Strong and weak laws of large numbers for Frechet sample means in bounded metric spaces

Cedric E. Ginestet

Department of Neuroimaging, King’s College London
Department of Mathematics and Statistics, Boston University
Motivating Example

Figure: A Space of Simple Graphs, $\mathcal{G}$ with $|V(G)| = 4$. 

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

Figure: A Space of Simple Graphs, $\mathcal{G}$ with $|V(G)| = 4$.

Hamming Distance

\[ d(G, G') := \sum_{i<j} I\{e_{ij} \neq e'_{ij}\}. \]
Non-unique Means

Figure: A Space of Simple Graphs, $\mathcal{G}$ with $|V(G)| = 4$.

Figure: Set of means is a superset of the sampled graphs.
The Frechet Mean

Barycentre as Average

- Given a set of points \( x_1, \ldots, x_n \in \mathbb{R}^k \);
- And a set of masses \( w_1, \ldots, w_n \in \mathbb{R} \);
- The barycentre is

\[
\bar{x} := \frac{1}{n} \sum_{i=1}^{n} w_i x_i.
\]

- When the distribution of mass is uniform, this is the centroid.
The Frechet Mean

Barycentre as Average

- Given a set of points $x_1, \ldots, x_n \in \mathbb{R}^k$;
- And a set of masses $w_1, \ldots, w_n \in \mathbb{R}$;
- The barycentre is
  \[ \bar{x} := \frac{1}{n} \sum_{i=1}^{n} w_i x_i. \]  
  \( 1 \)
- When the distribution of mass is uniform, this is the centroid.

Barycentre as Minimizer

- Take the $L^2$-metric on $\mathbb{R}^k$;
- The barycentre is a minimizer,
  \[ \bar{x} = \arg\min_{x' \in \mathbb{R}^k} \frac{1}{n} \sum_{i=1}^{n} w_i \| x_i - x' \|_2^2. \]  
  \( 2 \)
The Frechet Mean

The Most ‘Central’ Element

- Given a metric space \((\mathcal{X}, d)\),
- Let an abstract-valued r.v. \(X : (\Omega, \mathcal{F}, \mathbb{P}) \mapsto (\mathcal{X}, \mathcal{B})\),
- Fréchet (1948) defined the ‘mean’ value, for every \(r \geq 1\),

\[
\Theta^r := \arg\inf_{x' \in \mathcal{X}} \int_{\mathcal{X}} d(x, x')^r d\mu(x).
\]
The Frechet Mean

The Most ‘Central’ Element

- Given a metric space \((\mathcal{X}, d)\),
- Let an abstract-valued r.v. \(X : (\Omega, \mathcal{F}, \mathbb{P}) \mapsto (\mathcal{X}, \mathcal{B})\),
- Fréchet (1948) defined the ‘mean’ value, for every \(r \geq 1\),
\[
\Theta^r := \operatorname{arginf}_{x' \in \mathcal{X}} \int_{\mathcal{X}} d(x, x')^r d\mu(x).
\]

Frechet Sample Mean

- Given a family of abstract-valued r.v.s \(X_i\)’s, \(i = 1, \ldots, n\),
\[
\hat{\Theta}_n^r := \operatorname{arginf}_{x' \in \mathcal{X}} \frac{1}{n} \sum_{i=1}^{n} d(X_i, x')^r.
\]
- In general, \(\Theta^r, \hat{\Theta}_n^r\) will not be unique: i.e. \(\Theta^r, \hat{\Theta}_n^r \subseteq \mathcal{X}\).
Why the Frechet Mean?

‘All’ Statistics are Frechet

- For $x \in \mathbb{R}$, we obtain:
- If $d(x, x')^2 := |x - x'|^2$, then $\Theta$ is the arithmetic mean.
- If $d(x, x')^1 := |x - x'|^1$, then $\Theta$ is the median.
- If $d(x, x') := \mathcal{I}\{x = x'\}$, then $\Theta$ is the mode.

