Strong Laws of Large Numbers for Generalizations of Fréchet Mean Sets

Christof Schötz

math@christof-schoetz.de

A Fréchet mean of a random variable Y with values in a metric space $(\mathcal{Q}, d)$ is an element of the metric space that minimizes $q \mapsto \mathbb{E}[d(Y, q)^2]$. This minimizer may be non-unique. We study strong laws of large numbers for sets of generalized Fréchet means. Following generalizations are considered: the minimizers of $\mathbb{E}[d(Y, q)^\alpha]$ for $\alpha > 0$, the minimizers of $\mathbb{E}[H(d(Y, q))]$ for integrals H of non-decreasing functions, and the minimizers of $\mathbb{E}[\mathfrak{c}(Y, q)]$ for a quite unrestricted class of cost functions $\mathfrak{c}$. We show convergence of empirical versions of these sets in outer limit and in one-sided Hausdorff distance. The derived results require only minimal assumptions.

1. Fréchet Mean Sets

For a random variable Y with values in $\mathbb{R}^s$ and $\mathbb{E}[\|Y\|^2] < \infty$, it holds

$$\mathbb{E}[Y] = \arg\min_{q \in \mathbb{R}^s} \mathbb{E}[d(Y, q)^2],$$

where $d(x, y) = \|x - y\|$ is the Euclidean distance. We can also write

$$\mathbb{E}[Y] = \arg\min_{q \in \mathbb{R}^s} \mathbb{E}[d(Y, q)^2 - d(Y, 0)^2],$$

as we just add a constant term. In the latter equation, we only require Y to be once integrable, $\mathbb{E}[\|Y\|] < \infty$, instead of twice as $|\|y - q\|^2 - \|y\|^2| \leq 2\|y\|\|q\| + \|q\|^2$.

The concept of Fréchet mean, proposed in [Fréchet, 1948], builds upon this minimizing property of the Euclidean mean to generalize the expected value to random variables with values in a metric space. Let $(\mathcal{Q}, d)$ be a metric space. As a shorthand we may write $\overline{q, p}$ instead of $d(q, p)$. Let Y be a random variable with values in $\mathcal{Q}$. Fix an arbitrary element $o \in \mathcal{Q}$. The *Fréchet mean set* of Y is $M = \arg\min_{q \in \mathcal{Q}} \mathbb{E}[\overline{Y, q}^2 - \overline{Y, o}^2]$ assuming the expectations exist. This definition does not depend on o. The reason for subtracting

1

$\overline{Y,o}^2$ is the same as in the Euclidean case: We need to make less moment assumptions to obtain a meaningful value: The triangle inequality implies

$$\left|\overline{y,q}^2 - \overline{y,o}^2\right| \le \overline{o,q}\left(\overline{o,q} + 2\overline{y,o}\right)$$

for all $y, q, o \in \mathcal{Q}$. Thus, if $\mathbb{E}[\overline{Y,o}] < \infty$, then $\mathbb{E}[\overline{Y,q}^2 - \overline{Y,o}^2] < \infty$ for all $q \in \mathcal{Q}$.

In Euclidean spaces and other Hadamard spaces (metric spaces with nonpositive curvature), the Fréchet mean is always unique [Sturm, 2003, Proposition 4.3]. This is not true in general. On the circle, a uniform distribution on two antipodal points has two Fréchet means. For a deeper analysis of Fréchet means on the circle, see [Hotz and Huckemann, 2015]. Similarly, Fréchet means on many positively curved spaces like (hyper-)spheres may not be unique. For the metric space $\mathcal{Q} = \mathbb{R}$ with $d(q,p) = \sqrt{|q - p|}$, the Fréchet mean set is the set of medians, which may also be non-unique. These examples underline the importance of considering sets of minimizers in a general theory instead of assuming uniqueness.

The notion of *Fréchet mean* can be generalized to cases where the cost function to be minimized is not a squared metric, e.g. [Huckemann, 2011]. We will not explicitly write down measurablity conditions, but silently demand that all spaces have the necessary measurable structure and all functions are measurable when necessary. Let $(\mathcal{Q}, d)$ be a metric space and $\mathcal{Y}$ be a set. Let $\mathfrak{c} \colon \mathcal{Y} \times \mathcal{Q} \to \mathbb{R}$ be a function. Let Y be a random variable with values in $\mathcal{Y}$. Let $M := \arg\min_{q \in \mathcal{Q}} \mathbb{E}[\mathfrak{c}(Y, q)]$ assuming the expectations exist. In this context, $\mathfrak{c}$ is called *cost function*, $\mathcal{Y}$ is called *data space*, $\mathcal{Q}$ is called *descriptor space*, $q \mapsto \mathbb{E}[\mathfrak{c}(Y, q)]$ is called *objective function* (or *Fréchet function*), and M is called *generalized Fréchet mean set* or $\mathfrak{c}$-*Fréchet mean set*.

This general scenario contains the setting of general M-estimation. It includes many important statistical frameworks like maximum likelihood estimation, where $\mathcal{Q} = \Theta$ parameterizes a family of densities $(f_\vartheta)_{\vartheta \in \Theta}$ on $\mathcal{Y} = \mathbb{R}^p$ and $\mathfrak{c}(x, \vartheta) = -\log f_\vartheta(x)$, or linear regression, where $\mathcal{Q} = \mathbb{R}^{s+1}$, $\mathcal{Y} = (\{1\} \times \mathbb{R}^s) \times \mathbb{R}$, $\mathfrak{c}((x, y), \beta) = (y - \beta^\top x)^2$. It also includes nonstandard settings, e.g. [Huckemann, 2011], where geodesics in $\mathcal{Q}$ are fitted to points in $\mathcal{Y}$.

Fix an arbitrary element $o \in \mathcal{Q}$. We will use cost functions $\mathfrak{c}(y, q) = H(\overline{y,q}) - H(\overline{y,o})$, where $H(x) = \int_0^x h(t)\mathrm{d}t$ for a non-decreasing function h, and $\mathfrak{c}(y, q) = \overline{y,q}^\alpha - \overline{y,o}^\alpha$ with $\alpha > 0$. In both cases the set of minimizers does not depend on o. We call the minimizers of the former cost function *H-Fréchet means*. In the latter case, we call the minimizers *power Fréchet means* or α-*Fréchet means*. We can interpret the different exponents $\alpha = 2$, $\alpha = 1$, $\alpha \to 0$, $\alpha \to \infty$ as mean, median, mode, and circumcenter (or midrange), respectively, see [MacQueen, 1967]. The minimizers for $\alpha = 1$ are sometimes called *Fréchet median*, e.g. [Arnaudon et al., 2013]. If $\mathcal{Q}$ is a Banach space, then they are called *geometric* or *spatial median*, e.g. [Kemperman, 1987]. H-Fréchet means serve as a generalization of α-Fréchet means for $\alpha > 1$ as well as an intermediate result for proving strong laws of large numbers for α-Fréchet mean sets with $\alpha \in (0, 1]$.

For a function $f \colon \mathcal{Q} \to \mathbb{R}$ and $\epsilon \ge 0$, define

$$\epsilon\text{-}\arg\min_{q \in \mathcal{Q}} f(q) := \left\{q \in \mathcal{Q} \,\middle|\, f(q) \le \epsilon + \inf_{q \in \mathcal{Q}} f(q)\right\}.$$

Let $Y_1, \ldots, Y_n$ be independent random variables with the same distribution as Y. Choose $(\epsilon_n)_{n \in \mathbb{N}} \subseteq [0, \infty)$ with $\epsilon_n \xrightarrow{n \to \infty} 0$. Let $M_n := \epsilon_n\text{-}\arg\min_{q \in \mathcal{Q}} \frac{1}{n} \sum_{i=1}^n \mathfrak{c}(Y_i, q)$. Our goal is to show almost sure convergence of elements in M_n to elements in M.

Remark 1.1. Considering sets of elements that minimize the objective only up to ϵ_n makes the results more relevant to applications in which Fréchet mean sets are approximated numerically. Furthermore, it may allow us to find more elements of M in the limit of M_n than for $\epsilon_n = 0$, as discussed in Remark 2.5 below and appendix A.

There are different possibilities of how a convergence of sets M_n to a set M can be described.

Definition 1.2. Let $(\mathcal{Q}, d)$ be a metric space.

(i) Let $(B_n)_{n \in \mathbb{N}}$ with $B_n \subseteq \mathcal{Q}$ for all $n \in \mathbb{N}$. Then the *outer limit* of $(B_n)_{n \in \mathbb{N}}$ is

$$\lim_{n \to \infty} \overline{\sup} \, B_n := \bigcap_{n \in \mathbb{N}} \overline{\bigcup_{k \geq n} B_k} \,,$$

where $\overline{B}$ denotes the closure of the set B.

(ii) The *one-sided Hausdorff distance* between $B, B' \subseteq \mathcal{Q}$ is

$$d_{\subseteq}(B, B') := \sup_{x \in B} \inf_{x' \in B'} d(x, x') \,.$$

(iii) The *Hausdorff distance* between $B, B' \subseteq \mathcal{Q}$ is

$$d_{\mathsf{H}}(B, B') := \max(d_{\subseteq}(B, B'), d_{\subseteq}(B', B)) \,.$$

Remark 1.3.

(i) The outer limit is the set of all points of accumulation of all sequences $(x_n)_{n \in \mathbb{N}}$ with $x_n \in B_n$. We may write $\lim \overline{\sup}_{n \to \infty} B_n = \{q \in \mathcal{Q} \,|\, \liminf_{n \to \infty} d(B_n, q) = 0\}$, where $d(B, q) := \inf_{p \in B} d(p, q)$ for a subset $B \subseteq \mathcal{Q}$. The *inner limit* is dual to the outer limit. It is defined as $\mathrm{innerlim}_{n \to \infty} B_n := \{q \in \mathcal{Q} \,|\, \limsup_{n \to \infty} d(B_n, q) = 0\}$. Clearly, $\mathrm{innerlim}_{n \to \infty} B_n \subseteq \lim \overline{\sup}_{n \to \infty} B_n$. Thus, results of the form $\lim \overline{\sup}_{n \to \infty} B_n \subseteq B$, which we show below, are stronger than $\mathrm{innerlim}_{n \to \infty} B_n \subseteq B$.

(ii) It holds $d_{\subseteq}(B, B') = 0$ if and only if $B \subseteq \overline{B'}$, but $d_{\mathsf{H}}(B, B') = 0$ if and only if $\overline{B} = \overline{B'}$. The function d_{H} is a metric on the set of closed and bounded subsets of $\mathcal{Q}$.

