Fréchet Mean Set Estimation in the Hausdorff Metric, via Relaxation

Adam Quinn Jaffe

With Moïse Blanchard

I. Introduction

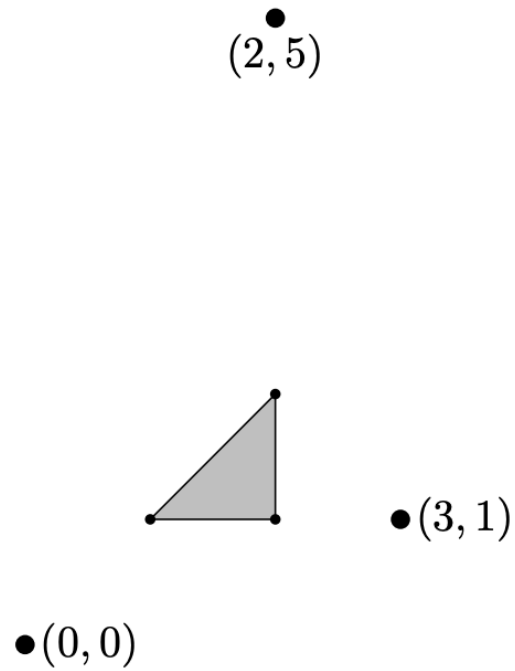
If the population Fréchet mean is unique, then questions of statistical Fréchet mean estimation are well-posed and can be studied.

However, the literature is divided on uniqueness:

- Assume it? (...)
- Deduce it from other assumptions? (Sturm 2003, Afsari 2011, Hotz-Huckemann 2015, Le Gouic-Loubes 2017, Cao-Monod 2022, etc.)
- Test for it? (Eltzner 2020)
- Give up on it? (Ziezold 1977, Bhattacharya-Patrangenaru 2003, Schötz 2022, Evans-AQJ 2024, etc.)

Even worse: In many examples, uniqueness is known to fail.

For example, Fréchet medians are typically non-unique in the *tropical projective space* (Lin-Yoshida 2018):



How can we do meaningful statistical inference without uniqueness?

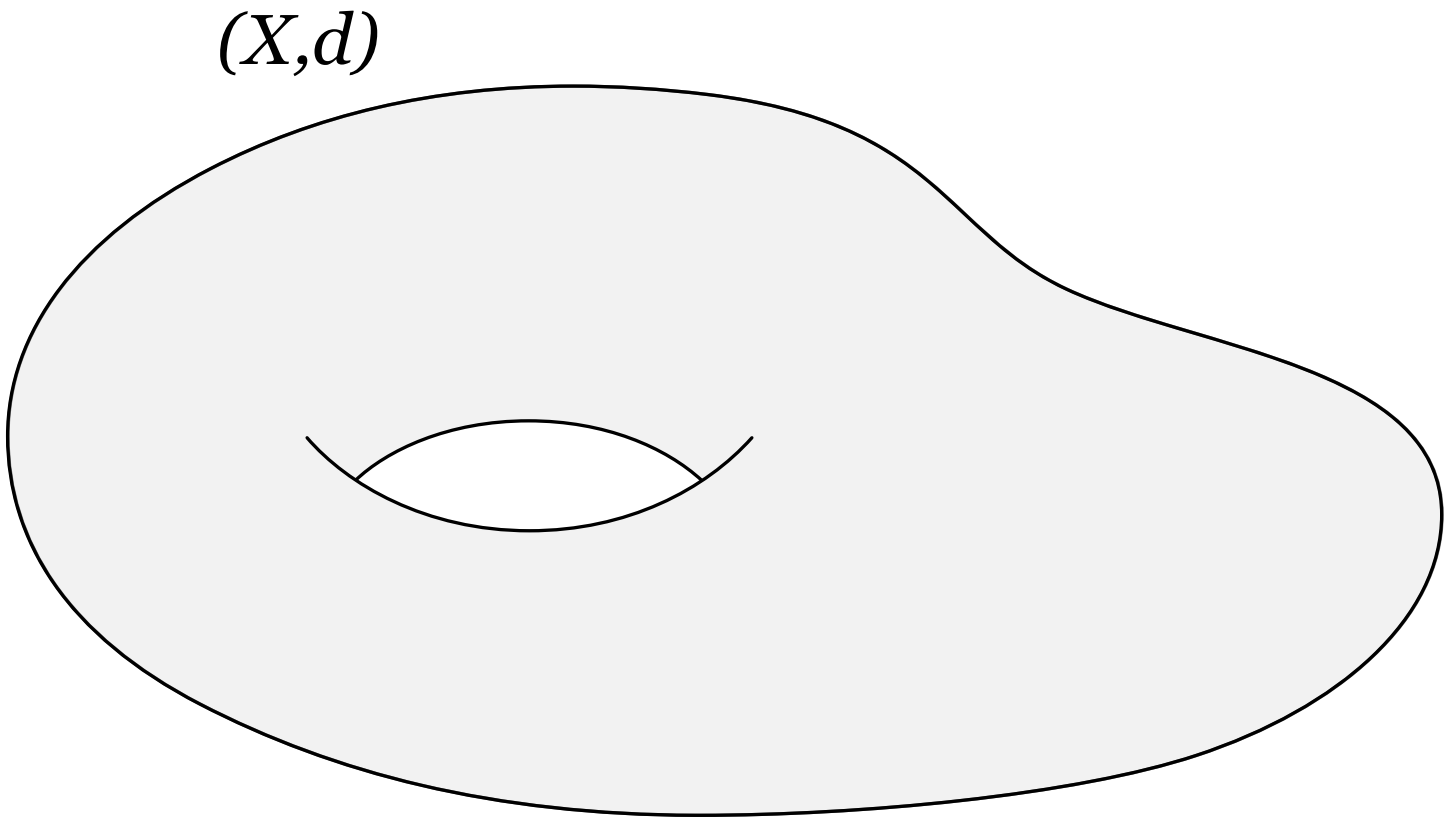
II. Problem Statement

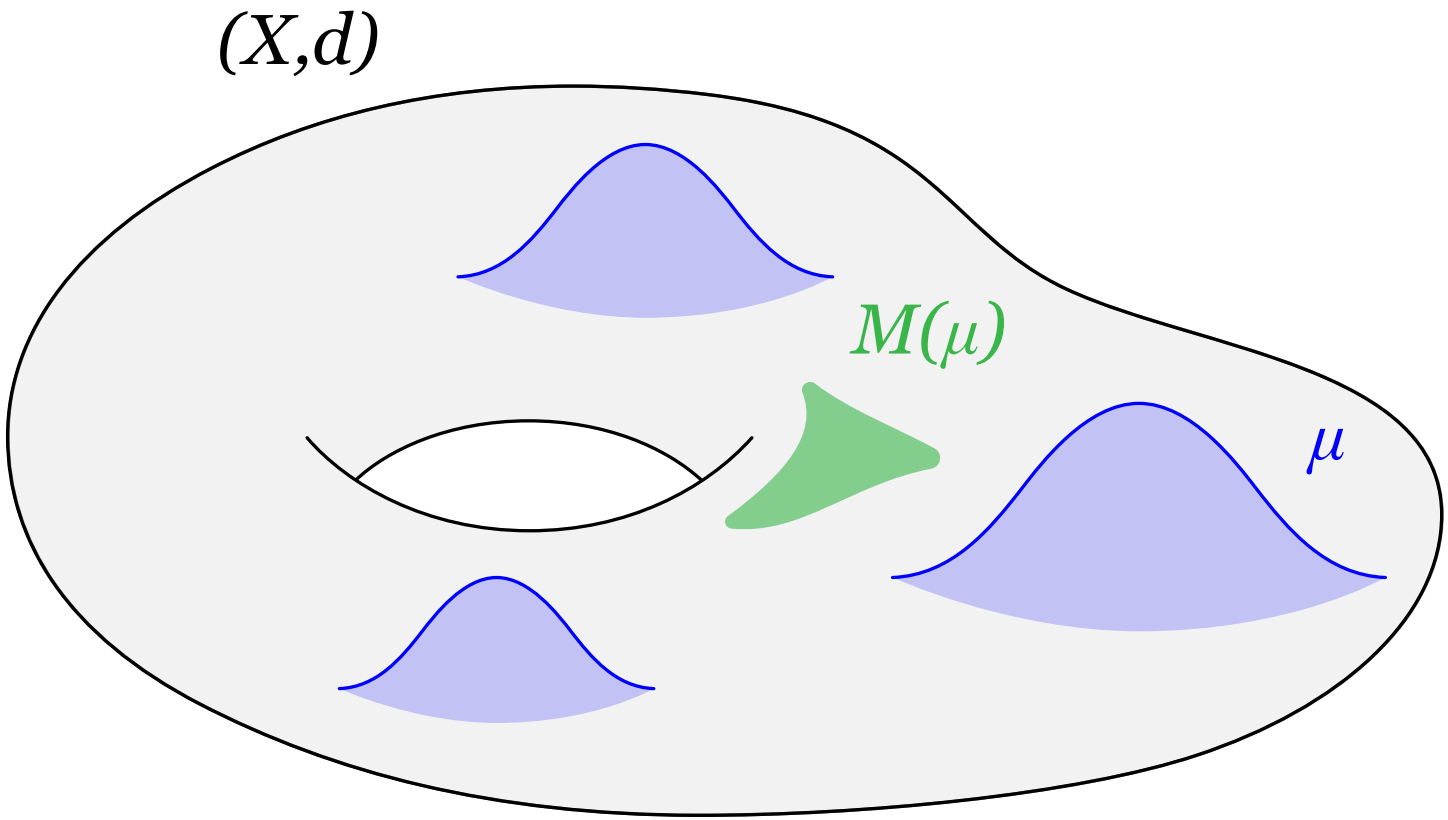
Let (X, d) be a metric space and μ a probability measure on X.

Define its *Fréchet mean set* as

$$M(\mu) := \arg\min_{x \in X} \int_X d^2(x, y)\, d\mu(y).$$

We take this to be the set of minimizers. In particular, it can be empty, it can be a singleton, or it can have more than one point.