Applications

- Hamming distance on finite alphabets (e.g. stretches of DNA).
- Hausdorff metric on spaces of images (e.g. neuroimaging).
- Procrustean metric on spaces of shapes (e.g. medical imaging).
- Geodesic distance on spaces of distributions (e.g. mach. learn.).
Questions and Assumptions

Convergence of Frechet Sample Mean

\[ \hat{\Theta}_n^r := \arg\inf_{x' \in \mathcal{X}} \frac{1}{n} \sum_{i=1}^{n} d(X_i, x')^r \rightarrow \arg\inf_{x' \in \mathcal{X}} \int_{\mathcal{X}} d(x, x')^r d\mu(x) =: \Theta^r, \]

- Ziezold (1977) in separable bounded metric spaces, for \( r = 2 \);
- Sverdrup-Thygeson (1981) in compact metric spaces, for \( r \geq 1 \).
Questions and Assumptions

Convergence of Frechet Sample Mean

\[ \hat{\Theta}_n^r := \arg\inf_{x' \in \mathcal{X}} \frac{1}{n} \sum_{i=1}^{n} d(X_i, x')^r \rightarrow \arg\inf_{x' \in \mathcal{X}} \int_{\mathcal{X}} d(x, x')^r d\mu(x) =: \Theta^r, \]

- Ziezold (1977) in separable bounded metric spaces, for \( r = 2 \);
- Sverdrup-Thygeson (1981) in compact metric spaces, for \( r \geq 1 \).

Topological and Measure-theoretic Assumptions

- Separable bounded \( (\mathcal{X}, d) \).
- Possibly empty \( \Theta_n^r, \Theta^r \).
- Possibly non-unique \( \Theta_n^r, \Theta^r \).
- Restricted Frechet sample mean.
Part I

Sequences of Frechet Sample Means
Set-valued Outer and Inner Limits

**Figure:** Outer and inner limits, defined wrt set inclusion on $\mathcal{X}$.

\[
\limsup_{n \to \infty} S_n := \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} S_n, \quad \text{and} \quad \liminf_{n \to \infty} S_n := \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} S_n.
\]
The closed interval $[-1, 1]$ is equipped with the Manhattan metric, and two point masses at $-1$ and $1$. For $r = 1$, the theoretical Fréchet mean is the median of $X$. But the sequence of Fréchet sample means diverges,

$$\limsup_{n \to \infty} \hat{\Theta}_n(\omega) = \{-1, 1\} \supset \liminf_{n \to \infty} \hat{\Theta}_n(\omega) = \emptyset.$$
Weaker Type of Convergence

Possible Solutions
Taken together, these two problems necessitate:

i. Study of asymptotic behavior of the outer limit of the $\hat{\Theta}_n$’s.

ii. Convergence of the Fréchet sample mean in terms of set inclusion.

Set-inclusion Convergence
We thus consider the following event,

$$\mathbb{P}\left[\left\{\omega \in \Omega : \limsup_{n \to \infty} \hat{\Theta}_n(\omega) \subseteq \Theta\right\}\right] = 1,$$

where recall that $\liminf S_n := \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} S_n$. 
The closed interval $[-1, 1]$ is equipped with the Manhattan metric, and two point masses at $-1$ and $1$, with $r = 2$. The theoretical Fréchet mean is the arithmetic mean. But the limsup and liminf of the Fréchet sample means are empty,

$$\limsup_{n \to \infty} \hat{\Theta}_n = \liminf_{n \to \infty} \hat{\Theta}_n = \emptyset.$$
Kuratowski Upper Limit

Equivalent Definitions
Given a sequence of subsets $A_n \subseteq \mathcal{X}$:

- Set of cluster points of sequences $a_n \in A_n$;
- The Kuratowski upper limit is

$$\limsup_{n \to \infty} A_n := \left\{ x \in \mathcal{X} : \liminf_{n \to \infty} d(x, A_n) = 0 \right\}.$$  

where liminf and Limsup are taken with respect to real numbers and subsets of $\mathcal{X}$, respectively.

