Fréchet Means in Infinite Dimensions

Adam Quinn Jaffe

What is the "mean" of samples $Y_1, \ldots, Y_n$ in a metric space $(\mathcal{X}, d)$?

If d is the metric given by a Hilbert space norm $\|\cdot\|$ on $\mathcal{X}$, then the sample mean $\frac{1}{n}(Y_1 + \cdots + Y_n)$ is the unique solution to:

$$\begin{cases} \text{minimize} & \frac{1}{n}\sum_{i=1}^{n} \|x - Y_i\|^2 \\ \text{over} & x \in \mathcal{X} \end{cases}$$

Depends only the metric structure of the Hilbert space!

What is the “mean” of a probability measure μ on $(\mathcal{X}, d)$?

If d is the metric given by a Hilbert space norm $\|\cdot\|$ on $\mathcal{X}$, then the expectation $\int_{\mathcal{X}} d^2(x,y)\,\mathrm{d}\mu(y)$ is the unique solution to:

$$\begin{cases} \text{minimize} & \int_{\mathcal{X}} \|x-y\|^2\,\mathrm{d}\mu(y) \\ \text{over} & x \in \mathcal{X} \end{cases}$$

Depends only the metric structure of the Hilbert space!

Define the *empirical Fréchet mean* as

$$\bar{M}_n := \operatorname*{arg\,min}_{x \in \mathcal{X}} \frac{1}{n} \sum_{i=1}^{n} d^2(x, Y_i)$$

and the *population Fréchet mean* as

$$M := \operatorname*{arg\,min}_{x \in \mathcal{X}} \int_{\mathcal{X}} d^2(x, y) \, \mathrm{d}\mu(y).$$

For this talk, assume uniquely achieved.

Do we have $\bar{M}_n \to M$ in a statistically meaningful sense?

If $\mathcal{X}$ is a Hilbert space, this follows from the classical limit theorems (SLLN, CLT, concentration inequalites, rates of convergence etc.)

Definition

We say that *the strong law of large numbers (SLLN) holds in* $(\mathcal{X}, d)$ if we have $d(\bar{M}_n, M) \to 0$ almost surely when $Y_1, \ldots, Y_n$ are IID samples.

In which metric spaces does the SLLN hold?

General results require some sort of “finite-dimensionality”:

We say that $(\mathcal{X}, d)$ is a *Heine-Borel space* if the closed balls $\bar{B}_r(x) := \{y \in \mathcal{X} : d(x, y) \leq r\}$ are compact for all $x \in \mathcal{X}, r \geq 0$.

Theorem (Schötz, 2022)

If $(\mathcal{X}, d)$ is a Heine-Borel space, then the SLLN holds in $(\mathcal{X}, d)$.

But finite-dimensionality is not necessary!

Theorem (Sturm 2003)

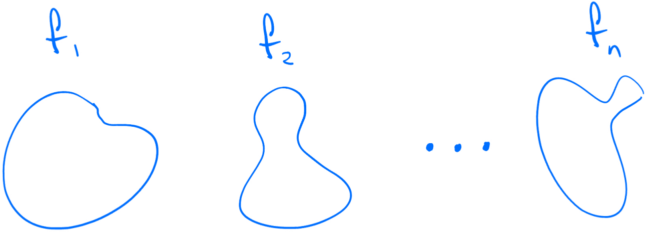
If $(\mathcal{X}, d)$ is a Hadamard space, then the SLLN holds in $(\mathcal{X}, d)$.

Theorem (Le Gouic-Loubes, 2017)

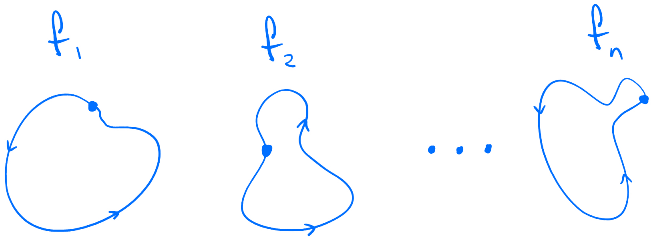
If (S, ρ) is a complete, locally compact, geodesic metric space, then the SLLN holds in the Wasserstein space $(\mathcal{P}_2(S, \rho), W_2)$.

Many other important examples of infinite-dimensional metric spaces where no asymptotic theory is known...

