Strong Consistency of Set-Valued
Fréchet Sample Means in Metric Spaces

Cedric E. Ginestet [*]

Abstract

The Fréchet mean generalizes the idea of averaging in spaces where pairwise addition is not well-defined. In general metric spaces, however, the Fréchet sample mean is not a consistent estimator of the theoretical Fréchet mean. For non-trivial examples, sequences of Fréchet sample mean sets may fail to converge in a set-analytical sense. Hence, it becomes necessary to consider other types of convergence. We show that a specific type of almost sure (a.s.) convergence for the Fréchet sample mean introduced by Ziezold (1977) is, in fact, equivalent to the consideration of the Kuratowski outer limit of a sequence of Fréchet sample means. Equipped with this outer limit, we prove different laws of large numbers for random variables taking values in separable (pseudo-)metric space with a bounded metric. In this setting, we describe strong laws of large numbers for the Fréchet sample mean. In particular, we demonstrate that all subsequences of Fréchet sample means converge to a subset of the theoretical mean. This result allows us to show that the Fréchet sample mean is metric squared error (MSE) consistent under the condition that their Kuratowski outer limits are non-empty. Convergence in probability and convergence in law of these sample estimators are also derived and the implications between these different modes of convergence are studied.

Key Words: Barycenter, Centroid, Consistency, Estimation theory, Equicontinuity, Fréchet mean, Graph-valued random variable, Karcher Mean, Point function

1. Introduction

All statistics are summaries. The epitome of these summaries is the sample mean, and its theoretical analog, the expected value. In an inspired monograph, Fréchet (1948) generalized this concept to any abstract metric space. He showed that the sole requirement for the definition of a mean element is the specification of a metric on the space of interest. Once this metric has been chosen and a probability measure has been defined on that metric space, the Fréchet mean is simply the element that minimizes the sum of the squared distances from all the elements in that space. The Fréchet mean generalizes other notions of means in abstract spaces, such as the centroid in Euclidean geometry, the barycenter or center of mass in physics, the Procrustean mean in shape spaces (Le, 1998), and the Karcher mean on Riemannian manifolds (Karcher, 1977). The sample version of the Fréchet mean can naturally be expressed using cumulative addition instead of the expectation, thereby producing a convex combination operator on metric spaces with both negative and positive Alexandrov curvature (Ginestet et al., 2012).

The object of this paper is to characterize the asymptotic behavior of the Fréchet sample mean in separable metric spaces with a bounded metric. We are here especially interested in metric spaces of simple graphs. Separability is a relatively mild topological assumption likely to be satisfied in most applications. The boundedness of the metric, however, is a more stringent condition. Nonetheless, there is a range of modern statistical applications for which the metric of interest is likely to

---

[*]Department of Mathematics and Statistics, College of Arts and Sciences, Boston University, 111 Cummington Mall, Boston, MA 02215

be bounded. In bioinformatics, the use of the Hamming (1950) distance on finite alphabets, such as stretches of DNA for instance, naturally gives rise to such assumptions (He et al., 2004). Similarly, the comparison of families of networks with a given number of nodes, as commonly done in neuroscience (Ginestet et al., 2011) may similarly generate bounded metric spaces; albeit the combinatorial nature of these metrics may lead to bounds that increase factorially with the number of nodes in these networks.

The asymptotic properties of the Fréchet sample mean have been studied by several authors. Ziezold (1977) proved a strong law of large numbers for Fréchet sample means defined in separable pseudo-metric spaces, where the metric is not assumed to satisfy the coincidence axiom. This a.s. convergence result has also been demonstrated for compact metric spaces by Sverdrup-Thygeson (1981). The perspectives adopted by these two authors are very different in nature. Given the fact that Sverdrup-Thygeson (1981) does not cite the work of Ziezold (1977), and because the work of the latter was published in a conference proceedings, it is probable that Sverdrup-Thygeson (1981) was not cognisant of Ziezold's proof technique.

The properties of sample Fréchet means on Riemannian manifolds have been particularly well-studied (Bhattacharya and Patrangenaru, 2002, 2005). When the Fréchet mean is assumed to be unique, the theorem of Sverdrup-Thygeson (1981) has been generalized by Bhattacharya and Patrangenaru (2003) for *proper* metric spaces. Recall that a metric space is proper, if and only if every bounded closed subsets of that space is compact. By the Hopf-Rinow theorem, every complete and connected Riemannian manifold is a proper metric space. Thus, Bhattacharya and Patrangenaru (2003) have weakened the compactness assumption made by Sverdrup-Thygeson (1981), and their strong law of large numbers apply to manifolds, under some very mild conditions. Recently, Kendall and Le (2011) have further generalized these results with a weak law of large numbers and a central limit theorem for sequences of Fréchet sample means based on non-iid random variables taking values on a Riemannian manifold. Here, we consider sequences of random variables taking values in separable pseudo-metric spaces with a bounded metric. Using boundedness, we provide a different proof of the strong consistency of the Fréchet sample mean from the one of Ziezold (1977).

Importantly, we also clarify previous results on the asymptotic consistency of the Fréchet sample mean, by showing that the modes of convergence studied by Ziezold (1977) and Sverdrup-Thygeson (1981) are, in fact, equivalent to the consideration of the Kuratowski outer limit of a sequence of Fréchet sample means. One of the core difficulties with the consideration of the asymptotic properties of Fréchet sample means is that such functions can be *multivalued*. That is, when the Fréchet sample mean is not unique, we obtain a random variable that is a set-valued function, which takes values in the power set of $\mathcal{X}$, or more precisely in the Borel σ-algebra of $\mathcal{X}$. It then becomes necessary to consider the convergence of multivalued functions. To this end, we resort to the tools of set-valued analysis, as described by Aubin and Frankowska (2009). This difficulty leads us to consider different 'types' of convergence, depending on whether we require the Fréchet sample mean to converge, or are simply interested in evaluating the asymptotic behavior of the outer limit of that sequence (see Molchanov, 2005, for an introduction to set-valued random variables).

The main innovation in this paper is our formal set-valued perspective. Note that our approach differs from the one of Bhattacharya and Bhattacharya (2012),
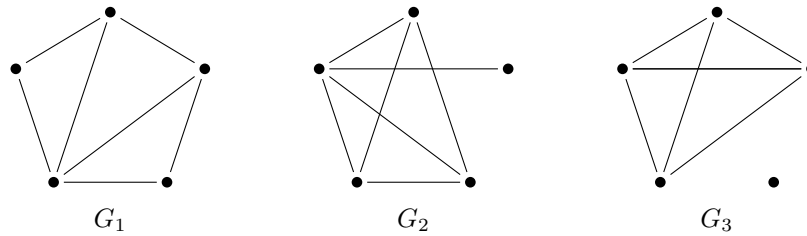
Figure 1. A sample of graphs, $G_i = (V, E)$, over five vertices, denoted by $G_i \in \mathcal{G}_5$.

since we have allowed the metric spaces of interest to be non-compact, and not necessarily equipped with a manifold structure. In particular, we identify the key role played by the Kuratowski outer limit when studying sequences of Fréchet sample means. This paper therefore constitutes an extension of the work of Ziezold (1977) and Sverdrup-Thygeson (1981) to Fréchet means of all orders, and to restricted Fréchet means. Moreover, we have emphasized the importance of point functions and of the Glivenko-Cantelli lemma.

