n}\sum_{i=1}^{n} \Gamma_n(\mathbf{Q}, \mathbf{X}_i)\right\| \leq C_{13}\delta_n^2, \tag{5.15}$$

for all sufficiently large n *almost surely*. Using the third inequality in Lemma 5.5, there exists a constant $C_{14} > 0$ such that $||\Delta_n(\mathbf{Q})|| \leq C_{14}||\mathbf{Q} - \mathbf{Q}_n(\mathbf{u})||^2$ for all appropriately large n. This along with (5.15) and the definitions of Γ_n and $\Delta_n(\mathbf{Q})$ yield

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

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

Hence, choosing $\mathbf{Q} = \overline{\mathbf{Q}}_n(\mathbf{u})$, and utilizing inequality (5.13) in the proof of Proposition 5.6, we get

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

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

Proof of Thm. 3.1.3. 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 5.2. We will show that $\sqrt{n}\{[\widetilde{J}_{n,\mathbf{Q}_n(\mathbf{u})}]^{-1}(\mathbf{U}_n) -$

$[\widetilde{J}_{\mathbf{Q}(\mathbf{u})}]^{-1}(\mathbf{U}_n)\} \to \mathbf{0}$ *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, $J_{n,\mathbf{Q}}$ and $\widetilde{J}_{n,\mathbf{Q}}$ 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 3.1. Further, the bound obtained in the second inequality in Lemma 5.5 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, $||\widetilde{J}_{n,\mathbf{Q}_n(\mathbf{u})} - \widetilde{J}_{n,\mathbf{Q}(\mathbf{u})}|| \leq B''||\mathbf{Q}_n(\mathbf{u}) - \mathbf{Q}(\mathbf{u})||$ 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 3.1 that $||\widetilde{J}_{n,\mathbf{Q}(\mathbf{u})} - \widetilde{J}_{\mathbf{Q}(\mathbf{u})}|| \to 0$ as $n \to \infty$. Since $\mathbf{Q}_n(\mathbf{u}) \to \mathbf{Q}(\mathbf{u})$, we now have $||\widetilde{J}_{n,\mathbf{Q}_n(\mathbf{u})} - \widetilde{J}_{\mathbf{Q}(\mathbf{u})}|| \to 0$ as $n \to \infty$. It follows from Proposition 2.1 in [8] that the linear operator $\widetilde{J}_{\mathbf{Q}(\mathbf{u})}$ 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

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

The convergence *in probability* asserted above holds because the operator norm of $\widetilde{J}_{n,\mathbf{Q}_n(\mathbf{u})}$ is uniformly bounded away from zero by Fact 5.4. The asymptotic Gaussianity of $\{\widetilde{J}_{\mathbf{Q}(\mathbf{u})}\}^{-1}(\sqrt{n}\mathbf{U}_n)$ follows from the central limit theorem for a triangular array of row-wise independent Hilbert space valued random elements (see, e.g., Corollary 7.8 in [1]). □

Proof of Thm. 4.1. The proof of the first statement follows directly from part (a) of Theorem 2.2 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 $S_{\mathbf{x}} \neq \mathbf{0}$. From the Fréchet differentiability of the norm in $\mathcal{X}^*$, we have $\widehat{SD}(\mathbf{x}) - SD(\mathbf{x}) = SGN_{S_{\mathbf{x}}}(\widehat{S}_{\mathbf{x}} - S_{\mathbf{x}}) + o(||\widehat{S}_{\mathbf{x}} - S_{\mathbf{x}}||)$. The central limit theorem for i.i.d. random elements in $\mathcal{X}^*$ (see, e.g., [1]) implies that $\sqrt{n}(\widehat{S}_{\mathbf{x}} - S_{\mathbf{x}})$ converges weakly to a zero mean Gaussian random element $\mathbf{W} \in \mathcal{X}^*$ as $n \to \infty$. In particular, $\sqrt{n}||\widehat{S}_{\mathbf{x}} - S_{\mathbf{x}}||$ is bounded *in probability* as $n \to \infty$. Since the map $SGN_{S_{\mathbf{x}}} : \mathcal{X}^* \to \mathbb{R}$ is continuous, we now have the result for $S_{\mathbf{x}} \neq \mathbf{0}$ using the continuous mapping theorem.

Now, we consider the case $S_{\mathbf{x}} = \mathbf{0}$. In this case, $\widehat{SD}(\mathbf{x}) - SD(\mathbf{x}) = -||\widehat{S}_{\mathbf{x}}||$. The central limit theorem for i.i.d. random elements in $\mathcal{X}^*$ yields that $\sqrt{n}\widehat{S}_{\mathbf{x}}$ 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. □

Acknowledgements

Research of the first author is partially supported by the SPM Fellowship of the Council of Scientific and Industrial Research, Government of India.

References

[1] Araujo, A. and Giné, E. (1980) *The central limit theorem for real and Banach valued random variables.* John Wiley & Sons.

[2] Asplund, E. (1968) Fréchet differentiability of convex functions. *Acta Math.* **121** 31–47.

[3] Borwein, J. M. and Vanderwerff, J. D. (2010) *Convex functions: constructions, characterizations and counterexamples* Cambridge University Press.

[4] Brown, B. M. (1983) Statistical uses of the spatial median. *J. Roy. Statist. Soc. Ser. B* **45** 25–30.