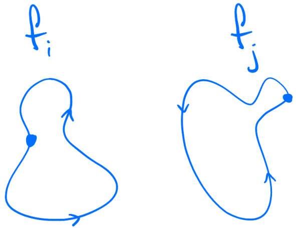
Example. spaces of planar loops in statistical shape analysis



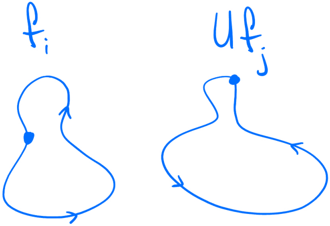
Example. spaces of planar loops in statistical shape analysis



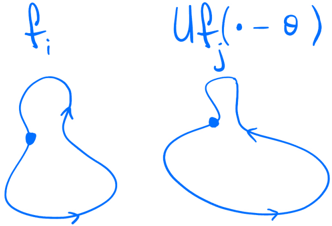
Example. spaces of planar loops in statistical shape analysis



Example. spaces of planar loops in statistical shape analysis



Example. spaces of planar loops in statistical shape analysis



Example. spaces of planar loops in statistical shape analysis

Let $\mathcal{X}_0$ denote the space of functions $C^{1,1}$ functions $f : \mathbb{S}^1 \to \mathbb{R}^2$ satisfying $\|f'(t)\| = 1$ for all $0 \le t \le 1$.

Let d_0 denote the following pseudometric on $\mathcal{X}_0$:

$$d_0(f,g) := \min_{\substack{\theta \in \mathbb{S}^1 \\ U \in \mathrm{SO}(2)}} \left(\int_0^{2\pi} \sum_{k=0}^{2} \|U f^{(k)}(t-\theta) - g^{(k)}(t)\|^2 \, \mathrm{d}t \right)^{1/2}$$

Then, consider quotient space $(\mathcal{X}, d)$ of $(\mathcal{X}_0, d_0)$.

One of many possible metrics on spaces of planar loops used in statistical shape analysis (Kurtek-Srivastava-Klassen-Ding 2012).

Definition

A topology w on $\mathcal{X}$ is called a *weak convergence for* $(\mathcal{X}, d)$ if:

Definition

A topology w on $\mathcal{X}$ is called a *weak convergence for* $(\mathcal{X}, d)$ if:

(W1) d-bounded sets are relatively w-compact.

Definition

A topology w on $\mathcal{X}$ is called a *weak convergence for* $(\mathcal{X}, d)$ if:

(W1) If $\{x_n\}_{n\in\mathbb{N}}$ and y in $\mathcal{X}$ satisfy $\sup_{n\in\mathbb{N}} d(x_n, y) < \infty$, then there exists a subsequence $\{n_k\}_{k\in\mathbb{N}}$ and a point $x \in \mathcal{X}$ satisfying $x_{n_k} \to x$ in w.

Definition

A topology w on $\mathcal{X}$ is called a *weak convergence for* $(\mathcal{X}, d)$ if:

(W1) If $\{x_n\}_{n\in\mathbb{N}}$ and y in $\mathcal{X}$ satisfy $\sup_{n\in\mathbb{N}} d(x_n, y) < \infty$, then there exists a subsequence $\{n_k\}_{k\in\mathbb{N}}$ and a point $x \in \mathcal{X}$ satisfying $x_{n_k} \to x$ in w.

(W2) d is w-lower-semicontinuous.

Definition

A topology w on $\mathcal{X}$ is called a *weak convergence for* $(\mathcal{X}, d)$ if:

(W1) If $\{x_n\}_{n\in\mathbb{N}}$ and y in $\mathcal{X}$ satisfy $\sup_{n\in\mathbb{N}} d(x_n, y) < \infty$, then there exists a subsequence $\{n_k\}_{k\in\mathbb{N}}$ and a point $x \in \mathcal{X}$ satisfying $x_{n_k} \to x$ in w.

(W2) If $\{x_n\}_{n\in\mathbb{N}}$ and x in $\mathcal{X}$ satisfy $x_n \to x$ in w, then for all $y \in \mathcal{X}$ we have $d(x, y) \leq \liminf_{n\to\infty} d(x_n, y)$.

Definition

A topology w on $\mathcal{X}$ is called a *weak convergence for* $(\mathcal{X}, d)$ if:

(W1) If $\{x_n\}_{n\in\mathbb{N}}$ and y in $\mathcal{X}$ satisfy $\sup_{n\in\mathbb{N}} d(x_n, y) < \infty$, then there exists a subsequence $\{n_k\}_{k\in\mathbb{N}}$ and a point $x \in \mathcal{X}$ satisfying $x_{n_k} \to x$ in w.

(W2) If $\{x_n\}_{n\in\mathbb{N}}$ and x in $\mathcal{X}$ satisfy $x_n \to x$ in w, then for all $y \in \mathcal{X}$ we have $d(x, y) \leq \liminf_{n\to\infty} d(x_n, y)$.