This paper is organized as follows. Firstly, we motivate this work with a counterintuitive example of a graph-valued mean set that includes its sample as a proper subset. This justifies our emphasis on set-valued convergence throughout the rest of the paper. In section 3, we then introduce and study different types of a.s. convergence for sequences of Fréchet sample mean sets, and show through counterexamples why the Kuratowski outer limit is adequate for this purpose. In section 4, we prove the strong consistency of the Fréchet sample mean sets in bounded metric spaces.

2. Motivating Example: Graph Means

We are here especially interested in spaces of simple graphs, $G_i := (V, E)$ with $i = 1, \ldots, n$, which have a fixed number of vertices, $N_v := |V(G_i)|$, but their edge set, $E(G_i)$ may vary. A graph is said to be simple, when it does not contain multiple edges, loops or weighted edges. Throughout this paper, we will assume that there exists a probability measure on the space of all such simple graphs. A sample of three such simple graphs for $N_v = 7$ is given in figure 1.

Statistically, one may be interested in computing the mean graph for this type of random variables. Such a mean quantity can be defined as the Fréchet mean of that variable with respect to some distance function on the space of interest. A standard distance function on spaces of graphs is the Hamming distance, which is defined as follows for any two graphs $G = (V, E)$ and $G' = (V, E')$ with N_v vertices,

$$d_H(G, G') := \sum_{i < j} \mathcal{I}\{e_{ij} \neq e'_{ij}\}.$$

We denote by $\mathcal{G}_{N_v}$ the space of all simple graphs with N_v vertices. Given a graph-valued random variable on $\mathcal{G}_{N_v}$, the mean value for a sample of n realizations is then given by the element in $\mathcal{G}_{N_v}$, which minimizes the squared distances to all the graphs in the sample considered. For general graph-valued random variables, however, such a mean element needs not be unique.

In figure 2, we consider a sample of $n = 2$ graphs S_1 and S_2 with $N_v = 4$ vertices. Using the Hamming distance, the Fréchet mean graphs are the following elements

(a) Sample of simple graphs, S, with n = 2.
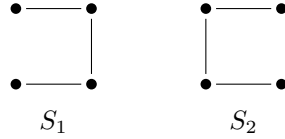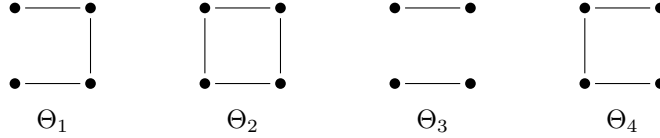

(b) Fréchet Mean, Θ, for this sample.

Figure 2. The sample of graphs in (a) is here a proper subset of the graph mean in (b), such that $S \subset \Theta$, where the Fréchet mean, Θ is computed with respect to the Hamming distance on the space of all simple graphs with $N_v = 4$ vertices.

of $\mathcal{G}_4$,

$$\Theta := \underset{G' \in \mathcal{G}_4}{\operatorname{argmin}} \sum_{l=1}^{n} \sum_{i<j} \mathcal{I}\{e_{ij}^{(l)} \neq e'_{ij}\}.$$

One can easily verify that the Fréchet mean is given by a set of four different simple graphs, as shown in figure 2. Hence, in this setting, we obtain the paradoxical result that the sample is a *proper* subset of the mean. This is somewhat counterintuitive, since we generally expect an average value to summarize information, and therefore to be more 'concentrated' than the sample values on which the mean is based.

Observe that the Hamming distance is here a bounded metric. In the sequel, we will consider the more general case of random variables taking values in separable metric spaces with bounded metrics, which encompasses graph-valued random variables, as a special case. Other popular choices of distance functions include the graph edit distance (Gao et al., 2010), and maximum common subgraph distance (Bunke, 1997).

3. Sequences of Fréchet Sample Means

3.1 Empirical and Theoretical Fréchet Means

A separable space $\mathcal{X}$ is endowed with a metric $d : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}^+$. This produces a metric space, $(\mathcal{X}, d)$, with elements x. Let a probability space be denoted by $(\Omega, \mathcal{F}, \mathbb{P})$, and define a random variable, X, on that space, which takes values in $(\mathcal{X}, \mathcal{B})$. Here, $\mathcal{B}$ is the Borel σ-algebra generated by the topology, τ on $\mathcal{X}$, induced by d. The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is assumed to be *complete*, in the sense that every subset of every null set is measurable. This is particularly convenient for constructing product spaces based on Ω that remain well-behaved. In addition, we define $\mu(B) := (\mathbb{P} \circ X^{-1})(B)$, for every $B \in \mathcal{B}$. Naturally, X is here assumed to be $(\mathcal{F}, \mathcal{B})$-measurable. Such a random variable will be termed an *abstract-valued* random variable, which will be contrasted with the more standard real-valued random variables.

In this setting, we compute the most '*central*' element. This is the element that has the smallest expected distance to all other elements in $\mathcal{X}$. This approach allows

us to define the following moments (Fréchet, 1948),

$$\Theta^r := \operatorname*{arginf}_{x' \in \mathcal{X}} \int_{\mathcal{X}} d(x, x')^r d\mu(x), \quad \text{and} \quad \sigma^r := \inf_{x' \in \mathcal{X}} \int_{\mathcal{X}} d(x, x')^r d\mu(x), \qquad (1)$$

for every $0 < r < \infty$, and where $\Theta^r \subseteq \mathcal{X}$. Observe that we are using the superscript r on the Fréchet variance as a simple marker of the order of the exponentiated metric. Thus, in general, it will not be true that $(\sigma^r)^{1/r}$ simplifies to σ^1.

These are commonly referred to as the Fréchet mean and variance when $r = 2$. For other choices of r, we will refer to these different Fréchet moments as Fréchet moments of *order r*. Note that if the *infimum* of $\mathbb{E}[d(x, x')^r]$ exists, then it is unique. However, the *argument of the infimum* may not necessarily exist and may not be unique. If such an argument does not exist, then $\Theta^r = \varnothing$. When the minimizer is not unique, the ensemble of minimizers is sometimes referred to as the *Fréchet mean set*. In particular, observe that if Θ is not a singleton, $\sigma^2 = \mathbb{E}[d(X, \theta)^2]$ for any $\theta \in \Theta$, will not, in general, be equivalent to $\mathbb{E}[d(X, \Theta)^2]$, where the distance between an element x and a non-empty subset A of $\mathcal{X}$ is defined as $d(x, A) := \inf\{d(x, y) : y \in A\}$, with $d(x, \varnothing) = \infty$. In this paper, Fréchet mean and Fréchet mean set will be used interchangeably. Observe that when $\mathcal{X}$ is a Hilbert space, endowed with the inner product metric, then there exists a unique global minimizer and Θ is therefore a singleton.