(iii) Elements from a sequence of sets might have sub-sequences that have no point of accumulation and are bounded away from the outer limit of the sequence of sets. That cannot happen with the one-sided Hausdorff limit. Here, every sub-sequence is eventually arbitrarily close to the limiting set. As an example, the outer limit of the sequence of sets $\{0, n\}$, $n \in \mathbb{N}$ on the Euclidean real line is $\{0\}$, but $d_{\subseteq}(\{0, n\}, \{0\}) \xrightarrow{n \to \infty} \infty$. Aside from an element with diverging distance

(n in the example), another cause for the two limits to not align may be non-compactness of bounded sets: Consider the space ℓ^2 of all sequences $(x_k)_{k \in \mathbb{N}} \subseteq \mathbb{R}$ with $\sum_{k=1}^{\infty} x_k^2 < \infty$ with distance $d((x_k)_{k \in \mathbb{N}}, (y_k)_{k \in \mathbb{N}}) = (\sum_{k=1}^{\infty}(x_k - y_k)^2)^{\frac{1}{2}}$. Let $\underline{0} \in \ell^2$ be the sequence with all entries equal to 0. Let $e^n := (e_k^n)_{k \in \mathbb{N}} \in \ell^2$ with $e_n^n = 1$ and $e_k^n = 0$ for all $k \neq n$. Then $d_{\subseteq}(\{\underline{0}, e_n\}, \{\underline{0}\}) = 1$ for all $n \in \mathbb{N}$, but $\lim \overline{\sup}_{n \to \infty} \{\underline{0}, e_n\} = \{\underline{0}\}$.

We will state conditions so that $\lim \overline{\sup}_{n \to \infty} M_n \subseteq M$ almost surely or $d_{\subseteq}(M_n, M) \xrightarrow{n \to \infty}_{\text{a.s.}} 0$, where the index a.s. indicates almost sure convergence. It is not easily possible to show $d_{\mathsf{H}}(M_n, M) \xrightarrow{n \to \infty}_{\text{a.s.}} 0$ if M is not a singleton, as discussed in Remark 2.5 below and appendix A. These limit theorems may be called strong laws of large numbers of the Fréchet mean set or (strong) consistency of the empirical Fréchet mean set. Notably, in [Evans and Jaffe, 2020] the connection to convergence in the sense of topology is made: If the set of closed subsets of $\mathcal{Q}$ is equipped with the *Kuratowski upper topology*, a sequence of closed subsets $(B_n)_{n \in \mathbb{N}}$ converges to a closed subset B if and only if $\lim \overline{\sup}_{n \to \infty} B_n \subseteq B$. If the set of nonempty compact subsets of $\mathcal{Q}$ is equipped with the *Hausdorff upper topology*, a sequence of nonempty compact subsets $(B_n)_{n \in \mathbb{N}}$ converges to a nonempty compact subset B if and only if $d_{\subseteq}(B_n, B) \xrightarrow{n \to \infty} 0$.

[Ziezold, 1977] shows a strong law in outer limit for Fréchet mean sets with a second moment condition. [Sverdrup-Thygeson, 1981] shows a strong law in outer limit for power Fréchet mean sets in compact spaces. [Bhattacharya and Patrangenaru, 2003] shows almost sure convergence of Fréchet mean sets in one-sided Hausdorff distance with a second moment condition. The independent parallel work [Evans and Jaffe, 2020] shows strong laws in outer limit and one-sided Hausdorff distance for α-Fréchet mean sets requiring $\mathbb{E}[\overline{Y,o}^{\alpha}] < \infty$, which is a second moment condition for the Fréchet mean. In contrast, we show strong laws of large numbers for power Fréchet mean sets in outer limit and in one-sided Hausdorff distance with less moment assumptions: For power $\alpha > 1$, we require $\mathbb{E}[\overline{Y,o}^{\alpha-1}] < \infty$, and for $\alpha \in (0, 1]$ no moment assumption is made, see Corollary 5.1 and Corollary 5.2. Thus, α-Fréchet means may be of interest in robust statistics. [Huckemann, 2011] shows almost sure convergence in one-side Hausdorff distance as well as in outer limit for generalized Fréchet means. Our results for $\mathfrak{c}$-Fréchet means require slightly less strict assumptions, see Theorem 3.2 and Theorem 3.5, which make them applicable in a larger class of settings and allows us to derive our results for H- and α-Fréchet means with minimal moment assumptions. Results in [Artstein and Wets, 1995, Korf and Wets, 2001, Choirat et al., 2003] imply strong laws and ergodic theorems in outer limit for generalized Fréchet means. We recite parts of these results to state Theorem 3.2. Furthermore, we show strong laws of large numbers for H-Fréchet means sets in outer limit, Corollary 4.3, and one-sided Hausdorff distance, Corollary 4.4. When M is singleton a quantitative version (rates of convergence) of the results presented in this article is given in [Schötz, 2019].

Before we consider the probabilistic setting, we present theory on convergence of minimizing sets for deterministic functions in section 2, where we partially follow [Rockafellar and Wets, 1998]. Thereafter, we derive strong laws of large numbers for $\mathfrak{c}$-Fréchet mean sets in section 3, for H-Fréchet mean sets in section 4, and for α-Fréchet mean sets in section 5. Ap-

pendix A uses the median as a simple example to illustrate some peculiarities when dealing with sets of Fréchet means. All strong laws in the main part of this article build upon [Korf and Wets, 2001, Theorem 1.1] – a deep convergence result for functions $\mathcal{Q} \to \mathbb{R}$. In appendix B, we show a different route to a strong law in one-side Hausdorff distance. This is illustrative, but requires slightly stricter assumptions. In appendix C some auxiliary results are stated and proven.

2. Convergence of Minimizer Sets of Deterministic Functions

Let $(\mathcal{Q}, d)$ be a metric space. The diameter of a set $B \subseteq \mathcal{Q}$ is defined as $\mathsf{diam}(B) = \sup_{q,p \in B} d(q, p)$. The following notion of convergence of functions will be useful to infer a convergence results of their minimizers.

Definition 2.1. Let $f, f_n \colon \mathcal{Q} \to \mathbb{R}$, $n \in \mathbb{N}$. The sequence $(f_n)_{n \in \mathbb{N}}$ *epi-converges* to f *at* $x \in \mathcal{Q}$ if and only if

$$\forall (x_n)_{n \in \mathbb{N}} \subseteq \mathcal{Q}, x_n \to x \colon \liminf_{n \to \infty} f_n(x_n) \geq f(x) \qquad \text{and}$$

$$\exists (y_n)_{n \in \mathbb{N}} \subseteq \mathcal{Q}, y_n \to x \colon \limsup_{n \to \infty} f_n(y_n) \leq f(x) \,.$$

The sequence $(f_n)_{n \in \mathbb{N}}$ *epi-converges* to f if and only if it epi-converges at all $x \in \mathcal{Q}$. We then write $f_n \xrightarrow{n \to \infty}_{\mathsf{epi}} f$.

We introduce some short notation. Let $f \colon \mathcal{Q} \to \mathbb{R}$ and $\epsilon \geq 0$. Denote $\inf f = \inf_{x \in \mathcal{Q}} f(x)$, $\arg\min f = \{x \in \mathcal{Q} \,|\, f(x) = \inf f\}$, $\epsilon\text{-}\arg\min f = \{x \in \mathcal{Q} \,|\, f(x) \leq \epsilon + \inf f\}$. Let $\delta > 0$ and $x_0 \in \mathcal{Q}$. Denote $\mathrm{B}_\delta(x_0) = \{x \in \mathcal{Q} \,|\, d(x, x_0) < \delta\}$. Furthermore, f is called *lower semi-continuous* if and only if $\liminf_{x \to x_0} f(x) \geq f(x_0)$ for all $x_0 \in \mathcal{Q}$.

To state convergence results for minimizing sets of deterministic functions, we need one final definition.

Definition 2.2. A sequence $(B_n)_{n \in \mathbb{N}}$ of sets $B_n \subseteq \mathcal{Q}$ is called *eventually precompact* if and only if there is $n \in \mathbb{N}$ such that the set $\bigcup_{k=n}^{\infty} B_k$ is precompact, i.e. its closure is compact.

The first theorem of this section relates epi-convergence of functions to convergence of their sets of minimizers in outer limit.

Theorem 2.3. *Let $f, f_n \colon \mathcal{Q} \to \mathbb{R}$. Let $(\epsilon_n)_{n \in \mathbb{N}} \subseteq [0, \infty)$ with $\epsilon_n \xrightarrow{n \to \infty} 0$. Assume $f_n \xrightarrow{n \to \infty}_{\mathsf{epi}} f$. Then*

$$\lim_{n \to \infty} \overline{\sup} \; \epsilon_n\text{-}\arg\min f_n \subseteq \arg\min f$$

and

$$\limsup_{n \to \infty} \inf f_n \leq \inf f \,.$$

Large parts of this theorem can be found e.g., in [Rockafellar and Wets, 1998, chapter 7]. To make this article more self-contained, we give a proof here.

Proof. Let $x \in \lim \overline{\sup}_{n \to \infty} \epsilon_n$-$\arg \min f_n$. Then there is a sequence $x_n \in \epsilon_n$-$\arg \min f_n$ with a subsequence converging to x, i.e., $x_{n_i} \xrightarrow{i \to \infty} x$, where $n_i \xrightarrow{i \to \infty} \infty$. Let $y \in \mathcal{Q}$ be arbitrary. As $f_n \xrightarrow{n \to \infty}_{\mathsf{epi}} f$, there is a sequence $(y_n)_{n \in \mathbb{N}} \subseteq \mathcal{Q}$ with $y_n \xrightarrow{n \to \infty} y$ and $\lim \sup_{n \to \infty} f_n(y_n) \leq f(y)$. It holds $f_{n_i}(x_{n_i}) \leq \epsilon_{n_i} + \inf f_{n_i} \leq \epsilon_{n_i} + f_{n_i}(y_{n_i})$. Thus, by the definition of epi-convergence and $\epsilon_n \xrightarrow{n \to \infty} 0$, we obtain

$$f(x) \leq \liminf_{i \to \infty} f_{n_i}(x_{n_i}) \leq \liminf_{i \to \infty} (\epsilon_{n_i} + f_{n_i}(y_{n_i})) \leq \limsup_{i \to \infty} f_{n_i}(y_{n_i}) \leq f(y) \,.$$

Thus, $x \in \arg \min f$. Next, we turn to the inequality of the infima. For $\epsilon > 0$ choose an arbitrary $x \in \epsilon$-$\arg \min f$. There is a sequence $(y_n)_{n \in \mathbb{N}} \subseteq \mathcal{Q}$ with $y_n \xrightarrow{n \to \infty} x$ and $f_n(y_n) \xrightarrow{n \to \infty} f(x)$. Thus,

$$\limsup_{n \to \infty} \inf f_n \leq \limsup_{n \to \infty} f_n(y_n) \leq \inf f + \epsilon \,.$$

$\square$

It is illustrative to compare this result with Theorem B.2, which shows that a stronger notion of convergence for functions – convergences uniformly on bounded sets – yields convergence of sets of minimizers in one-sided Hausdorff distance, which is a stronger notion of convergence of sets as the next theorem shows.