Lemma

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

$$\limsup_{n \to \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m.$$
Part II

Almost Sure Consistency of Frechet Sample Mean
Almost Sure Consistency of Fréchet Sample Mean

Main Result

▸ In a separable bounded metric space \((X, d)\),

▸ For any \(X_1, \ldots, X_n\) be an iid sequence,

▸ Convergence of the infima:

\[
\hat{\sigma}_n^r := \inf_{x' \in X} \frac{1}{n} \sum_{i=1}^{n} d(X_i, x')^r \rightarrow \inf_{x' \in X} \int_X d(x, x')^r dx =: \sigma^r \quad \text{a.s.,}
\]

▸ Convergence of the arginf's:

\[
\limsup_{n \to \infty} \hat{\Theta}_n^r \subseteq \Theta^r \quad \text{a.s.}
\]

▸ For every \(r > 0\).
Strong Law of Large Numbers

Pointwise Convergence

- For every $x' \in \mathcal{X}$, each $\int d(x, x') r \, d\mu_n(x) \in \mathbb{R}$.
- By the strong law of large numbers, we have,

$$\lim_{n \to \infty} \left| \int_{\mathcal{X}} d_{\mathcal{Z}}^r \, d\mu_n - \int_{\mathcal{X}} d_{\mathcal{Z}}^r \, d\mu \right| = 0 \quad \text{a.s.,}$$

for every $z \in \mathcal{X}$, where $\mu_n := \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}$. 

The value of $\hat{\Theta}_n$ depends on the whole of $f_{\mathcal{Y}}: \mathcal{X} \mapsto \frac{1}{n} \sum_{i=1}^{n} d(X_i, \cdot)$. We need functional convergence.
Strong Law of Large Numbers

Pointwise Convergence

- For every $x' \in X$, each $\int d(x, x')^r d\mu_n(x) \in \mathbb{R}$.
- By the strong law of large numbers, we have,

$$
\lim_{n \to \infty} \left| \int_X d^r_z d\mu_n - \int_X d^r_z d\mu \right| = 0 \quad \text{a.s.,}
$$

for every $z \in X$, where $\mu_n := \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$.
- The value of $\hat{\Theta}_n^r$ depends on the whole of

$$
f^r_n : X \mapsto \frac{1}{n} \sum_{i=1}^n d(X_i, \cdot)^r.
$$

- We need functional convergence.
Almost Sure Uniform Weak Convergence

Glivenko-Cantelli Lemma (Rao, 1962)

i. Let $\mathcal{F}(X)$ be a class of real-valued functions on separable $X$, 
ii. Let a sequence of finite measures $\mu_n$, and $\mu$, 
iii. $\mathcal{F}(X)$ is dominated by a continuous integrable function on $X$, 
iv. $\mathcal{F}(X)$ is equicontinuous, and 
v. $\mu_n \Rightarrow \mu$, a.s.;

then we obtain \textit{uniform} a.s. weak convergence,

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

Definition
For some \( z \in \mathcal{X} \), the \( z \)-point function is

\[
d_z(x) := d(z, x),
\]
for every \( x \in \mathcal{X} \). The class of point functions on \((\mathcal{X}, d)\) is then denoted by

\[
\mathcal{D}^r(\mathcal{X}) := \{d^r_z : \forall z \in \mathcal{X}\}.
\]
for every \( r \geq 1 \).

Lemma
If \((\mathcal{X}, d)\) is a bounded metric space, then for every \( r \geq 1 \), \( \mathcal{D}^r(\mathcal{X}) \) is

i. Uniformly bounded;

ii. Uniformly equicontinuous.
Almost Sure Consistency of Frechet Sample Mean

Strengthening of a.s. Weak Convergence

Using the Glivenko-Cantelli lemma, we obtain for every $r \geq 1$,

$$\lim_{n \to \infty} \sup_{z \in D^r} \left| \int_{\mathcal{X}} d^r_z d\mu_n - \int_{\mathcal{X}} d^r_z d\mu \right| = 0 \quad \text{a.s.}$$

where $\mu_n := \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}$.
Almost Sure Consistency of Frechet Sample Mean

Strengthening of a.s. Weak Convergence
Using the Glivenko-Cantelli lemma, we obtain for every \( r \geq 1 \),

\[
\lim_{n \to \infty} \sup_{z \in \mathcal{D}^r} \left| \int_{\mathcal{X}} d^r_z \, d\mu_n - \int_{\mathcal{X}} d^r_z \, d\mu \right| = 0 \quad \text{a.s..}
\]

where \( \mu_n := \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i} \).

