Inference for Fréchet Means:

Empirical Likelihood Methods versus CLT

Huiling Le
University of Nottingham

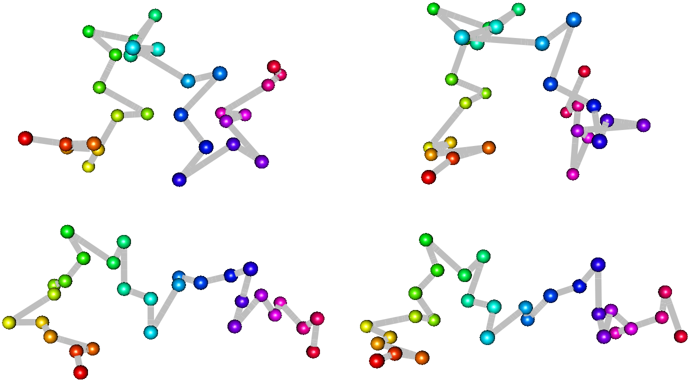
This work is variously joint with Thomas Hotz, Andy Wood & Yan Xi

Content

1. Introduction/motivation: two different types of non-Euclidean data
   - Data on manifolds
   - Data on stratified spaces
2. Fréchet means
   - CLT on manifolds
   - CLT on stratified spaces
3. Empirical likelihood (EL) approach
   - EL method on Euclidean spaces
   - EL for Fréchet means on manifolds
   - EL for Fréchet means on stratified spaces

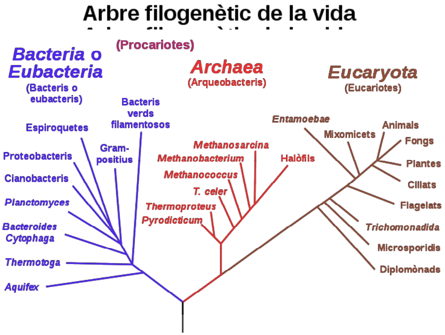
Introduction: data on manifolds

- Directional data (on sphere).
- Shape analysis of configurations with a finite number of landmarks



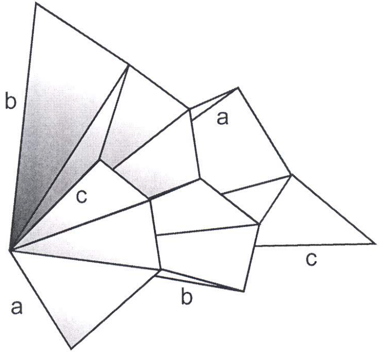
Introduction: data on stratified spaces

Phylogenetic trees:



Introduction: data on stratified spaces

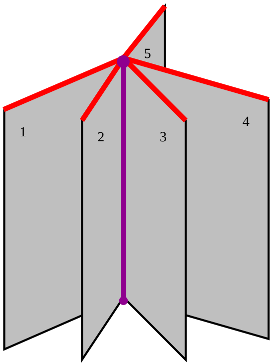
BHV space of phylogenetic trees (Billera *et al.* (2001)):



$\boldsymbol{T}_4$: *space of phylogenetic trees with four leaves*

Introduction: data on stratified spaces

A much simpler stratified space: open book



Fréchet means

Definition (Fréchet (1948))

A point x_0 in a metric space $(\boldsymbol{M}, \rho)$ is called a Fréchet mean of a (finite) measure μ on $\boldsymbol{M}$ if the Fréchet function

$$F_\mu(x) = \frac{1}{2} \int_{\boldsymbol{M}} \rho(x, y)^2 d\mu(y)$$

achieves its *global* minimum at x_0.

We shall always assume that F_μ is finite at least at one point to ensure the existence of Fréchet means and that μ has a unique Fréchet mean.

Fréchet means

- Fréchet means generalize Euclidean means.
- SLLN holds for Fréchet means (Ziezold (1977)).

Fréchet means

- Fréchet means generalize Euclidean means.
- SLLN holds for Fréchet means (Ziezold (1977)).