Theorem 2.4. *Let $(B_n)_{n \in \mathbb{N}}$ with $B_n \subseteq \mathcal{Q}$ for all $n \in \mathbb{N}$. Let $B \subseteq \mathcal{Q}$.*

(i) *If $d_\subseteq(B_n, B) \xrightarrow{n \to \infty} 0$ then $\lim \overline{\sup}_{n \to \infty} B_n \subseteq \overline{B}$.*

(ii) *Assume $(B_n)_{n \in \mathbb{N}}$ is eventually precompact. If $\lim \overline{\sup}_{n \to \infty} B_n \subseteq \overline{B}$ then $d_\subseteq(B_n, B) \xrightarrow{n \to \infty} 0$.*

Proof.

(i) Assume $d_\subseteq(B_n, B) \xrightarrow{n \to \infty} 0$. Let $x_\infty \in \lim \overline{\sup}_{n \to \infty} B_n$, i.e., there is a sequence $(x_{n_k})_{k \in \mathbb{N}} \subseteq \mathcal{Q}$ with $n_1 < n_2 < \ldots$ and $x_{n_k} \in B_{n_k}$ such that $x_{n_k} \xrightarrow{k \to \infty} x_\infty$. Thus,

$$\inf_{x \in B} d(x_\infty, x) \leq d(x_\infty, x_{n_k}) + \inf_{x \in B} d(x_{n_k}, x) \xrightarrow{k \to \infty} 0 \,.$$

This shows $\inf_{x \in B} d(x_\infty, x) = 0$. Hence,

$$\lim_{n \to \infty} \overline{\sup} B_n \subseteq \{x_\infty \in \mathcal{Q} \mid \inf_{x \in B} d(x_\infty, x) = 0\} = \overline{B} \,.$$

(ii) Assume $\lim \overline{\sup}_{n \to \infty} B_n \subseteq \overline{B}$. Further assume the existence of $\epsilon > 0$ and a sequence $(x_{n_k})_{k \in \mathbb{N}} \subseteq \mathcal{Q}$ with $n_1 < n_2 < \ldots$ and $x_{n_k} \in B_{n_k}$ such that $\inf_{x \in B} d(x_{n_k}, x) \geq \epsilon$. As $(B_{n_k})_{k \in \mathbb{N}}$ is eventually precompact, the sequence $(x_{n_k})_{k \in \mathbb{N}}$ has an accumulation point x_∞ in $\overline{\bigcup_{k \geq k_0} B_{n_k}}$ for some $k_0 \in \mathbb{N}$ with $\inf_{x \in B} d(x_\infty, x) \geq \epsilon$. In particular, $x_\infty \notin \overline{B}$, which contradicts the first assumption in the proof. Thus, a sequence $(x_{n_k})_{k \in \mathbb{N}}$ with these properties cannot exist, which implies $d_\subseteq(B_n, B) \xrightarrow{n \to \infty} 0$.

6

$\square$

Note that the argument for the second part is essentially the same as in [Huckemann, 2011, proof of Theorem A.4].

Remark 2.5. Together Theorem 2.3 and Theorem 2.4 may yield convergence of minimizers in one-sided Hausdorff distance. But even if $d_\subseteq(\epsilon_n\text{-}\arg\min f_n, \arg\min f) \xrightarrow{n\to\infty} 0$, $d_{\mathsf{H}}(\epsilon_n\text{-}\arg\min f_n, \arg\min f)$ does not necessarily vanish unless $\arg\min f$ is a singleton. Similarly, for an arbitrary sequence $\epsilon_n \xrightarrow{n\to\infty} 0$, the outer limit of $\epsilon_n\text{-}\arg\min f_n$ may be a strict subset of $\arg\min f$. But according to [Rockafellar and Wets, 1998, Theorem 7.31 (c)], there exists a sequence $(\epsilon_n)_{n\in\mathbb{N}}$ with $\epsilon_n \xrightarrow{n\to\infty} 0$ slow enough such that $\lim\overline{\sup}_{n\to\infty} \epsilon_n\text{-}\arg\min f_n = \arg\min f$. An explicit example of this phenomenon is presented in appendix A.

3. Strong Laws for $\mathfrak{c}$-Fréchet Mean Sets

Let $(\mathcal{Q}, d)$ be a metric space, the descriptor space. Let $\mathcal{Y}$ be a set, the data space. Let $\mathfrak{c}\colon \mathcal{Y} \times \mathcal{Q} \to \mathbb{R}$ be a function, the cost function. Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space that is silently underlying all random variables in this section. Let Y be a random variable with values in $\mathcal{Y}$. Denote the $\mathfrak{c}$-Fréchet mean set of Y as $M = \arg\min_{q\in\mathcal{Q}} \mathbb{E}[\mathfrak{c}(Y,q)]$. Let $Y_1, \ldots, Y_n$ be independent random variables with the same distribution as Y. Choose $(\epsilon_n)_{n\in\mathbb{N}} \subseteq [0, \infty)$ with $\epsilon_n \xrightarrow{n\to\infty} 0$. Set $M_n = \epsilon_n\text{-}\arg\min_{q\in\mathcal{Q}} \frac{1}{n}\sum_{i=1}^n \mathfrak{c}(Y_i, q)$.

Assumptions 3.1.

- POLISH: $(\mathcal{Q}, d)$ is separable and complete.

- LOWERSEMICONTINUITY: $q \mapsto \mathfrak{c}(y, q)$ is lower semi-continuous.

- INTEGRABLE: $\mathbb{E}[|\mathfrak{c}(Y,q)|] < \infty$ for all $q \in \mathcal{Q}$.

- INTEGRABLEINF: $\mathbb{E}[\inf_{q\in\mathcal{Q}} \mathfrak{c}(Y,q)] > -\infty$.

Theorem 3.2. *Assume POLISH, LOWERSEMICONTINUITY, INTEGRABLE, and INTEGRABLEINF. Then, almost surely,*

$$\lim_{n\to\infty}\overline{\sup}\ M_n \subseteq M\,.$$

Proof. Define $F(q) = \mathbb{E}[\mathfrak{c}(Y,q)]$, $F_n(q) = \frac{1}{n}\sum_{i=1}^n \mathfrak{c}(Y_i, q)$. By INTEGRABLE, $F(q) < \infty$. [Korf and Wets, 2001, Theorem 1.1] states that $F_n \xrightarrow[\mathsf{epi}]{n\to\infty} F$ almost surely if POLISH, LOWERSEMICONTINUITY, and INTEGRABLEINF are true. Theorem 2.3 then implies $\lim\overline{\sup}_{n\to\infty} M_n \subseteq M$ almost surely. $\square$

Assumptions 3.3.

- HEINEBOREL: Every closed bounded set in $\mathcal{Q}$ is compact.

- SAMPLEHEINEBOREL: Almost surely following is true: There is $N_0 \in \mathbb{N}$ such that every closed and bounded subset of $\bigcup_{n \geq N_0} M_n$ is compact.

- UPPERBOUND: $\mathbb{E}[\sup_{q \in B} |\mathfrak{c}(Y, q)|] < \infty$ for all bounded sets $B \subseteq \mathcal{Q}$.

- LOWERBOUND: There are $o \in \mathcal{Q}$, $\psi^+, \psi^- \colon [0, \infty) \to [0, \infty)$, $\mathfrak{a}^+, \mathfrak{a}^- \in (0, \infty)$, and $\sigma(Y_1, \ldots, Y_n)$-measurable random variables $\mathfrak{a}_n^+, \mathfrak{a}_n^- \in [0, \infty)$ such that

$$\mathfrak{a}^+ \psi^+(\overline{q,o}) - \mathfrak{a}^- \psi^-(\overline{q,o}) \leq \mathbb{E}[\mathfrak{c}(Y, q)],$$

$$\mathfrak{a}_n^+ \psi^+(\overline{q,o}) - \mathfrak{a}_n^- \psi^-(\overline{q,o}) \leq \frac{1}{n} \sum_{i=1}^{n} \mathfrak{c}(Y_i, q)$$

for all $q \in \mathcal{Q}$. Furthermore, $\mathfrak{a}_n^+ \xrightarrow{n \to \infty}_{\text{a.s.}} \mathfrak{a}^+$ and $\mathfrak{a}_n^- \xrightarrow{n \to \infty}_{\text{a.s.}} \mathfrak{a}^-$. Lastly, $\psi^+(\delta)/\psi^-(\delta) \xrightarrow{\delta \to \infty} \infty$.

Remark 3.4.

- Following implications hold:

$$\text{HEINEBOREL} \Rightarrow \text{POLISH},$$
$$\text{HEINEBOREL} \Rightarrow \text{SAMPLEHEINEBOREL},$$
$$\text{UPPERBOUND} \Rightarrow \text{INTEGRABLE}.$$

- On HEINEBOREL: A space enjoying this property is also called *boundedly compact* or *proper* metric space. The Euclidean spaces $\mathbb{R}^s$, finite dimensional Riemannian manifolds, as well as $\mathcal{C}^\infty(U)$ for open subsets $U \subseteq \mathbb{R}^s$ fulfill HEINE–BOREL [Edwards, 1995, section 8.4.7]. See [Williamson and Janos, 1987] for a construction of further spaces where HEINE–BOREL is true.

- On SAMPLEHEINEBOREL and infinite dimension: If $M_n = \{m_n\}$ and $M = \{m\}$ are singleton sets and $m_n \xrightarrow{n \to \infty} m$ almost surely, then SAMPLEHEINEBOREL holds. It is less strict than HEINEBOREL: In separable Hilbert spaces of infinite dimension HEINEBOREL does not hold. But with the metric d induced by the inner product and $\mathfrak{c} = d^2$, $\mathfrak{c}$-Fréchet means are unique and equal to the usual notion of mean. Furthermore, strong laws of large numbers in Hilbert spaces are well-known, see e.g. [Kawabe, 1986]. Thus, SAMPLEHEINEBOREL is true. Let it be noted that proving SAMPLEHEINEBOREL in a space where HEINEBOREL is false may be of similar difficulty as showing convergence of Fréchet means directly. In the case of infinite dimensional Banach spaces, results on strong laws of large numbers for a different notion of mean – the Bochner integral – are well established, see e.g. [Hoffmann-Jørgensen and Pisier, 1976].

- On LOWERBOUND: We illustrate this condition in the linear regression setting with $\mathcal{Q} = \mathbb{R}^{s+1}$, $\mathcal{Y} = (\{1\} \times \mathbb{R}^s) \times \mathbb{R}$, $\mathfrak{c}((x, y), \beta) = (y - \beta^\top x)^2 - y^2 = -2\beta^\top xy + \beta^\top xx^\top \beta$. Let (X, Y) be random variables with values in $\mathcal{Y}$. Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be

independent with the same distribution as (X, Y). We can set $o = 0 \in \mathbb{R}^{s+1}$, $\mathfrak{a}^+ = \lambda_{\mathsf{min}}(\mathbb{E}[XX^\top])$, where λ_{min} denotes the smallest eigenvalue, $\mathfrak{a}^- = 2\|\mathbb{E}[XY]\|$, $\mathfrak{a}_n^+ = \lambda_{\mathsf{min}}(\frac{1}{n}\sum_{i=1}^n X_i X_i^\top)$, $\mathfrak{a}_n^- = 2\|\frac{1}{n}\sum_{i=1}^n X_i Y_i\|$, $\psi^+(\delta) = \delta^2$ and $\psi^-(\delta) = \delta$. If $\lambda_{\mathsf{min}}(\mathbb{E}[XX^\top]) > 0$, the largest eigenvalue $\lambda_{\mathsf{max}}(\mathbb{E}[XX^\top]) < \infty$, and $\mathbb{E}[\|XY\|] < \infty$, all conditions are fulfilled.