Analogously, for a given sequence of abstract-valued random variables $X_i : \Omega \mapsto \mathcal{X}$, for every $i = 1, \ldots, n$, one may define the following Fréchet sample moments of the rth order

$$\widehat{\Theta}_n^r := \operatorname*{arginf}_{x' \in \mathcal{X}} \frac{1}{n} \sum_{i=1}^{n} d(X_i, x')^r \quad \text{and} \quad \widehat{\sigma}_n^r := \inf_{x' \in \mathcal{X}} \frac{1}{n} \sum_{i=1}^{n} d(X_i, x')^r. \qquad (2)$$

Observe that, even for the sample versions of the Fréchet moments, these infima need not be attained, and therefore these quantities may be empty for each n. When there is no ambiguity as to the order of $\widehat{\Theta}_n^r$, we will simply refer to this quantity as $\widehat{\Theta}_n$, and similarly for Θ. In the sequel, an element of Θ and an element of $\widehat{\Theta}_n$ will be respectively denoted by θ and $\hat{\theta}_n$. Our interest will mainly lie in considering Fréchet moments of the second order, albeit some examples will also be studied where $r = 1$. It is easy to see that the Fréchet mean and Fréchet sample mean are closed subsets of $\mathcal{X}$, if $\mathcal{X}$ is Polish.

Lemma 1. *For any space $(\mathcal{X}, d)$, Θ^r and the $\widehat{\Theta}_n^r$'s are closed in $\mathcal{X}$, for every $r \geq 1$.*

Proof. Clearly, if $\Theta^r = \varnothing$, then $\operatorname{cl}(\Theta^r) = \Theta^r$ and similarly for the $\widehat{\Theta}_n^r$'s. Now, fix $r = 1$, and consider the Fréchet mean set $\Theta \subseteq \mathcal{X}$. Recall that the boundary of Θ is defined as $\partial(\Theta) := \{x \in \mathcal{X} : d(\Theta, x) = d(\Theta^C, x) = 0\}$, where $\Theta^C := \mathcal{X} \setminus \Theta$. We proceed by contradiction. Assume that $\theta_0 \in \partial(\Theta)$ and $\theta_0 \notin \Theta$, then it follows that there exists $\theta \in \Theta$, such that by the triangle inequality, $d(\theta_0, X) \leq d(\theta_0, \theta) + d(\theta, X)$, for every $X \in \mathcal{X}$. Taking the expectation, this gives

$$\mathbb{E}[d(\theta_0, X)] \leq d(\theta_0, \theta) + \mathbb{E}[d(\theta, X)] = \inf_{x' \in \mathcal{X}} \mathbb{E}[d(X, x')],$$

since $d(\theta_0, \Theta) = 0$, and using the definition of Θ in equation (1). Thus, θ_0 is optimal with respect to the infimum over $\mathcal{X}$. However, we have assumed that $\theta_0 \notin \Theta$, which leads to a contradiction, and therefore $\partial(\Theta) \subseteq \Theta$.

Next, consider the case of $r > 1$. Through a classical result on metric spaces (see, for instance Fréchet, 1948, p.229), we have

$$\left(\mathbb{E}[d(\theta_0, X)^r]\right)^{1/r} \leq \left(\mathbb{E}[d(\theta_0, \theta)^r]\right)^{1/r} + \left(\mathbb{E}[d(\theta, X)^r]\right)^{1/r},$$

for every $r > 1$, and the result immediately follows, using the same argument. The proof is identical for the $\widehat{\Theta}_n^r$'s. $\square$

3.2 Convergence of Fréchet Sample Mean Sets

In this section, we study and compare different modes of convergence for set-valued random variables. In particular, note that our chosen modes of convergence differ from the ones used by Bhattacharya and Bhattacharya (2012), since we are not here assuming the compactness of the underlying metric space $\mathcal{X}$. Moreover, the target Fréchet mean set is also allowed to be empty, thereby making it difficult to implement the methods of Bhattacharya and Bhattacharya (2012).

For the Fréchet sample mean and its theoretical analogue, a.s. convergence could be defined in $(\mathcal{X}, d)$ using sequences of random sets as follows,

$$\mathbb{P}\left[\left\{\omega \in \Omega : \widehat{\Theta}_n(\omega) \to \Theta\right\}\right] = 1, \tag{3}$$

where observe that Θ is here treated as a fixed subset of $\mathcal{X}$. The event in equation (3) will have probability one if the sequence of random sets, denoted $\widehat{\Theta}_n$, converges a.s. in a set-theoretical sense such that

$$\liminf_{n \to \infty} \widehat{\Theta}_n(\omega) = \limsup_{n \to \infty} \widehat{\Theta}_n(\omega) = \Theta, \tag{4}$$

for almost every $\omega \in \Omega$, and where $\liminf S_n := \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} S_n$, and $\limsup S_n := \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} S_n$ denote the standard inner and outer limits of a sequence of subsets of $\mathcal{X}$. For most purposes, however, this type of convergence is too strong. In fact, this criterion does not hold for Fréchet sample means defined with respect to general abstract-valued random variables. There are many non-trivial examples of sequences of Fréchet sample means that diverge. Consider the following example adapted from the three-dimensional case described by Sverdrup-Thygeson (1981).

Example 1. Let the interval, $\mathcal{X} := [-1, 1] \subset \mathbb{R}$, and equip this set with the usual Manhattan distance, defined as $d(x, y) := |x - y|$ for every $x, y \in \mathcal{X}$. Additionally, let the random variable X, which takes values in $\mathcal{X}$, and which satisfies the following $\mathbb{P}[X = -1] = \mathbb{P}[X = 1] = 1/2$. This construction is illustrated in panel (a) of figure 3. The theoretical Fréchet mean of order $r = 1$ can be readily found as

$$\Theta^1 = \operatorname*{arginf}_{x' \in \mathcal{X}} \sum_{x \in \{-1, 1\}} d(x, x')\mathbb{P}[x] = \mathcal{X},$$

since the energy function satisfies $\mathcal{E}(x') := \sum d(x, x')\mathbb{P}[x] = 1$ for every $x' \in \mathcal{X}$. Here, the Fréchet mean defined with respect to the Manhattan distance coincides with the *median* of the real-valued random variable X (Feldman and Tucker, 1966).