[5] Boyd, J. P. (1984) Asymptotic coefficients of Hermite function series. *J. Comput. Phys.* **54** 382–410.

[6] Bugni, F. A., Hall, P., Horowitz, J. L. and Neumann, G. R. (2009) Goodness-of-fit tests for functional data. *Econom. J.* **12** S1–S18.

[7] Cadre, B. (2001) Convergent estimators for the L_1-median of Banach valued random variable. *Statistics* **35** 509–521.

[8] Cardot, H., Cénac, P. and Zitt, P-A. Efficient and fast estimation of the geometric median in Hilbert spaces with an averaged stochastic gradient algorithm. *Bernoulli* **19** 18–43.

[9] Chakraborty, A. and Chaudhuri, P. On data depth in infinite dimensional spaces. Published online in *Ann. Inst. Stat. Math.* (2013) `http://dx.doi.org/10.1007/s10463-013-0416-y`

[10] Chakraborty, A. and Chaudhuri, P. A Wilcoxon–Mann–Whitney type test for infinite dimensional data. Technical Report No. R5/2013, Theoretical Statistics and Mathematics Unit. Indian Statistical Institute, Kolkata, India.

[11] Chaouch, M. and Goga, C. (2012) Using complex surveys to estimate the L_1-median of a functional variable: application to electricity load curves. *Int. Statist. Rev.* **80** 40–59.

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

[13] Donoho, D. L. and Gasko, M. (1992) Breakdown properties of location estimates based on half-space depth and projected outlyingness. *Ann. Statist.* **20** 1803–1827.

[14] Fabian, M., Habala, P., Hájek, P., Santalucía, V. M., Pelant, J. and Zizler, Vaclav (2001) *Functional analysis and infinite-dimensional geometry.* Springer-Verlag.

[15] Fraiman, R. and Pateiro-López, B. (2012) Quantiles for finite and infinite dimensional data. *J. Multivariate Anal.* **108** 1–14.

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

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

[18] Kemperman, J. H. B. (1987) The median of a finite measure on a Banach space. *Statistical data analysis based on the L_1-norm and related methods (Neuchâtel, 1987)* 217–230.

[19] Kolmogorov, A. N. and Tihomirov, V. M. (1961) ε-entropy and ε-capacity of sets in functional spaces. *Amer. Math. Soc. Transl. 2* **17** 277–364.

[20] Koltchinskii, V. I. (1997) M-estimation, convexity and quantiles. *Ann. Statist.* **25** 435–477.

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

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

[23] Liu, R. Y., Parelius, J. M. and Singh, K. (1999) Multivariate analysis by data depth: descriptive statistics, graphics and inference. *Ann. Statist.* **27** 783–858.

[24] Lopez-Pintado, S. and Romo, J. (2009) On the concept of depth for functional data. *J. Amer. Statist. Assoc.* **104** 718–734.

[25] Lopez-Pintado, S. and Romo, J. (2011) A half-region depth for functional data. *Comput. Statist. Data Anal.* **55** 1679–1695.

[26] Möttönen, J., Oja, H. and Tienari, J. (1997) On the efficiency of multivariate spatial sign and rank tests. *Ann. Statist.* **25** 542–552.

[27] Oja, H. (1983) Descriptive statistics for multivariate distributions. *Statist. Probab. Lett.* **1** 327–332.

[28] Rasmussen, C. E. and Williams, C. K. I. (2006) *Gaussian processes for machine learning.* MIT Press.

[29] Serfling, R. (2002) A depth function and a scale curve based on spatial quantiles. *Statistical data analysis based on the L_1-norm and related methods (Neuchâtel, 2002)* 25–38.

[30] Small, C. G. (1990) A survey of multidimensional medians. *Int. Statist. Rev.* **58** 263–277.

[31] Sun, Y. and Genton, M. G. (2011) Functional boxplots. *J. Comput. Graph. Statist.* **20** 316–334.

[32] Trefethen, L. N. (2008) Is Gauss quadrature better than Clenshaw-Curtis? *SIAM Rev.* **50** 67–87.

[33] Valadier, M. (1984) La mulltiapplication medianes conditionelles. *Z. Wahrsch. Verw. Gebiete* **67** 279–282.

[34] Van der Vaart, A. W. and Wellner, J. (1996) *Weak Convergence and Empirical Processes*. Springer Verlag.

[35] Vardi, Y. and Zhang, C-H. (2000) The multivariate L^1-median and associated data depth. *Proc. Natl. Acad. Sci. USA* **97** 1423–1423.

[36] Vretblad, A. (2003) *Fourier analysis and its applications*. Springer Verlag.

[37] Wang, H. and Xiang, S. (2012) On the convergence rates of Legendre approximation. *Math. Comp.* **81** 861–877.

[38] Yu, S., Tresp, V. and Yu, K. (2007) Robust multi-task learning with t-processes. *Proceedings of the 24th International Conference on Machine Learning (Oregon, 2007)* 1103–1110.

[39] Yurinskiĭ, V. V. (1976) Exponential inequalities for sums of random vectors. *J. Multivariate Anal.* **6** 473–499.

[40] Zuo, Y. and Serfling, R. (2000a) General notions of statistical depth function. *Ann. Statist.* **28** 461–482.