For a further application of LOWERBOUND, see the proof of Corollary 4.4 in the next section.

Theorem 3.5. *Assume* POLISH, LOWERSEMICONTINUITY, INTEGRABLEINF, SAMPLE-HEINEBOREL, UPPERBOUND, *and* LOWERBOUND. *Then*

$$d_\subseteq(M_n, M) \xrightarrow{n\to\infty}_{\mathsf{a.s.}} 0 \,.$$

Proof. The proof consists of following steps:

1. Apply Theorem 3.2.

2. Reduction to a bounded set.

3. Show that M_n is eventually precompact almost surely.

4. Apply Theorem 2.4.

Step 1. POLISH, LOWERSEMICONTINUITY, and INTEGRABLEINF are assumptions. UPPERBOUND implies INTEGRABLE. Thus, Theorem 3.2 yields $\lim \overline{\sup}_{n\to\infty} M_n \subseteq M$ almost surely.
Step 2. Define $F(q) = \mathbb{E}[\mathfrak{c}(Y, q)]$, $F_n(q) = \frac{1}{n}\sum_{i=1}^n \mathfrak{c}(Y_i, q)$. We want to show that there is a bounded set $B_1 \subseteq \mathcal{Q}$ such that $F(q) \geq F(m) + 1$ and $F_n(q) \geq F_n(m) + 1$ for all $q \in \mathcal{Q} \setminus B_1$ and $m \in M$. If $\mathcal{Q}$ is bounded, we can take $B_1 = \mathcal{Q}$. Assume $\mathcal{Q}$ is not bounded.

Let $m \in M$. By UPPERBOUND, $F(m) < \infty$. Let $o \in \mathcal{Q}$ from LOWERBOUND. Due to LOWERBOUND, $F(q) \geq \mathfrak{a}^+\psi^+(\delta) - \mathfrak{a}^-\psi^-(\delta) \geq F(m) + 2$ for all $q \in \mathcal{Q} \setminus \mathrm{B}_\delta(o)$ and δ large enough. This holds for all $m \in M$ as $F(m)$ does not change with m. We set $B_1 = \mathrm{B}_\delta(o)$. For F_n, it holds $F_n(m) \xrightarrow{n\to\infty}_{\mathsf{a.s.}} F(m)$ and $\inf_{q \in \mathcal{Q} \setminus B_1} F_n(q) \geq \mathfrak{a}_n^+\psi^+(\delta) - \mathfrak{a}_n^-\psi^-(\delta)$ with $\mathfrak{a}_n^+ \xrightarrow{n\to\infty}_{\mathsf{a.s.}} \mathfrak{a}^+$ and $\mathfrak{a}_n^- \xrightarrow{n\to\infty}_{\mathsf{a.s.}} \mathfrak{a}^-$. Thus, there is a random variable N_1 such that almost surely $F_n(q) \geq F_n(m) + 1$ for all $n \geq N_1$, $q \in \mathcal{Q} \setminus B_1$, and $m \in M$.
Step 3. Take N_0 from SAMPLEHEINEBOREL. Choose $N_2 \geq \max(N_0, N_1)$ such that $\epsilon_n < 1$ for all $n \geq N_2$. Then $M_n \subseteq B_1$ for all $n \geq N_2$. Thus, $\bigcup_{n \geq N_2} M_n$ is bounded and – due to SAMPLEHEINEBOREL, POLISH, and Lemma C.5 – precompact almost surely.
Step 4. Finally, step 1 and 3 together with Theorem 2.4 yield $d_\subseteq(M_n, M) \xrightarrow{n\to\infty}_{\mathsf{a.s.}} 0$. □

4. Strong Laws for H-Fréchet Mean Sets

Let $(\mathcal{Q}, d)$ be a metric space. Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space that is silently underlying all random variables in this section. Let Y be a random variable with values in $\mathcal{Q}$. Let $h \colon [0, \infty) \to [0, \infty)$ be a non-decreasing function. Define $H \colon [0, \infty) \to [0, \infty), x \mapsto$

$\int_0^x h(t)\mathrm{d}t$. Fix an arbitrary element $o \in \mathcal{Q}$. Denote the H-Fréchet mean set of Y as $M = \arg\min_{q\in\mathcal{Q}} \mathbb{E}[H(\overline{Y,q}) - H(\overline{Y,o})]$. Let $Y_1, \ldots, Y_n$ be independent random variables with the same distribution as Y. Choose $(\epsilon_n)_{n\in\mathbb{N}} \subseteq [0, \infty)$ with $\epsilon_n \xrightarrow{n\to\infty} 0$. Set $M_n = \epsilon_n$-$\arg\min_{q\in\mathcal{Q}} \frac{1}{n} \sum_{i=1}^n (H(\overline{Y_i,q}) - H(\overline{Y_i,o}))$.

Assumptions 4.1.

- INFINITEINCREASE: $h(x) \xrightarrow{x\to\infty} \infty$.

- ADDITIVITY: There is $b \in [1, \infty)$ such that $h(2x) \le bh(x)$ for all $x \ge 0$.

- h-MOMENT: $\mathbb{E}[h(\overline{Y,o})] < \infty$.

Remark 4.2.

- On ADDITIVITY: This implies $h(x + y) \le b(h(x) + h(y))$ for all $x, y \ge 0$, see Lemma C.2 (appendix). If h is concave, ADDITIVITY holds with $b = 2$ and we even have $h(x + y) \le h(x) + h(y)$. This condition is not very restrictive, but it excludes functions that grow exponentially.

Corollary 4.3. *Assume* POLISH, ADDITIVITY, *and* h-MOMENT. *Then, almost surely,*

$$\limsup_{n\to\infty} M_n \subseteq M\,.$$

Proof. We check the conditions of Theorem 3.2. POLISH is an assumption. LOWERSEMICONTINUITY is fulfilled as $(q, p) \mapsto d(q, p)$ and $x \mapsto H(x)$ are continuous. For INTEGRABLE, we note that H is non-decreasing and apply Lemma C.2 (i),

$$
\begin{aligned}
|H(\overline{y,q}) - H(\overline{y,o})| &\le |\overline{y,q} - \overline{y,o}|\, h(\max(\overline{y,q}, \overline{y,o})) \\
&\le \overline{q,o}\, h(\overline{q,o} + \overline{y,o}) \\
&\le b\, \overline{q,o}\, (h(\overline{q,o}) + h(\overline{y,o}))\,,
\end{aligned}
$$

where the last inequality follows from Lemma C.2 (ii) using ADDITIVITY. Thus, h-MOMENT implies INTEGRABLE. To show INTEGRABLEINF, we note that H is non-decreasing and apply Lemma C.2 (iii),

$$
\begin{aligned}
H(\overline{y,q}) - H(\overline{y,o}) &\ge H(|\overline{y,o} - \overline{q,o}|) - H(\overline{y,o}) \\
&\ge b^{-1} H(\overline{q,o}) - 2\, \overline{q,o}\, h(\overline{y,o})
\end{aligned}
$$

due to ADDITIVITY. Furthermore, $H(\delta) = \int_0^\delta h(x)\mathrm{d}x \ge \frac{1}{2}\delta h(\frac{1}{2}\delta)$. With that, h-MOMENT implies INTEGRABLEINF. Thus, Theorem 3.2 can be applied. $\square$

Corollary 4.4. *Assume* SAMPLEHEINEBOREL, POLISH, ADDITIVITY, INFINITEINCREASE, *and* h-MOMENT. *Then*

$$d_\subseteq(M_n, M) \xrightarrow[\text{a.s.}]{n\to\infty} 0\,.$$

Proof. We check the conditions of Theorem 3.5. SAMPLEHEINEBOREL and POLISH are assumptions of the corollary. LOWERSEMICONTINUITY and INTEGRABLEINF are shown in the proof of Corollary 4.3. Following that proof, we find, due to ADDITIVITY,

$$|H(\overline{y,q}) - H(\overline{y,o})| \leq b\,\overline{q,o}\,(h(\overline{q,o}) + h(\overline{y,o}))\,,$$
$$H(\overline{y,q}) - H(\overline{y,o}) \geq b^{-1}H(\overline{q,o}) - 2\,\overline{q,o}\,h(\overline{y,o})\,,$$
$$H(\delta) \geq \frac{1}{2}\delta h\left(\frac{1}{2}\delta\right)\,.$$

The first inequality together with h-MOMENT implies UPPERBOUND. For LOWER-BOUND, we use the second inequality. We set $\psi^+(\delta) = b^{-1}H(\delta)$, $\psi^-(\delta) = 2\delta$, $\mathfrak{a}^+ = \mathfrak{a}_n^+ = 1$, $\mathfrak{a}^- = \mathbb{E}[h(\overline{Y,o})]$, and $\mathfrak{a}_n^- = \frac{1}{n}\sum_{i=1}^n h(\overline{Y_i,o})$ with $\mathfrak{a}_n^- \xrightarrow{n\to\infty}_{\text{a.s.}} \mathfrak{a}^-$ due to h-MOMENT. Because of the third inequality, $\psi^+(\delta)/\psi^-(\delta) \geq \frac{1}{4}b^{-1}h(\frac{1}{2}\delta) \xrightarrow{\delta\to\infty} \infty$ by INFINITEINCREASE. $\qquad\square$

5. Strong Laws for α-Fréchet Mean Sets

Let $(\mathcal{Q}, d)$ be a metric space. Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space that is silently underlying all random variables in this section. Let Y be a random variable with values in $\mathcal{Y}$. Let $\alpha > 0$. Fix an arbitrary element $o \in \mathcal{Q}$. Denote the α-Fréchet mean set of Y as $M = \arg\min_{q\in\mathcal{Q}} \mathbb{E}[\overline{Y,q}^\alpha - \overline{Y,o}^\alpha]$. Let $Y_1, \ldots, Y_n$ be independent random variables with the same distribution as Y. Choose $(\epsilon_n)_{n\in\mathbb{N}} \subseteq [0, \infty)$ with $\epsilon_n \xrightarrow{n\to\infty} 0$. Set $M_n = \epsilon_n\text{-}\arg\min_{q\in\mathcal{Q}} \frac{1}{n}\sum_{i=1}^n(\overline{Y_i,q}^\alpha - \overline{Y_i,o}^\alpha)$.