For the empirical Fréchet mean, $\widehat{\Theta}_n^1$, first compute $S_n := \sum_{i=1}^{n} X_i$. Clearly, the S_n's are integer-valued. Observe the correspondence between the values of S_n and the values taken by the Fréchet sample mean. If the event $\{S_n = 0\}$ occurs, then
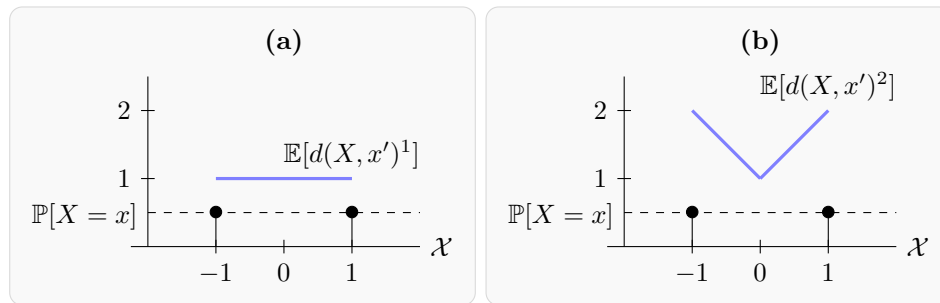
Figure 3. Metric and measure spaces considered in examples 1 and 2. In both panels, the closed interval $[-1, 1]$ is equipped with the Manhattan (or taxicab) metric, and two point masses are specified at -1 and 1. Different Fréchean inferences are conducted by taking $r = 1$ and $r = 2$ in panels (a) and (b), respectively. In the first case, the theoretical Fréchet mean coincides with the median of X, whereas in panel (b), the theoretical Fréchet mean coincides with the arithmetic mean. However, the sequence of Fréchet *sample* means diverge in both cases, when convergence is evaluated using set-valued liminf and limsup, as described in equation (4).

it can easily be seen that $\widehat{\Theta}_n$ is equal to $\mathcal{X}$. Similarly, $\{S_n \geq 1\}$, and $\{S_n \leq -1\}$ respectively imply that $\hat{\theta}_n = 1$ and $\hat{\theta}_n = -1$. Now,

$$\mathbb{P}\left[\{S_{2n} = 0\}\right] = \binom{2n}{n}\left(\frac{1}{2}\right)^{2n} \approx (n\pi)^{-1/2},$$

for every n, using Stirling's approximation. Since $\mathbb{P}\left[\{S_n = 0\}\right]$ is null, when n is odd, it follows that $\sum_{n=1}^{\infty} \mathbb{P}\left[\{S_n = 0\}\right] < \infty$, and therefore by the Borel-Cantelli lemma, we have $\mathbb{P}\left[\{S_n = 0\} \text{ i.o.}\right] = 0$, where i.o. means infinitely often. This implies that $\mathbb{P}[\{\widehat{\Theta}_n = \mathcal{X}\} \text{ i.o.}] = 0$, and hence $\limsup \widehat{\Theta}_n \neq \mathcal{X}$.

By using a similar argument, one can observe that $\mathbb{P}[\{S_n \leq -1\} \text{ i.o.}] = \mathbb{P}[\{S_n \geq 1\} \text{ i.o.}] = 1$, which implies that $\mathbb{P}[\{\hat{\theta}_n = -1\} \text{ i.o.}] = \mathbb{P}[\{\hat{\theta}_n = 1\} \text{ i.o.}] = 1$, and therefore $\{-1, 1\}$ is the limit superior of the sequence of Fréchet mean sets. By contrast, there does not exist an $N > 0$, such that $\hat{\theta}_n = 1$, for every $n \geq N$. An identical statement holds for $\hat{\theta}_n = -1$, and therefore the limit inferior of $\widehat{\Theta}_n$ is empty. Thus,

$$\limsup_{n \to \infty} \widehat{\Theta}_n(\omega) = \{-1, 1\} \supset \liminf_{n \to \infty} \widehat{\Theta}_n(\omega) = \varnothing,$$

and the sequence of Fréchet sample means diverges, as criterion (4) is not satisfied.

The preceding example highlights two important aspects of the asymptotic behavior of the Fréchet sample mean set. Firstly, the Fréchet sample mean will in general fail to converge in the sense that its outer and inner limits need not be identical. In such cases, the sequence of Fréchet sample means exhibit an oscillatory property (see Feldman and Tucker, 1966). Secondly, the limit superior of a sequence of Fréchet sample means may solely represent a subset of the theoretical Fréchet mean. Taken together, these two problems necessitate (i) the study of the asymptotic behavior of the *outer limit* of the $\widehat{\Theta}_n$'s, and (ii) the consideration of the convergence of the Fréchet sample mean in terms of *set inclusion*, as a subset of the theoretical Fréchet mean. The passage from equations to inclusions is a natural step in the generalization of singleton-valued analysis to set-valued analysis.

Example 1 leads to the formulation of a weaker type of convergence, which can be expressed as the probability of the following event,

$$\left\{\omega \in \Omega : \limsup_{n \to \infty} \widehat{\Theta}_n(\omega) \subseteq \Theta\right\}. \tag{5}$$

However, we here encounter a slightly different problem than the one highlighted in our first example. This second issue can be illustrated through another counterexample, which shows that this particular type of a.s. convergence does not agree with the analogous real-valued a.s. convergence. That is, the reformulation of a given real-valued random variable into an abstract-valued setting, equipped with the same topology produces a divergent Fréchet sample mean in terms of equation (5). As a result, we obtain the somewhat counterintuitive result that the arithmetic sample mean differs from the corresponding Fréchet sample mean.

Example 2. Consider the same setting described in example 1, where now $r = 2$ (see panel (b) of figure 3). One can immediately see that the theoretical Fréchet mean is a singleton set,

$$\Theta^2 = \underset{x' \in \mathcal{X}}{\arg\inf} \sum_{x \in \{-1,1\}} d(x, x')^2 \mathbb{P}[x] = 0,$$

which coincides with the expected value of the real-valued random variable X. For the Fréchet sample mean, we know from example 1 that $\mathbb{P}[\{S_n = 0\} \text{ i.o.}] = 0$ and therefore the probability of the sequence of empirical Fréchet means including $\mathbb{E}[X]$ infinitely often is null. That is, for $r = 2$, we have $\mathbb{P}[\{\hat{\theta}_n = 0\} \text{ i.o.}] = 0$. Observe that the same is true for any other specific sequence of realizations of X. Consider the case of $S_{3n} = nx_1 + 2nx_2$, where $x_1 = -1$ and $x_2 = 1$. For this subsequence, there exists a unique infimum, which is $\hat{\theta}_n = 1/3$. The probability of this event occurring is as follows,

$$\mathbb{P}\left[\{S_{3n} = nx_1 + 2nx_2\}\right] = \binom{3n}{n}\left(\frac{1}{2}\right)^{3n} \approx (1/2)^{5n},$$

which was approximated using Stirling's formulae. Clearly, all possible values of the Fréchet sample mean of X can be represented as a formulae of the form $nx_1 + \alpha nx_2$, for some $\alpha \in \mathbb{N}$. Using the Borel-Cantelli lemma, it therefore follows that there does not exist a point in $[-1, 1]$ that $\hat{\theta}_n$ will visit infinitely often, and hence $\limsup \widehat{\Theta}_n = \liminf \widehat{\Theta}_n = \varnothing$. By contrast, the arithmetic sample mean, $\bar{X}_n := n^{-1} \sum_{i=1}^n X_i$ trivially converges to the expected value of X a.s., since for every $\epsilon > 0$, there exists an $N > 1$, for which $d(\bar{X}_n(\omega), \mathbb{E}[X]) < \epsilon$, for every $n \geq N$, for almost every $\omega \in \Omega$. Thus, for this example, we reach the counterintuitive conclusion that $\bar{X}_n \notin \limsup \widehat{\Theta}_n$, for every n.