(X,d)

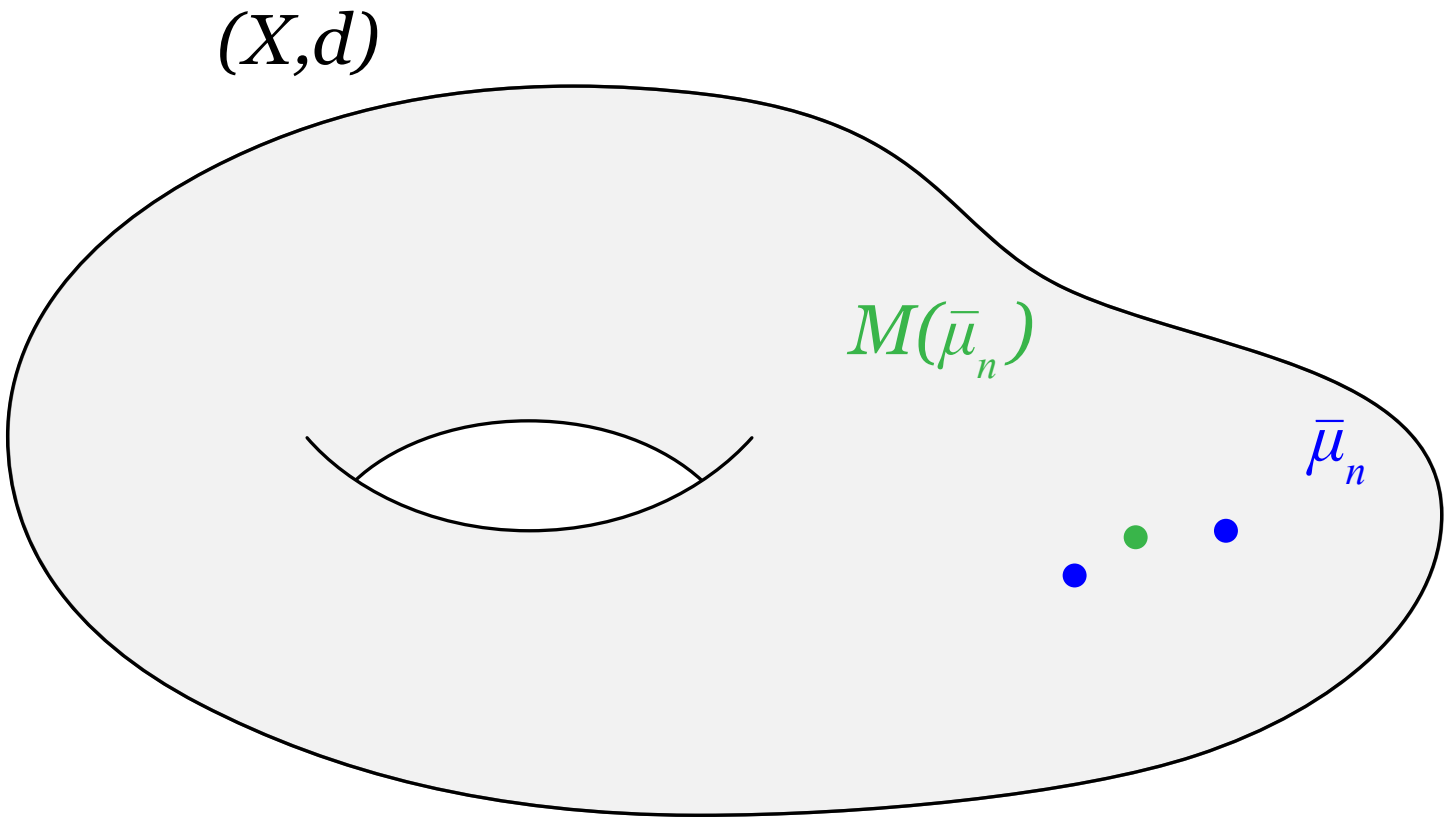

(X,d)
M(μ)
μ

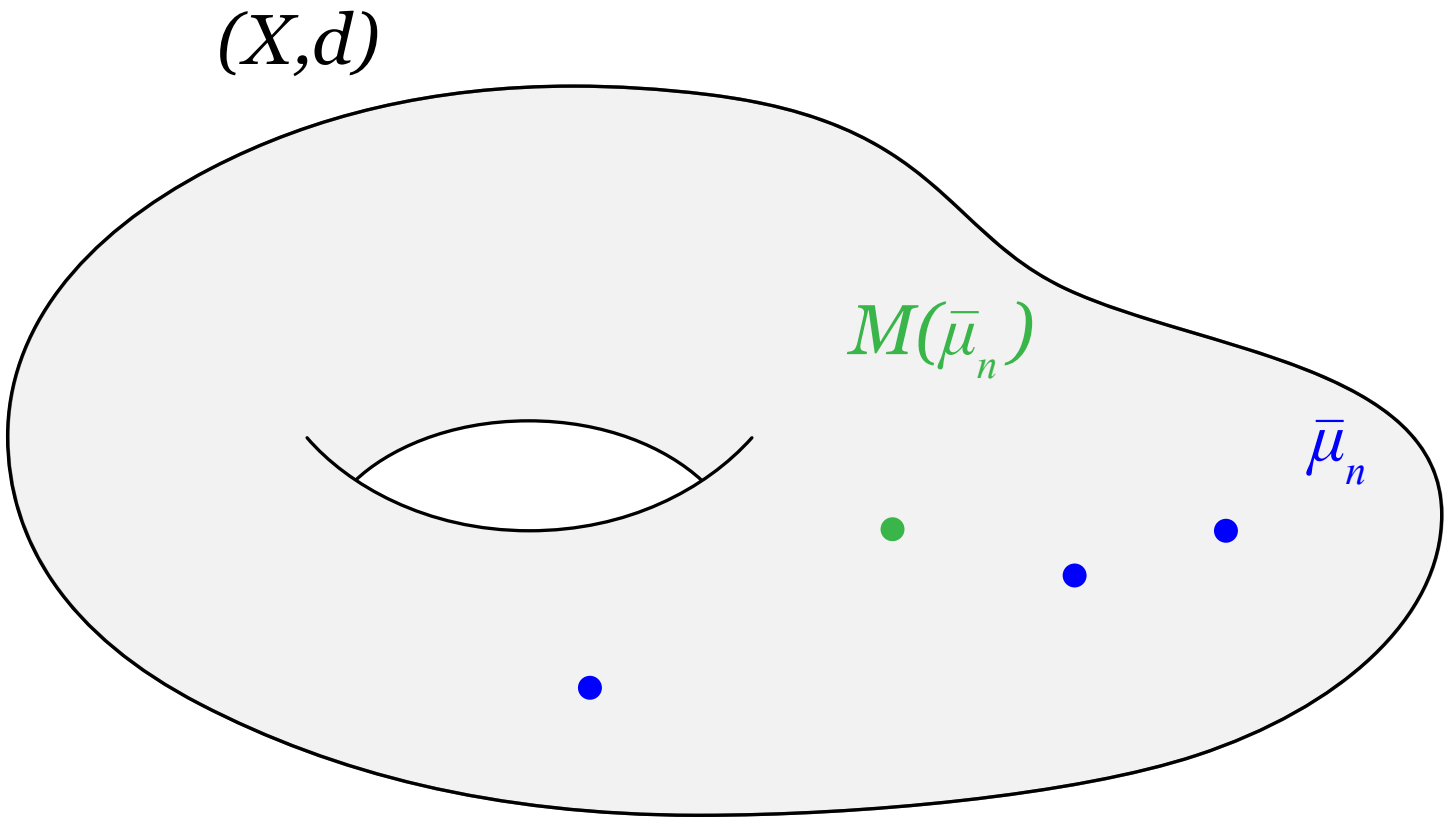
Suppose μ is unknown but we have independent, identically-distributed samples $Y_1, Y_2, \ldots$ from μ. How can we estimate the set $M(\mu)$?

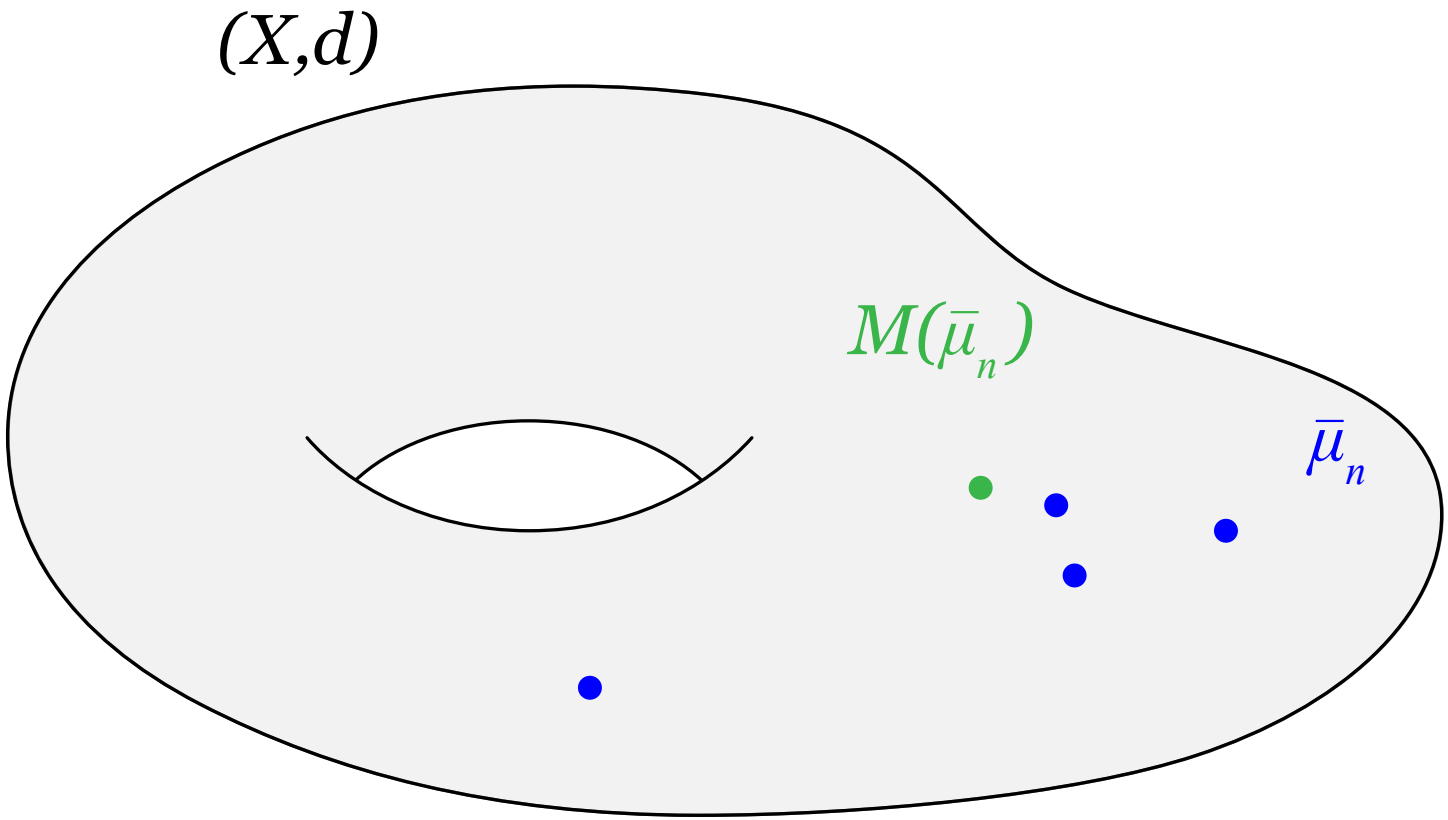
Natural idea is to consider the *empirical Fréchet mean set:*