Corollary 5.1. *Let $\alpha > 1$. Assume $\mathbb{E}[\overline{Y,o}^{\alpha-1}] < \infty$ and* POLISH.

  (i) *Then $\lim\overline{\sup}_{n\to\infty} M_n \subseteq M$ almost surely.*

  (ii) *Additionally, assume* SAMPLEHEINEBOREL. *Then $d_\subseteq(M_n, M) \xrightarrow{n\to\infty}_{\text{a.s.}} 0$.*

Proof. Set $h(x) = \alpha x^{\alpha-1}$. This function is non-decreasing, fulfills ADDITIVITY with $b = 2^{\alpha-1}$ and INFINITEINCREASE. Due to $\mathbb{E}[\overline{Y,o}^{\alpha-1}] < \infty$, h-MOMENT is fulfilled. Furthermore, $H(x) = x^\alpha$. Thus, Corollary 4.4 and Corollary 4.3 imply the claims. $\qquad\square$

Corollary 5.2. *Let $\alpha \in (0, 1]$. Assume* POLISH.

  (i) *Then $\lim\overline{\sup}_{n\to\infty} M_n \subseteq M$ almost surely.*

  (ii) *Additionally, assume* SAMPLEHEINEBOREL. *Then $d_\subseteq(M_n, M) \xrightarrow{n\to\infty}_{\text{a.s.}} 0$.*

Proof. First, consider the case $\alpha = 1$. Apply Lemma C.3 (appendix) on $\overline{Y,o}$ to obtain a function $h\colon [0, \infty) \to [0, \infty)$ which is strictly increasing, continuous, concave, fulfills INFINITEINCREASE, and $\mathbb{E}[h(\overline{Y,o})] < \infty$. Concavity implies ADDITIVITY with $b = 2$. As its derivative is strictly increasing, $H(x) = \int_0^x h(t)\mathrm{d}t$ is convex and strictly increasing. Thus, H has an inverse H^{-1} and H^{-1} is concave. This implies that $d_H(q, p) = H^{-1}(\overline{q,p})$ is a metric.

As H^{-1} is concave, there are $u_0, u_1 \in [0, \infty)$ such that $H^{-1}(x) \leq u_0 + u_1 x$ for all $x \geq 0$. As h is concave, there are $v_0, v_1 \in [0, \infty)$ such that $h(u_0 + u_1 x) \leq v_0 + v_1 h(x)$ for all $x \geq 0$. Thus, $\mathbb{E}[h(d_H(Y, o))] = \mathbb{E}[h(H^{-1}(\overline{Y,o}))] \leq v_0 + v_1 \mathbb{E}[h(\overline{Y,o})] < \infty$. Hence, h-MOMENT is true for the metric d_H.

Moreover, POLISH and HEINEBOREL-type properties of $(\mathcal{Q}, d)$ are preserved in $(\mathcal{Q}, d_H)$, as H^{-1} is strictly increasing, concave, and continuous, with $H^{-1}(0) = 0$ and $H^{-1}(\delta) \xrightarrow{\delta \to \infty} \infty$, and thus, the properties boundedness, compactness, separability, and completeness coincide for d and d_H. Applying Corollary 4.3 and Corollary 4.4 on the minimizers of $\mathbb{E}[H(d_H(Y, q)) - H(d_H(Y, o))] = \mathbb{E}[\overline{Y, q} - \overline{Y, o}]$ now yields the claims for $\alpha = 1$.

For $\alpha \in (0, 1)$ just note, that $\tilde{d}(q, p) = d(q, p)^\alpha$ is a metric, which preserves POLISH and HEINEBOREL-type properties, and apply the result for $\alpha = 1$ on $\tilde{d}$. $\qquad\square$

Remark 5.3. For convergence, we need the $(\alpha - 1)$-moment to be finite in the case of $\alpha \geq 1$. [Schötz, 2019, Corollary 5] shows that in metric spaces with nonnegative curvature the typical parametric rate of convergence $n^{-\frac{1}{2}}$ is obtained for α-Fréchet means assuming the $2(\alpha - 1)$-moment to be finite in the case of $\alpha \in [1, 2]$ under some further conditions.

A. Example: The Set of Medians

Let $s \in \mathbb{N}$. Consider the metric space $(\mathbb{R}^s, d_1)$, where $d_1(q, p) = \|q - p\|_1 = \sum_{j=1}^s |q_j - p_j|$. The power Fréchet mean with $\alpha = 1$ in this space is equivalent to the standard $(\alpha = 2)$ Fréchet mean in $(\mathbb{R}^s, d_1^{\frac{1}{2}})$. For $s = 1$ it is equal to the median. Let $Y = (Y^1, \ldots, Y^s)$ be a random vector in $\mathbb{R}^s$ such that $\mathbb{P}(Y^k = 0) = \mathbb{P}(Y^k = 1) = \frac{1}{2}$ for $k = 1, \ldots, s$ and $Y^1, \ldots, Y^s$ are independent. Let $Y_1, Y_2, \ldots$ be independent and identically distributed copies of Y. Let $M = \arg\min_{q \in \mathbb{R}^s} \mathbb{E}[d_1(Y, q)]$ be the Fréchet mean set of Y and $M_n = \epsilon_n\text{-}\arg\min_{q \in \mathbb{R}^s} \frac{1}{n} \sum_{i=1}^n d_1(Y_i, q)$ its sample version.

A.1. No Convergence in Hausdorff Distance

First consider the case $s = 1$ and $\epsilon_n = 0$. As $s = 1$, $M = \arg\min_{q \in \mathbb{R}} \mathbb{E}[|Y - q|]$ is the median of Y, which is $M = [0, 1]$ as $2\mathbb{E}[|Y - q|] = |1 - q| + |q|$ achieves its minimal value 1 precisely for all $q \in [0, 1]$. Define $p_n := \frac{1}{n} \sum_{i=1}^n Y_i$. Then the empirical objective function is $F_n(q) := p_n |1 - q| + (1 - p_n) |q|$, i.e., the sample Fréchet mean set is $M_n = \arg\min_{q \in \mathbb{R}} F_n(q)$. If n is odd, then either $M_n = \{0\}$ or $M_n = \{1\}$ holds. The same is true for an even value of n except when $p_n = \frac{1}{2}$, in which case $M_n = [0, 1]$. Thus, $d_\subseteq(M_n, M) = 0$, but $d_{\mathsf{H}}(M_n, M)$ does not converge almost surely.

A.2. The Outer Limit as a Strict Subset

Next, we keep $\epsilon_n = 0$, but consider the value of $\lim \overline{\sup}_{n \to \infty} M_n$ in a multi-dimensional setting, i.e., $s \in \mathbb{N}$, as this yields a potentially surprising result: By Lemma C.4, M is just the Cartesian product of the median sets in each dimension, i.e., $M = [0, 1]^s$ (this is

not to be confused with the geometric median, which is the Fréchet mean with respect to the square root of the Euclidean norm). Similarly, $M_n = \bigtimes_{k=1}^{s} M_n^k$ decomposes into the sample Fréchet mean sets M_n^k of each dimension $k = 1, \ldots, s$. It holds $M_n^k = [0, 1]$ if and only if the respective value of $p_n^k := \frac{1}{n} \sum_{i=1}^{n} Y_i^k$ is equal to $\frac{1}{2}$, i.e., if and only if the symmetric simple random walk $S_n^k := \sum_{i=1}^{n} (2Y_i^k - 1)$ hits 0. Let $N = \#\{n \in \mathbb{N} \,|\, S_n^1 = \cdots = S_n^s = 0\}$. Let $A \subseteq \mathbb{R}$ and $B_n \subseteq \mathbb{R}$ for all $n \in \mathbb{N}$. If $A \subseteq B_n$ for infinitely many n, then $A \subseteq \lim \overline{\sup}_{n \to \infty} B_n$. Thus, we want to know whether N is finite or infinite. This is answered by *Pólya's Recurrence Theorem* [Pólya, 1921]. It implies that for $s \in \{1, 2\}$, $N = \infty$ almost surely. Furthermore, if $s \geq 3$, then $N < \infty$ almost surely. To find which points are not element of the outer limit of sample Fréchet mean sets, note following fact: For an open subset $A \subseteq \mathbb{R}$, if $A \subseteq \mathbb{R} \setminus B_n$ for all but finitely many n, then $A \subseteq \mathbb{R} \setminus \lim \overline{\sup}_{n \to \infty} B_n$. We conclude, that in general a vector $x \in [0, 1]^s$ is an element of $\lim \overline{\sup}_{n \to \infty} M_n$ if and only if at most two entries are not in $\{0, 1\}$, i.e., almost surely

$$\lim_{n \to \infty} \overline{\sup} \, M_n = \left\{ (x_1, \ldots, x_s) \in [0, 1]^s \,\middle|\, \#\{k \in \{1, \ldots, s\} | x_k \in (0, 1)\} \leq 2 \right\}.$$

Thus, $\lim \overline{\sup}_{n \to \infty} M_n = M$ for $s \in \{1, 2\}$ and $\lim \overline{\sup}_{n \to \infty} M_n \subsetneq M$ for $s \geq 3$.

A.3. Convergence in Hausdorff Distance

Lastly, we use the setting $s = 1$ and $\epsilon_n \in [0, \infty)$, where we want to find ϵ_n such that $[0, 1] \subseteq M_n$. At least one of 0 and 1 is a minimizer of $F_n(q) = p_n \,|1 - q| + (1 - p_n) \,|q|$ and the $F_n(q)$ is linear on $[0, 1]$. Thus, $[0, 1] \subseteq M_n$ if and only if $\epsilon_n \geq |F_n(0) - F_n(1)|$. This is equivalent to $|p_n - \frac{1}{2}| < \frac{1}{2} \epsilon_n$. By Markov's inequality

$$\mathbb{P}\left(\left| p_n - \frac{1}{2} \right| \geq \frac{1}{2} \epsilon_n \right) \leq \frac{n^{-3} \mathbb{E}\left[(Y - \frac{1}{2})^4 \right]}{2^{-4} \epsilon_n^4}.$$

For $\epsilon_n = n^{-\frac{1}{4}}$, we obtain

$$\sum_{n=1}^{\infty} \mathbb{P}\left(\left| p_n - \frac{1}{2} \right| \geq \frac{1}{2} \epsilon_n \right) \leq \sum_{n=1}^{\infty} n^{-2} < \infty.$$

The Borel–Cantelli lemma implies that almost surely and for all n large enough, $|p_n - \frac{1}{2}| < \frac{1}{2} \epsilon_n$ and thus, $[0, 1] \subseteq M_n$. Together with Corollary 5.2, we obtain $d_{\mathsf{H}}(M_n, M) \xrightarrow[\text{a.s.}]{n \to \infty} 0$.