This paradoxical disagreement between the divergence of the Fréchet sample mean and the classical convergence of the arithmetic sample mean in such a simple example requires a strengthening of our definition of the a.s. convergence of $\widehat{\Theta}_n$. This particular problem seemed to have been implicitly identified by Ziezold (1977), as this author proposed the following type of convergence, which specializes the event presented in equation (5),

$$\left\{ \omega \in \Omega : \bigcap_{n=1}^{\infty} \overline{\bigcup_{m=n}^{\infty} \widehat{\Theta}_m(\omega)} \subseteq \Theta \right\}, \tag{6}$$

where $\overline{A}$ indicates the *closure* of set A in $\mathcal{X}$. For convenience, this particular type of convergence will be denoted by $\lim\overline{\sup} \, \widehat{\Theta}_n \subseteq \Theta$, a.s., where the limsup operator is here defined with respect to set inclusion on the power set of $\mathcal{X}$. It is easy to see why definition (6) resolves the issue illustrated in example 2. By taking the closure

of $\bigcup_{m=n}^{\infty} \widehat{\Theta}_m$, we include all the elements for which there exists a sequence of $\hat{\theta}_n$'s converging to $\mathbb{E}[X]$, and therefore for real-valued random variables,

$$\mathbb{E}[X] \in \overline{\bigcup_{m=n}^{\infty} \widehat{\Theta}_m},$$

for every n, which implies that $\lim\overline{\sup} \, \widehat{\Theta}_n = \{\mathbb{E}[X]\}$, as desired, thereby ensuring complete agreement between the classical and Fréchet inferential approaches for this particular example. Note that these issues are neither related to the completeness of the underlying space of interest, nor associated to the question of the non-emptiness of Θ.

Since Sverdrup-Thygeson (1981) assumed that $\mathcal{X}$ is compact, it follows that Θ and $\widehat{\Theta}_n$ are non-empty, in this case. The separability of $\mathcal{X}$ is not sufficient to ensure that Θ and the $\widehat{\Theta}_n$'s are non-empty. Nonetheless, observe that if $\widehat{\Theta}_n = \varnothing$, then the events in equations (5) and (6) are trivially almost certain, since $\varnothing \subseteq A$, for all $A \subseteq \mathcal{X}$.

3.3 Kuratowski Upper Limit

It can easily be shown that the type of convergence envisaged by Ziezold (1977) is, in fact, equivalent to the celebrated upper limit introduced by Kuratowski (1966), which has been adopted as the preferred type of convergence in set-valued analysis (see Aubin and Frankowska, 2009). The Kuratowski upper limit is defined over a metric space $(\mathcal{X}, d)$, for some sequence of subsets $A_n \subseteq \mathcal{X}$, as follows

$$\begin{aligned} \mathrm{Limsup}_{n\to\infty} A_n &:= \left\{ x \in \mathcal{X} : \liminf_{n\to\infty} d(x, A_n) = 0 \right\} \\ &= \left\{ x \in \mathcal{X} : \{A_n \cap N_\epsilon(x) \neq \varnothing\} \text{ i.o., } \forall \, \epsilon > 0 \right\}, \end{aligned} \tag{7}$$

where liminf and Limsup are taken with respect to real numbers and subsets of $\mathcal{X}$, respectively, and with $N_\epsilon(x) := \{x' \in \mathcal{X} : d(x, x') < \epsilon\}$. The second formulation of Limsup in equation (7) immediately follows from the positivity of the metric. Also, observe that the Kuratowski upper limit is equivalent to the set of *cluster points* of the sequences, $x_n \in A_n$ (Aubin and Frankowska, 2009). Clearly, the Kuratowski upper limit of any sequence of sets is *closed*, and moreover, it contains the conventional set-theoretical upper limit, such that for any sequence of random sets A_n,

$$\limsup_{n\to\infty} A_n \subseteq \mathrm{Limsup}_{n\to\infty} A_n.$$

Importantly, it can be easily shown that the Kuratowski upper limit and the quantity studied by Ziezold (1977) are equivalent, as stated in the following lemma.

Lemma 2. *Given a metric space $(\mathcal{X}, d)$, for any sequence of sets $A_n \subseteq \mathcal{X}$,*

$$\lim\overline{\sup}_{n\to\infty} A_n = \mathrm{Limsup}_{n\to\infty} A_n.$$

Proof. Clearly, $\lim\overline{\sup} A_n = \varnothing$, if and only if, $\mathrm{Limsup} \, A_n = \varnothing$. Thus, assume that these two outer limits are non-empty, and choose $x_0 \in \lim\overline{\sup} A_n$. Then, $x_0 \in \overline{\bigcup_{m=N}^{\infty} A_m}$ for every N and there exists a subsequence x_k such that $x_k \in A_{n_k}$, for every k, which satisfies $x_k \to x_0$. Hence, we have $\liminf d(x_0, A_n) = 0$, and by definition (7), $\lim\overline{\sup} A_n \subseteq \mathrm{Limsup} \, A_n$.

Conversely, choose $x_0 \in \text{Limsup}\, A_n$. Then, there exists a subsequence x_k such that $x_k \in A_{n_k} \cap N_\epsilon(x_0)$, for every k and for every $\epsilon > 0$, which satisfies $x_k \to x_0$, as $k \to \infty$. This implies that $x_0 \in \bigcap_{N=1}^\infty \overline{\bigcup_{m=N}^\infty A_m}$, and therefore $\lim\overline{\sup}\, A_n \supseteq \text{Limsup}\, A_n$, which completes the proof. □

Observe that $\text{Limsup}\, A_n$ can be empty. Consider the following diverging sequence of sets, $A_n := [n-1, n+1]$, for every n. It is immediate that $\text{Limsup}\, A_n = \varnothing$. Throughout the rest of the paper, we will neither assume the existence nor the uniqueness of Θ^r and the Θ_n^r's. In particular, in the sequel, Θ^r may be empty, a subset of $\mathcal{X}$, or a singleton set.

4. Almost Sure Consistency of Fréchet Sample Mean

In this section, we prove a strong law of large numbers for sample Fréchet means in spaces having a bounded metric. This result can be regarded as an adaptation of Ziezold's (1977) original result to spaces equipped with a bounded metric. This new proof also allows us to re-formulate Ziezold's theorem using the Kuratowski upper limit.