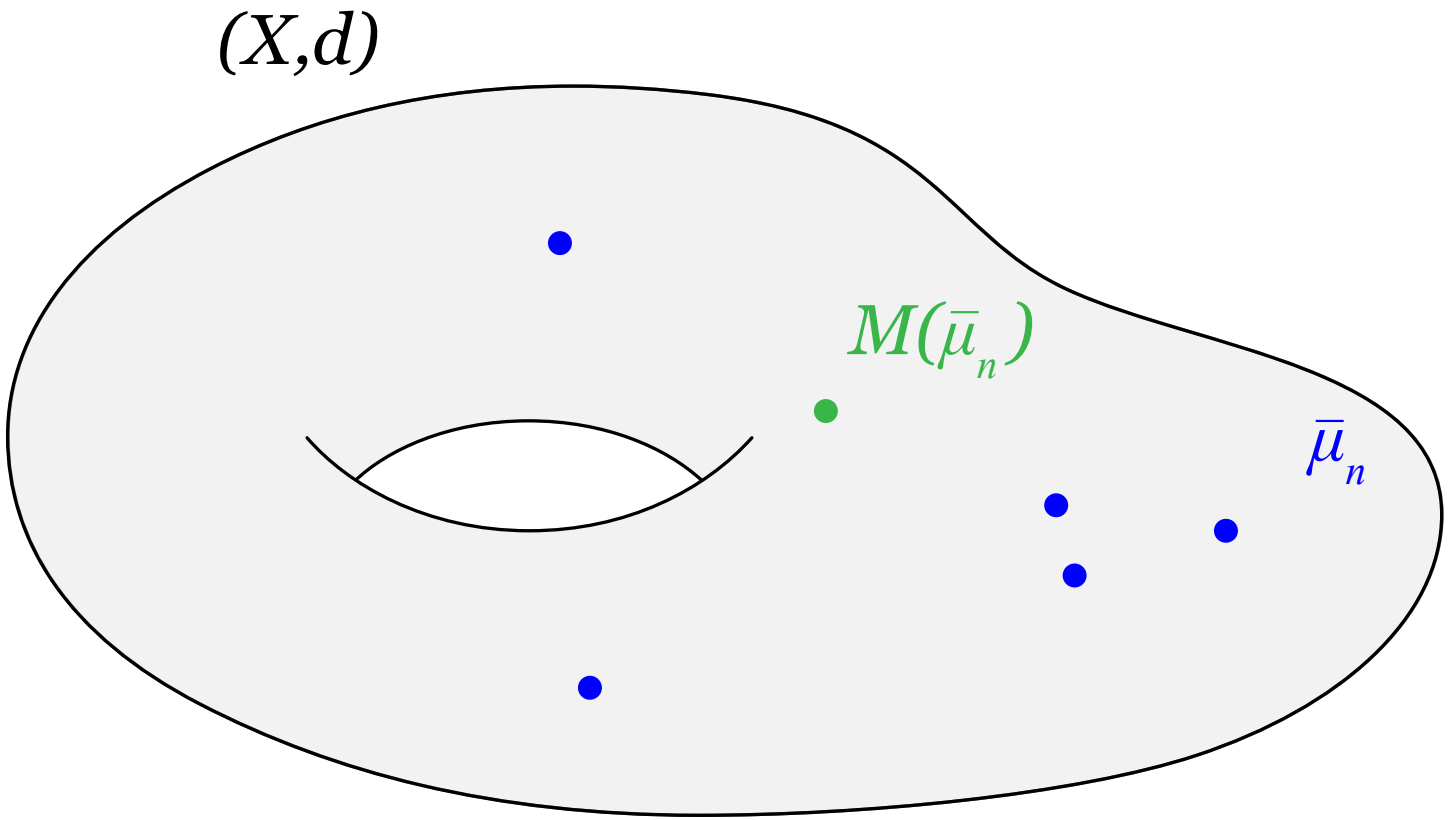
$$M(\bar{\mu}_n) := \arg\min_{x \in X} \frac{1}{n} \sum_{i=1}^{n} d^2(x, Y_i).$$

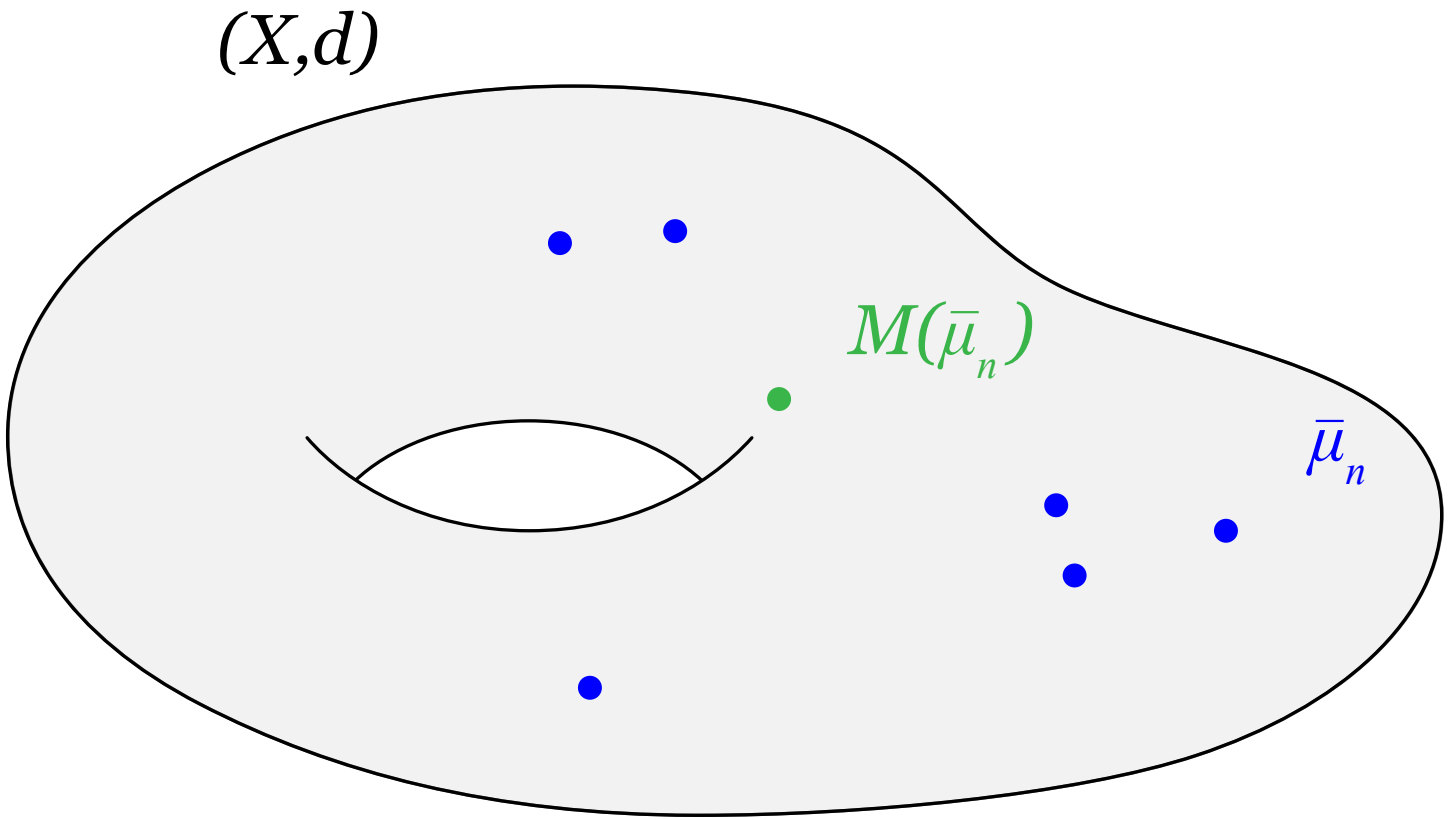
Here, $\bar{\mu}_n := \frac{1}{n} \sum_{i=1}^{n} \delta_{Y_i}$ is the *empirical distribution.*