- There is generally no closed form for Fréchet means: they are implicitly defined.
- In practice, they are usually estimated by iterative algorithms.
- The iterative algorithms can be very slow, e.g. in the case of phylogenetic trees.

Fréchet means

Given $X_1, \cdots, X_n$: *iid* random variables on $\boldsymbol{M}$ with sample Fréchet mean $\hat{X}_n$ and (common) distribution μ, which has Fréchet mean x_0.

Inference on means is often linked with CLT:

In the case $\boldsymbol{M} = \mathbb{R}^m$, if

$$\sqrt{n}\{\hat{X}_n - x_0\} \xrightarrow{\text{d}} N(0, \Gamma),$$

then $\Gamma = \text{cov}(X_1)$ and

$$n\|\Gamma^{-1/2}(\hat{X}_n - x_0)\|^2 \xrightarrow{\text{d}} \chi^2(m).$$

Fréchet means: CLT on manifolds

The case when $\boldsymbol{M}$ is a Riemannian manifold:

Theorem

Under certain technical conditions,

$$\sqrt{n}\exp_{x_0}^{-1}(\hat{X}_n) \xrightarrow{d} N(0, E[H_{x_0,X_1}]^{-1}\Gamma E[H_{x_0,X_1}]^{-\top})$$

as $n \to \infty$, where

$$\Gamma = \mathrm{cov}(\exp_{x_0}^{-1}(X_1)),$$

and $H_{x,y}$ is a map on the tangent space at x, defined by

$$H_{x,y} : v \mapsto -D_v \exp_x^{-1}(y).$$

(Cf. Bhattacharya & Patrangenaru (2005); Bhattacharya & Bhattacharya (2008); Kendall & Le (2011).)

Fréchet means: CLT on manifolds

- As $n \to \infty$,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \exp_{x_0}^{-1}(X_i) \xrightarrow{\text{d}} N(0, \Gamma).$$

- For fixed y, $H_{x,y}$ is closely linked to the Hessian of $\rho(x, y)^2$.

Fréchet means: CLT on manifolds

Difficulties in using the CLT for inference of Fréchet means:

- needs the estimation of $E[H_{x_0,X_1}]$;
- involves the 'technical' condition:

$$\lim_{r\to 0} \mathrm{E}\left[\sup_{x\in\mathrm{B}(x_0,r)} \|H_{x,X_1} - H_{x_0,X_1}\|\right] = 0,$$

arising due to the 'delta' method used in the proof.

Fréchet means: CLT on manifolds

Difficulties in using the CLT for inference of Fréchet means:

- needs the estimation of $E[H_{x_0,X_1}]$;
- involves the 'technical' condition:

$$\lim_{r \to 0} \mathrm{E}\left[\sup_{x \in \mathrm{B}(x_0,r)} \|H_{x,X_1} - H_{x_0,X_1}\|\right] = 0,$$

  arising due to the 'delta' method used in the proof.

If the support of the distribution of X_1 is *not* disjoint from $\mathcal{C}(x_0)$ (the cut locus of x_0), the 'technical' condition does not generally hold.

The von Mises distribution on the circle is such an example.

Fréchet means: CLT on manifolds

BUT, this 'technical' condition can be removed at a price:

Theorem

If $\boldsymbol{M}$ is a compact manifold and if the distribution of X_1 has a continuous density f_{X_1} (w.r.t. to volume measure) in a neighbourhood of $\mathcal{C}(x_0)$, then

$$\sqrt{n}\exp_{x_0}^{-1}(\hat{X}_n) \xrightarrow{d} N(0, \mathrm{E}\left[\Phi_{x_0,X_1}\right]^{-1} \Gamma\, \mathrm{E}\left[\Phi_{x_0,X_1}\right]^{-\top})$$

as $n \to \infty$, where

$$\Phi_{x_0,X_1} = \mathrm{E}\left[H_{x_0,X_1}\right] + J_{x_0,X_1}.$$

J is an integral of the distribution over a 'nice' *co-dimension one* subset of $\mathcal{C}(x_0)$, where the integrand depends on the nature of the relationship between x_0 and $\mathcal{C}(x_0)$.

