On Some Statistical Challenges Coming with Non-Euclidean Data

Stephan F. Huckemann and Benjamin Eltzner

University of Göttingen,
Felix Bernstein Institute for Mathematical Statistics in the Biosciences

Feb. 20, 2018

TAGS - Linking Topology to Algebraic Geometry and Statistics (Feb. 19 – 23, 2018) Max-Planck-Institut Leipzig

supported by the

Niedersachsen Vorab of the Volkswagen Foundation,

and the DFG SFB 755 + HU 1575/4

What is this about?

I am interested in non-Euclidean data:

- e.g. 3D structure of RNA molecules $\rightarrow$ shape spaces,
- phylogenetic descendence trees $\rightarrow$ spaces of trees,
- data on trees, graphs,
- (toy) example: data on spheres

What is this about?

I am interested in non-Euclidean data:

- e.g. 3D structure of RNA molecules $\to$ shape spaces,
- phylogenetic descendence trees $\to$ spaces of trees,
- data on trees, graphs,
- (toy) example: data on spheres

What do statisticians do with data?

- find simple descriptors,
- compare datasets via descriptors,
- inference: with confidence test for equality of data.

What is this about?

I am interested in non-Euclidean data:

- e.g. 3D structure of RNA molecules $\to$ shape spaces,
- phylogenetic descendence trees $\to$ spaces of trees,
- data on trees, graphs,
- (toy) example: data on spheres

What do statisticians do with data?

- find simple descriptors,
- compare datasets via descriptors,
- inference: with confidence test for equality of data.

How?

- with exact distributions, or
- with asymptotic central limit theorems (CLTs).

Euclidean SLLN and CLT

Let $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} X \in \mathbb{R}^m$, $m \in \mathbb{N}$, and $\bar{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j$

Theorem (Strong Law of Large Numbers)

$E[\|X\|] < \infty \Rightarrow \bar{X}_n \overset{a.s.}{\to} \mathbb{E}[X]$

Euclidean SLLN and CLT

Let $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} X \in \mathbb{R}^m$, $m \in \mathbb{N}$, and $\bar{X}_n = \frac{1}{n}\sum_{j=1}^n X_j$

Theorem (Strong Law of Large Numbers)

$E[\|X\|] < \infty \Rightarrow \bar{X}_n \overset{a.s.}{\to} \mathbb{E}[X]$

Theorem (Central Limit Theorem)

$E[\|X\|^2] < \infty \Rightarrow \sqrt{n}(\bar{X}_n - \mathbb{E}[X]) \overset{\mathcal{D}}{\to} \mathcal{N}(0, \text{cov}[X])$

Euclidean SLLN and CLT

Let $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} X \in \mathbb{R}^m$, $m \in \mathbb{N}$, and $\bar{X}_n = \frac{1}{n}\sum_{j=1}^n X_j$

Theorem (Strong Law of Large Numbers)

$E[\|X\|] < \infty \Rightarrow \bar{X}_n \overset{a.s.}{\to} \mathbb{E}[X]$

Theorem (Central Limit Theorem)

$E[\|X\|^2] < \infty \Rightarrow \sqrt{n}(\bar{X}_n - \mathbb{E}[X]) \overset{\mathcal{D}}{\to} \mathcal{N}(0, \text{cov}[X])$

Let $\text{cov}[X_1, \ldots, X_n] = \frac{1}{n}\sum_{j=1}^n (X_j - \bar{X}_n)(X_j - \bar{X}_n)^T$. Then

Theorem (One-Sample Test)

$E[\|X\|^2] < \infty \Rightarrow$

$n\,\frac{n-m}{m}(\bar{X}_n - \mathbb{E}[X])^T \text{cov}[X_1, \ldots, X_n]^{-1}(\bar{X}_n - \mathbb{E}[X]) \overset{\mathcal{D}}{\to} \mathcal{F}_{m,n-m}$

Euclidean SLLN and CLT

Let $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} X \in \mathbb{R}^m$, $m \in \mathbb{N}$, and $\bar{X}_n = \frac{1}{n}\sum_{j=1}^n X_j$

Theorem (Strong Law of Large Numbers)

$E[\|X\|] < \infty \Rightarrow \bar{X}_n \overset{a.s.}{\to} \mathbb{E}[X]$

Theorem (Central Limit Theorem)

$E[\|X\|^2] < \infty \Rightarrow \sqrt{n}(\bar{X}_n - \mathbb{E}[X]) \overset{\mathcal{D}}{\to} \mathcal{N}(0, \text{cov}[X])$

Let $\text{cov}[X_1, \ldots, X_n] = \frac{1}{n}\sum_{j=1}^n (X_j - \bar{X}_n)(X_j - \bar{X}_n)^T$. Then

Theorem (One-Sample Test)

$E[\|X\|^2] < \infty \Rightarrow$

$n\,\frac{n-m}{m}(\bar{X}_n - \mathbb{E}[X])^T \text{cov}[X_1, \ldots, X_n]^{-1}(\bar{X}_n - \mathbb{E}[X]) \overset{\mathcal{D}}{\to} \mathcal{F}_{m,n-m}$

Here $Y_n \overset{\mathcal{D}}{\to} Z_n \Leftrightarrow \mathbb{E}[f(Y_n)] - \mathbb{E}[f(Z_n)] \to 0\ \forall$ testfunctions f

Euclidean SLLN and CLT

Let $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} X \in \mathbb{R}^m$, $m \in \mathbb{N}$, and $\bar{X}_n = \frac{1}{n}\sum_{j=1}^n X_j$

Theorem (Strong Law of Large Numbers)

$E[\|X\|] < \infty \Rightarrow \bar{X}_n \overset{a.s.}{\to} \mathbb{E}[X]$

Theorem (Central Limit Theorem)

$E[\|X\|^2] < \infty \Rightarrow \sqrt{n}(\bar{X}_n - \mathbb{E}[X]) \overset{\mathcal{D}}{\to} \mathcal{N}(0, \mathrm{cov}[X])$

Let $\mathrm{cov}[X_1, \ldots, X_n] = \frac{1}{n}\sum_{j=1}^n (X_j - \bar{X}_n)(X_j - \bar{X}_n)^T$. Then

Theorem (One-Sample Test)

$E[\|X\|^2] < \infty \Rightarrow$

$n\,\frac{n-m}{m}(\bar{X}_n - \mathbb{E}[X])^T \mathrm{cov}[X_1, \ldots, X_n]^{-1}(\bar{X}_n - \mathbb{E}[X]) \overset{\mathcal{D}}{\to} \mathcal{F}_{m,n-m}$

Here $Y_n \overset{\mathcal{D}}{\to} Z_n \Leftrightarrow \mathbb{E}[f(Y_n)] - \mathbb{E}[f(Z_n)] \to 0 \ \forall$ testfunctions f

$\exists$ more involved Two-Sample Tests.

Principal Component Analysis (PCA)

Spectral decomposition giving main modes of variation

- $\mathrm{cov}[X] = \Gamma\Lambda\Gamma^T$, $\mathrm{cov}[X_1, \ldots, X_n] = \Gamma(n)\Lambda(n)\Gamma(n)^T$.
- With eigenvectors $\Gamma = (\gamma_1, \ldots, \gamma_m)$, $\Gamma(n) = (\gamma_1(n), \ldots, \gamma_m(n))$ to
- eigenvalues $\lambda_1 \geq \ldots \geq \lambda_m \geq 0$, $\lambda_1(n) \geq \ldots \geq \lambda_m(n) \geq 0$, resp.,

Theorem (Asymptotic PCA, Anderson (1963); Watson (1983))

$E[\|X\|^4] < \infty, \lambda_k$ *simple,* $\langle \gamma_k, \gamma_k(n)\rangle \geq 0$

$$\Rightarrow \sqrt{n}(\gamma_k(n) - \gamma_k) \xrightarrow{\mathbb{P}} \mathcal{N}\left(0, \sum_{k\neq j=1}^m \frac{\gamma_j\gamma_j^T \,\mathrm{cov}[XX']\gamma_k}{\lambda_k-\lambda_j}\right)$$

Principal Component Analysis (PCA)

Spectral decomposition giving main modes of variation

- $\text{cov}[X] = \Gamma\Lambda\Gamma^T$, $\text{cov}[X_1, \ldots, X_n] = \Gamma(n)\Lambda(n)\Gamma(n)^T$.
- With eigenvectors
  $\Gamma = (\gamma_1, \ldots, \gamma_m)$, $\Gamma(n) = (\gamma_1(n), \ldots, \gamma_m(n))$ to
- eigenvalues
  $\lambda_1 \geq \ldots \geq \lambda_m \geq 0$, $\lambda_1(n) \geq \ldots \geq \lambda_m(n) \geq 0$, resp.,

Theorem (Asymptotic PCA, Anderson (1963); Watson (1983))

$E[\|X\|^4] < \infty, \lambda_k$ *simple*, $\langle \gamma_k, \gamma_k(n) \rangle \geq 0$

$$\Rightarrow \sqrt{n}(\gamma_k(n) - \gamma_k) \xrightarrow{\mathbb{P}} \mathcal{N}\left(0, \sum_{k \neq j=1}^m \frac{\gamma_j \gamma_j^T \,\text{cov}[XX']\gamma_k}{\lambda_k - \lambda_j}\right)$$

Note, $\gamma_k \in \mathbb{S}^{m-1}$. Actually in $\mathbb{RP}^{m-1}$.
And, limiting distribution in $T_{\gamma_k}\mathbb{S}^{m-1} \cong T_{\pm\gamma_k}\mathbb{RP}^{m-1}$.