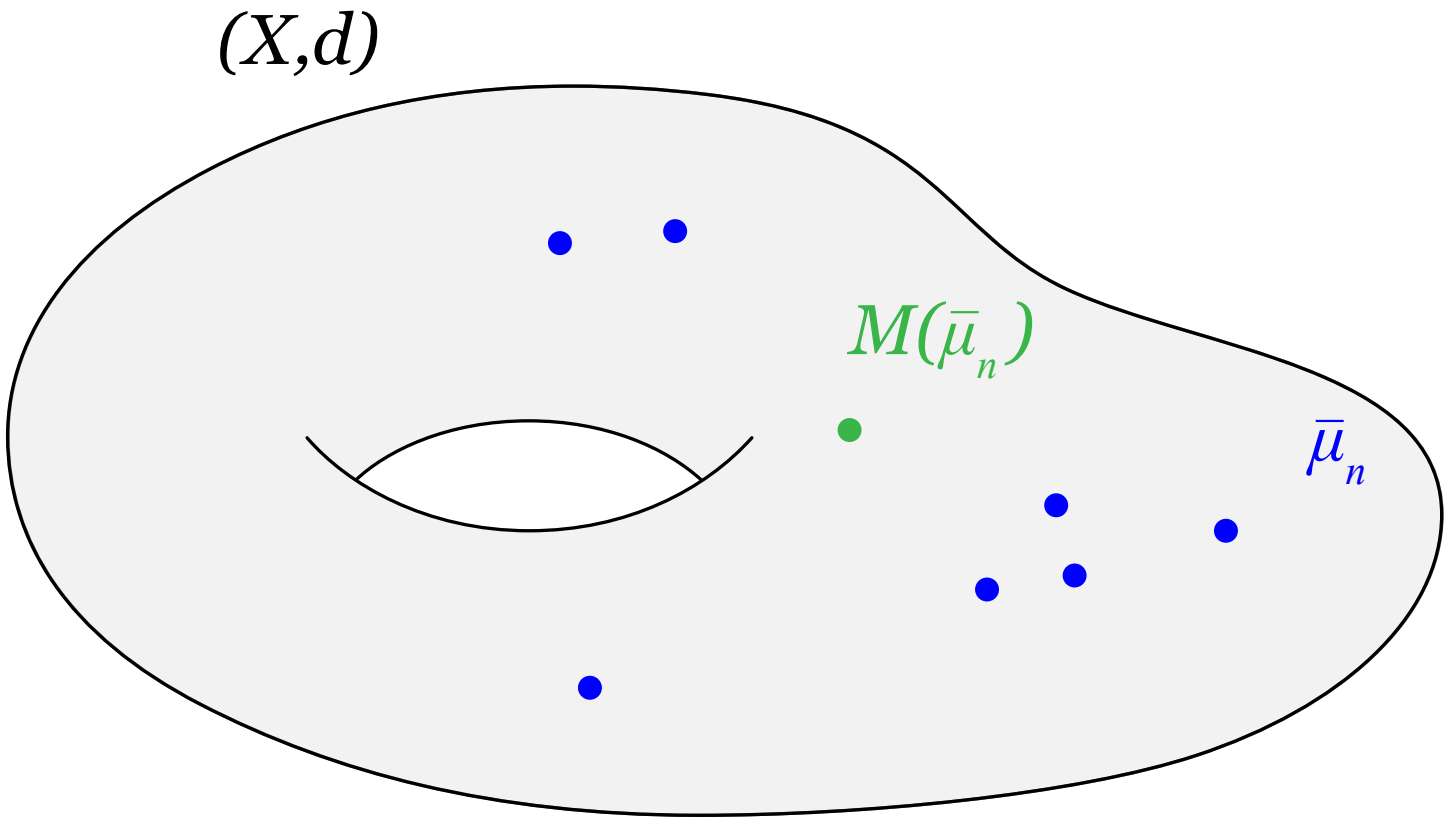

(X,d)
$M(\bar{\mu}_n)$
$\bar{\mu}_n$

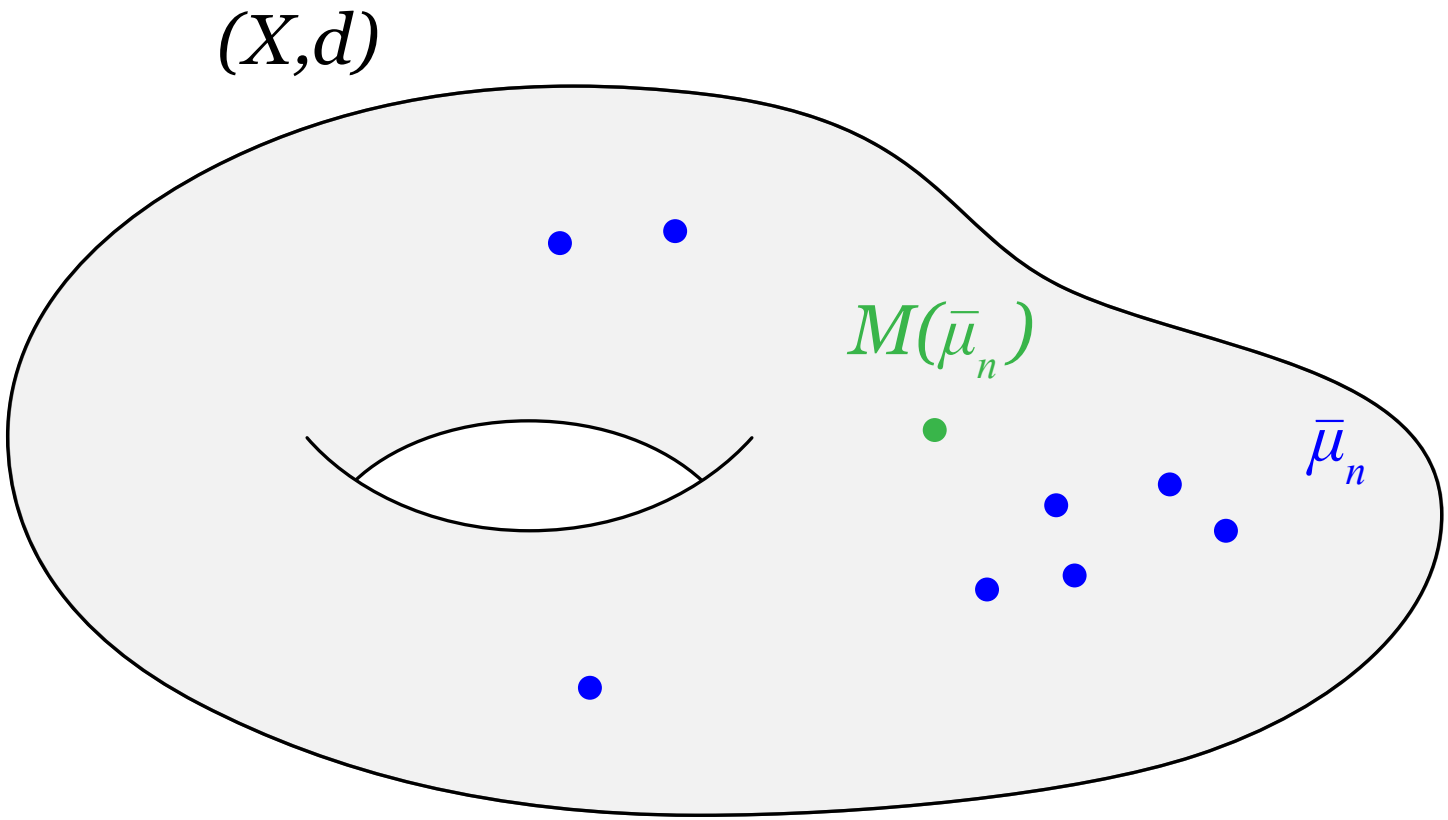

(X,d)
$M(\bar{\mu}_n)$
$\bar{\mu}_n$


(X,d)
$M(\bar{\mu}_n)$
$\bar{\mu}_n$


(X,d)
$M(\bar{\mu}_n)$
$\bar{\mu}_n$


(X,d)
$M(\bar{\mu}_n)$
$\bar{\mu}_n$


(X,d)
$M(\bar{\mu}_n)$
$\bar{\mu}_n$


(X,d)
M(μ̄_n)
μ̄_n

How does $M(\bar{\mu}_n)$ converge to $M(\mu)$?

One result is the "no false positives" property:

Theorem (Evans-AQJ 2024, Schötz 2022)

If (X, d) is finite-dimensional, then every empirical Fréchet mean is close to some population Fréchet mean.

How does $M(\bar{\mu}_n)$ converge to $M(\mu)$?

One result is the "no false positives" property:

Theorem (Evans-AQJ 2024, Schötz 2022)

If (X, d) is finite-dimensional, then we have

$$\max_{\bar{x}_n \in M(\bar{\mu}_n)} \min_{x \in M(\mu)} d(\bar{x}_n, x) \to 0 \qquad \text{(noFP)}$$

almost surely.

Do we have an analogous "no false negatives" property? No:

Theorem (Evans-AQJ 2024)

If (X, d) is finite and $\#M(\mu) > 1$, then there exists a population Fréchet mean which is not close to any empirical Fréchet mean.

Do we have an analogous "no false negatives" property? No:

Theorem (Evans-AQJ 2024)

If (X, d) is finite and $\#M(\mu) > 1$, then

$$\max_{x \in M(\mu)} \min_{\bar{x}_n \in M(\bar{\mu}_n)} d(\bar{x}_n, x) \to 0 \qquad \text{(noFN)}$$

occurs with probability zero.

Is this bad news? Both properties of "no false positives" and "no false negatives" are natural and desirable.

However, the empirical Fréchet mean is not the only estimator!

Question

Can we construct an estimator $\hat{M}_n = \hat{M}_n(Y_1, \ldots, Y_n)$ *of* $M(\mu)$ *which satisfies both "no false positives" and "no false negatives"?*

Is this bad news? Both properties of "no false positives" and "no false negatives" are natural and desirable.

However, the empirical Fréchet mean is not the only estimator!

Question

Can we construct an estimator $\hat{M}_n = \hat{M}_n(Y_1, \ldots, Y_n)$ *of* $M(\mu)$ *which satisfies both*

$$\max_{\hat{x}_n \in \hat{M}_n} \min_{x \in M(\mu)} d(\hat{x}_n, x) \to 0 \qquad \text{(noFP)}$$

and

$$\max_{x \in M(\mu)} \min_{\hat{x}_n \in \hat{M}} d(\hat{x}_n, x) \to 0 \qquad \text{(noFN)}$$

almost surely?