Fréchet means: CLT on manifolds

Three examples:

1. $\boldsymbol{M} = S^1$ (Hotz & Huckemann (2015)):

$$J_{0,X_1} = -2\pi\, f_{X_1}(\pi).$$

2. $\boldsymbol{M} = S^1 \times S^1$ (the flat torus):

$$J_{(0,0),X_1} = -2\pi \begin{pmatrix} \int_{-\pi}^{\pi} f_{X_1}(y_1, \pi)\,\mathrm{d}\,y_1 & 0 \\ 0 & \int_{-\pi}^{\pi} f_{X_1}(\pi, y_2)\,\mathrm{d}\,y_2 \end{pmatrix}.$$

Fréchet means: CLT on manifolds

3. $\boldsymbol{M} = RP^2$:

$$J_{(0,0),X_1} = -\pi \int_0^{\pi} \begin{pmatrix} \cos^2\theta & \sin\theta\cos\theta \\ \sin\theta\cos\theta & \sin^2\theta \end{pmatrix} f_{X_1}(\pi/2, \theta)\,\mathrm{d}\,\theta.$$

- $J_{x_0,X_1} = 0$ on S^m if $m > 1$.
- $J_{x_0,X_1} = 0$ if $f_{X_1} = 0$ on $\mathcal{C}(x_0)$.

Fréchet means: CLT on manifolds

In summary:

using the CLT for inference of Fréchet means
requires a lot of work
estimating the relevant covariance matrix.

Fréchet means: CLT on stratified spaces

Assume that $\boldsymbol{M}$ is an open book.

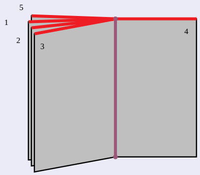
Case 1: the Fréchet mean x_0 lies within a page (say page 1).

Then, the CLT resembles that on manifolds:

Theorem (Hotz et.al. (2013))

$$\sqrt{n}\{\hat{X}_n - x_0\} \xrightarrow{d} N(0, \Gamma)$$

as $n \to \infty$*, where* $\Gamma = \mathrm{cov}(F_1(X_i))$ *and* F_k *is the so-called folding map:*



Fréchet means: CLT on stratified spaces

- For fixed k, $F_k(X_i)$ are Euclidean r.v.'s.
- For general stratified spaces,
  - F_k are replaced by the *log map*;
  - for the covariance matrix in the CLT, Γ needs to be multiplied on both sides by another component, arising from the derivative of the log map.

Fréchet means: CLT on stratified spaces

Case 2: the Fréchet mean x_0 lies on the spine.

Then, it is either *sticky* i.e.

$$\hat{X}_n \text{ lies on the spine for all sufficiently large } n,$$

or *partly sticky* i.e.

$$\hat{X}_n \text{ lies either on the spine or within a particular page}$$
$$\text{for all sufficiently large } n$$

(cf. Hotz et.al. (2013)).

Fréchet means: CLT on stratified spaces

- For the sticky case, the limiting distribution of $\sqrt{n}\{\hat{X}_n - x_0\}$ is Gaussian with support on the spine.
- For the partly sticky case,
  - $$P(\hat{X}_n \text{ lies on the spine}) \longrightarrow \frac{1}{2};$$
  - the limiting distribution of $\sqrt{n}\{\hat{X}_n - x_0\}$ is no longer Gaussian, but still related to a Gaussian distribution.

Fréchet means: CLT on stratified spaces

In summary:

in addition to the issues highlighted for manifolds,
Fréchet means behave in a non-classical way
if they are on the spine (co-dimension one subspace).

EL method on Euclidean space

Empirical likelihood (cf. Owen (2001))

- is a non-parametric method of statistical inference;
- combines non-parametric methods with likelihood approach;
- uses mainly optimization algorithms for computation.