Theorem 1. *Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a separable bounded metric space $(\mathcal{X}, d)$, let $X_1, \ldots, X_n$ be a sequence of independent and identically distributed (iid) abstract-valued random variables, such that $X_i : \Omega \mapsto \mathcal{X}$, for every X_i. Then,*

$$\widehat{\sigma}_n^r \to \sigma^r \quad a.s., \qquad and \qquad \underset{n\to\infty}{\text{Limsup}}\ \widehat{\Theta}_n^r \subseteq \Theta^r \quad a.s.,$$

for every finite $r \geq 1$, and where Limsup is defined as in equation (7).

The particular mode of convergence of the Fréchet sample mean used in theorem 1 will sometimes be denoted by $X_n \overset{a.s.}{\to} X$, which implies that $\text{Limsup}\, X_n \subseteq X$ with probability one. Observe that the integrability of the rth order metric is implied by the finiteness of both d and μ. Since $d(x, y) \leq M$, for every $x, y \in \mathcal{X}$, we have for any arbitrary $\alpha \in \mathcal{X}$ and finite $r \geq 1$,

$$\mathbb{E}[d(X, \alpha)^r] = \int_\mathcal{X} |d(x, \alpha)|^r d\mu(x) \leq \int_\mathcal{X} M^r d\mu(x) = M^r \mu(\mathcal{X}) < \infty,$$

by the linearity of the Lebesgue integral, and the fact that μ is a probability measure. The integrability of the exponentiated metric was not explicitly assumed by Sverdrup-Thygeson (1981). This author, however, assumed that $\mathcal{X}$ is compact, which implies that d^r is integrable for any finite $r \geq 1$.

The key to the proof of theorem 1 is based on a classical result, due to Rao (1962), which stipulates the conditions under which the weak convergence of a probability measure is equivalent to the uniform convergence of a probability measure, in a sense made clear in theorem 2. This can be seen as a generalization of the Glivenko-Cantelli lemma to random variables taking values in separable metric spaces (see also Parthasarathy, 1967, chap. 2). In this result, we will need to define a class of functions on the separable space $\mathcal{X}$, which we will denote by $\mathcal{F} := \mathcal{F}(\mathcal{X})$, whereby every $f \in \mathcal{F}$ is a real-valued continuous function that satisfies $f : \mathcal{X} \mapsto \mathbb{R}$. Such a class of functions is said to be *uniformly bounded* when for every $f \in \mathcal{F}$, and every $x \in \mathcal{X}$, there exists an $M \in \mathbb{R}$, such that $f(x) \leq M$. In addition, $\mathcal{F}$ is *equicontinuous at a point* $x_0 \in \mathcal{X}$, if for every $\epsilon > 0$, there exists $\delta(x_0) > 0$, such that for every $u \in N_\delta(x_0) := \{u \in \mathcal{X} : d(x_0, u) < \delta\}$, we have $|f(x) - f(u)| < \epsilon$, for

every $f \in \mathcal{F}$. The class $\mathcal{F}$ is said to be equicontinuous if it is equicontinuous for every $x \in \mathcal{X}$. Finally, $\mathcal{F}$ is said to be *uniformly equicontinuous* if δ does not depend on x_0. We will denote the collection of all finite measures on $\mathcal{B}$ by $\mathcal{M}(\mathcal{B})$, and $\Rightarrow$ will indicate weak convergence.

Theorem 2 (Rao, 1962, p.672). *Let $\mathcal{F}(\mathcal{X})$ be a class of real-valued functions on a separable space $\mathcal{X}$, and assume that $\mathcal{F}(\mathcal{X})$ is (i) dominated by a continuous integrable function on $\mathcal{X}$, and that (ii) $\mathcal{F}(\mathcal{X})$ is equicontinuous. If, for some sequence of measures $\mu_n \in \mathcal{M}(\mathcal{B})$, and $\mu \in \mathcal{M}(\mathcal{B})$, we have $\mu_n \Rightarrow \mu$, a.s., then*

$$\lim_{n \to \infty} \sup_{f \in \mathcal{F}} \left| \int f d\mu_n - \int f d\mu \right| = 0, \quad a.s..$$

The following lemma will be used in the proof of theorem 1. This result links the properties of a bounded metric space with the conditions required in Rao's (1962) theorem. For this purpose, we will require the following classes of *point functions* on a metric space (see Searcóid, 2007).

Definition 1. *For any metric space $(\mathcal{X}, d)$, the z-point function is defined as $d_z(x) := d(z,x)$ for every $x \in \mathcal{X}$. The class of point functions on $(\mathcal{X}, d)$ is then denoted by $\mathcal{D}(\mathcal{X}) := \{d_z : \forall z \in \mathcal{X}\}$. Similarly, we will make use of the class of exponentiated point functions, defined as follows,*

$$\mathcal{D}^r(\mathcal{X}) := \{d_z^r : \forall z \in \mathcal{X}\},$$

for every finite $r \geq 1$, and where elements in either $\mathcal{D}$ or $\mathcal{D}^r$ will be denoted by d_z, and d_z^r, respectively.

Lemma 3. *If $(\mathcal{X}, d)$ is a bounded metric space, then $\mathcal{D}^r(\mathcal{X})$ is uniformly bounded and uniformly equicontinuous for every finite $r \geq 1$.*

Proof. By the boundedness of $(\mathcal{X}, d)$, there exists an $M \in \mathbb{R}$, such that $d(x,y) \leq M$, for every $x, y \in \mathcal{X}$. Therefore, $d_z(x) \leq M$, for every $x \in \mathcal{X}$, for every $d_z \in \mathcal{D}$, and thus $\mathcal{D}$ is uniformly bounded. Moreover, since $d_z^r(x) \leq M^r < \infty$, for every finite $r \geq 1$, it follows that each $\mathcal{D}^r$ also forms a uniformly bounded class of functions. Next, by the reverse triangle inequality, we have $|d_z(x) - d_z(x_0)| \leq d(x, x_0)$, for all $x, x_0, z \in \mathcal{X}$, thereby proving the (uniform) equicontinuity of the class $\mathcal{D}$ on $\mathcal{X}$. For the case of $r \geq 1$, we consider the exponentiated version of the triangle inequality. Using the binomial expansion,

$$d(z,x)^r \leq \Big(d(z,x_0) + d(x_0,x) \Big)^r$$

$$= d(z,x_0)^r + \sum_{k=1}^{r-1} \binom{r}{k} d(z,x_0)^{r-k} d(x_0,x)^k + d(x_0,x)^r.$$