B. Alternative Route to One-Sided Hausdorff Convergence

In this section, we show an alternative proof of a strong law of large numbers for generalized Fréchet mean sets in one-sided Hausdorff distance. Although the final result, Theorem B.5, is weaker than Theorem 3.5, it is very illustrative to follow this line of proof: In contrast to the arguments in the main part of the article, it does not rely on the powerful result [Korf and Wets, 2001, Theorem 1.1], which seems to be rather complex

to prove. Instead our reasoning here is simpler and more self-contained. Furthermore, a comparison between convergence in outer limit and in one-sided Hausdorff distance seems more natural in view of the deterministic results Theorem 2.3 and Theorem B.2, and the stochastic results Theorem 3.2 and Theorem B.5.

B.1. Convergence of Minimizer Sets of Deterministic Functions

Let $(\mathcal{Q}, d)$ be a metric space. For $A \subseteq \mathcal{Q}$ and $\delta > 0$, denote $\mathrm{B}_\delta(A) = \bigcup_{x \in A} \mathrm{B}_\delta(x)$.

Definition B.1.

(i) Let $f, f_n \colon \mathcal{Q} \to \mathbb{R}$, $n \in \mathbb{N}$. The sequence $(f_n)_{n \in \mathbb{N}}$ *converges* to f *uniformly on bounded sets* if and only if for every $B \subseteq \mathcal{Q}$ with $\mathsf{diam}(B) < \infty$,

$$\lim_{n \to \infty} \sup_{x \in B} |f_n(x) - f(x)| = 0\,.$$

We then write $f_n \xrightarrow{n \to \infty}_{\mathsf{ubs}} f$.

(ii) A sequence $(B_n)_{n \in \mathbb{N}}$ of sets $B_n \subseteq \mathcal{Q}$ is called *eventually bounded* if and only if

$$\limsup_{n \to \infty} \mathsf{diam}\left(\bigcup_{k=n}^{\infty} B_k\right) < \infty\,.$$

(iii) A function f has *approachable minimizers* if and only if for all $\epsilon > 0$ there is a $\delta > 0$ such that $\delta\text{-}\arg\min f \subseteq B_\epsilon(\arg\min f)$.

The last definition directly implies that $d_\subseteq(\delta\text{-}\arg\min f, \arg\min f) \xrightarrow{\delta \to 0} 0$ is equivalent to f having approachable minimizers. Furthermore, if f has approachable minimizers, then $\arg\min f \neq \emptyset$, as for every $\delta > 0$ the set $\delta\text{-}\arg\min f$ is non-empty, but $\mathrm{B}_\epsilon(\emptyset) = \emptyset$.

Theorem B.2. *Let $f, f_n \colon \mathcal{Q} \to \mathbb{R}$. Let $(\epsilon_n)_{n \in \mathbb{N}} \subseteq [0, \infty)$ with $\epsilon_n \xrightarrow{n \to \infty} 0$. Assume f has approachable minimizers, $f_n \xrightarrow{n \to \infty}_{\mathsf{ubs}} f$, and $(\epsilon_n\text{-}\arg\min f_n)_{n \in \mathbb{N}}$ is eventually bounded. Then*

$$d_\subseteq(\epsilon_n\text{-}\arg\min f_n, \arg\min f) \xrightarrow{n \to \infty} 0$$

and

$$\inf f_n \xrightarrow{n \to \infty} \inf f\,.$$

Proof. Let $\epsilon > 0$. As f has approachable minimizers, there is $\delta > 0$ such that $(3\delta)\text{-}\arg\min f \subseteq \mathrm{B}_\epsilon(\arg\min f)$; also $\arg\min f \neq \emptyset$. Let $y \in \arg\min f$. As $f_n(y) \xrightarrow{n \to \infty} f(y)$, there is $n_1 \in \mathbb{N}$ such that $\inf f_n \leq \inf f + \delta$ for all $n \geq n_1$. As $\epsilon_n \xrightarrow{n \to \infty} 0$, there is $n_2 \in \mathbb{N}$ such that $\epsilon_n \leq \delta$ for all $n \geq n_2$. As $(\epsilon_n\text{-}\arg\min f_n)_{n \in \mathbb{N}}$ is eventually bounded, there is $n_3 \in \mathbb{N}$ such that $\mathsf{diam}(B) < \infty$ for $B = \bigcup_{n \geq n_3} \epsilon_n\text{-}\arg\min f_n$. As $f_n \xrightarrow{n \to \infty}_{\mathsf{ubs}} f$ there is n_4 such that $\sup_{x \in B} |f_n(x) - f(x)| \leq \delta$. Let $n \geq \max(n_1, n_2, n_3, n_4)$ and $x \in \epsilon_n\text{-}\arg\min f_n$. Then

$$f(x) \leq f_n(x) + \delta \leq \inf f_n + 2\delta \leq \inf f + 3\delta\,.$$

Thus, $x \in (3\delta)$- $\arg\min f$. By the choice of ϵ and δ, we obtain ϵ_n- $\arg\min f_n \subseteq B_\epsilon(\arg\min f)$ or equivalently $d_\subseteq(\epsilon_n$- $\arg\min f_n, \arg\min f) \leq \epsilon$.

Finally, we show the convergence of the infima. We already know $\inf f_n \leq \inf f + \epsilon$ for all $\epsilon > 0$ and n large enough. If $\inf f_n \xrightarrow{n\to\infty} \inf f$ does not hold, there is a sequence $x_n \in \epsilon_n$- $\arg\min f_n$ and $\epsilon > 0$ such that $f_n(x_n) < \inf f - \epsilon$ for all n large enough. As before, because of eventual boundedness and uniform convergence on bounded sets, we have $\sup_{k\in\mathbb{N}}|f_n(x_k) - f(x_k)| \xrightarrow{n\to\infty} 0$. Therefore, for all $\epsilon > 0$ we have $f(x_n) \leq f_n(x_n)+\epsilon$ for n large enough, which contradicts $f_n(x_n) < \inf f - \epsilon$. $\qquad\square$

In the following, we construct examples to show that none of the conditions for one-sided Hausdorff convergence can be dropped.

Example B.3.

(i) Let $f, f_n\colon \mathbb{N}_0 \to \mathbb{R}$, $f_n = 1 - \mathbb{1}_{\{0,n\}}$, $f = 1 - \mathbb{1}_{\{0\}}$, $d(i,j) = 1$ for $i \neq j$. It holds that f is continuous and has approachable minimizers, and the sequence of nonempty sets $\arg\min f_n = \{0,n\}$ is eventually bounded, as $\mathsf{diam}(A) \leq 1$ for every $A \subseteq \mathbb{N}_0$. Furthermore, f_n converges to f uniformly on compact sets, which are exactly the finite subsets of $\mathbb{N}_0$, but not uniformly on bounded sets like $\mathbb{N}_0$ itself. There is a subsequence of minimizers $x_n = n \in \arg\min f_n$ that is always bounded away from 0, the minimizer of f. This shows that uniform convergence on compact sets (instead of bounded sets) is not enough.

(ii) As above, let $f, f_n\colon \mathbb{N}_0 \to \mathbb{R}$, $f_n = 1 - \mathbb{1}_{\{0,n\}}$, $f = 1 - \mathbb{1}_{\{0\}}$, but define $d(i,j) = |i-j|$. It holds that f is continuous and has approachable minimizers, and $f_n \xrightarrow{n\to\infty}_{\mathsf{ubs}} f$, but the sequence of nonempty sets $\arg\min f_n = \{0,n\}$ is not eventually bounded. Again, there is a subsequence of minimizers $x_n = n \in \arg\min f_n$ that is always bounded away from 0, the minimizer of f. This shows that eventual boundedness of minimizer sets cannot be dropped.

(iii) Let $f, f_n\colon \mathbb{N}_0 \to \mathbb{R}$, $f(0) = 0$, $f(i) = \frac{1}{i}$, $f_n(i) = f(i)\mathbb{1}_{\{i<n\}}$, and set $d(i,j) = 1$ for $i \neq j$. It holds that f is continuous, but f does not have approachable minimizers. The sequence of nonempty sets $\arg\min f_n = \{0,n,n+1,\dots\}$ is eventually bounded and $f_n \xrightarrow{n\to\infty}_{\mathsf{ubs}} f$. There is a subsequence of minimizers $x_n = n \in \arg\min f_n$ that is always bounded away from 0, the minimizer of f. This shows that approachability of minimizers of f cannot be dropped.

B.2. Strong Laws for $\mathfrak{c}$-Fréchet Mean Sets

Let $(\mathcal{Q}, d)$ be a metric space, the descriptor space. Let $\mathcal{Y}$ be a set, the data space. Let $\mathfrak{c}\colon \mathcal{Y} \times \mathcal{Q} \to \mathbb{R}$ be a function, the cost function. Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space that is silently underlying all random variables in this section. Let Y be a random variable with values in $\mathcal{Y}$. Denote the $\mathfrak{c}$-Fréchet mean set of Y as $M = \arg\min_{q\in\mathcal{Q}} \mathbb{E}[\mathfrak{c}(Y,q)]$. Let $Y_1,\dots,Y_n$ be independent random variables with the same distribution as Y. Choose $(\epsilon_n)_{n\in\mathbb{N}} \subseteq [0,\infty)$ with $\epsilon_n \xrightarrow{n\to\infty} 0$. Set $M_n = \epsilon_n$- $\arg\min_{q\in\mathcal{Q}} \frac{1}{n}\sum_{i=1}^n \mathfrak{c}(Y_i,q)$.

Assumptions B.4.

- CONTINUITY: The function $q \mapsto \mathfrak{c}(Y, q)$ is continuous almost surely.