EL method on Euclidean space

Empirical likelihood (cf. Owen (2001))

- is a non-parametric method of statistical inference;
- combines non-parametric methods with likelihood approach;
- uses mainly optimization algorithms for computation.

For a given set of *i.i.d* data $X_1, \cdots, X_n$ in $\mathbb{R}^m$, the non-parametric likelihood $L(\mu)$ for a distribution μ on $\mathbb{R}^m$ is defined as

$$L(\mu) = \prod_{i=1}^{n} \mu(\{X_i\}).$$

Then, use $L(\mu)$ to define

$$R_n(\mu) = \frac{L(\mu)}{L(\mu_n)},$$

where μ_n denotes the empirical distribution of $X_1, \cdots, X_n$.

EL method on Euclidean space

Finally, for a given family of distributions $\mathcal{P}$, the empirical likelihood for means is defined as

$$\mathcal{R}_n(x) = \sup\{R_n(\mu) \mid \text{ mean of } \mu = x,\ \mu \in \mathcal{P}\}.$$

If we choose $\mathcal{P}$ to consist of all distributions taking values in $\{X_1, \cdots, X_n\}$, then $\log(\mathcal{R}_n(x))$ for $x \in \mathbb{R}^m$ is the solution of the following optimization problem:

$$\max \sum_{i=1}^{n} \log p_i \qquad \text{subject to } \begin{cases} \sum_{i=1}^{n} p_i = 1,\ p_i \geqslant 0; \\ \sum_{i=1}^{n} p_i X_i = x. \end{cases}$$

EL method on Euclidean space

Empirical likelihood hypothesis tests reject H_0 : mean of $\mu_0 = x_0$, when $\mathcal{R}_n(x_0) < r_0$ for some threshold value r_0.

The threshold r_0 may be chosen using an empirical likelihood theorem, a non-parametric analogue of Wilks' theorem (cf. Owen (2001)):

Let $X_1, \cdots, X_n$ be i.i.d. random variables on $\mathbb{R}^m$. Let $x_0 = \mathrm{E}[X_1]$, and suppose that $\mathrm{cov}(X_1)$ is positive definite. Then

$$-2\log(\mathcal{R}_n(x_0)) \xrightarrow{\mathrm{d}} \chi^2(m)$$

as $n \to \infty$.

EL for Fréchet means on manifolds

When $\boldsymbol{M}$ is a Riemannian manifold:

The Fréchet mean x_0 of a distribution μ on $\boldsymbol{M}$ must satisfy

$$\int_{\boldsymbol{M}} \exp_{x_0}^{-1}(y)\, d\mu(y) = 0.$$

- This gives us a link between Fréchet means of X and Euclidean means of $\exp_{x_0}^{-1}(X)$:

  if X on $\boldsymbol{M}$ has Fréchet mean x_0, then the Euclidean random variable $\exp_{x_0}^{-1}(X)$ has its Euclidean mean at the origin of the tangent space at x_0.

- If $\boldsymbol{M}$ is complete, simply-connected and of non-positive curvature, the above is also sufficient for x_0 to be the Fréchet mean of μ.

EL for Fréchet means on manifolds

The above link with Euclidean r.v.'s leads us to consider the EL for the critical points of Fréchet functions:

$$\mathcal{R}_n(x) = \sup\left\{ R_n(\mu) \mid \int_{\boldsymbol{M}} \exp_{x_0}^{-1}(y)\, d\mu(y) = 0,\ \mu \in \mathcal{P} \right\}.$$

Treat this as the EL for Fréchet means.