Outline

Non-Euclidean Descriptors

- Fréchet means
  - intrinsic (Kobayashi and Nomizu (1969); Bhattacharya and Patrangenaru (2003))
  - extrinsic (Hendriks and Landsman (1996); Bhattacharya and Patrangenaru (2003))
  - residual (Jupp (1988))
  - Procrustes (Gower (1975))
  - Ziezold (Ziezold (1994))
  - ⋮
- principal geodesics (Fletcher and Joshi (2004); H. et al 2010)
- principal submanifolds
  - (almost) totally geodesic (Jung et al. (2012): PN(G)S)
  - horizontal subspaces (Sommer (2016))
  - geodesic flows (Panaretos et al. (2014))
  - barycentric subspaces (Pennec (2017); Nye et al. (2016))
- flags of principal submanifolds (Pennec (2017))
- ...

The CLT for Intrinsic Means on Manifolds

- M: m-dimensional Riemannian C^2 manifold
- d: intrinsic geodesic distance
- $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} X \in M$: random variables
- Fréchet (population) mean set:
  $E[X] = \text{argmin}_{\mu \in M} \mathbb{E}[d(\mu, X)^2]$
- Fréchet (sample) mean set:
  $E_n[X_1, \dots, X_n] = \text{argmin}_{\mu \in M} \sum_{j=1}^n d(\mu, X_j)^2$
- $E[X] = \{\mu\}, \mu_n \in E_n[X_1, \dots, X_n]$ measurable
- $\phi : M \to \mathbb{R}^m$ local C^2 chart near μ

Theorem (Bhattacharya and Patrangenaru (2005))

Under some additional regularity conditions

$$\sqrt{n}(\phi(\mu_n) - \phi(\mu)) \xrightarrow{\mathbb{P}} \mathcal{N}(0, \Sigma)$$

with suitable $\Sigma \geq 0$.

Idea of Proof

- W.l.o.g $\phi(\mu) = 0$, $\phi(\mu_n) = x_n$.

Idea of Proof

- W.l.o.g $\phi(\mu) = 0$, $\phi(\mu_n) = x_n$.
- SLLN by Ziezold (1977); Bhattacharya and Patrangenaru (2003): $x_n \stackrel{a.s.}{\to} 0$.

Idea of Proof

- W.l.o.g $\phi(\mu) = 0$, $\phi(\mu_n) = x_n$.
- SLLN by Ziezold (1977); Bhattacharya and Patrangenaru (2003): $x_n \stackrel{\text{a.s.}}{\to} 0$.
- Fréchet functions:

$$F_n(x) = \frac{1}{2n} \sum_{j=1}^{n} d(X_j, \phi^{-1}(x))^2, \; F(x) = \frac{1}{2}\, \mathbb{E}[d(X, \phi^{-1}(x))^2]$$

Idea of Proof

- W.l.o.g $\phi(\mu) = 0$, $\phi(\mu_n) = x_n$.
- SLLN by Ziezold (1977); Bhattacharya and Patrangenaru (2003): $x_n \stackrel{a.s.}{\to} 0$.
- Fréchet functions:

$$F_n(x) = \frac{1}{2n}\sum_{j=1}^{n} d(X_j, \phi^{-1}(x))^2,\; F(x) = \frac{1}{2}\,\mathbb{E}[d(X, \phi^{-1}(x))^2]$$

- Taylor expansion (with suitable $\widetilde{x}$ between 0 and x_0):

$$\sqrt{n}\,\mathrm{grad}|_{x=x_0} F_n(x) \quad = \quad \sqrt{n}\,\mathrm{grad}|_{x=0} F_n(x) + \mathrm{Hess}|_{x=\widetilde{x}} F_n(x)\sqrt{n}x_0\,,$$

Idea of Proof

- W.l.o.g $\phi(\mu) = 0$, $\phi(\mu_n) = x_n$.
- SLLN by Ziezold (1977); Bhattacharya and Patrangenaru (2003): $x_n \overset{a.s.}{\to} 0$.
- Fréchet functions:

$$F_n(x) = \frac{1}{2n}\sum_{j=1}^{n} d(X_j, \phi^{-1}(x))^2, \; F(x) = \frac{1}{2}\,\mathbb{E}[d(X, \phi^{-1}(x))^2]$$

- Taylor expansion (with suitable $\widetilde{x}$ between 0 and x_0):

$$\sqrt{n}\,\mathrm{grad}|_{x=x_0} F_n(x) \quad = \quad \sqrt{n}\,\mathrm{grad}|_{x=0} F_n(x) + \mathrm{Hess}|_{x=\widetilde{x}} F_n(x)\sqrt{n}x_0\,,$$

If generalized weak law ($n \to \infty$ and $x_0 \to 0$)

$$\mathrm{Hess}|_{x=\widetilde{x}} F_n(x) \overset{\mathbb{P}}{\to} \mathrm{Hess}|_{x=0} F(x)\,,$$

Idea of Proof

- W.l.o.g $\phi(\mu) = 0$, $\phi(\mu_n) = x_n$.
- SLLN by Ziezold (1977); Bhattacharya and Patrangenaru (2003): $x_n \stackrel{a.s.}{\to} 0$.
- Fréchet functions:

$$F_n(x) = \frac{1}{2n}\sum_{j=1}^{n} d(X_j, \phi^{-1}(x))^2, \; F(x) = \frac{1}{2}\,\mathbb{E}[d(X, \phi^{-1}(x))^2]$$

- Taylor expansion (with suitable $\widetilde{x}$ between 0 and x_0):

$$\sqrt{n}\,\mathrm{grad}|_{x=x_0} F_n(x) \quad = \quad \sqrt{n}\,\mathrm{grad}|_{x=0} F_n(x) + \mathrm{Hess}|_{x=\widetilde{x}} F_n(x)\sqrt{n}x_0 \,,$$

If generalized weak law ($n \to \infty$ and $x_0 \to 0$)

$$\mathrm{Hess}|_{x=\widetilde{x}} F_n(x) \stackrel{\mathbb{P}}{\to} \mathrm{Hess}|_{x=0} F(x) \,,$$

holds also for random $x_0 = x_n$, and if $\mathrm{Hess}|_{x=0} F(x) > 0$

Idea of Proof

- W.l.o.g $\phi(\mu) = 0$, $\phi(\mu_n) = x_n$.
- SLLN by Ziezold (1977); Bhattacharya and Patrangenaru (2003): $x_n \stackrel{a.s.}{\to} 0$.
- Fréchet functions:

$$F_n(x) = \frac{1}{2n}\sum_{j=1}^{n} d(X_j, \phi^{-1}(x))^2, \; F(x) = \frac{1}{2}\,\mathbb{E}[d(X, \phi^{-1}(x))^2]$$

- Taylor expansion (with suitable $\widetilde{x}$ between 0 and x_0):

$$\sqrt{n}\,\mathrm{grad}|_{x=x_0} F_n(x) \quad = \quad \sqrt{n}\,\mathrm{grad}|_{x=0} F_n(x) + \mathrm{Hess}|_{x=\widetilde{x}} F_n(x) \sqrt{n} x_0 \,,$$

If generalized weak law ($n \to \infty$ and $x_0 \to 0$)

$$\mathrm{Hess}|_{x=\widetilde{x}} F_n(x) \stackrel{\mathbb{P}}{\to} \mathrm{Hess}|_{x=0} F(x)\,,$$

holds also for random $x_0 = x_n$, and if $\mathrm{Hess}|_{x=0} F(x) > 0$ $\Rightarrow$ BP-CLT.

Make a Mental Note

For the BP-CLT to hold, we need

(i) a C^2 manifold structure with C^2 distance^2 near

(ii) a unique population mean μ,

(iii) for all random $\hat{x}_n \overset{\text{a.s.}}{\to} 0$,

$$\mathrm{Hess}|_{x=\hat{x}_n} F_n(x) \overset{\mathbb{P}}{\to} \mathrm{Hess}|_{x=0} F(x)\,,$$

(iv) $\mathrm{Hess}|_{x=0} F(x) > 0\,.$

Make a Mental Note

For the BP-CLT to hold, we need

(i) a C^2 manifold structure with C^2 distance^2 near

(ii) a unique population mean μ,

(iii) for all random $\hat{x}_n \stackrel{\text{a.s.}}{\to} 0$,

$$\mathrm{Hess}|_{x=\hat{x}_n} F_n(x) \stackrel{\mathbb{P}}{\to} \mathrm{Hess}|_{x=0} F(x)\,,$$

(iv) $\mathrm{Hess}|_{x=0} F(x) > 0\,.$

Now a CLT for geodesics or more general subspaces?

Abstract Setup

- Random elements $X_1, \ldots, X_n \sim X$ ($n \in \mathbb{N}$) on a topological data space Q
- linked via a continous "distance" $\rho : Q \times P \to [0, \infty)$ to a topological descriptor space P, with continous $d : P \times P \to [0, \infty)$ vanishing exactly on diagonal,
- giving in P generalized Fréchet means

$$\begin{array}{lrcl} \text{population:} & E & = & \operatorname{argmin}_{p \in P} \mathbb{E}\big(\rho(X, p)^2\big) \\ \text{sample:} & E_n & = & \operatorname{argmin}_{p \in P} \sum_{j=1}^{n} \rho\big(X_j, p\big)^2 \end{array} .$$

Abstract Setup

- Random elements $X_1, \ldots, X_n \sim X$ ($n \in \mathbb{N}$) on a topological data space Q
- linked via a continous "distance" $\rho : Q \times P \to [0, \infty)$ to a topological descriptor space P, with continous $d : P \times P \to [0, \infty)$ vanishing exactly on diagonal,
- giving in P generalized Fréchet means

$$\begin{array}{lrcl} \text{population:} & E & = & \operatorname{argmin}_{p \in P} \mathbb{E}\big(\rho(X, p)^2\big) \\ \text{sample:} & E_n & = & \operatorname{argmin}_{p \in P} \sum_{j=1}^{n} \rho\big(X_j, p\big)^2 \end{array}.$$

- (ρ, d) is a uniform link if

$$\forall p \in P, \epsilon > 0 \exists \delta = \delta(\epsilon, p) > 0 \text{ such that}$$
$$|\rho(x, p') - \rho(x, p)| < \epsilon \forall x \in Q, p' \in P \text{ with } d(p, p') < \delta.$$

Abstract Setup

- Random elements $X_1, \ldots, X_n \sim X$ ($n \in \mathbb{N}$) on a topological data space Q
- linked via a continous "distance" $\rho : Q \times P \to [0, \infty)$ to a topological descriptor space P, with continous $d : P \times P \to [0, \infty)$ vanishing exactly on diagonal,
- giving in P generalized Fréchet means

$$\begin{array}{lrcl} \text{population:} & E & = & \operatorname{argmin}_{p \in P} \mathbb{E}\big(\rho(X,p)^2\big) \\ \text{sample:} & E_n & = & \operatorname{argmin}_{p \in P} \sum_{j=1}^n \rho\big(X_j,p\big)^2 \end{array}.$$

- (ρ, d) is a uniform link if

$$\forall p \in P, \epsilon > 0 \exists \delta = \delta(\epsilon, p) > 0 \text{ such that}$$
$$|\rho(x,p') - \rho(x,p)| < \epsilon \forall x \in Q, p' \in P \text{ with } d(p,p') < \delta.$$

- Is the case if Q is compact.