Theorem B.5. *Assume HEINEBOREL, CONTINUITY, UPPERBOUND, and LOWERBOUND. Then*

$$d_\subseteq(M_n, M) \xrightarrow{n \to \infty}_{\text{a.s.}} 0.$$

Proof. Define $F(q) = \mathbb{E}[\mathfrak{c}(Y, q)]$, $F_n(q) = \frac{1}{n} \sum_{i=1}^n \mathfrak{c}(Y_i, q)$. The proof consists of following steps:

1. Show that $F_n \xrightarrow{n \to \infty}_{\text{ubs}} F$ almost surely.

2. Reduction to a bounded set.

3. Show that F has approachable minimizers.

4. Show that M_n is eventually bounded.

5. Apply Theorem B.2.

<u>Step 1.</u> To show uniform convergence on bounded sets, we will use the uniform law of large numbers, Theorem C.1 below. Let $B \subseteq \mathcal{Q}$ be a bounded set. By HEINEBOREL, $\overline{B}$ is compact. By CONTINUITY, $q \mapsto \mathfrak{c}(Y, q)$ is almost surely continuous. By UPPERBOUND, $\mathbb{E}[\sup_{q \in B} |\mathfrak{c}(Y, q)|] < \infty$. Thus, Theorem C.1 implies that $q \mapsto F(q)$ is continuous and

$$\sup_{q \in B} |F_n(q) - F(q)| \xrightarrow{n \to \infty}_{\text{a.s.}} 0.$$

Fix an arbitrary element $o \in \mathcal{Q}$. For all bounded sets B, there is $\delta \in \mathbb{N}$ such that $B \subseteq \mathrm{B}_\delta(o)$. By the previous considerations, uniform convergence holds almost surely for all $(\mathrm{B}_\delta(o))_{\delta \in \mathbb{N}}$. Thus, $F_n \xrightarrow{n \to \infty}_{\text{ubs}} F$ almost surely.
<u>Step 2.</u> Find $B_1 \subseteq \mathcal{Q}$ and a random variable $N_1 \in \mathbb{N}$ as in step 2 in the proof of Theorem 3.5.
<u>Step 3.</u> Clearly, $M \subseteq B_1$ is bounded. Furthermore, for all $\epsilon > 0$ small enough the set $D_\epsilon = \overline{B_1 \setminus \mathrm{B}_\epsilon(M)}$ is not empty (if it is, increase δ), does not contain any element of M and, by HEINEBOREL, is compact. Thus, the continuous function $q \mapsto F(q)$ attains its infimum on D_ϵ where $\inf_{q \in D_\epsilon} F(q) > \inf_{q \in \mathcal{Q}} F(q)$. Take $\zeta = \min(1, \frac{1}{2}(\inf_{q \in D_\epsilon} F(q) - \inf_{q \in \mathcal{Q}} F(q)))$. Then ζ-$\arg\min_{q \in \mathcal{Q}} F(q) \subseteq \mathrm{B}_\epsilon(M)$, i.e., F has approachable minimizers.
<u>Step 4.</u> For $\epsilon_n < 1$ and $n \geq N_1$, it holds $M_n \subseteq B_1$. Thus, $(M_n)_{n \in \mathbb{N}}$ is eventually bounded almost surely.
<u>Step 5.</u> Finally, Theorem B.2 implies $d_\subseteq(M_n, M) \xrightarrow{n \to \infty}_{\text{a.s.}} 0$. $\qquad\square$

C. Auxiliary Results

There are many versions of uniform laws of large numbers in the literature. We state and prove one version that is tailored to our needs.

Theorem C.1. *Let $(\mathcal{Y}, \Sigma_{\mathcal{Y}})$ be a measurable space and Y be a random variable with values in $\mathcal{Y}$. Let $Y_1, \ldots, Y_n$ be independent and have the same distribution as Y. Let $(\mathcal{Q}, d)$ be a metric space and $B \subseteq \mathcal{Q}$ compact. Let $f : \mathcal{Y} \times B \to \mathbb{R}$ be such that $q \mapsto f(Y, q)$ is almost surely continuous. Assume there is a random variable Z such that $|f(Y, q)| \leq Z$ for all $q \in B$ with $\mathbb{E}[Z] < \infty$. Then $q \mapsto \mathbb{E}[f(Y, q)]$ is continuous and*

$$\sup_{q \in B} \left| \frac{1}{n} \sum_{i=1}^{n} f(Y_i, q) - \mathbb{E}[f(Y, q)] \right| \xrightarrow[\text{a.s.}]{n \to \infty} 0 \,.$$

Proof. Let $\epsilon > 0$. As B is compact, there is a finite set $\{q_1, \ldots, q_k\} \subseteq \mathcal{Q}$ such that $B \subseteq \bigcup_{\ell=1}^{k} \mathrm{B}_\epsilon(q_\ell)$. We split the supremum,

$$\sup_{q \in B} \left| \frac{1}{n} \sum_{i=1}^{n} f(Y_i, q) - \mathbb{E}[f(Y, q)] \right|$$

$$\leq \sup_{\ell \in \{1, \ldots, k\}} \sup_{q \in \mathrm{B}_\epsilon(q_\ell)} \left| \frac{1}{n} \sum_{i=1}^{n} (f(Y_i, q) - f(Y_i, q_\ell)) - \mathbb{E}[f(Y, q) - f(Y, q_\ell)] \right|$$

$$+ \sup_{\ell \in \{1, \ldots, k\}} \left| \frac{1}{n} \sum_{i=1}^{n} f(Y_i, q_\ell) - \mathbb{E}[f(Y, q_\ell)] \right| \,.$$

For the second summand, by the standard strong law of large numbers applied to each $\ell \in \{1, \ldots, k\}$ with $\mathbb{E}[Z] < \infty$,

$$\sup_{\ell \in \{1, \ldots, k\}} \left| \frac{1}{n} \sum_{i=1}^{n} f(Y_i, q_\ell) - \mathbb{E}[f(Y, q_\ell)] \right| \xrightarrow[\text{a.s.}]{n \to \infty} 0 \,.$$

For the first summand,

$$\sup_{\ell \in \{1, \ldots, k\}} \sup_{q \in \mathrm{B}_\epsilon(q_\ell)} \left| \frac{1}{n} \sum_{i=1}^{n} (f(Y_i, q) - f(Y_i, q_\ell)) - \mathbb{E}[f(Y, q) - f(Y, q_\ell)] \right|$$

$$\leq \frac{1}{n} \sum_{i=1}^{n} \sup_{q,p \in B, \overline{q,p} \leq \epsilon} |f(Y_i, q) - f(Y_i, p)| + \mathbb{E}\left[\sup_{q,p \in B, \overline{q,p} \leq \epsilon} |f(Y, q) - f(Y, p)| \right] \,.$$

By the standard strong law of large numbers with $\mathbb{E}[Z] < \infty$,

$$\frac{1}{n} \sum_{i=1}^{n} \sup_{q,p \in B, \overline{q,p} \leq \epsilon} |f(Y_i, q) - f(Y_i, p)| \xrightarrow[\text{a.s.}]{n \to \infty} \mathbb{E}\left[\sup_{q,p \in B, \overline{q,p} \leq \epsilon} |f(Y, q) - f(Y, p)| \right] \,.$$

Thus,

$$\mathbb{P}\left(\limsup_{n \to \infty} \sup_{q \in B} \left| \frac{1}{n} \sum_{i=1}^{n} f(Y_i, q) - \mathbb{E}[f(Y, q)] \right| \leq a_\epsilon \right) = 1 \,, \tag{1}$$

where $a_\epsilon = 2\mathbb{E}\left[\sup_{q,p \in B, \overline{q,p} \leq \epsilon} |f(Y, q) - f(Y, p)| \right]$. As $q \mapsto f(Y, q)$ is almost surely continuous and B is compact, $q \mapsto f(Y, q)$ is almost surely uniformly continuous, i.e., for

all $\delta > 0$ there is $\varepsilon(\delta, Y) > 0$ such that $|f(Y, q) - f(Y, p)| \leq \delta$ for all $\overline{q,p} \leq \varepsilon(\delta, Y)$. As $\mathbb{E}[Z] < \infty$, we can use dominated convergence to obtain

$$\lim_{\epsilon \searrow 0} \mathbb{E}\left[\sup_{q,p \in B, \overline{q,p} \leq \epsilon} |f(Y, p) - f(Y, p)|\right] = \mathbb{E}\left[\lim_{\epsilon \searrow 0} \sup_{q,p \in B, \overline{q,p} \leq \epsilon} |f(Y, p) - f(Y, p)|\right] = 0\,.$$

Thus, $a_\epsilon \xrightarrow{\epsilon \searrow 0} 0$. Together with (1), this implies

$$\sup_{q \in B}\left|\frac{1}{n}\sum_{i=1}^{n} f(Y_i, q) - \mathbb{E}[f(Y, q)]\right| \xrightarrow[\text{a.s.}]{n \to \infty} 0\,.$$

We have also shown that $q \mapsto \mathbb{E}[f(Y, q)]$ is continuous, as $|\mathbb{E}[f(Y, q)] - \mathbb{E}[f(Y, p)]| \leq a_{\overline{q,p}}$. $\square$

Lemma C.2. *Let $h\colon [0, \infty) \to [0, \infty)$ be a non-decreasing function. Define $H\colon [0, \infty) \to [0, \infty), x \mapsto \int_0^x h(t)\mathrm{d}t$. Let $x, y \geq 0$. Then*

(i) $|H(x) - H(y)| \leq |x - y|\, h(\max(x, y))$.

Assume, there is $b \in [1, \infty)$ such that $h(2u) \leq bh(u)$ for all $u \geq 0$. Then

(ii) $\frac{1}{2}h(x) + \frac{1}{2}h(y) \leq h(x + y) \leq b\left(h(x) + h(y)\right)$,

(iii) $H(|x - y|) - H(x) \geq b^{-1}H(y) - 2yh(x)$.

Proof. (i) This is a direct consequence of the mean value theorem.

(ii) As h is non-decreasing, $\max(h(x), h(y)) \leq h(x + y) \leq \max(h(2x), h(2y))$. By the definition of b and with $\frac{1}{2}(u + v) \leq \max(u, v) \leq u + v$ for $u, v \geq 0$ the claim follows.

(iii) First, consider the case $x \geq y$. Define $f(x, y) = H(x - y) - H(x) - b^{-1}H(y) + 2yh(x)$. We want to show $f(x, y) \geq 0$. The derivative of f with respect to y is

$$\partial_y f(x, y) = -h(x - y) - b^{-1}h(y) + 2h(x)\,.$$

By applying the first inequality of (ii) to $h(x) = h((x - y) + y)$, we obtain $\partial_y f(x, y) \geq 0$ as $b^{-1} \leq 1$. Hence, $f(x, y) \geq f(x, 0) = 0$, as $H(y) = 0$.

Now, consider the case $x \leq y$. Set $g(x, y) = H(y - x) - H(x) - b^{-1}H(y) + 2yh(x)$, which yields

$$\partial_y g(x, y) = h(y - x) - b^{-1}h(y) + 2h(x)\,.$$

By applying the second inequality of (ii) to $h(y) = h((y - x) + x)$, we obtain $\partial_y g(x, y) \geq 0$ as $b^{-1} \leq 1$. Thus, $g(x, y) \geq g(x, x) = -(1 + b^{-1})H(x) + 2xh(x)$ as $H(0) = 0$. By the definition of H, as h is non-decreasing, $H(x) \leq xh(x)$. Hence, $g(x, y) \geq 0$ as $1 + b^{-1} \leq 2$.

Together, we have shown $H(|x - y|) - H(x) - b^{-1}H(y) + 2yh(x) \geq 0$ for all $x, y \geq 0$. $\square$

Lemma C.3. *Let X be a random variable with values in $[0, \infty)$. Then there is a strictly increasing, continuous, and concave function $h\colon [0, \infty) \to [0, \infty)$ with $h(\delta) \xrightarrow{\delta \to \infty} \infty$ such that $\mathbb{E}[h(X)] < \infty$.*

Proof. If there is $K > 0$ such that $\mathbb{P}(X < K) = 1$ take $h(x) = x$. Now, assume that X is not almost surely bounded. We first construct a non-decreasing function $\tilde{h}\colon [0, \infty) \to [0, \infty)$ such that $\tilde{h}(x) \xrightarrow{x \to \infty} \infty$ with $\mathbb{E}[\tilde{h}(X)] < \infty$. Then we construct a function h from $\tilde{h}$ with all desired properties.