Almost Sure Convergence of \( \hat{\sigma}^r_n \)

- In order to show that \( \hat{\sigma}^r_n \to \sigma^r \) a.s.,
- We also need the following ‘sandwich’ relationship,

\[
T_n^r \leq \hat{\sigma}^r_n - \sigma^r \leq T_n^r.
\]
Almost Sure Consistency of Frechet Sample Mean

Sandwich Argument I

By the \textit{minimality} of $\theta \in \Theta^2$,

$$T_n(\hat{\theta}_n) := \frac{1}{n} \sum_{i=1}^{n} d^2_{\hat{\theta}_n} (X_i) - \int_{X} d^2_{\hat{\theta}_n} (x) d\mu(x)$$

$$\leq \frac{1}{n} \sum_{i=1}^{n} d^2_{\theta} (X_i) - \int_{X} d^2_{\theta} (x) d\mu(x) =: T^*_n(\hat{\theta}_n).$$
Almost Sure Consistency of Frechet Sample Mean

Sandwich Argument I

By the minimality of $\theta \in \Theta^2$,

$$T_n(\hat{\theta}_n) := \frac{1}{n} \sum_{i=1}^{n} d^2_{\hat{\theta}_n}(X_i) - \int_{\mathcal{X}} d^2_{\hat{\theta}_n}(x) d\mu(x)$$

$$\leq \frac{1}{n} \sum_{i=1}^{n} d^2(X_i) - \int_{\mathcal{X}} d^2_{\theta}(x) d\mu(x) =: T_n^*(\hat{\theta}_n).$$

Sandwich Argument II

By the minimality of $\hat{\theta}_n \in \hat{\Theta}_n^2$,

$$T_n^*(\hat{\theta}_n) := \frac{1}{n} \sum_{i=1}^{n} d^2_{\hat{\theta}_n}(X_i) - \int_{\mathcal{X}} d^2_{\hat{\theta}_n}(x) d\mu(x)$$

$$\leq \frac{1}{n} \sum_{i=1}^{n} d^2(X_i) - \int_{\mathcal{X}} d^2_{\theta}(x) d\mu(x) =: T_n(\theta).$$
From Real Line Back to Metric Space

$\phi(\cdot)$ is a bijection.

$\phi(\cdot)$ is bicontinuous.

$Limsup\hat{\Theta} \subseteq \Theta$ a.s.
From Real Line Back to Metric Space

Homeomorphism from $(\mathbb{R}, d_E)$ to $(\mathcal{X}, d)$:

- $\varphi(\cdot)$ is a bijection.
- $\varphi(\cdot)$ is bicontinuous.
- Limsup $\hat{\Theta}_n^r \subseteq \Theta^r$ a.s.
Homeomorphism through Quotient Space

$\phi(X, d) \rightarrow (X/\sim, d_\sim) \rightarrow g(\mathbb{R}^+, d_E)$

Canonical decomposition (Bourbaki, 1989)

- $(X/\sim, d_\sim)$ is the quotient space induced by $\sim$.
- With the equivalence relationship:
- $x \sim x'$ if, and only if, $\mathbb{E}[d(X, x)^r] = \mathbb{E}[d(X, x')^r]$. 
Sample Restricted Frechet Mean

When minimization is infeasible, define

$$
\hat{\Theta}^\ast, r_n := \arg\min_{x' \in X} \frac{1}{n} \sum_{i=1}^{n} d(X_i, x')^r \quad \text{and} \quad \hat{\sigma}^\ast, r_n := \min_{x' \in X} \frac{1}{n} \sum_{i=1}^{n} d(X_i, x')^r,
$$

where $X := \{X_1, \ldots, X_n\} \subseteq \mathcal{X}$ denotes the set of sampled variables.