Abstract Setup

- Random elements $X_1, \dots, X_n \sim X$ ($n \in \mathbb{N}$) on a topological data space Q
- linked via a continous "distance" $\rho : Q \times P \to [0, \infty)$ to a topological descriptor space P, with continous $d : P \times P \to [0, \infty)$ vanishing exactly on diagonal,
- giving in P generalized Fréchet means

$$\begin{array}{lrcl} \text{population:} & E & = & \operatorname{argmin}_{p \in P} \mathbb{E}\big(\rho(X,p)^2\big) \\ \text{sample:} & E_n & = & \operatorname{argmin}_{p \in P} \sum_{j=1}^n \rho\big(X_j,p\big)^2 \end{array} .$$

- (ρ, d) is a coercive link if $\exists p_0 \in P, C > 0$ such that $\forall p', p'_n, p_n \in P$ with $d(p_0, p_n) \to \infty \leftarrow d(p', p'_n)$

$$\rho(x, p_n) \to \infty \forall x \in Q \text{ with } \rho(x, p_0) < C;$$
$$d(p_0, p'_n) \to \infty.$$

Abstract Setup

- Random elements $X_1, \dots, X_n \sim X$ ($n \in \mathbb{N}$) on a topological data space Q
- linked via a continous "distance" $\rho : Q \times P \to [0, \infty)$ to a topological descriptor space P, with continous $d : P \times P \to [0, \infty)$ vanishing exactly on diagonal,
- giving in P generalized Fréchet means

$$\begin{array}{lrcl} \text{population:} & E & = & \operatorname{argmin}_{p \in P} \mathbb{E}\big(\rho(X, p)^2\big) \\ \text{sample:} & E_n & = & \operatorname{argmin}_{p \in P} \sum_{j=1}^n \rho\big(X_j, p\big)^2 \end{array} .$$

- (ρ, d) is a coercive link if $\exists p_0 \in P, C > 0$ such that $\forall p', p'_n, p_n \in P$ with $d(p_0, p_n) \to \infty \leftarrow d(p', p'_n)$

$$\rho(x, p_n) \to \infty \forall x \in Q \text{ with } \rho(x, p_0) < C;$$
$$d(p_0, p'_n) \to \infty.$$

- Is the case if Q and P are compact.

The Two Strong Laws

Theorem (H. 2011b)

Ziezold Strong Consistency (cf. Ziezold (1977)) holds i.e.

$$\bigcap_{n=1}^{\infty} \overline{\bigcup_{k=n}^{\infty}} E_k \subset E \text{ a.s. } ,$$

if $\mathbb{E}\big(\rho(X,p)^2\big) < \infty \forall p \in P$, *Q separable,* (ρ, d) *uniform.*

Bhattacharya-Patrangenaru strong consistency (cf. Bhattacharya and Patrangenaru (2003)) holds if additionally $E \neq \emptyset$, (ρ, d) *coercive and* (P, d) *is Heine-Borel, i.e.* $\forall \epsilon > 0, \omega \in \Omega$ *a.s.* $\exists n = n(\epsilon, \omega) \in \mathbb{N}$ *such that*

$$\bigcup_{k=n}^{\infty} E_k \subset \{p \in P : d(E,p) \le \epsilon\} .$$

The BP-CLT for Generalized Fréchet Means

- With smooth chart ϕ of P near unique population mean $p^* = \phi(0)$, ρ^2 smooth,

The BP-CLT for Generalized Fréchet Means

- With smooth chart ϕ of P near unique population mean $p^* = \phi(0)$, ρ^2 smooth,
- Fréchet functions

$$F_n(x) = \frac{1}{2n}\sum_{j=1}^{n}\rho(X_j,\phi^{-1}(x))^2,\ F(x) = \frac{1}{2}\,\mathbb{E}[\rho(X,\phi^{-1}(x))^2]$$

The BP-CLT for Generalized Fréchet Means

- With smooth chart ϕ of P near unique population mean $p^* = \phi(0)$, ρ^2 smooth,
- Fréchet functions

$$F_n(x) = \frac{1}{2n} \sum_{j=1}^{n} \rho(X_j, \phi^{-1}(x))^2, \; F(x) = \frac{1}{2} \mathbb{E}[\rho(X, \phi^{-1}(x))^2]$$

- Taylor expansion (with suitable $\widetilde{x}$ between 0 and x_0)

$$\sqrt{n}\,\mathrm{grad}|_{x=x_0} F_n(x) \quad = \quad \sqrt{n}\,\mathrm{grad}|_{x=0} F_n(x) + \mathrm{Hess}|_{x=\widetilde{x}} F_n(x) \sqrt{n} x_0 \,,$$

The BP-CLT for Generalized Fréchet Means

- With smooth chart ϕ of P near unique population mean $p^* = \phi(0)$, ρ^2 smooth,
- Fréchet functions

$$F_n(x) = \frac{1}{2n}\sum_{j=1}^{n} \rho(X_j, \phi^{-1}(x))^2, \; F(x) = \frac{1}{2}\,\mathbb{E}[\rho(X, \phi^{-1}(x))^2]$$

- Taylor expansion (with suitable $\widetilde{x}$ between 0 and x_0)

$$\sqrt{n}\,\mathrm{grad}|_{x=x_0} F_n(x) \quad = \quad \sqrt{n}\,\mathrm{grad}|_{x=0} F_n(x) + \mathrm{Hess}|_{x=\widetilde{x}} F_n(x)\sqrt{n}x_0\,,$$

- If the generalized weak law ($n \to \infty$ and $x_0 \to 0$)

$$\mathrm{Hess}|_{x=\widetilde{x}} F_n(x) \xrightarrow{\mathbb{P}} \mathrm{Hess}|_{x=0} F(x)\,,$$

The BP-CLT for Generalized Fréchet Means

- With smooth chart ϕ of P near unique population mean $p^* = \phi(0)$, ρ^2 smooth,
- Fréchet functions

$$F_n(x) = \frac{1}{2n} \sum_{j=1}^{n} \rho(X_j, \phi^{-1}(x))^2, \; F(x) = \frac{1}{2} \mathbb{E}[\rho(X, \phi^{-1}(x))^2]$$

- Taylor expansion (with suitable $\widetilde{x}$ between 0 and x_0)

$$\sqrt{n}\,\mathrm{grad}|_{x=x_0} F_n(x) \quad = \quad \sqrt{n}\,\mathrm{grad}|_{x=0} F_n(x) + \mathrm{Hess}|_{x=\widetilde{x}} F_n(x) \sqrt{n} x_0 \,,$$

- If the generalized weak law ($n \to \infty$ and $x_0 \to 0$)

$$\mathrm{Hess}|_{x=\widetilde{x}} F_n(x) \stackrel{\mathbb{P}}{\to} \mathrm{Hess}|_{x=0} F(x) \,,$$

- holds also for random $\phi(p_n) = x_0$, $p_n \in E_n$ measurable selection, and if $\mathrm{Hess}|_{x=0} F(x) > 0$

The BP-CLT for Generalized Fréchet Means

- With smooth chart ϕ of P near unique population mean $p^* = \phi(0)$, ρ^2 smooth,
- Fréchet functions

$$F_n(x) = \frac{1}{2n} \sum_{j=1}^{n} \rho(X_j, \phi^{-1}(x))^2, \; F(x) = \frac{1}{2} \mathbb{E}[\rho(X, \phi^{-1}(x))^2]$$

- Taylor expansion (with suitable $\widetilde{x}$ between 0 and x_0)

$$\sqrt{n}\,\mathrm{grad}|_{x=x_0} F_n(x) \quad = \quad \sqrt{n}\,\mathrm{grad}|_{x=0} F_n(x) + \mathrm{Hess}|_{x=\widetilde{x}} F_n(x) \sqrt{n} x_0 \,,$$

- If the generalized weak law ($n \to \infty$ and $x_0 \to 0$)

$$\mathrm{Hess}|_{x=\widetilde{x}} F_n(x) \stackrel{\mathbb{P}}{\to} \mathrm{Hess}|_{x=0} F(x) \,,$$

- holds also for random $\phi(p_n) = x_0$, $p_n \in E_n$ measurable selection, and if $\mathrm{Hess}|_{x=0} F(x) > 0$

The BP-CLT for Generalized Fréchet Means

- With smooth chart ϕ of P near unique population mean $p^* = \phi(0)$, ρ^2 smooth,
- Fréchet functions

$$F_n(x) = \frac{1}{2n} \sum_{j=1}^{n} \rho(X_j, \phi^{-1}(x))^2, \; F(x) = \frac{1}{2} \mathbb{E}[\rho(X, \phi^{-1}(x))^2]$$

- Taylor expansion (with suitable $\widetilde{x}$ between 0 and x_0)

$$\sqrt{n}\,\mathrm{grad}|_{x=x_0} F_n(x) = \sqrt{n}\,\mathrm{grad}|_{x=0} F_n(x) + \mathrm{Hess}|_{x=\widetilde{x}} F_n(x) \sqrt{n} x_0 \,,$$

- If the generalized weak law ($n \to \infty$ and $x_0 \to 0$)

$$\mathrm{Hess}|_{x=\widetilde{x}} F_n(x) \stackrel{\mathbb{P}}{\to} \mathrm{Hess}|_{x=0} F(x) \,,$$

- holds also for random $\phi(p_n) = x_0$, $p_n \in E_n$ measurable selection, and if $\mathrm{Hess}|_{x=0} F(x) > 0$

Theorem (H. 2011a)

$\sqrt{n}\,\phi(\hat{p}_n) \stackrel{\mathcal{D}}{\to} \mathcal{N}(0, \Sigma)$ *with suitable* $\Sigma > 0$.