Then, on $\boldsymbol{M} \setminus \bigcup_{i=1}^{n} \mathcal{C}(X_i)$, it corresponds to the solution of the optimization problem:

$$\log(\mathcal{R}_n(x)) = \max \sum_{i=1}^{n} \log p_i,$$

$$\text{subject to } \begin{cases} \sum_{i=1}^{n} p_i = 1,\ p_i \geqslant 0; \\ \sum_{i=1}^{n} p_i \exp_x^{-1}(X_i) = 0. \end{cases}$$

EL for Fréchet means on manifolds

Theorem

Assume that $X_1, \cdots, X_n$ are i.i.d. random variables on $\boldsymbol{M}$ with Fréchet mean x_0. Then,

$$-2\log(\mathcal{R}_n(x_0)) \xrightarrow{d} \chi^2(m)$$

as $n \to \infty$.

Since this only involves Euclidean random variables, we have avoided the need for estimation of the extra terms in the covariance in CLT.

Care is required:

- $\log(\mathcal{R}_n)$ is not necessarily a concave function on $\boldsymbol{M}$.

EL for Fréchet means on stratified spaces

For simplicity, we concentrate on open-books.

Without loss of generality, we assume that the support of μ has non-empty intersection with all pages.

EL for Fréchet means on stratified spaces

For simplicity, we concentrate on open-books.

Without loss of generality, we assume that the support of μ has non-empty intersection with all pages.

Case 1: the Fréchet mean x_0 of μ lies within a page. Then

$$\begin{aligned} & x_0 \text{ is the Fréchet mean of } \mu \\ \Leftrightarrow \quad & x_0 \text{ is the critical point of } F_\mu. \end{aligned}$$

EL for Fréchet means on stratified spaces

For simplicity, we concentrate on open-books.

Without loss of generality, we assume that the support of μ has non-empty intersection with all pages.

Case 1: the Fréchet mean x_0 of μ lies within a page. Then

$$\begin{aligned} & x_0 \text{ is the Fréchet mean of } \mu \\ \Leftrightarrow \quad & x_0 \text{ is the critical point of } F_\mu. \end{aligned}$$

This leads us to define, for x lying within a page, the EL for Fréchet means to be

$$\mathcal{R}_n(x) = \sup \left\{ R_n(\mu) \mid x \text{ is the critical point of } F_\mu,\ \mu \in \mathcal{P} \right\}.$$

EL for Fréchet means on stratified spaces

Case 1: the Fréchet mean x_0 of μ lies within a page, say page 1.

Then, x_0 is characterized by

$$\int_{\boldsymbol{M}} F_1(y)\, d\mu(y) = x_0.$$

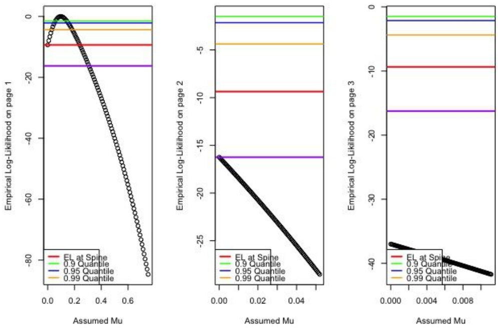
Thus, the EL for Fréchet means on page k satisfies

$$\log(\mathcal{R}_n(x)) = \max \sum_{i=1}^{n} \log p_i$$

$$\text{subject to} \begin{cases} \sum_{i=1}^{n} p_i = 1,\ p_i \geqslant 0; \\ \sum_{i=1}^{n} p_i F_k(X_i) = x. \end{cases}$$

EL for Fréchet means on stratified spaces

$\log(\mathcal{R}_n)$ is a concave function on each page, but is generally discontinuous when crossing the spine:



EL for Fréchet means on stratified spaces

Theorem

Assume that $X_1, \cdots, X_n$ are i.i.d with distribution μ and that the Fréchet mean x_0 of μ is not on the spine. Then

$$-2\log(\mathcal{R}_n(x_0)) \xrightarrow{d} \chi^2(m)$$

as $n \to \infty$.

EL for Fréchet means on stratified spaces

Case 2: the Fréchet mean x_0 is on the spine.