Theoretical Restricted Frechet Mean

$$
\Theta^\ast, r := \arg\min_{x' \in W} \int_{\mathcal{X}} d(x, x')^r d\mu(x), \quad \text{and} \quad \sigma^\ast, r := \min_{x' \in W} \int_{\mathcal{X}} d(x, x')^r d\mu(x),
$$

where $W$ is the support of $\mu$, which is assumed to be closed.
Generalized MSE Decomposition

MSE Definition
The case of $r = 2$ will be referred to as **metric squared error (MSE)** convergence, and the MSE is defined as follows,

$$\text{MSE}_d(\hat{\Theta}_n) := \mathbb{E}[d(\hat{\Theta}_n, \Theta)^2].$$

MSE Decomposition
Given a sample estimator $\hat{\Theta}_n$ of $\Theta$, we have

$$\text{MSE}_d(\hat{\Theta}_n) \leq 2 \text{Var}_d(\hat{\Theta}_n) + 2b_d^2(\hat{\Theta}_n),$$

where $\text{Var}_d(\cdot)$ and $b_d^2(\cdot)$ are the variance and bias induced by the metric $d$. 
Part III
Statistical Inference on Self-organizing Maps (SOMs) in Neuroimaging
Self-organizing Maps

SOMs

- Kohonen’s (2001) maps.
- Unsupervised artificial neural network.
- Produce a (typically) planar layer of neurons.
- Projection of the inputs into a two-dimensional grid.

Neuroimaging Data

- Projection of individual spatio-temporal patterns.
- Families of subject-specific SOMs.
- Drawing inference on group differences.
- What are the group means?
Self-organizing Maps

SOMs

- Kohonen’s (2001) maps.
- Unsupervised artificial neural network.
- Produce a (typically) planar layer of neurons.
- Projection of the inputs into a two-dimensional grid.

Neuroimaging Data

- Projection of individual spatio-temporal patterns.
- Families of subject-specific SOMs.
- Drawing inference on group differences.
- What are the group means?
Inference on SOMs

Distance Function on Spaces of SOMs

- Sum of Minimum Distances, SMD($M_x, M_y$), is

$$\frac{1}{2V} \left( \sum_{w_x \in M_x} \min_{w_y \in M_y} d_e(w_x, w_y) + \sum_{w_y \in M_y} \min_{w_x \in M_x} d_e(w_y, w_x) \right).$$

- SMD is not a metric: Use the restricted Frechet mean.
- Variants of SMD can be constructed.
Inference on SOMs

Distance Function on Spaces of SOMs

- Sum of Minimum Distances, $\text{SMD}(M_x, M_y)$, is
  $$\frac{1}{2V} \left( \sum_{w_x \in M_x} \min_{w_y \in M_y} d_e(w_x, w_y) + \sum_{w_y \in M_y} \min_{w_x \in M_x} d_e(w_y, w_x) \right).$$

- $\text{SMD}$ is not a metric: Use the restricted Frechet mean.
- Variants of $\text{SMD}$ can be constructed.

Generalized $t$-test

- $H_0 : d(\mu_1, \mu_2) = \delta_0$, we may use the following Frechet $t$-statistic,
  $$t_F = \frac{d(\overline{M}_1, \overline{M}_2) - \delta_0}{S_p \left(1/n_1 + 1/n_2\right)^{1/2}}.$$

- The denominator, $S_p$, is the classical pooled sample variance.
SC1, SC2 and SC3 correspond to three different scenarios.

Spatio-temporal (SC1), temporal (SC2), and spatial (SC3) differences.

Different SMD functions captures different aspects of the SOMs.
Conclusion and Extensions

Consistency of Frechet Sample Mean

- Under separable $\mathcal{X}$ and bounded $d$.
- For non-unique means.
- For every $r \geq 1$.

Possible Extensions

- Distance functions without the triangle inequality.
- Non-iid random variables.
- Rate of convergence.
- Statistical inference.
Acknowledgements & Funding Agencies

Paper


Acknowledgements

- Eric D. Kolaczyk (Boston University)
- Wilfrid S. Kendall (Warwick University)
- Arnaud Fournel (Université de Lyon)
- Dino Sejdinovic and Bharath Sriperumbudur (Gatsby)

Funding Agencies

[Logo of National Institute for Health Research]
[Logo of International Neuroinformatics Coordinating Facility]