Backward Nested Families of Descriptors

Q (topological, separable = ts): Data space

Backward Nested Families of Descriptors

Q (topological, separable = ts): Data space

(i) $\exists\{P_j\}_{j=0}^m$ (ts) with continuous $d_j : P_j \times P_j \to [0, \infty)$ vanishing exactly on the diagonal, $P_m = \{Q\}$;

Backward Nested Families of Descriptors

Q (topological, separable = ts): Data space

(i) $\exists \{P_j\}_{j=0}^m$ (ts) with continuous $d_j : P_j \times P_j \to [0, \infty)$ vanishing exactly on the diagonal, $P_m = \{Q\}$;

(ii) every $p \in P_j$ ($j = 1, \ldots, m$) is itself a topological space giving rise to a topological space $\emptyset \neq S_p \subseteq P_{j-1}$ with

$$\rho_p : p \times S_p \to [0, \infty) \text{ , continuous ;}$$

Backward Nested Families of Descriptors

Q (topological, separable = ts): Data space

(i) $\exists \{P_j\}_{j=0}^m$ (ts) with continuous $d_j : P_j \times P_j \to [0, \infty)$ vanishing exactly on the diagonal, $P_m = \{Q\}$;

(ii) every $p \in P_j$ ($j = 1, \ldots, m$) is itself a topological space giving rise to a topological space $\emptyset \neq S_p \subseteq P_{j-1}$ with

$$\rho_p : p \times S_p \to [0, \infty) \text{ , continuous ;}$$

(iii) $\forall\, p \in P_j$ ($j = 1, \ldots, m$) and $s \in S_p$ $\exists$ “projection”

$$\pi_{p,s} : p \to s \text{ , measurable .}$$

Backward Nested Families of Descriptors

Q (topological, separable = ts): Data space

(i) $\exists \{P_j\}_{j=0}^m$ (ts) with continuous $d_j : P_j \times P_j \to [0,\infty)$ vanishing exactly on the diagonal, $P_m = \{Q\}$;

(ii) every $p \in P_j$ ($j = 1, \dots, m$) is itself a topological space giving rise to a topological space $\emptyset \neq S_p \subseteq P_{j-1}$ with

$$\rho_p : p \times S_p \to [0,\infty)\text{ , continuous ;}$$

(iii) $\forall\, p \in P_j$ ($j = 1, \dots, m$) and $s \in S_p$ $\exists$ "projection"

$$\pi_{p,s} : p \to s\text{ , measurable .}$$

For $j \in \{1, \dots, m\}$ and $k \in \{1, \dots, j\}$,

$$f = \{p^j, \dots, p^{j-k}\},\ \text{with } p^{l-1} \in S_{p^l}, l = j-k+1, \dots, j$$

is BNFD from P_j to P_{j-k} from the space

$$T_{j,k} = \left\{ f = \{p^{j-l}\}_{l=0}^k : p^{l-1} \in S_{p^l}, l = j-k+1, \dots, j \right\} ,$$

Backward Nested Families of Descriptors

Q (topological, separable = ts): Data space

(i) $\exists \{P_j\}_{j=0}^m$ (ts) with continuous $d_j : P_j \times P_j \to [0, \infty)$ vanishing exactly on the diagonal, $P_m = \{Q\}$;

(ii) every $p \in P_j$ ($j = 1, \ldots, m$) is itself a topological space giving rise to a topological space $\emptyset \neq S_p \subseteq P_{j-1}$ with

$$\rho_p : p \times S_p \to [0, \infty)\text{ , continuous ;}$$

(iii) $\forall\, p \in P_j$ ($j = 1, \ldots, m$) and $s \in S_p$ $\exists$ "projection"

$$\pi_{p,s} : p \to s\text{ , measurable .}$$

For $j \in \{1, \ldots, m\}$ and $k \in \{1, \ldots, j\}$,

$$f = \{p^j, \ldots, p^{j-k}\},\ \text{with } p^{l-1} \in S_{p^l}, l = j - k + 1, \ldots, j$$

is BNFD from P_j to P_{j-k} from the space

$$T_{j,k} = \left\{ f = \{p^{j-l}\}_{l=0}^k : p^{l-1} \in S_{p^l}, l = j - k + 1, \ldots, j \right\}\ ,$$

with projection along each descriptor

$$\pi_f = \pi_{p^{m-k+1}, p^{m-k}} \circ \ldots \circ \pi_{p^m, p^{m-1}} : p^m \to p^{m-k}$$

Backward Nested Families of Descriptors

For another BNFD $f' = \{p'^{j-l}\}_{l=0}^{k} \in T_{j,k}$ set

$$d^j(f, f') = \sqrt{\sum_{l=0}^{k} d_j(p^{j-l}, p'^{j-l})^2}\,.$$

Backward Nested Fréchet Means

Random elements $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} X$ on a data space Q admitting BNFDs give rise to backward nested population and sample means (BN means) recursively defined via $f^m = \{Q\} = f_n^m$, i.e. $p^m = Q = p_n^m$ and for $j = m, \dots, 1$,

$$p^{j-1} \in \operatorname*{argmin}_{s \in S_{p^j}} \mathbb{E}[\rho_{p^j}(\pi_{f^j} \circ X, s)^2], \qquad f^{j-1} = (p^k)_{k=j-1}^m$$

$$p_n^{j-1} \in \operatorname*{argmin}_{s \in S_{p_n^j}} \sum_{i=1}^n \rho_{p_n^j}(\pi_{f_n^j} \circ X_i, s)^2, \qquad f_n^{j-1} = (p_n^k)_{k=j-1}^m .$$

Backward Nested Fréchet Means

Random elements $X_1, \ldots, X_n \stackrel{\text{i.i.d.}}{\sim} X$ on a data space Q admitting BNFDs give rise to backward nested population and sample means (BN means) recursively defined via $f^m = \{Q\} = f_n^m$, i.e. $p^m = Q = p_n^m$ and for $j = m, \ldots, 1$,

$$p^{j-1} \in \operatorname*{argmin}_{s \in S_{p^j}} \mathbb{E}[\rho_{p^j}(\pi_{f^j} \circ X, s)^2], \qquad f^{j-1} = (p^k)_{k=j-1}^m$$

$$p_n^{j-1} \in \operatorname*{argmin}_{s \in S_{p_n^j}} \sum_{i=1}^n \rho_{p_n^j}(\pi_{f_n^j} \circ X_i, s)^2, \qquad f_n^{j-1} = (p_n^k)_{k=j-1}^m .$$

If all of the population minimizers are unique, we speak of unique BN means.

Strong Law

Theorem (H. and Eltzner (2017))

If the BN population means $f = (p^m, \dots, p^{m-k})$ are unique and $f_n = (p_n^m, \dots, p_n^{m-k})$ is a measurable selection of BN sample means then under "reasonable" assumptions

$$f_n \to f \text{ a.s.}$$

i.e. $\exists \Omega' \subseteq \Omega$ *m'ble with* $\mathbb{P}(\Omega') = 1$ *such that* $\forall \epsilon > 0$ *and* $\omega \in \Omega'$, $\exists N(\epsilon, \omega) \in \mathbb{N}$

$$d(f_n, f) < \epsilon \quad \forall n \geq N(\epsilon, \omega)\,.$$

The Joint CLT [H. and Eltzner (2017)]

With local chart $\eta \stackrel{\psi^{-1}}{\mapsto} f^{j-1} \mapsto \rho_{p^j}(\pi_{f^j} \circ X, p^{j-1})^2 := \tau^j(\eta, X)$:

$$\sqrt{n} H_\psi \big(\psi(f_n^{j-1}) - \psi(f'^{j-1})\big) \quad \to \quad \mathcal{N}(0, B_\psi)\,.$$

The Joint CLT [H. and Eltzner (2017)]

With local chart $\eta \overset{\psi^{-1}}{\mapsto} f^{j-1} \mapsto \rho_{p^j}(\pi_{f^j} \circ X, p^{j-1})^2 := \tau^j(\eta, X)$:

$$\sqrt{n} H_\psi \left(\psi(f_n^{j-1}) - \psi(f'^{j-1})\right) \quad \to \quad \mathcal{N}(0, B_\psi)\,.$$

Idea of proof:

$$0 \quad = \quad \operatorname{grad}_\eta \sum_{k=1}^{n} \tau^j(\eta_n, X_k) + \sum_{l=j+1}^{m} \lambda_n^l \operatorname{grad}_\eta \sum_{k=1}^{n} \tau^l(\eta_n, X_k)$$

The Joint CLT [H. and Eltzner (2017)]

With local chart $\eta \overset{\psi^{-1}}{\mapsto} f^{j-1} \mapsto \rho_{p^j}(\pi_{f^j} \circ X, p^{j-1})^2 := \tau^j(\eta, X)$:

$$\sqrt{n} H_\psi \left(\psi(f_n^{j-1}) - \psi(f'^{j-1})\right) \quad \to \quad \mathcal{N}(0, B_\psi) \,.$$

Idea of proof:

$$\begin{aligned}
0 &= \operatorname{grad}_\eta \sum_{k=1}^{n} \tau^j(\eta_n, X_k) + \sum_{l=j+1}^{m} \lambda_n^l \operatorname{grad}_\eta \sum_{k=1}^{n} \tau^l(\eta_n, X_k) \\
&= \operatorname{grad}_\eta \sum_{k=1}^{n} \tau^j(\eta', X_k) + \sum_{l=j+1}^{m} \lambda_n^l \operatorname{grad}_\eta \sum_{k=1}^{n} \tau^l(\eta', X_k) \\
&\quad + \left(\operatorname{Hess}_\eta \sum_{k=1}^{n} \tau^j(\widetilde{\eta}_n, X_k) + \sum_{l=j+1}^{m} \lambda_n^l \operatorname{Hess}_\eta \sum_{k=1}^{n} \tau^l(\widetilde{\eta}_n, X_k)\right) \\
&\qquad \cdot (\eta' - \eta_n)
\end{aligned}$$

with $\widetilde{\eta}_n$ between η' and η_n.