Then, x_0 is characterized by

$$\int_M P_S(x)\, d\mu(x) = P_S(x_0)(= x_0) \quad \text{and}$$
$$\int_M \langle F_j(x), e_j \rangle\, d\mu(x) \leqslant 0, \quad j = 1, \cdots, \ell,$$

where P_S denotes the projection to the spine and e_j denotes the 'outward' unit vector that is tangent to page j and is orthogonal to the spine.

Under our assumption, the inequalities above can include at most one equality.

EL for Fréchet means on stratified spaces

This characterization implies that the logarithm of the corresponding EL optimization problem on the spine becomes more complicated:

$$\log(\mathcal{R}_n(x)) = \max \sum_{i=1}^{n} \log p_i,$$

$$\text{subject to} \begin{cases} \sum_{i=1}^{n} p_i = 1, \;\; p_i \geqslant 0; \\ \sum_{i=1}^{n} p_i P_S(X_i) = x; \\ \sum_{i=1}^{n} p_i \langle F_k(X_i), e_k \rangle \leqslant 0, \quad k = 1, \cdots, \ell. \end{cases}$$

EL for Fréchet means on stratified spaces

However, this can be simplified using knowledge of the sample Euclidean means of $F_k(X_i)$.

It is known that

either

(i) there is (only one) k say $k = 1$, such that the sample mean of $F_1(X_1), \cdots, F_1(X_n)$ is on page 1;

or

(ii) otherwise.

EL for Fréchet means on stratified spaces

Proposition

For case (i) *with* $k = 1$*, the constraint optimization problem associated with the EL for Fréchet means on the spine is equivalent to the following*

$$\max \sum_{i=1}^{n} \log p_i \qquad \textit{subject to} \begin{cases} \sum_{i=1}^{n} p_i = 1, \ p_i \geqslant 0; \\ \sum_{i=1}^{n} p_i F_1(X_i) = x. \end{cases}$$

EL for Fréchet means on stratified spaces

Proposition

If x is on the spine and is sufficiently close to $\hat{X}_n$ then, for case (ii), the constraint optimization problem associated with the EL for Fréchet means is equivalent to the following

$$\max \sum_{i=1}^{n} \log p_i \qquad \textit{subject to} \begin{cases} \sum_{i=1}^{n} p_i = 1, \ p_i \geqslant 0; \\ \sum_{i=1}^{n} p_i P_S(X_i) = x. \end{cases}$$

EL for Fréchet means on stratified spaces

Theorem

Assume that the Fréchet mean x_0 of μ is on the spine. Then, as $n \to \infty$,

$$-2\log(\mathcal{R}_n(x_0)) \xrightarrow{d} \begin{cases} \frac{1}{2}\{\chi^2(m) + \chi^2(m-1)\} & \text{if } x_0 \text{ is partly sticky} \\ \chi^2(m-1) & \text{if } x_0 \text{ is sticky.} \end{cases}$$

Comparison of the two methods

- Using the CLT for inference of Fréchet means requires a lot of work estimating the relevant covariance matrix;
- in addition, on stratified spaces, Fréchet means behave in a non-classical way if they are on co-dimension one subspaces.

Comparison of the two methods

- Using the CLT for inference of Fréchet means requires a lot of work estimating the relevant covariance matrix;
- in addition, on stratified spaces, Fréchet means behave in a non-classical way if they are on co-dimension one subspaces.

While the proposed EL approach for Fréchet means

- involves (standard, Euclidean) optimization procedures;
- requires only finding the inverse exponential (or logarithmic) images of data, avoiding the need for estimation of covariances in the CLT.

Comparison of the two methods

- Using the CLT for inference of Fréchet means requires a lot of work estimating the relevant covariance matrix;
- in addition, on stratified spaces, Fréchet means behave in a non-classical way if they are on co-dimension one subspaces.

While the proposed EL approach for Fréchet means

- involves (standard, Euclidean) optimization procedures;
- requires only finding the inverse exponential (or logarithmic) images of data, avoiding the need for estimation of covariances in the CLT.

Moreover, the EL approach can also be considered for estimation. Then, it transforms iterative algorithms into maximization problems for the relevant Euclidean random variables.