Is this bad news? Both properties of "no false positives" and "no false negatives" are natural and desirable.

However, the empirical Fréchet mean is not the only estimator!

Question

Can we construct an estimator $\hat{M}_n = \hat{M}_n(Y_1, \ldots, Y_n)$ *of* $M(\mu)$ *which satisfies*

$$d_{\mathrm{H}}(\hat{M}_n, M(\mu)) \to 0$$

almost surely, where d_{H} *denotes the Hausdorff metric?*

In this case we say that $\hat{M}_n$ is d_{H}*-consistent.*

I. Introduction

II. Problem Statement

III. Main Results

IV. Phylogenetic Application

V. Future Work

III. Main Results

The empirical Fréchet mean $M(\bar{\mu}_n)$ is "not large enough" to capture all points of the population Fréchet mean $M(\mu)$.

So, let's enlarge it.

For $\varepsilon \geq 0$, define

$$M(\mu; \varepsilon) := \left\{ x \in X : \int_X d^2(x, y)\, d\mu(y) \leq \min_{z \in X} \int_X d^2(z, y)\, d\mu(y) + \varepsilon \right\}$$

called the *ε-relaxed Fréchet mean.* Note $M(\mu; 0) = M(\mu)$.

Idea is to use $\hat{M}_n := M(\bar{\mu}_n, \varepsilon_n)$ for some carefully-chosen relaxation ε_n.

Know that $\varepsilon_n = 0$ leads to "no false positives". More generally, $\varepsilon_n \to 0$ leads to "no false positives" (Schötz 2022).

Can we get "no false negatives" by choosing $\varepsilon_n \to 0$ sufficiently slowly? In some simple examples, it is known that $n^{-1/4}$ is d_{H}-consistent (Schötz 2022).

Is it always possible to find some sufficiently slow ε_n?

If yes, what is the fastest possible sufficiently slow ε_n?

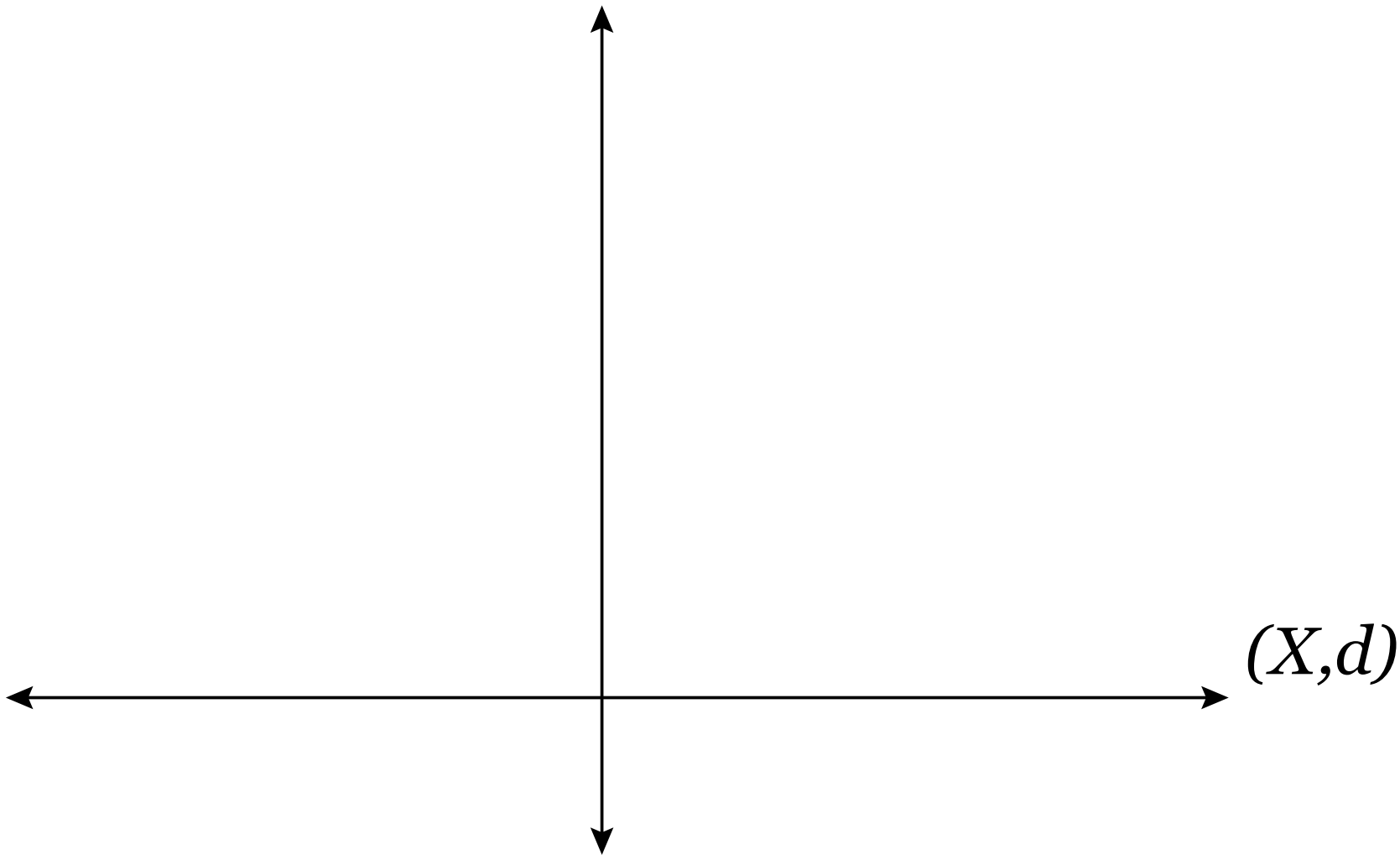
Define the *Fréchet functional* and the *empirical Fréchet functional* via

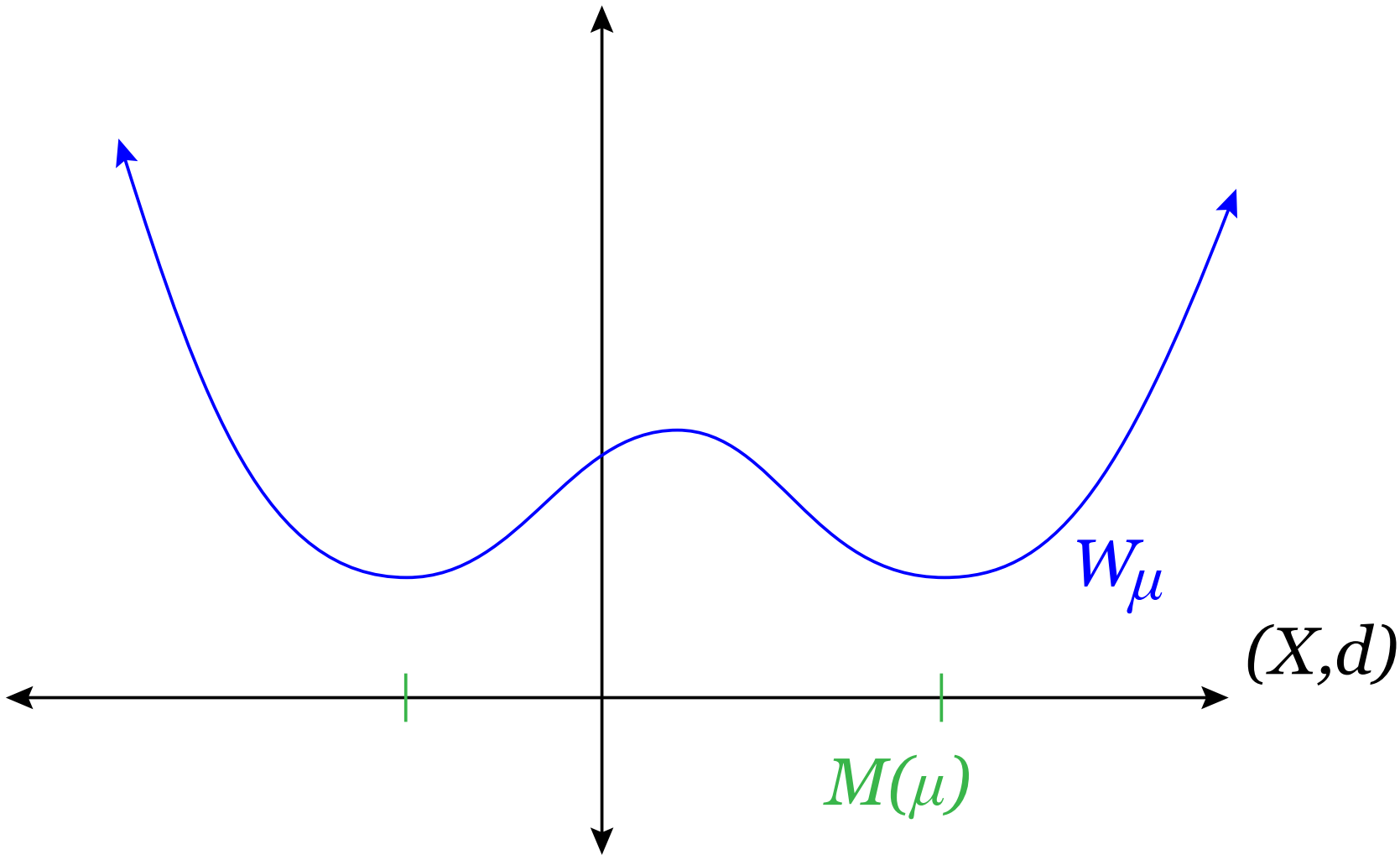
$$W_\mu(x) := \int_X d^2(x,y)\, d\mu(y) \qquad \text{and} \qquad W_{\bar{\mu}_n}(x) := \frac{1}{n}\sum_{i=1}^{n} d^2(x, Y_i)$$