Let F be the distribution function of X, $F(x) = \mathbb{P}(X \le x)$. Let $z_1 = 0$ and $z_{n+1} = \inf\left\{x \ge z_n + 1 \mid 1 - F(x) \le \frac{1}{n}\right\}$. As $F(x) \xrightarrow{x \to \infty} 1$, $z_n < \infty$. Furthermore, $z_{n+1} - z_n \ge 1$. Moreover, as X is not almost surely bounded, $1 - F(x) > 0$ for all $x \ge 0$. Set

$$g(x) = \sum_{n=1}^{\infty} (z_{n+1} - z_n)^{-1} n^{-2} \mathbb{1}_{[z_n, z_{n+1})}(x) \,,$$

$$\tilde{h}(x) = \int_0^x \frac{g(t)}{1 - F(t)} \mathrm{d}t \,.$$

Then

$$\lim_{x \to \infty} \tilde{h}(x) = \int_0^\infty \frac{g(t)}{1 - F(t)} \mathrm{d}t \ge \sum_{n=1}^{\infty} n^{-1} = \infty \,.$$

Moreover, $\tilde{h}(x)$ is strictly increasing, as $g(t) \ge 0$ and $1 - F(t) \ge 0$. The function $\tilde{h}$ is continuously differentiable everywhere except at point z_n, $n \in \mathbb{N}$. Thus,

$$
\begin{aligned}
\mathbb{E}[\tilde{h}(X)] &= \int_0^\infty \mathbb{P}\Big(\tilde{h}(X) > t\Big) \mathrm{d}t \\
&= \int_0^\infty \mathbb{P}\Big(X > \tilde{h}^{-1}(t)\Big) \mathrm{d}t \\
&= \int_0^\infty \tilde{h}'(t) \mathbb{P}(X > t) \mathrm{d}t \\
&= \int_0^\infty g(t) \mathrm{d}t \\
&= \sum_{n=1}^{\infty} n^{-2} < \infty \,.
\end{aligned}
$$

Let $a_0 = 1$, $x_0 = 0$, $x_{n+1} = \inf\left\{x \ge x_n + a_n^{-1} \mid \tilde{h}(x) \ge n + 1\right\}$ and $a_{n+1} = (x_{n+1} - x_n)^{-1}$. Let $h\colon [0, \infty) \to [0, \infty)$ be the linear interpolation of $(x_n, n)_{n \in \mathbb{N}_0}$. As $\tilde{h}(x) \xrightarrow{x \to \infty} \infty$, all x_n are finite. Hence, $h(x) \xrightarrow{x \to \infty} \infty$. Because of $a_n > 0$, h is strictly increasing. Furthermore, $a_{n+1} \le a_n$ as $x_{n+1} \ge x_n + a_n^{-1}$. As h is continuous and a_n is the derivative of h in the interval (x_n, x_{n+1}), h is concave. Lastly, $h(x) \le \tilde{h}(x) + 1$. Thus, $\mathbb{E}[h(X)] < \infty$. $\qquad \square$

Lemma C.4 (Fréchet means in product spaces)**.** *Let $K \in \mathbb{N}$. Let $(\mathcal{Q}_1, d_1), \dots, (\mathcal{Q}_K, d_K)$ be metric spaces. Let $\alpha \ge 1$. Set $\mathcal{Q} := \bigtimes_{k=1}^K \mathcal{Q}_k$ and $d\colon \mathcal{Q} \times \mathcal{Q} \to [0, \infty)$, $d(q, p) :=$*

$(\sum_{k=1}^{K} d_k(q_k, p_k)^\alpha)^{\frac{1}{\alpha}}$. *Then $(\mathcal{Q}, d)$ is a metric space. Let $Y = (Y^1, \ldots, Y^K)$ be a tuple of random variables such that Y^k has values in $\mathcal{Q}_k$ and $\mathbb{E}[d(Y, o)^{\alpha-1}] < \infty$ for an element $o \in \mathcal{Q}$. Let M be the α-Fréchet mean set of Y in $(\mathcal{Q}, d)$, and M^k be the α-Fréchet mean set of Y^k in $(\mathcal{Q}_k, d_k)$. Then M is the Cartesian product of the sets M^k, i.e.,*

$$M = \underset{k=1}{\overset{K}{\bigtimes}} M^k.$$

Lemma C.4 can be proven by straight forward calculations.

Lemma C.5. *Let $(\mathcal{Q}, d)$ be a complete metric space and $A \subseteq \mathcal{Q}$. Assume that all closed subsets $B \subseteq A$ are compact. Then $\overline{A}$ is compact.*

Proof. Let $(a_k)_{k\in\mathbb{N}} \subseteq \overline{A}$. We show that $(a_k)_{k\in\mathbb{N}}$ has a converging subsequence with limit $a \in \overline{A}$, which implies compactness of $\overline{A}$.

Let $\delta_n = 2^{-n}$. Define $B_n := \mathcal{Q} \setminus \mathrm{B}_{\delta_n}(\mathcal{Q} \setminus A)$. The sets B_n are closed and subsets of A. Thus, they are compact. Let $b_k^n \in \arg\min_{b\in B_n} d(b, a_k)$. Such an element exists as B_n is compact. Furthermore, $d(b_k^n, a_k) \leq \delta_n$. Define the subindex sequences $(k(n, \ell))_{\ell\in\mathbb{N}} \subseteq \mathbb{N}$ such that $k(0, \ell) = \ell$ and $(b_{k(n,\ell)}^n)_{\ell\in\mathbb{N}}$ is a converging subsequence of $(b_{k(n-1,\ell)}^n)_{\ell\in\mathbb{N}}$ with limit $b_{k(n,\ell)}^n \xrightarrow{\ell\to\infty} b_\infty^n$ and $d(b_{k(n,\ell)}^n, b_\infty^n) \leq \delta_\ell$. By the triangle inequality $d(b_k^{n_1}, b_k^{n_2}) \leq d(b_k^{n_1}, a_k) + d(b_k^{n_2}, a_k) \leq \delta_{n_1} + \delta_{n_2}$. Thus, $d(b_\infty^{n_1}, b_\infty^{n_2}) \leq \delta_{n_1} + \delta_{n_2}$, which makes $(b_\infty^n)_{n\in\mathbb{N}}$ a Cauchy-sequence. Define a as its limit, i.e., $b_\infty^n \xrightarrow{n\to\infty} a$. As $b_\infty^n \in B_n \subseteq \overline{A}$, also $a \in \overline{A}$. Finally, the triangle inequality yields

$$d(a_{k(n,n)}, a) \leq d(a_{k(n,n)}, b_{k(n,n)}^n) + d(b_{k(n,n)}^n, b_\infty^n) + d(b_\infty^n, a) \xrightarrow{n\to\infty} 0,$$

i.e., $(a_{k(n,n)})_{n\in\mathbb{N}}$ is subsequence of $(a_k)_{k\in\mathbb{N}}$ which converges in $\overline{A}$. $\square$

References

[Arnaudon et al., 2013] Arnaudon, M., Barbaresco, F., and Yang, L. (2013). Medians and means in Riemannian geometry: existence, uniqueness and computation. In *Matrix information geometry*, pages 169–197. Springer, Heidelberg.

[Artstein and Wets, 1995] Artstein, Z. and Wets, R. J.-B. (1995). Consistency of minimizers and the SLLN for stochastic programs. *J. Convex Anal.*, 2(1-2):1–17.

[Bhattacharya and Patrangenaru, 2003] Bhattacharya, R. and Patrangenaru, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds. I. *Ann. Statist.*, 31(1):1–29.

[Choirat et al., 2003] Choirat, C., Hess, C., and Seri, R. (2003). A functional version of the birkhoff ergodic theorem for a normal integrand: A variational approach. *Ann. Probab.*, 31(1):63–92.

[Edwards, 1995] Edwards, R. E. (1995). *Functional analysis.* Dover Publications, Inc., New York. Theory and applications, Corrected reprint of the 1965 original.

[Evans and Jaffe, 2020] Evans, S. N. and Jaffe, A. Q. (2020). Strong laws of large numbers for Fréchet means. arXiv preprint.

[Fréchet, 1948] Fréchet, M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. *Ann. Inst. H. Poincaré*, 10:215–310.

[Hoffmann-Jørgensen and Pisier, 1976] Hoffmann-Jørgensen, J. and Pisier, G. (1976). The law of large numbers and the central limit theorem in Banach spaces. *Ann. Probability*, 4(4):587–599.

[Hotz and Huckemann, 2015] Hotz, T. and Huckemann, S. (2015). Intrinsic means on the circle: uniqueness, locus and asymptotics. *Ann. Inst. Statist. Math.*, 67(1):177–193.

[Huckemann, 2011] Huckemann, S. F. (2011). Intrinsic inference on the mean geodesic of planar shapes and tree discrimination by leaf growth. *Ann. Statist.*, 39(2):1098–1124.

[Kawabe, 1986] Kawabe, J. (1986). Characterization of Hilbert spaces by the strong law of large numbers. *J. Multivariate Anal.*, 20(1):155–160.

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

[Korf and Wets, 2001] Korf, L. A. and Wets, R. J.-B. (2001). Random lsc functions: An ergodic theorem. *Mathematics of Operations Research*, 26(2):421–445.

[MacQueen, 1967] MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, pages 281–297, Berkeley, Calif. University of California Press.

[Pólya, 1921] Pólya, G. (1921). Über eine Aufgabe der Wahrscheinlichkeitsrechnung betreffend die Irrfahrt im Straßennetz. *Math. Ann.*, 84(1-2):149–160.

[Rockafellar and Wets, 1998] Rockafellar, R. and Wets, R. J.-B. (1998). *Variational Analysis.* Springer Verlag, Heidelberg, Berlin, New York.

[Schötz, 2019] Schötz, C. (2019). Convergence rates for the generalized Fréchet mean via the quadruple inequality. *Electron. J. Stat.*, 13(2):4280–4345.

[Sturm, 2003] Sturm, K.-T. (2003). Probability measures on metric spaces of nonpositive curvature. In *Heat kernels and analysis on manifolds, graphs, and metric spaces (Paris, 2002)*, volume 338 of *Contemp. Math.*, pages 357–390. Amer. Math. Soc., Providence, RI.

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

[Williamson and Janos, 1987] Williamson, R. and Janos, L. (1987). Constructing metrics with the Heine-Borel property. *Proc. Amer. Math. Soc.*, 100(3):567–573.

[Ziezold, 1977] Ziezold, H. (1977). On expected figures and a strong law of large numbers for random elements in quasi-metric spaces. In *Transactions of the Seventh Prague Conference on Information Theory, Statistical Decision Functions, Random Processes and of the Eighth European Meeting of Statisticians (Tech. Univ. Prague, Prague, 1974), Vol. A*, pages 591–602.