(W3) d is w-continuous only where it is d-continuous.

Definition

A topology w on $\mathcal{X}$ is called a *weak convergence for* $(\mathcal{X}, d)$ if:

(W1) If $\{x_n\}_{n\in\mathbb{N}}$ and y in $\mathcal{X}$ satisfy $\sup_{n\in\mathbb{N}} d(x_n, y) < \infty$, then there exists a subsequence $\{n_k\}_{k\in\mathbb{N}}$ and a point $x \in \mathcal{X}$ satisfying $x_{n_k} \to x$ in w.

(W2) If $\{x_n\}_{n\in\mathbb{N}}$ and x in $\mathcal{X}$ satisfy $x_n \to x$ in w, then for all $y \in \mathcal{X}$ we have $d(x, y) \le \liminf_{n\to\infty} d(x_n, y)$.

(W3) If $\{x_n\}_{n\in\mathbb{N}}$ and x in $\mathcal{X}$ satisfy $x_n \to x$ in w and satisfy $d(x_n, y) \to d(x, y)$ for some $y \in \mathcal{X}$, then we have $x_n \to x$ in d.

Definition

A topology w on $\mathcal{X}$ is called a *weak convergence for* $(\mathcal{X}, d)$ if:

(W1) If $\{x_n\}_{n\in\mathbb{N}}$ and y in $\mathcal{X}$ satisfy $\sup_{n\in\mathbb{N}} d(x_n, y) < \infty$, then there exists a subsequence $\{n_k\}_{k\in\mathbb{N}}$ and a point $x \in \mathcal{X}$ satisfying $x_{n_k} \to x$ in w.

(W2) If $\{x_n\}_{n\in\mathbb{N}}$ and x in $\mathcal{X}$ satisfy $x_n \to x$ in w, then for all $y \in \mathcal{X}$ we have $d(x, y) \leq \liminf_{n\to\infty} d(x_n, y)$.

(W3) If $\{x_n\}_{n\in\mathbb{N}}$ and x in $\mathcal{X}$ satisfy $x_n \to x$ in w and satisfy $d(x_n, y) \to d(x, y)$ for some $y \in \mathcal{X}$, then we have $x_n \to x$ in d.

We say that $(\mathcal{X}, d)$ *admits a weak convergence* if there exists a weak convergence w for $(\mathcal{X}, d)$.

Theorem (AQJ 2024)

If $(\mathcal{X}, d)$ admits a weak convergence, then the SLLN holds in $(\mathcal{X}, d)$.

Notice that the conclusion refers only to the metric d.

The existence of a weak convergence is a *geometric* property of $(\mathcal{X}, d)$.

I. Introduction

II. On Weak Convergence

III. The Continuity Theorem

IV. Future Work

II. On Weak Convergence

Basic examples:

Recall topology w on $\mathcal{X}$ is called a *weak convergence for* $(\mathcal{X}, d)$ if:

(W1) d-bounded sets are relatively w-compact,

(W2) d is w-lower-semicontinuous, and

(W3) d is w-continuous only where it is d-continuous.

metric space	weak convergence	(W1)	(W2)	(W3)
Heine-Borel space	metric convergence	definition	trivial	trivial
Hadamard space	Jost's convergence	Jost's Banach-Alaoglu theorem	weak lower semi-continuity of metric	Kadec-Klee property
Wasserstein space	weak convergence of probability measures	Prokhorov theorem	Fatou lemma	definition
uniformly convex Banach space	weak topology	Milman-Pettis theorem	weak lower semi-continuity of norm	Kadec-Klee property

Example.

Let $\mathbb{H}$ be a (possibly infinite-dimensional) Hilbert space

Let $\mathbb{K}$ denote the space of covariance operators on $\mathbb{H}$

Set $\Pi(\Sigma, \Sigma') := W_2(\mathcal{N}(0, \Sigma), \mathcal{N}(0, \Sigma'))$, the Bures-Wasserstein metric

For $\{\Sigma_n\}_{n\in\mathbb{N}}$ and Σ in $\mathbb{K}$, say $\Sigma_n \to \Sigma$ in w if

$$\int_{\mathbb{H}} \phi \, \mathrm{d}\mathcal{N}(0, \Sigma_n) \to \int_{\mathbb{H}} \phi \, \mathrm{d}\mathcal{N}(0, \Sigma)$$

for all weakly continuous $\phi : \mathbb{H} \to \mathbb{R}$.

Similar to existing considerations in the theory of gradient flows (Ambrosio-Gigli-Savaré 2008), and recovers known results for Bures-Wasserstein barycenters (Masarotto-Panaretos-Zemel 2019).