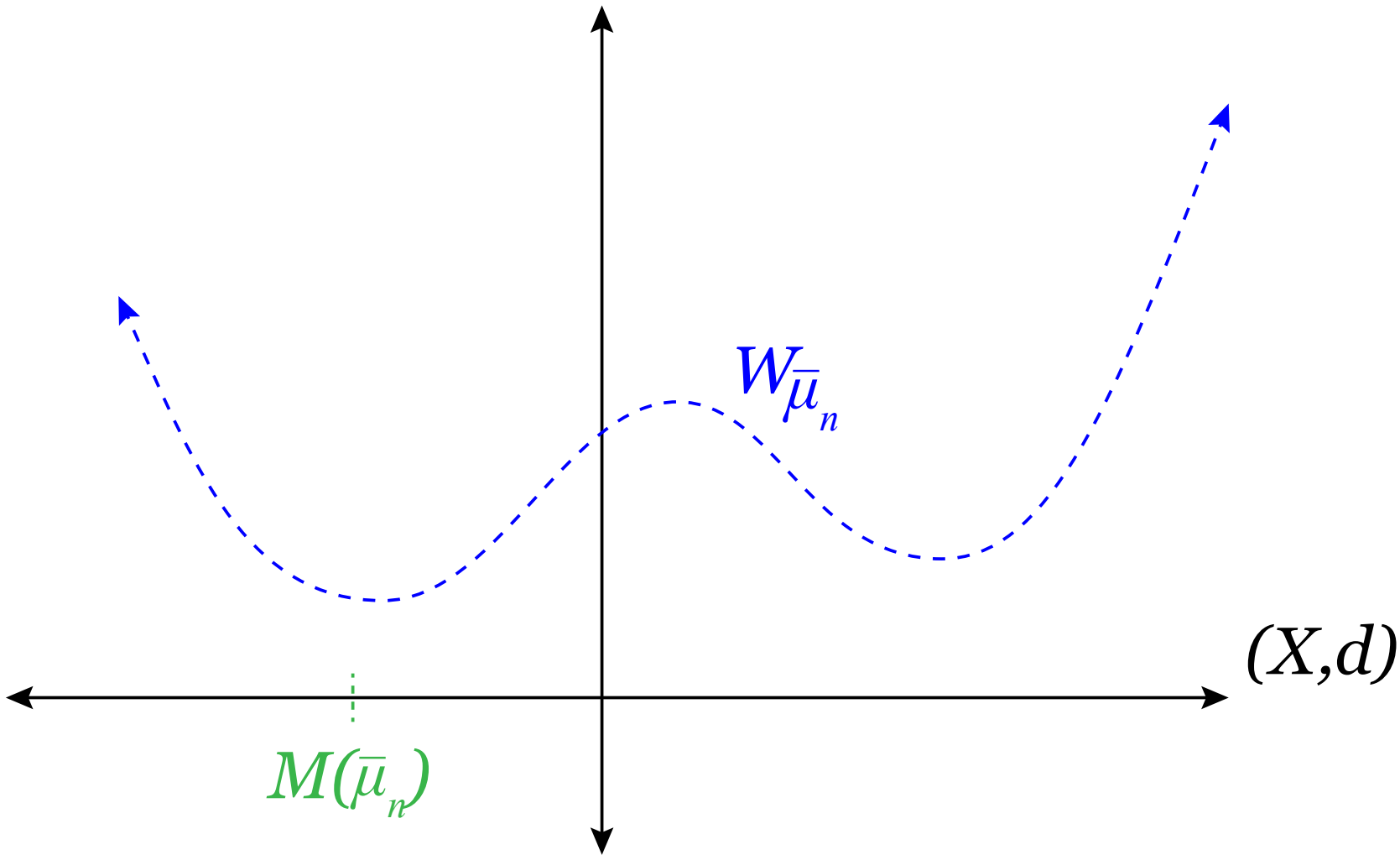
for $x \in X$.

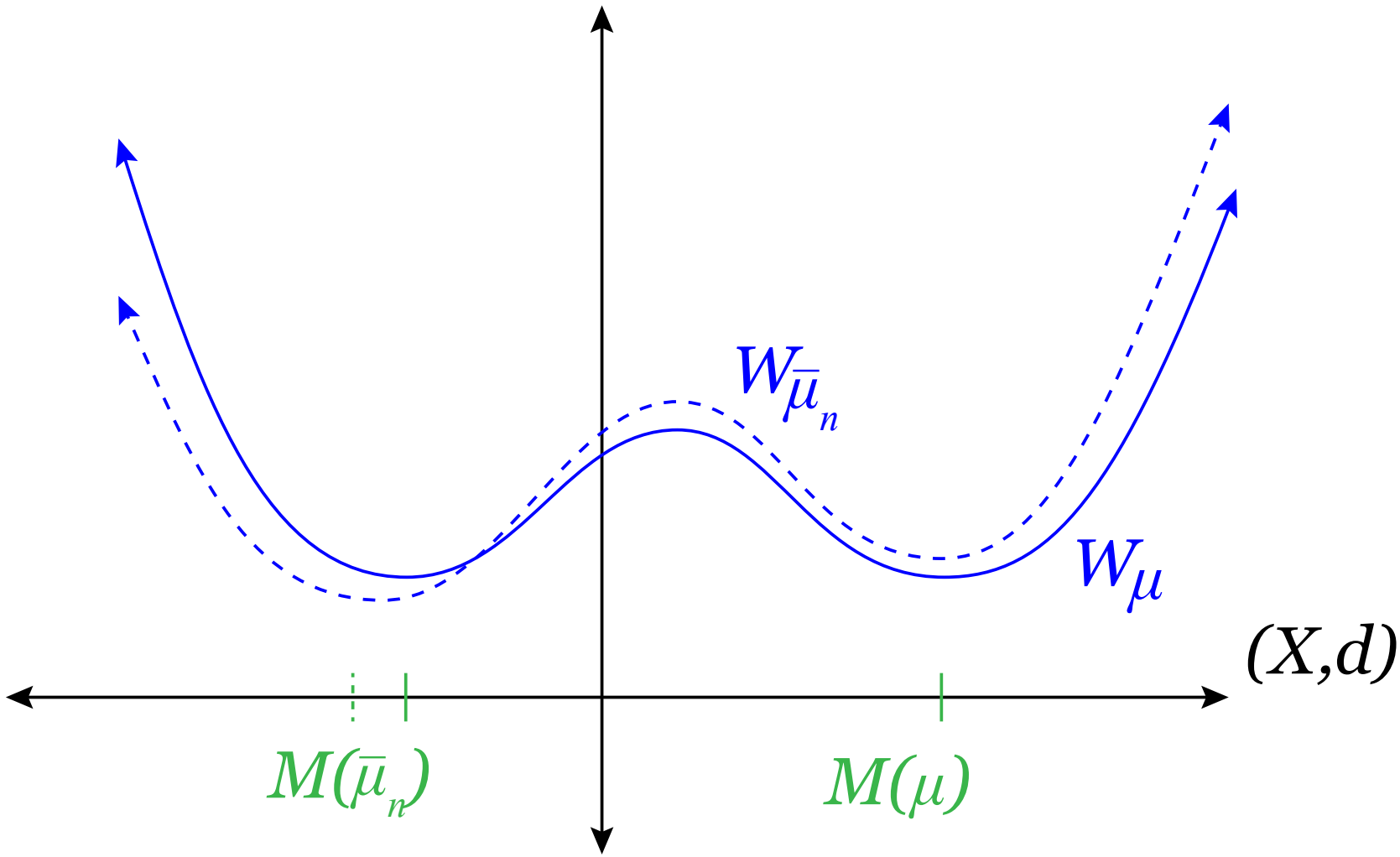
For each $x \in X$ we have $W_{\bar{\mu}_n}(x) \to W_\mu(x)$ almost surely, by the SLLN.

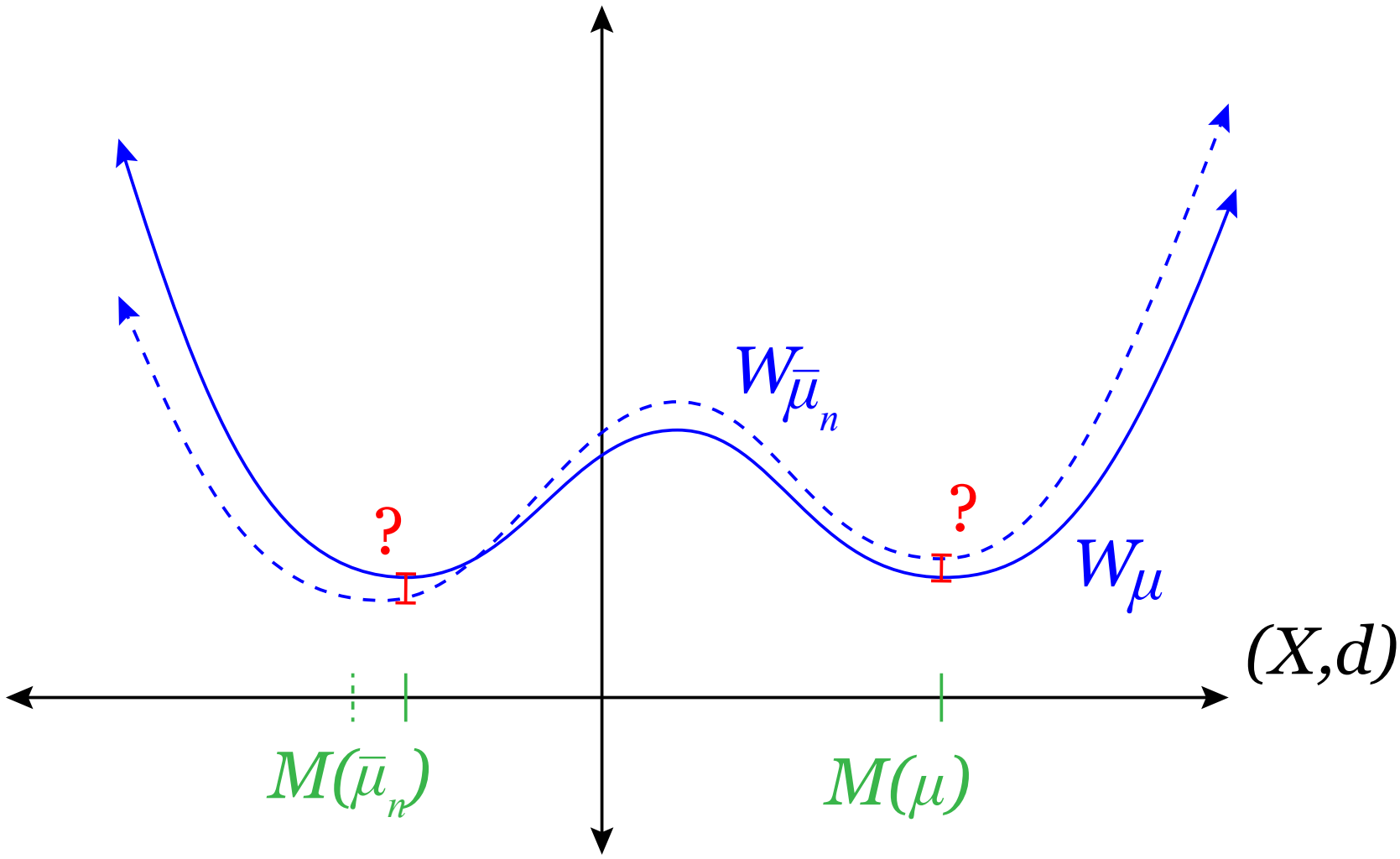
But we need $W_{\bar{\mu}_n} \to W_\mu$ in some stronger sense in order to understand how $M(\bar{\mu}_n)$ converges to $M(\mu)$.