The Joint CLT [H. and Eltzner (2017)]

With local chart $\eta \overset{\psi^{-1}}{\mapsto} f^{j-1} \mapsto \rho_{p^j}(\pi_{f^j} \circ X, p^{j-1})^2 := \tau^j(\eta, X)$:

$$\sqrt{n} H_\psi \left(\psi(f_n^{j-1}) - \psi(f'^{j-1})\right) \quad \to \quad \mathcal{N}(0, B_\psi)\,.$$

Idea of proof:

$$\begin{aligned}
0 \quad &= \quad \mathrm{grad}_\eta \sum_{k=1}^{n} \tau^j(\eta_n, X_k) + \sum_{l=j+1}^{m} \lambda_n^l \,\mathrm{grad}_\eta \sum_{k=1}^{n} \tau^l(\eta_n, X_k) \\
&= \quad \mathrm{grad}_\eta \sum_{k=1}^{n} \tau^j(\eta', X_k) + \sum_{l=j+1}^{m} \lambda_n^l \,\mathrm{grad}_\eta \sum_{k=1}^{n} \tau^l(\eta', X_k) \\
&\qquad + \left(\mathrm{Hess}_\eta \sum_{k=1}^{n} \tau^j(\widetilde{\eta}_n, X_k) + \sum_{l=j+1}^{m} \lambda_n^l \,\mathrm{Hess}_\eta \sum_{k=1}^{n} \tau^l(\widetilde{\eta}_n, X_k)\right) \\
&\qquad\quad \cdot(\eta' - \eta_n)
\end{aligned}$$

with $\widetilde{\eta}_n$ between η' and η_n. N.B.: $\lambda_n^l \overset{\mathbb{P}}{\to} \lambda^l$.

The Joint Central Limit Theorem

With local chart $\eta \stackrel{\psi^{-1}}{\mapsto} f^{j-1} \mapsto \rho_{p^j}(\pi_{f^j} \circ X, p^{j-1})^2 := \tau^j(\eta, X)$:

$$\sqrt{n} H_\psi \left(\psi(f_n^{j-1}) - \psi(f'^{j-1}) \right) \quad \to \quad \mathcal{N}(0, B_\psi)$$

and typical regularity conditions, where

$$H_\psi \quad = \quad \mathbb{E} \left[\mathrm{Hess}_\eta \tau^j(\eta', X) + \sum_{l=j+1}^{m} \lambda^l \, \mathrm{Hess}_\eta \tau^l(\eta', X) \right] \text{ and}$$

$$B_\psi \quad = \quad \mathrm{cov} \left[\mathrm{grad}_\eta \tau^j(\eta', X) + \sum_{l=j+1}^{m} \lambda^l \, \mathrm{grad}_\eta \tau^l(\eta', X) \right] .$$

and $\lambda_{j+1}, \ldots \lambda_m \in \mathbb{R}$ are suitable such that

$$\mathrm{grad}_\eta \, \mathbb{E}\left[\tau^j(\eta, X)\right] + \sum_{l=j+1}^{m} \lambda^l \, \mathrm{grad}_\eta \, \mathbb{E}\left[\tau^l(\eta, X)\right]$$

vanishes at $\eta = \eta'$.

Factoring Charts

If the following diagram commutes we say the chart factors

$$\begin{array}{ccccccccc} T_{m-1,j-1} & \ni & f^{j-1} & = & (f^j, p^{j-1}) & \xrightarrow{\psi} & \eta & = & (\theta, \xi) \\ & & & & \downarrow \pi^{P_{j-1}} & & & & \downarrow \pi^{\mathbb{R}^{\dim(\theta)}} \\ P_{j-1} & \ni & & & p^{j-1} & \xrightarrow{\phi} & & & \theta \end{array}$$

Factoring Charts

If the following diagram commutes we say the chart factors

$$\begin{array}{cccccccc} T_{m-1,j-1} & \ni & f^{j-1} & = & (f^j, p^{j-1}) & \stackrel{\psi}{\to} & \eta & = & (\theta, \xi) \\ & & & & \downarrow \pi^{P_{j-1}} & & & & \downarrow \pi^{\mathbb{R}^{\dim(\theta)}} \\ P_{j-1} & \ni & & & p^{j-1} & \stackrel{\phi}{\to} & & & \theta \end{array}$$

Then

$$\begin{aligned} \eta = (\theta, \xi) \stackrel{\psi^{-1}}{\mapsto} f^{j-1} \quad &\mapsto \quad \rho_{p^j}(\pi_{f^j} \circ X, p^{j-1})^2 \\ &= \quad \rho_{\pi^{P_j} \circ \psi_2^{-1}(\xi)}\left(\pi_{\psi_2^{-1}(\xi)} \circ X, \psi_1^{-1}(\theta)\right)^2 \\ &=: \quad \tau^j(\theta, \xi, X)\,, \end{aligned}$$

Factoring Charts

If the following diagram commutes we say the chart factors

$$\begin{array}{ccccccc} T_{m-1,j-1} & \ni & f^{j-1} & = & (f^j, p^{j-1}) & \overset{\psi}{\to} & \eta & = & (\theta,\xi) \\ & & & & \downarrow \pi^{P_{j-1}} & & & & \downarrow \pi^{\mathbb{R}^{\dim(\theta)}} \\ P_{j-1} & \ni & & & p^{j-1} & \overset{\phi}{\to} & & & \theta \end{array}$$

Then

$$\begin{aligned} \eta = (\theta,\xi) \overset{\psi^{-1}}{\mapsto} f^{j-1} \quad \mapsto & \quad \rho_{p^j}(\pi_{f^j} \circ X, p^{j-1})^2 \\ = & \quad \rho_{\pi^{P_j} \circ \psi_2^{-1}(\xi)} \left(\pi_{\psi_2^{-1}(\xi)} \circ X, \psi_1^{-1}(\theta) \right)^2 \\ =: & \quad \tau^j(\theta,\xi,X)\,, \end{aligned}$$

Taylor expansion at $\eta' = (\theta', \xi')$ gives a joint Gaussian CLT,

$$\sqrt{n} H_\psi(\eta_n - \eta') = \sqrt{n} H_\psi \begin{pmatrix} \theta_n - \theta' \\ \xi_n - \xi' \end{pmatrix} \quad \to \quad \mathcal{N}(0, B_\psi)$$

and projection to the θ coordinate preserves Gaussianity.

BNFD CLT Incl. Factoring Charts

Holds for (Eltzner and H. 2017)

- PNS,PNGS.
- 1st geodesic PC on manifolds including intrinsic mean on 1st PC,
- 1st geodesic PC on Kendall shape spaces (notably not a manifold beginning with dim 3) including intrinsic mean on 1st PC,
- working on barycentric subspaces by Pennec (2017),
- ?

BNFD CLT Incl. Factoring Charts

Holds for (Eltzner and H. 2017)

- PNS,PNGS.
- 1st geodesic PC on manifolds including intrinsic mean on 1st PC,
- 1st geodesic PC on Kendall shape spaces (notably not a manifold beginning with dim 3) including intrinsic mean on 1st PC,
- working on barycentric subspaces by Pennec (2017),
- ?

Practioner's advice:

- For a two-sample test, need empirical covariances.
- Suitably bootstrap data (Eltzner and H. 2017).