Similarly, for any given $x_0 \in \mathcal{X}$, $d(z,x_0)^r \leq d(z,x)^r + \sum_{k=1}^{r-1} \binom{r}{k} d(z,x)^{r-k} d(x,x_0)^k + d(x,x_0)^r$. Combining these two inequalities and invoking the symmetry of d, we have

$$|d(z,x)^r - d(z,x_0)^r| \leq d(x_0,x)^r + d(x_0,x) M^{r-1} \sum_{k=1}^{r-1} \binom{r}{k}$$

$$\leq d(x_0,x) M^{r-1} \left(1 + \sum_{k=1}^{r-1} \binom{r}{k} \right),$$

where M is the uniform bound on the class $\mathcal{D}$. Now, choose $\delta = \epsilon/\gamma M^{r-1}$, where $\gamma := 1 + \sum_{k=1}^{r-1} \binom{r}{k}$, such that if $d(x, x_0) < \delta$, then $|d_z^r(x) - d_z^r(x_0)| < \gamma \delta M^{r-1} = \epsilon$, for every $x \in N_\delta(x_0)$, for every $d_z^r \in \mathcal{D}^r$, thence proving the equicontinuity of $\mathcal{D}^r$ at x_0. Since δ did not depend on the choice of x_0, it follows that $\mathcal{D}^r$ is also uniformly equicontinuous. $\square$

Proof of Theorem 1. Observe that the theorem is trivially verified if $\mathrm{Limsup}\,\widehat{\Theta}_n^r = \varnothing$. Thus, assume that $\mathrm{Limsup}\,\widehat{\Theta}_n^r$ is non-empty. We here adopt the line of argument followed by Sverdrup-Thygeson (1981). However, since we are not assuming compactness, there are several aspects of Sverdrup-Thygeson's proof that becomes somewhat delicate. In the sequel, we will make use of the following quantities formulated with respect to the class of point functions described in definition 1. For every $z \in \mathcal{X}$, let

$$T_n(z) := \frac{1}{n}\sum_{i=1}^{n} d_z^r(X_i) - \int_{\mathcal{X}} d_z^r(x)d\mu(x), \tag{8}$$

and similarly,

$$T_n^*(z) := \frac{1}{n}\sum_{i=1}^{n} d_z^r(X_i) - \int_{\mathcal{X}} d_\theta^r(x)d\mu(x). \tag{9}$$

Since $T_n(x)$ is real-valued, one can invoke the strong law of large numbers for real-valued random variables, which gives

$$T_n(z) \to 0, \quad \text{a.s.,} \quad \forall\, z \in \mathcal{X}. \tag{10}$$

Note, however, that since we have used infima in the definitions of the Fréchet theoretical and sample means in equations (1) and (2), it follows that the convergence of $T_n(z) \to 0$ is not assured when z is an element of Θ or an element of $\widehat{\Theta}_n$. However, as established in lemma 3, the class of point functions, $\mathcal{D}^r(\mathcal{X})$, is uniformly bounded and (uniformly) equicontinuous. Moreover, we have seen that the finiteness of $\mathbb{E}[d_z^r(X)]$ is implied by the boundedness of d, such that $\mathbb{E}[d_z^r(X)] \leq M^r\mu(\mathcal{X})$. Thus, it follows that there exists a continuous integrable function, i.e. $f(x) := M^r$, dominating every $d_z^r \in \mathcal{D}^r$. Moreover, a classical result on the convergence of empirical measures based on iid random variables taking values in separable metric spaces (see Parthasarathy, 1967, theorem 7.1, p.53) implies that

$$\mu_n \Rightarrow \mu, \quad \text{a.s.,} \tag{11}$$

where $\mu_n := n^{-1}\sum_{i=1}^{n} \delta_{X_i}$, is the empirical measure on $\mathcal{X}$. Therefore, we are in a position to apply theorem 2, which shows that the empirical measure, μ_n, converges uniformly with probability 1. That is,

$$\mathbb{P}\left[\sup_{z \in \mathcal{D}^r}\left|\frac{1}{n}\sum_{i=1}^{n} d_z^r(X_i) - \int_{\mathcal{X}} d_z^r(x)d\mu(x)\right| \to 0\right] = 1,$$

which may be re-written as

$$\sup_{z \in \mathcal{D}^r}\left|T_n(z)\right| = \sup_{z \in \mathcal{X}}\left|T_n(z)\right| \to 0, \quad \text{a.s..} \tag{12}$$

Consequently, $T_n(\hat{\theta}_n) \to 0$, a.s., and $T_n(\theta) \to 0$, a.s., for every $\hat{\theta}_n \in \widehat{\Theta}_n$ and every $\theta \in \Theta$, respectively.

Further, from the definition of $\hat{\theta}_n$ and θ, we can 'sandwich' $T_n^*(\hat{\theta}_n)$ in the following manner. Firstly, observe that by the minimality of the θ's,

$$
\begin{aligned}
T_n(\hat{\theta}_n) &= \frac{1}{n} \sum_{i=1}^{n} d_{\hat{\theta}_n}^r(X_i) - \int_{\mathcal{X}} d_{\hat{\theta}_n}^r(x) d\mu(x) \\
&\leq \frac{1}{n} \sum_{i=1}^{n} d_{\hat{\theta}_n}^r(X_i) - \int_{\mathcal{X}} d_{\theta}^r(x) d\mu(x) = T_n^*(\hat{\theta}_n).
\end{aligned}
\tag{13}
$$

Secondly, by the minimality of the $\hat{\theta}_n$'s, we similarly have,

$$
\begin{aligned}
T_n^*(\hat{\theta}_n) &= \frac{1}{n} \sum_{i=1}^{n} d_{\hat{\theta}_n}^r(X_i) - \int_{\mathcal{X}} d_{\theta}^r(x) d\mu(x) \\
&\leq \frac{1}{n} \sum_{i=1}^{n} d_{\theta}^r(X_i) - \int_{\mathcal{X}} d_{\theta}^r(x) d\mu(x) = T_n(\theta).
\end{aligned}
\tag{14}
$$

Thence, combining equations (13) and (14), we obtain,

$$
T_n(\hat{\theta}_n) \leq T_n^*(\hat{\theta}_n) \leq T_n(\theta),
$$

such that, using equation (12),

$$
|T_n^*(\hat{\theta}_n)| \leq \max\{|T_n(\hat{\theta}_n)|, |T_n(\theta)|\} \to 0, \quad \text{a.s.}, \tag{15}
$$

which proves the a.s. convergence of $\widehat{\sigma}_n^r$ to σ^r.

We now turn to the convergence properties of the Fréchet sample mean of the rth order, $\widehat{\Theta}_n^r$. Here, we generalize Ziezold's (1977) proof strategy to Fréchet sample means of any order (see also Molchanov, 2005, p.185). Choosing

$$
\hat{\theta} \in \underset{n \to \infty}{\text{Limsup}} \ \widehat{\Theta}_n^r,
$$

it then suffices to show that $\hat{\theta} \in \Theta^r$, which is verified if $\mathbb{E}[d(X, \hat{\theta})^r] \leq \mathbb{E}[d(X, x')^r]$, for every $x' \in \mathcal{X}$. We proceed by constructing the following subsequence of natural numbers.