(X,d)


W_μ
(X,d)
$M(\mu)$


$W_{\bar{\mu}_n}$
(X,d)
$M(\bar{\mu}_n)$


$W_{\bar{\mu}_n}$
W_μ
(X,d)
$M(\bar{\mu}_n)$
$M(\mu)$


$W_{\bar{\mu}_n}$
?
?
W_μ
(X,d)
$M(\bar{\mu}_n)$
$M(\mu)$

Theorem (Blanchard-AQJ)

If (X, d) is finite-dimensional, then there exists a number $\sigma(\mu) \in [0, \infty)$ such that $\varepsilon_n = \sigma(\mu) n^{-1/2} (\log \log n)^{1/2}$ is the cutoff between d_{H}-consistency and d_{H}-inconsistency:

- *Any relaxation slower than ε_n gives a d_{H}-consistent estimator.*
- *Any relaxation faster than ε_n gives a d_{H}-inconsistent estimator.*

Theorem (Blanchard-AQJ)

If (X, d) is finite-dimensional, then there exists a number $\sigma(\mu) \in [0, \infty)$ such that the relaxation scale $\varepsilon_n = cn^{-1/2}(\log\log n)^{1/2}$ for $c > 0$ satisfies:

- *If $c > \sigma(\mu)$, then $M_n(\bar{\mu}_n, \varepsilon_n)$ is d_{H}-consistent.*
- *If $c < \sigma(\mu)$, then $M_n(\bar{\mu}_n, \varepsilon_n)$ is not d_{H}-consistent.*

The critical parameter $\sigma(\mu)$ represents the maximal scale of the difference of the fluctuations of $W_{\bar{\mu}_n}$ on $M(\mu)$:

$$\sigma(\mu) := \sqrt{2} \cdot \sup_{x,x' \in M(\mu)} \sqrt{\mathrm{Var}(d^2(x, Y_1) - d^2(x', Y_1))}.$$

This depends on the population distribution, so it is not known.

Some ways around this:

- Use any asymptotically slower rate, like $\varepsilon_n \propto n^{-1/2}(\log n)^{1/2}$.
- Upper bound $\sigma(\mu)$ using additional moment information.
- Estimate $\sigma(\mu)$ from the data.

Can we just replace $n^{-1/2}(\log\log n)^{1/2}$ with $n^{-1/2}$?

Theorem (Blanchard-AQJ)

If (X, d) is finite-dimensional, then the relaxation $\varepsilon_n = cn^{-1/2}$ leads to estimator whose error probability decays like a Gaussian as a function of $c > 0$.

Can we just replace $n^{-1/2}(\log\log n)^{1/2}$ with $n^{-1/2}$?

Theorem (Blanchard-AQJ)

If (X, d) is finite-dimensional, then there exist numbers $m(\mu), \sigma(\mu) \in [0, \infty)$ such that the relaxation $\varepsilon_n = cn^{-1/2}$ for $c \geq m(\mu)$ satisfies

$$\sup_{\delta>0} \lim_{n\to\infty} \mathbb{P}\left(d_{\mathrm{H}}(M(\bar{\mu}_n, \varepsilon_n), M(\mu)) \geq \delta\right) \leq \exp\left(-\frac{(c - m(\mu))^2}{\sigma^2(\mu)}\right).$$

How do these estimators work in practice?

Phylogenetic Application

In computational phylogenetics, one has data in the form of *trees*.

Classical geometry on the space is trees is the *BHV metric* (Billera-Holmes-Vogtmann 2001), which has good and bad properties:

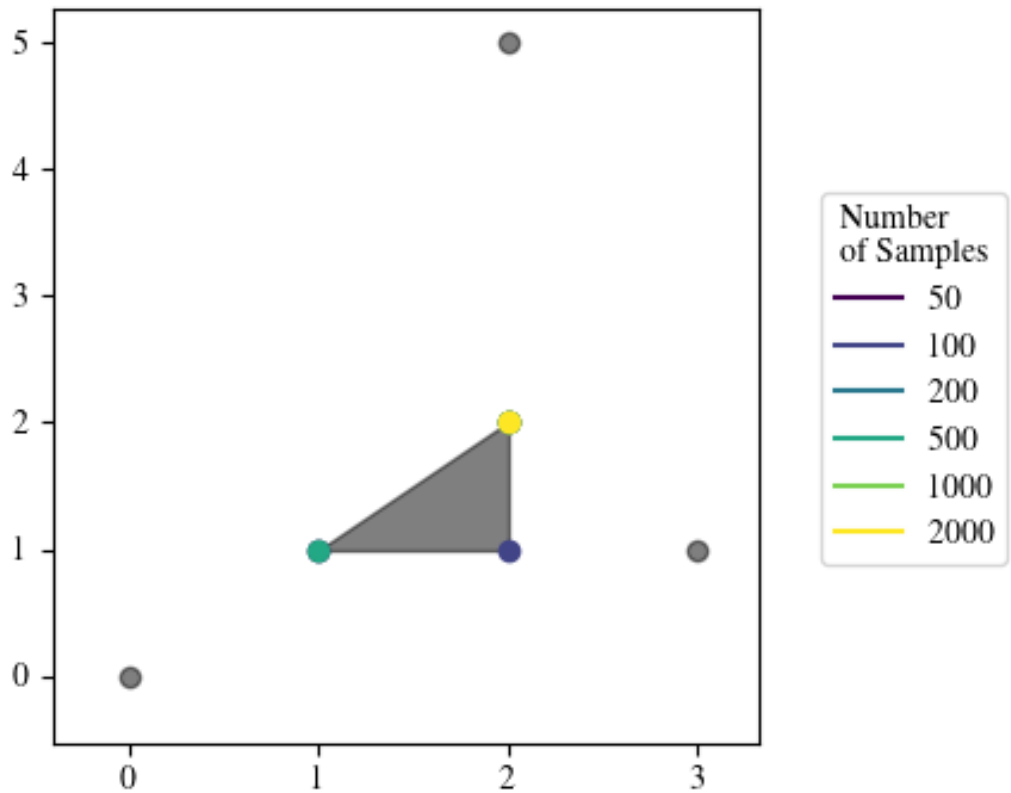
- Non-positive curvature, so Fréchet means are unique
- Stratified space, so locally Euclidean except at some singularities
- Hard to compute Fréchet means and geodesics
- Geodesics typically pass through the origin
- "Stickiness"

Alternative geometry is the *tropical projective metric* (Lin-Yoshida 2018), in which Fréchet medians are the object of interest. Different properties than the BHV treespace:

- Not a stratified space, and no non-positive curvature
- Fréchet medians and geodesics are easily computable
- Geodesics typically do not pass through origin
- Fréchet medians typically non-unique

The tropical projective metric is a non-Euclidean metric on $\mathbb{R}^m$, and each tree on N leaves is embedded in $\mathbb{R}^{\binom{N}{2}-1}$ as its distance matrix.

Unrelaxed Fréchet medians are bad estimators:



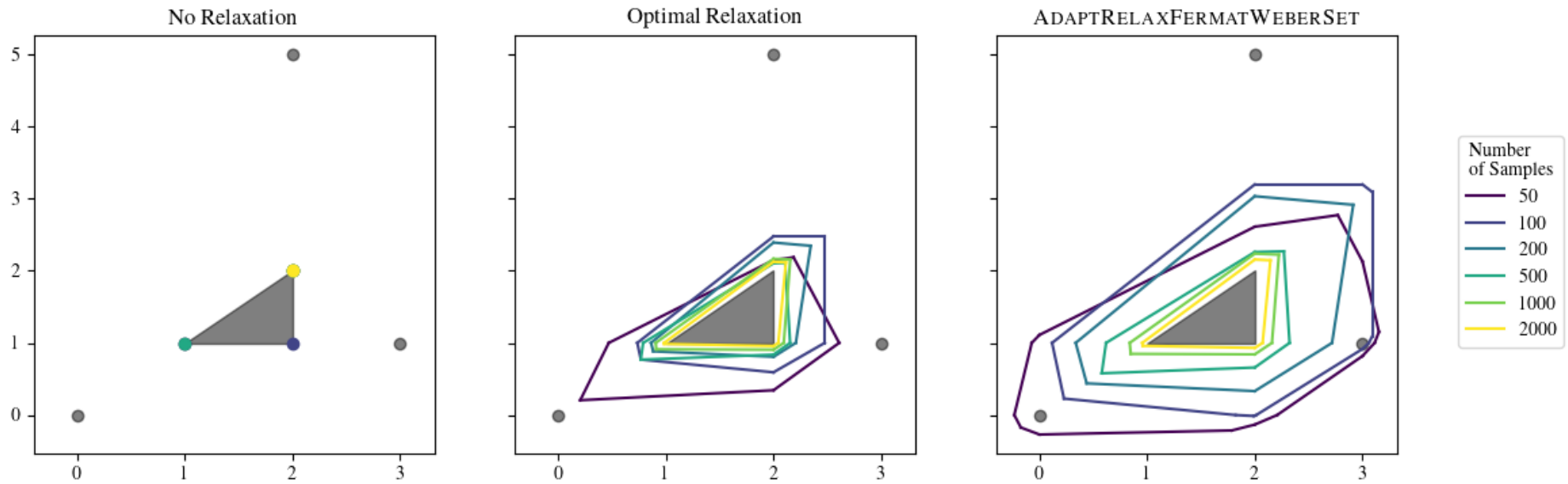
In this setting of tropical projective treespace, we can exactly implement an adaptive Fréchet median set estimation algorithm:

AdaptRelaxFermatWeberSet

We can implement this algorithm with standard convex optimization and polyhedral geometry software.