Revisiting "Typical Regularity Conditions" I

Recall conditions

(i) a C^2 manifold structure with C^2 distance^2 near

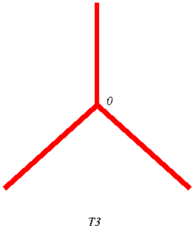
(ii) a unique population mean μ,

Revisiting "Typical Regularity Conditions" I

Recall conditions

(i) a C^2 manifold structure with C^2 distance^2 near

(ii) a unique population mean μ,

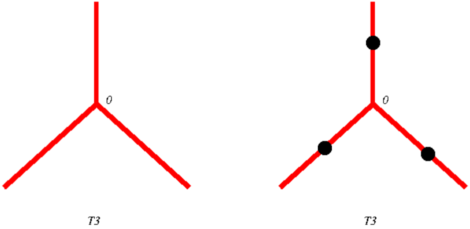


Revisiting “Typical Regularity Conditions” I

Recall conditions

(i) a C^2 manifold structure with C^2 distance^2 near

(ii) a unique population mean μ,

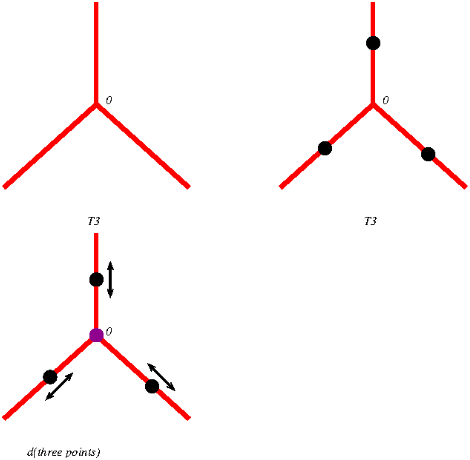


Revisiting “Typical Regularity Conditions” I

Recall conditions

(i) a C^2 manifold structure with C^2 distance^2 near

(ii) a unique population mean μ,

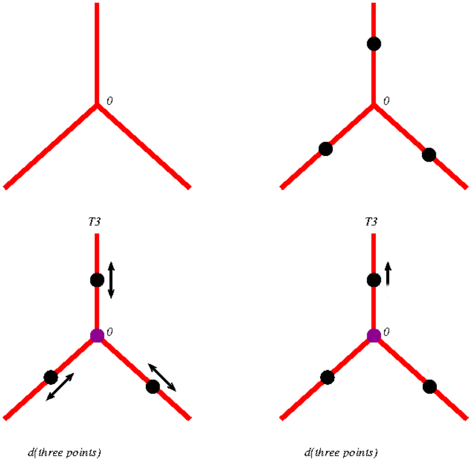


Revisiting “Typical Regularity Conditions” I

Recall conditions

(i) a C^2 manifold structure with C^2 distance^2 near

(ii) a unique population mean μ,



Hotz et al. (2013)

A General Definition for Stickiness

Let $\mathcal{M}$ be a set of measures on a metric space (Q, ρ). Assume $\mathcal{M}$ has a given topology. A mean is a continuous map

$$\mathcal{M} \to \{\text{closed subsets of } Q\}\,.$$

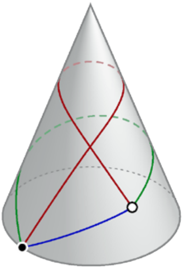
A measure μ sticks to a closed subset $C \subset Q$ if every neighborhood of μ in $\mathcal{M}$ contains a nonempty open subset consisting of measures whose mean sets are contained in C.

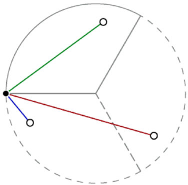
Typical topology by Wasserstein metric

$$\rho(\mu,\nu) = \sup_{f \in \mathrm{Lip}_1(\mathcal{K},\mathbb{R})} \left(\int f\, d\mu - \int f\, d\nu \right),$$

(H. et al. 2015).

Example: The Cone





Exercise

Unless $X =$ *cone point a.s.,* $E^\rho \neq$ *cone point.*

The Hyperbolic Cone

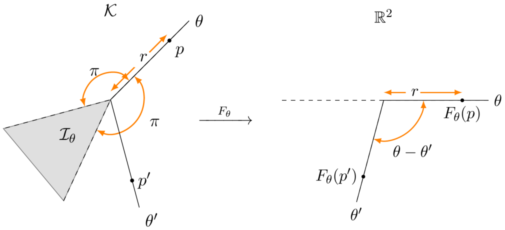
- opening angle $\alpha > 2\pi$
- contains way more ice cream

The Hyperbolic Cone

- opening angle $\alpha > 2\pi$
- contains way more ice cream
- can be embedded in $\mathbb{R}^3$ only non-isometrically, say, as a kale

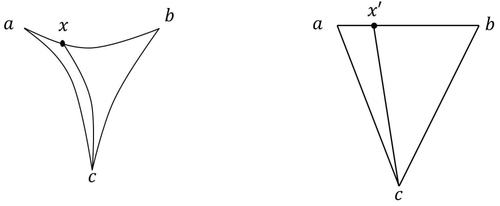
$$\mathcal{K} = ([0, \infty) \times [0, \alpha])/\sim$$

- polar coordinates $p = (r, \theta) \in [0, \infty) \times [0, \alpha]/\sim$
- folding map F_θ



Uniqueness of Fréchet Means Under Non-Positive Curvature

A metric space (Q, ρ) is NPC if every ρ-triangle mapped to $\mathbb{R}^2$ is more skinny, i.e., ρ-distances accross are smaller than corresponding Euclidean distances



Theorem (Sturm (2003))

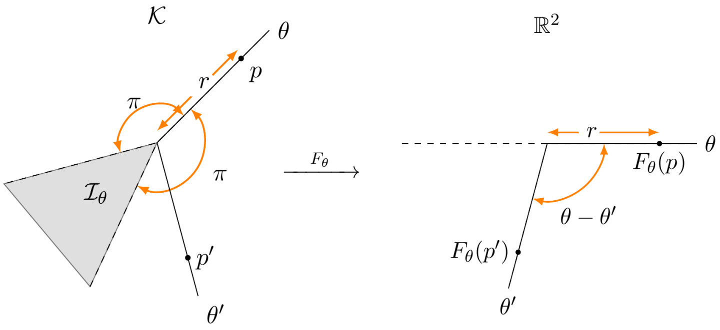
On a complete NPC metric space, Fréchet means are unique.

Notation: $\{\mu_n\} := E_n^\rho$, $\{\mu\} := E^\rho$.

Folded Moments

Recall the folding map

$$F_\theta(r',\theta') = \begin{cases} \mathbf{0} & \text{if } r' = 0 \\ \big(r'\cos(\theta'-\theta), r'\sin(\theta'-\theta)\big) & \text{if } \big|\theta'-\theta\big| < \pi \text{ and } r' > 0 \\ (-r', 0) & \text{if } \big|\theta'-\theta\big| \geq \pi \text{ and } r' > 0. \end{cases}$$



Folded Moments

Recall the folding map

$$F_\theta(r',\theta') = \begin{cases} \mathbf{0} & \text{if } r' = 0 \\ \big(r'\cos(\theta'-\theta), r'\sin(\theta'-\theta)\big) & \text{if } \left|\theta'-\theta\right| < \pi \text{ and } r' > 0 \\ (-r', 0) & \text{if } \left|\theta'-\theta\right| \geq \pi \text{ and } r' > 0. \end{cases}$$

in conjunction with a measure $\mathbb{P}^X$ on $\mathcal{K}$, giving rise to folded moments

$$m_\theta = \int_{\mathcal{K}} F_\theta(p)\, d\,\mathbb{P}^X(p)$$

Key feature: Under integrability $\int_{\mathcal{K}} \rho(\mathbf{0}, p)\, d\,\mathbb{P}^X(p) < \infty$,

$$\frac{d}{d\theta} m_{\theta,1} = m_{\theta,2}$$

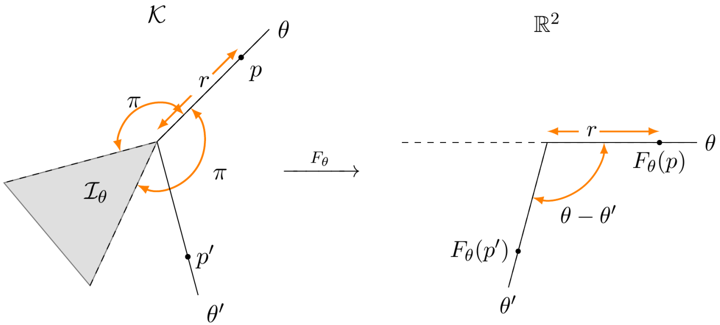
(derivative is zero on the shadow and so is $m_{\theta,2}$).

The Shadow's (Boundary) Effect

$$D_\theta^\pm \frac{dm_{\theta,1}}{d\theta} = D_\theta^\pm m_{\theta,2} = -m_{\theta,1} + \int_{\mathcal{I}_\theta^\mp} \left(-\rho(\mathbf{0},p)\right) d\,\mathbb{P}^X(p)$$
$$\leq -m_{\theta,1} - \int_{\mathcal{I}_\theta} \rho(\mathbf{0},p)\, d\,\mathbb{P}^X(p)$$

with

$$\mathcal{I}_\theta^+ = \mathcal{K} \setminus \{(r,\theta') \mid r > 0 \text{ and } -\pi \leq \theta' - \theta < \pi\},$$
$$\mathcal{I}_\theta^- = \mathcal{K} \setminus \{(r,\theta') \mid r > 0 \text{ and } -\pi < \theta' - \theta \leq \pi\}.$$

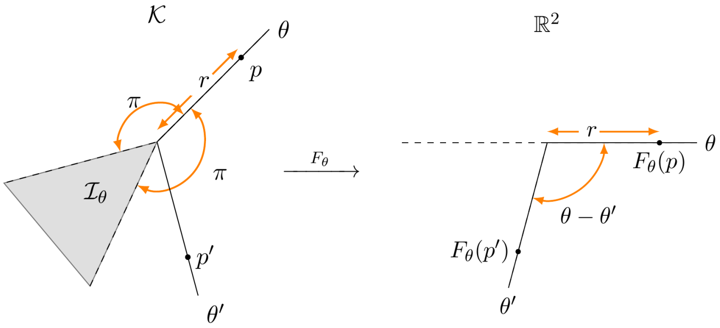


The Shadow's (Boundary) Effect

$$D_\theta^\pm \frac{dm_{\theta,1}}{d\theta} = D_\theta^\pm m_{\theta,2} = -m_{\theta,1} - \int_{\mathcal{I}_\theta^\mp} \rho(\mathbf{0},p)\, d\,\mathbb{P}^X(p)$$
$$\leq -m_{\theta,1} - \int_{\mathcal{I}_\theta} \rho(\mathbf{0},p)\, d\,\mathbb{P}^X(p)$$

with

$$\mathcal{I}_\theta^+ = \mathcal{K} \setminus \{(r,\theta') \mid r > 0 \text{ and } -\pi \leq \theta' - \theta < \pi\},$$
$$\mathcal{I}_\theta^- = \mathcal{K} \setminus \{(r,\theta') \mid r > 0 \text{ and } -\pi < \theta' - \theta \leq \pi\}.$$