Closure properties:

Want to provide a "calculus" for constructing further examples of interest, as in Grenander's pattern theory (Mumford 2003).

metric space operation	inputs	output	weak convergence
restriction	$(\mathcal{X}, d)$ with weak convergence w, w-closed subset $\mathcal{X}' \subseteq \mathcal{X}$	$(\mathcal{X}', d)$	w
product	$(\mathcal{X}_1, d_1)$ with weak convergence w_1, $(\mathcal{X}_2, d_2)$ with weak convergence w_2	$(\mathcal{X}_1 \times \mathcal{X}_2, d_1 \otimes_q d_2)$	$w_1 \otimes w_2$
quotient	$(\mathcal{X}, d)$ with weak convergence w, compact group G acting isometrically	$(\mathcal{X}/G, d_{\mathcal{X}/G})$	Γ-convergence of orbits under w
regularization	$(\mathcal{X}, d)$ with weak convergence w, proper metric group (G, ρ) acting isometrically	$(\mathcal{X}, d_{G,\rho})$	w

Example.

Let $\mathcal{X}_0$ denote the space of functions $C^{1,1}$ functions $f : \mathbb{S}^1 \to \mathbb{R}^2$ satisfying $\|f'(t)\| = 1$ for all $0 \leq t \leq 1$.

Let d_0 denote the following pseudometric on $\mathcal{X}_0$:

$$d_0(f,g) := \min_{\substack{\theta \in \mathbb{S}^1 \\ U \in \mathrm{SO}(2)}} \left(\int_0^{2\pi} \sum_{k=0}^{2} \|U f^{(k)}(t-\theta) - g^{(k)}(t)\|^2 \, \mathrm{d}t \right)^{1/2}$$

Then, consider quotient space $(\mathcal{X}, d)$ of $(\mathcal{X}_0, d_0)$.

One of many possible metrics on spaces of planar loops used in statistical shape analysis (Kurtek-Srivastava-Klassen-Ding, 2012).

III. The Continuity Theorem

Want a general asymptotic theory for Fréchet means which...

- ☑ holds in many infinite-dimensional metric spaces of interest
- ☐ implies many limit theorems of interest
- ☐ requires the minimal moment assumptions
- ☐ requires no uniqueness assumptions

Many limit theorems at once?

Notice that the empirical Fréchet mean depends only on the empirical measure $\bar{\mu}_n := \frac{1}{n}\sum_{i=1}^n \delta_{Y_i}$:

$$M = \operatorname*{arg\,min}_{x\in\mathcal{X}} \frac{1}{n}\sum_{i=1}^{n} d^2(x, Y_i)$$

So, define

$$M(\mu) := \operatorname*{arg\,min}_{x\in\mathcal{X}} \int_{\mathcal{X}} d^2(x, y)\,\mathrm{d}\mu(y).$$

Since we have $\mu_n \to \mu$ in many senses, we can deduce limit theorems for Fréchet mean sets from limit theorem from empirical measures.

Minimal moment assumption?

Write $\mathcal{P}_p(\mathcal{X})$ for the set of μ with $\int_{\mathcal{X}} d^p(x,y)\,\mathrm{d}\mu(y) < \infty$ for all $x \in \mathcal{X}$.

Define the map $M : \mathcal{P}_2(\mathcal{X}) \to 2^{\mathcal{X}}$ via

$$M(\mu) := \operatorname*{arg\,min}_{x \in \mathcal{X}} \int_{\mathcal{X}} d^2(x,y)\,\mathrm{d}\mu(y)$$

Minimal moment assumption?

Write $\mathcal{P}_p(\mathcal{X})$ for the set of μ with $\int_{\mathcal{X}} d^p(x,y)\,\mathrm{d}\mu(y) < \infty$ for all $x \in \mathcal{X}$.

Define the map $M : \mathcal{P}_1(\mathcal{X}) \to 2^{\mathcal{X}}$ via

$$M(\mu) := \arg\min_{x\in\mathcal{X}} \int_{\mathcal{X}} (d^2(x,y) - d^2(o,y))\,\mathrm{d}\mu(y)$$

for arbitrary $o \in \mathcal{X}$.

This makes sense because

$$|d^2(x,y) - d^2(o,y)| \leq d(x,o)(d(x,y) + d(o,y))$$

in any metric space!

Convergence without uniqueness?

For non-empty bounded subsets $\{M_n\}_{n\in\mathbb{N}}$ and M of $\mathcal{X}$, consider

$$\max_{x_n\in M_n}\min_{x\in M} d(x_n, x) \to 0.$$

Plainly, every element of M_n is close to some element of M.