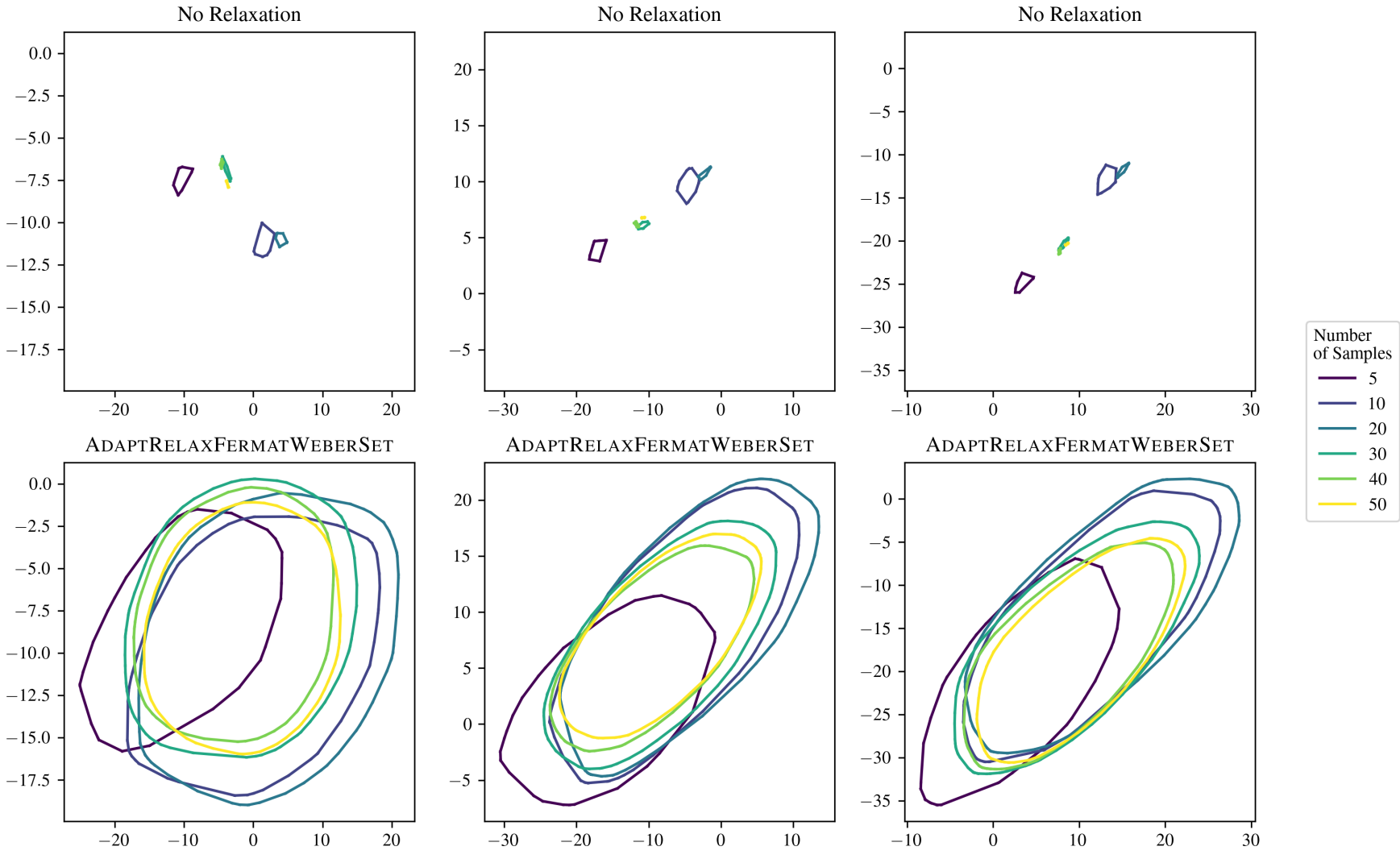
A simulated 3-leaf data set (dimension $\binom{3}{2} - 1 = 2$).

Exact plots of the estimated regions:



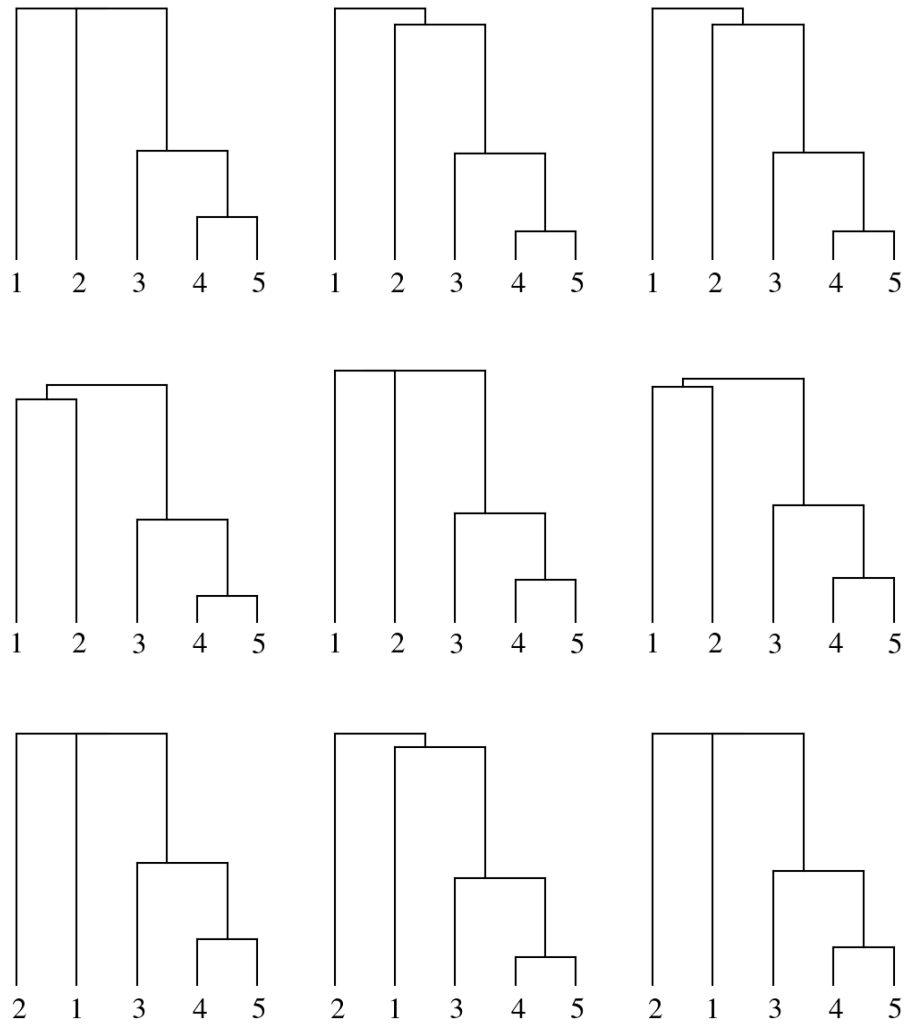
A real 4-leaf data set of influenza evolution (dimension $\binom{4}{2} - 1 = 5$).

We plot projections of the estimated regions onto random 2-dimensional subspaces:



A real 5-leaf data set of influenza evolution (dimension $\binom{5}{2} - 1 = 9$).

We uniformly some points from the interior of the estimated region:



Practical takeaways:

- Unrelaxed Fréchet means may be missing information.
- Relaxation methods provide a very conservative outer estimate.
- Computational difficulty is high.

Future Work

Unrelaxed Fréchet means are too small, and relaxed Fr'echet means are too big. How to balance these effects? Can we set up a rigorous hypothesis testing framework?

Can we efficiently implement relaxed Fréchet mean set estimators in other applications of interest?

Extend this theory to general ill-posed M-estimation problems?

What happens at the critical relaxation $\varepsilon_n = \sigma(\mu)n^{-1/2}(\log\log n)^{1/2}$?

Concentration inequalities for the slow relaxation $\varepsilon_n \propto n^{-1/2}(\log n)^{1/2}$?

Thank you!

References

Afsari, B. (2011). Riemmanian L^p center of mass: existence, uniqueness, and convexity. *Proc. Amer. Math. Soc.* **139** 655-673.

Bhattacharya, R. and Patrangenaru, V. (2003). Large sample theory of intrinsice and extrinsic sample means on manifolds. I. *Ann. Statist.* **31** 1-29.

Billera, L.J., Holmes, S. P., and Vogtmann, K. (2001). Geometry of the space of phylogenetic trees. *Adv. in Appl. Math.* **27**. 733-767

Cao, Y. and Monod, A. (2022). A geometric condition for uniqueness of Fréchet means of persistence diagrams. *arxiv pre-print.*

Eltzner, B. (2020). Testing for uniqueness of estimators. *arxiv pre-print.*

Evans, S. N. and Jaffe, A. Q. (2024). Limit theorems for Fréchet mean sets. *Bernoulli.* **30**(1). 419-447.

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

Le Gouic, T. and Loubes, J.-M. (2017). Existence and consistency of Wasserstein barycenters. *Probab. Theory Related Fields.* **168**(3). 901-917.

Lin, B. and Yoshida, R. (2018). Tropical Fermat-Weber points. *SIAM J. Discrete Math.* **32**. 1229-1245.

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

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). Contemp. Math.* **338**. 357-390.

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*. 591-602.