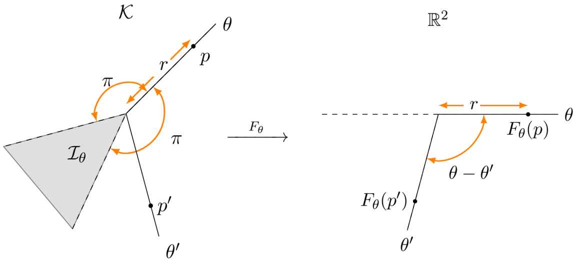


The Shadow's (Boundary) Effect

$$D_\theta^\pm \frac{dm_{\theta,1}}{d\theta} = D_\theta^\pm m_{\theta,2} = -m_{\theta,1} - \int_{\mathcal{I}_\theta^\mp} \rho(\mathbf{0},p)\, d\,\mathbb{P}^X(p) \leq -m_{\theta,1} - \int_{\mathcal{I}_\theta} \rho(\mathbf{0},p)\, d\,\mathbb{P}^X(p)$$

with

$$\mathcal{I}_\theta^+ = \mathcal{K} \setminus \{(r,\theta') \mid r > 0 \text{ and } -\pi \leq \theta' - \theta < \pi\},$$
$$\mathcal{I}_\theta^- = \mathcal{K} \setminus \{(r,\theta') \mid r > 0 \text{ and } -\pi < \theta' - \theta \leq \pi\}.$$



$$m_{\theta,1} \leq -m_{\theta',1} \quad \forall (1,\theta') \in \mathcal{I}_\theta\,.$$

The Shadow's (Boundary) Effect

$$D^{\pm}_{\theta} \frac{dm_{\theta,1}}{d\theta} = D^{\pm}_{\theta} m_{\theta,2} = -m_{\theta,1} - \int_{\mathcal{I}^{\mp}_{\theta}} \rho(\mathbf{0}, p)\, d\,\mathbb{P}^X(p)$$
$$\leq -m_{\theta,1} - \int_{\mathcal{I}_{\theta}} \rho(\mathbf{0}, p)\, d\,\mathbb{P}^X(p)$$

with

$$\mathcal{I}^{+}_{\theta} = \mathcal{K} \setminus \{(r, \theta') \mid r > 0 \text{ and } -\pi \leq \theta' - \theta < \pi\},$$
$$\mathcal{I}^{-}_{\theta} = \mathcal{K} \setminus \{(r, \theta') \mid r > 0 \text{ and } -\pi < \theta' - \theta \leq \pi\}.$$

$$m_{\theta,1} \leq -m_{\theta',1} \quad \forall (1, \theta') \in \mathcal{I}_{\theta}\,.$$

$$\Downarrow$$

Lemma

Let $A \neq B$ and $m_{\theta,1} \geq 0$ on $\theta \in [A, B]$. Then $|A - B| \leq \pi$.

- *if $m_{\theta,1} = 0\ \forall \theta \in [A, B] \Rightarrow \mathbb{P}^X(\mathcal{I}_{\theta}) = 0\ \forall \theta \in [A, B]$.*
- *if $m_{\theta,1} > 0$ then it's concave there.*

The Strong Law

Theorem

Assuming integrability and nondegeneracy,

$$1_{\{m_{\theta,1} \geq 0\}} \subset [0, \alpha]/\sim \text{ is a closed interval, or empty}$$

that is exactly one of the following:

(fully sticky) *empty, then* $\mu = \mathbf{0}$ *and* $\exists n^*(\omega) \in \mathbb{N}$ *such that* $\mu_n(\omega) = \mathbf{0}$ *for all* $n \geq n^*(\omega)$*, a.s.*

The Strong Law

Theorem

Assuming integrability and nondegeneracy,

$$1_{\{m_{\theta,1}\geq 0\}} \subset [0,\alpha]/\sim \text{ is a closed interval, or empty}$$

that is exactly one of the following:

(fully sticky) *empty, then* $\mu = \mathbf{0}$ *and* $\exists n^*(\omega) \in \mathbb{N}$ *such that* $\mu_n(\omega) = \mathbf{0}$ *for all* $n \geq n^*(\omega)$, *a.s.*

(partly sticky) *of length* $< \pi$, *with* $m_{\theta,1} = 0$ *on its entirety such that* $\mu = \mathbf{0}$ *and* $\mu_n(\omega) \to \mathbf{0}$ *a.s. Furthermore, if* $1_{\{m_{\theta,1}\geq 0\}} \subset (A,B)$ $\Rightarrow \exists n^*(\omega) \in \mathbb{N}$ *such that* $\mu_n(\omega) \in \overline{C_{(A,B)}}$ *for all* $n \geq n^*(\omega)$ *a.s.*

The Strong Law

Theorem

Assuming integrability and nondegeneracy,

$$1_{\{m_{\theta,1}\geq 0\}} \subset [0,\alpha]/\sim \text{ is a closed interval, or empty}$$

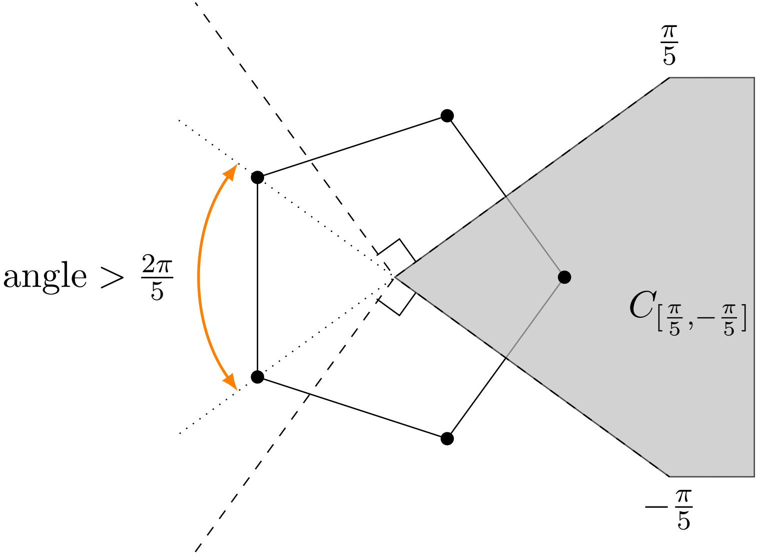
that is exactly one of the following:

(fully sticky) *empty, then* $\mu = \mathbf{0}$ *and* $\exists n^*(\omega) \in \mathbb{N}$ *such that* $\mu_n(\omega) = \mathbf{0}$ *for all* $n \geq n^*(\omega)$, *a.s.*

(partly sticky) *of length* $< \pi$, *with* $m_{\theta,1} = 0$ *on its entirety such that* $\mu = \mathbf{0}$ *and* $\mu_n(\omega) \to \mathbf{0}$ *a.s. Furthermore, if* $1_{\{m_{\theta,1}\geq 0\}} \subset (A,B)$ $\Rightarrow \exists n^*(\omega) \in \mathbb{N}$ *such that* $\mu_n(\omega) \in \overline{C_{(A,B)}}$ *for all* $n \geq n^*(\omega)$ *a.s.*

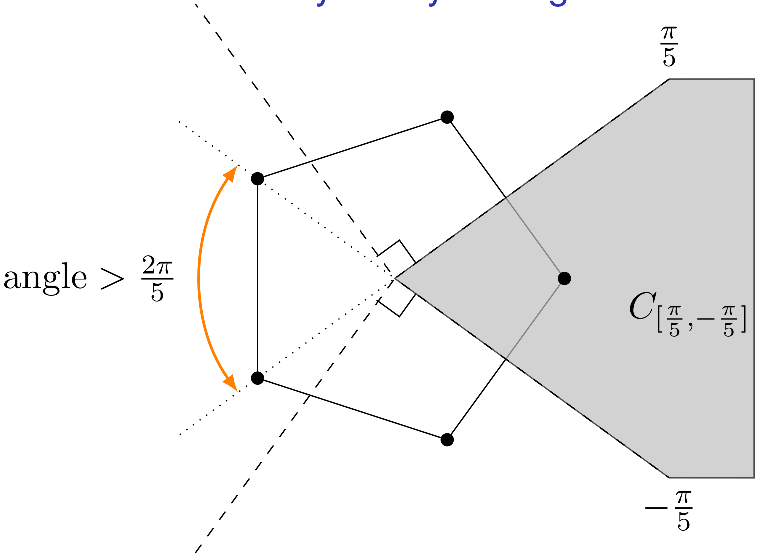
(nonsticky) *of length* $\leq \pi$, *with* $m_{\theta,1}$ *strictly concave (and hence strictly positive) on its interior.* $\mu_n(\omega) \to \mu \neq \mathbf{0}$ *a.s.*

The Partly Sticky Strong Law



Uniformly sampling from a pentagon's vertices

The Partly Sticky Strong Law



For the uniform on $\{(r,\theta) : -\pi < \theta < \pi\}$ the fluctuation is only on $[0,\infty)$

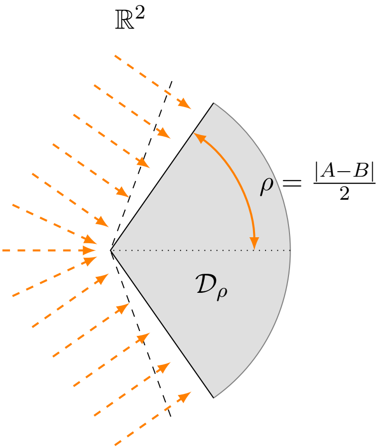
Sticky CLTs on the Kale

1. Fully sticky $\Rightarrow$ trivial CLT $\checkmark$
2. Partly sticky $\Rightarrow$??
3. Nonsticky $\Rightarrow$ BP-CLT (classical $\sqrt{n}$ Gaussian)?