Observe that from the definition of the Kuratowski upper limit and the equivalence relation reported in lemma 2, it follows that $\hat{\theta} \in \text{Cl}(\bigcup_{m=n}^{\infty} \widehat{\Theta}_m^r)$, for every n, where $\text{Cl}(\cdot)$ denotes the closure of a set. Thus, one can construct a subsequence, $\{n_k : k \in \mathbb{N}\}$, such that for every k, there exists an element $\hat{\theta}_k \in \bigcup_{m=k}^{\infty} \widehat{\Theta}_m^r$, which satisfies $d(\hat{\theta}_k, \hat{\theta}) \leq 1/k$. Moreover, we can define $n_k := \min\{n \in \mathbb{N} : n \geq k, \hat{\theta}_k \in \widehat{\Theta}_n^r\}$. Now, after an application of the triangle inequality, followed by the Minkowski inequality, we have

$$
\left(\frac{1}{n_k} \sum_{i=1}^{n_k} d(X_i, \hat{\theta})^r \right)^{1/r} \leq \left(\frac{1}{n_k} \sum_{i=1}^{n_k} d(X_i, \hat{\theta}_k)^r \right)^{1/r} + \left(\frac{1}{n_k} \sum_{i=1}^{n_k} d(\hat{\theta}_k, \hat{\theta})^r \right)^{1/r},
$$

which gives

$$
\left(\frac{1}{n_k} \sum_{i=1}^{n_k} d(X_i, \hat{\theta})^r \right)^{1/r} \leq \left(\frac{1}{n_k} \sum_{i=1}^{n_k} d(X_i, \hat{\theta}_k)^r \right)^{1/r} + \frac{1}{k}.
$$

As $k \to \infty$, it then follows from equation (12) that since $(n_k)_{k \in \mathbb{N}}$ is a subsequence of $(n)_{n \in \mathbb{N}}$, we obtain

$$\left(\mathbb{E}[d(X, \hat{\theta})^r] \right)^{1/r} \leq \liminf_{k \to \infty} \left(\frac{1}{n_k} \sum_{i=1}^{n_k} d(X_i, \hat{\theta}_k)^r \right)^{1/r}, \tag{16}$$

where liminf is here taken with respect to non-negative real numbers. Moreover, by construction, each $\hat{\theta}_k$ is minimal with respect to any element $x' \in \mathcal{X}$, such that

$$\frac{1}{n_k} \sum_{i=1}^{n_k} d(X_i, \hat{\theta}_k)^r \leq \frac{1}{n_k} \sum_{i=1}^{n_k} d(X_i, x')^r, \tag{17}$$

for every $x' \in \mathcal{X}$ and $k \in \mathbb{N}$. Observe that given the continuity and monotonicity of $g(x) := x^{1/r}$ on positive real numbers, we have $\liminf g(x_n) = g(\liminf x_n)$, for every sequence satisfying $x_n \in \mathbb{R}^+$. Therefore, it suffices to combine equations (16) and (17) in order to obtain $\mathbb{E}[d(X, \hat{\theta})^r] \leq \mathbb{E}[d(X, x')^r]$, for every $x' \in \mathcal{X}$, as required. Thence, $\hat{\theta} \in \Theta^r$ a.s., but since $\hat{\theta}$ was arbitrary, we have $\text{Limsup} \, \widehat{\Theta}_n^r \subseteq \Theta^r$ a.s., as required. $\qquad \square$

References

Aubin, J.P. and Frankowska, H. (2009). *Set-Valued Analysis*. Birkhäuser, Boston.

Bhattacharya, A. and Bhattacharya, R. (2012). *Nonparametric Inference on Manifolds with Applications to Shape Spaces*. Cambridge University Press, New York.

Bhattacharya, R. and Patrangenaru, V. (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds. II. *The Annals of Statistics*, **33**(3), 1225–1259.

Bhattacharya, R. and Patrangenaru, V. (2002). Nonparametic estimation of location and dispersion on Riemannian manifolds. *Journal of Statistical Planning and Inference*, **108**(12), 23–35.

Bhattacharya, R. and Patrangenaru, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds. I. *The Annals of Statistics*, **31**(1), 1–29.

Bunke, H. (1997). On a relation between graph edit distance and maximum common subgraph. *Pattern Recognition Letters*, **18**(8), 689–694.

Feldman, D. and Tucker, H. (1966). Estimation of non-unique quantiles. *Annals of Mathematical Statistics*, **37**(2), 451–457.

Fréchet, M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. *Annales de L'Institut Henri Poincaré*, **10(4)**, 215–310.

Gao, X., Xiao, B., Tao, D., and Li, X. (2010). A survey of graph edit distance. *Pattern Analysis and Applications*, **13**(1), 113–129.

Ginestet, C.E., Nichols, T.E., Bullmore, E.T., and Simmons, A. (2011). Brain network analysis: Separating differences in cost from differences in topology. *PLoS ONE*, **6(7)**, e21570.

Ginestet, C.E., Simmons, A., and Kolaczyk, E.D. (2012). Weighted frechet means as convex combinations in metric spaces: Properties and generalized median inequalities. *Statistics and Probability Letters*, **82**(10), 1859–1863.

Hamming, R. (1950). Error detecting and error correcting codes. *Bell System Technical Journal*, **29**, 147–160.

He, M., Petoukhov, S., and Ricci, P. (2004). Genetic code, hamming distance and stochastic matrices. *Bulletin of Mathematical Biology*, **66**(5), 1405–1421.

Karcher, H. (1977). Riemannian center of mass and mollifier smoothing. *Communication in Pure and Applied Mathematics*, **30**, 509–541.

Kendall, W. and Le, H. (2011). Limit theorems for empirical fréchet means of independent and non-identically distributed manifold-valued random variables. *Brazilian Journal of Probability and Statistics*, **25**(3), 323–352.

Kuratowski, K. (1966). *Topology. Volume I. New Edition, Revised and Augmented.* Academic Press, New York.

Le, H. (1998). On the consistency of procrustean mean shapes. *Advances in Applied Probability*, **30**(1), 53–63.

Molchanov, I. (2005). *Theory of Random Sets.* Springer, London.

Parthasarathy, K. (1967). *Probability Measures on Metric Spaces.* American Mathematical Society, London.

Rao, R.R. (1962). Relations between weak and uniform convergence of measures with applications. *The Annals of Mathematical Statistics*, **33**(2), 659–680.

Searcóid, M. (2007). *Metric spaces.* Springer, London.

Sverdrup-Thygeson, H. (1981). Strong law of large numbers for measures of central tendency and dispersion of random variables in compact metric spaces. *The Annals of Statistics*, **9**(1), 141–145.

Ziezold, H. (1977). On expected figures and a strong law of large numbers for random elements in quasi-metric spaces. *Transactions of the Seventh Prague Conference on Information Theory, Statistical Decision Functions, Random Processes and of the 1974 European Meeting of Statisticians.*