If $\{M_n\}_{n\in\mathbb{N}}$ and M are singletons, then equivalent to $d(M_n, M) \to 0$.

Let W_1 denote the 1-Wasserstein metric (i.e., Kantorovich-Rubinstein metric) on $\mathcal{P}_1(\mathcal{X})$.

Theorem (AQJ 2024)

If $(\mathcal{X}, d)$ is separable and admits a weak convergence and $\{\mu_n\}_{n\in\mathbb{N}}$ and μ in $\mathcal{P}_1(\mathcal{X})$ satisfy $W_1(\mu_n, \mu) \to 0$, then

$$\max_{x_n \in M(\mu_n)} \min_{x \in M(\mu)} d(x_n, x) \to 0$$

as $n \to \infty$.

This is a purely analytic result, but we will later take $\mu_n := \bar{\mu}_n$.

Proof Outline.

Take $x_n \in M(\mu_n)$ for all $n \in \mathbb{N}$, then:

(S0) Show that $\{x_n\}_{n\in\mathbb{N}}$ is d-bounded.

(S1) Get $\{n_k\}_{k\in\mathbb{N}}$ and $x_\infty \in \mathcal{X}$ with $x_{n_k} \to x_\infty$ in w.

(S2) Show $x_\infty \in M(\mu)$.

(S3) Show $x_{n_k} \to x_\infty$ in d.

Definition

A topology w on $\mathcal{X}$ is called a *weak convergence for* $(\mathcal{X}, d)$ if:

(W1) d-bounded sets are relatively w-compact

(W2) d is w-lower-semicontinuous.

(W3) d is w-continuous only where it is d-continuous.

Earlier works (Thorpe-Theil-Johansen-Cade 2015, Le Gouic-Loubes 2017, Schötz 2022) outline same method of proof.

IV. Future Work

Verify existence of weak convergence for other metric spaces:

- large deformation diffeomorphic metric mapping (LDDMM) (Bauer-Bruveris-Michor 2014, Younes 2019)
- invariant L^2 metric on graphons (Kolaczyk-Lin-Rosenberg-Walters-Xu 2020)

Non-examples:

- Easy case $\mathcal{X} = [-1, 0) \cup (0, 1]$, with $\mu = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$
- What if $\mathcal{X}$ is complete?

Refining other limit theorems:

- Central limit theorem requries some further differentiable structure
- Large deviations theory requires some additional steps

Thank you!

References

L. Ambrosio, N. Gigli, and G. Savaré. *Gradient Flows in Metric Spaces and in the Space of Probability Measures.* Lectures in Mathematics (ETH Zürich Birkhäuser), 2nd ed., 2008.

M. Bauer, M. Bruveris, and P. Michor. Overview of the geometries of shape spaces and diffeomorphism groups. *J. Math. Imag. Vis.,* 50:60-97, 2014

T. Le Gouic and J.-M. Loubes. Existence and consistency of Wasserstein barycenters. *Probab. Theory Related Fields,* 168:901-917, 2017.

E. D. Kolaczyk, L. Lin, S. Rosenberg, J. Walters, and J. Xu. Averages of unlabeled networks: geomeric characterization and asymptotic behavior. *Ann. Statist.,* 48:514-538, 2020.

S. Kurtek, A. Srivastava, E. Klassen, and Z. Ding, Statistical modeling of curves using shapes and related features. *J. Amer. Statist. Assoc.*, 107:1152-1165, 2012.

V. Masarotto, V. M. Panaretos, and Y. Zemel. Procrustes metrics on covariance operators and optimal transportation of Gaussian processes. *Sankhya A*, 8:172-213, 2019.

M. Miller and L. Younes. Group actions, homeomorphisms, and matching: a general framework. *Internat. J. Comp. Vis.,* 41:61-84, 2001.

D. Mumford. *Pattern theory: a unifying perspective.* Fields Medallists' lectures (World Scientific Publishing River Edge New Jersey), 2003.

C. Schötz. Strong laws of large numbers for generalizations of Fréchet mean sets. *Statistics,* 56(1):34-52, 2022.

K.-T. Sturm. Probability Measures on Metric Spaces of Nonpositive Curvature. *Comtemp. Math.,* 338:357-390, 2003.

M. Thorpe, F. Theil, A. M. Johansen, and N. Cade. Convergence of the k-means minimization problem using Γ-convergence. *SIAM J. Appl. Math.*, 75:2444-2474, 2015.

L. Younes. *Shapes and Diffeomorphisms.* Applied Mathematical Sciences (Springer Berlin), 2nd ed., 2019.