The Partly Sticky CLT

In case of square integrability and $1_{\{m_{\theta,1}\geq 0\}} = [A,B] = 1_{\{m_{\theta,1}=0\}}$ *(recall length* $< \pi$*), with center* θ^**, decompose suitable Gaussian in* $\mathbb{R}^2$ *centered at* $\mathbf{0}$ *into three parts:*

- $\mathcal{G}_1$ *in cone* D_ρ*,*
- $\mathcal{G}_2$ *in the two adjacent cones with* 90^0 *opening,*
- $\mathcal{G}_3$ *in the rest.*



The limiting distribution of

$$\sqrt{n}\big(F_{\theta^*}(\mu_n) - \mathbf{0}\big)$$

is

$$\mathcal{G}_1 + \pi_{C_A\cup C_B}\circ\mathcal{G}_2 + \pi_{\mathbf{0}}\circ\mathcal{G}_3\,.$$

The Nonsticky CLT

In case of square integrability and $\mu = (r^*, \theta^*)$, $r^* > 0$, *define*

$$\kappa(\omega) = \begin{cases} \dfrac{\int_{\mathcal{I}^+_{\theta^*}} \rho(\mathbf{0}, p)\, d\,\mathbb{P}^X(p)}{r^*} & \text{if } e_2 \cdot F_{\theta^*}\mu_n(\omega) < 0 \\[2ex] \dfrac{\int_{\mathcal{I}^-_{\theta^*}} \rho(\mathbf{0}, p)\, d\,\mathbb{P}^X(p)}{r^*} & \text{if } e_2 \cdot F_{\theta^*}\mu_n(\omega) > 0, \end{cases}$$

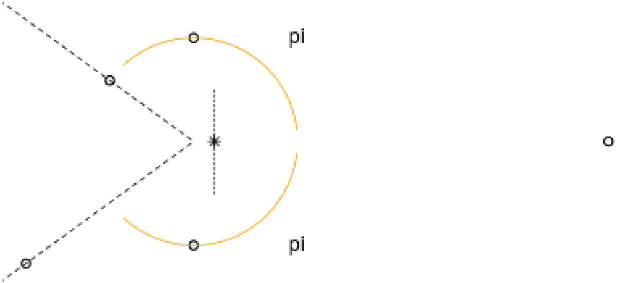
and

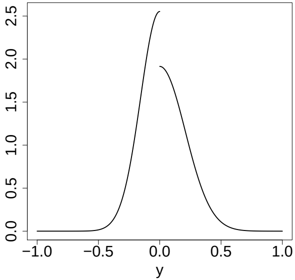
$$Q_n(W) = \mathbb{P}\left\{\left(\sqrt{n}(e_1 \cdot F_{\theta^*}\mu_n - r^*, (1+\kappa)e_2 \cdot F_{\theta^*}\mu_n\right) \in W)\right\}.$$

Then, $Q_n \to \mathcal{G}$ *weakly where* $\mathcal{G}$ *is a suitable Gaussian in* $\mathbb{R}^2$ *centered at* $(r^*, 0)$ *with covariance*

$$\int_{\mathbb{R}^2} (y - F_{\theta^*}\mu)(y - F_{\theta^*}\mu)^T\, d\,\mathbb{P}^X \circ F^{-1}_{\theta^*}(y)\,.$$

Example of Non-Gaussian Nonsticky CLT



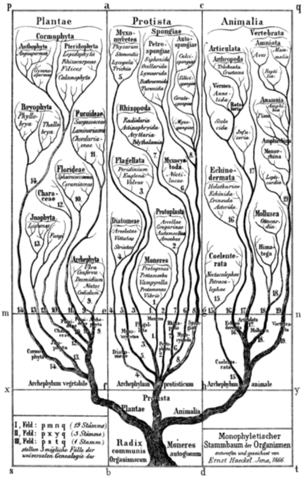


Curiosities and Motivation

- The fully sticky CLT only requires integrability.
- Even the nonsticky CLT may be non-Gaussian.

Curiosities and Motivation

- The fully sticky CLT only requires integrability.
- Even the nonsticky CLT may be non-Gaussian.
- This research has been motivated by statistical analysis of phylogenetic trees, the famous BHV tree space (Billera et al. (2001)) has a hyperbolic singularity at the cone point = star tree.



Tree of Evolution by Haeckel (1879)

Revisiting “Typical Regularity Conditions” II

Recall conditions

(iii) for all random $\hat{x}_n \stackrel{\text{a.s.}}{\to} 0$, $\mathrm{Hess}|_{x=\hat{x}_n} F_n(x) \stackrel{\mathbb{P}}{\to} \mathrm{Hess}|_{x=0} F(x)$,

(iv) $\mathrm{Hess}|_{x=0} F(x) > 0\,.$

Revisiting “Typical Regularity Conditions” II

Recall conditions

(iii) for all random $\hat{x}_n \stackrel{\text{a.s.}}{\to} 0$, $\mathrm{Hess}|_{x=\hat{x}_n} F_n(x) \stackrel{\mathbb{P}}{\to} \mathrm{Hess}|_{x=0} F(x)$,

(iv) $\mathrm{Hess}|_{x=0} F(x) > 0\,.$

Consider (McKilliam et al. (2012), Hotz and H. 2015):

- $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} X \in \mathbb{S}^1 = [-\pi, \pi]/\sim$
- Fréchet means 0 (population), x_n (sample)
- f local density near $-\pi \cong \pi$, w.l.o.g. $x \geq 0$

Revisiting "Typical Regularity Conditions" II

Recall conditions

(iii) for all random $\hat{x}_n \stackrel{\text{a.s.}}{\to} 0$, $\mathrm{Hess}|_{x=\hat{x}_n} F_n(x) \stackrel{\mathbb{P}}{\to} \mathrm{Hess}|_{x=0} F(x)$,

(iv) $\mathrm{Hess}|_{x=0} F(x) > 0$.

Consider (McKilliam et al. (2012), Hotz and H. 2015):

- $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} X \in \mathbb{S}^1 = [-\pi, \pi]/\sim$
- Fréchet means 0 (population), x_n (sample)
- f local density near $-\pi \cong \pi$, w.l.o.g. $x \geq 0$

$$2nF_n(x) = \sum_{x-\pi \leq X_j} (X_j - x)^2 + \sum_{X_j < x-\pi} (X_j + 2\pi - x)^2$$

$$= \sum_{j=1}^{n} (X_j - x)^2 + 4\pi \sum_{X_j < x-\pi} (X_j - x + \pi)$$

Revisiting "Typical Regularity Conditions" II

Recall conditions

(iii) for all random $\hat{x}_n \stackrel{\text{a.s.}}{\to} 0$, $\mathrm{Hess}|_{x=\hat{x}_n} F_n(x) \stackrel{\mathbb{P}}{\to} \mathrm{Hess}|_{x=0} F(x)$,

(iv) $\mathrm{Hess}|_{x=0} F(x) > 0$.

Consider (McKilliam et al. (2012), Hotz and H. 2015):

- $X_1, \ldots, X_n \stackrel{\text{i.i.d.}}{\sim} X \in \mathbb{S}^1 = [-\pi, \pi]/\sim$
- Fréchet means 0 (population), x_n (sample)
- f local density near $-\pi \cong \pi$, w.l.o.g. $x \geq 0$

$$2nF_n(x) = \sum_{x-\pi \leq X_j} (X_j - x)^2 + \sum_{X_j < x-\pi} (X_j + 2\pi - x)^2$$

$$= \sum_{j=1}^{n} (X_j - x)^2 + 4\pi \sum_{X_j < x-\pi} (X_j - x + \pi)$$

$\mathrm{Hess}|_x F_n(x) = 1$ a.s., $\mathrm{Hess}|_{x=0} F(x) = 1 - 2\pi f(-\pi)$.

Revisiting “Typical Regularity Conditions” II

Recall conditions

(iii) for all random $\hat{x}_n \stackrel{\text{a.s.}}{\to} 0$, $\mathrm{Hess}|_{x=\hat{x}_n} F_n(x) \stackrel{\mathbb{P}}{\to} \mathrm{Hess}|_{x=0} F(x)$,

(iv) $\mathrm{Hess}|_{x=0} F(x) > 0$.

Consider (McKilliam et al. (2012), Hotz and H. 2015):

- $X_1, \ldots, X_n \stackrel{\text{i.i.d.}}{\sim} X \in \mathbb{S}^1 = [-\pi, \pi]/\sim$
- Fréchet means 0 (population), x_n (sample)
- f local density near $-\pi \cong \pi$, w.l.o.g. $x \geq 0$

$$2nF_n(x) = \sum_{x-\pi \leq X_j} (X_j - x)^2 + \sum_{X_j < x-\pi} (X_j + 2\pi - x)^2$$

$$= \sum_{j=1}^{n} (X_j - x)^2 + 4\pi \sum_{X_j < x-\pi} (X_j - x + \pi)$$

$\mathrm{Hess}|_x F_n(x) = 1$ a.s., $\mathrm{Hess}|_{x=0} F(x) = 1 - 2\pi f(-\pi)$.
$f(-\pi) > 0$ possible!

Revisiting "Typical Regularity Conditions" II

Recall conditions

(iii) for all random $\hat{x}_n \stackrel{\text{a.s.}}{\to} 0$, $\mathrm{Hess}|_{x=\hat{x}_n} F_n(x) \stackrel{\mathbb{P}}{\to} \mathrm{Hess}|_{x=0} F(x)$,

(iv) $\mathrm{Hess}|_{x=0} F(x) > 0$.

Consider (McKilliam et al. (2012), Hotz and H. 2015):

- $X_1, \ldots, X_n \stackrel{\text{i.i.d.}}{\sim} X \in \mathbb{S}^1 = [-\pi, \pi]/\sim$
- Fréchet means 0 (population), x_n (sample)
- f local density near $-\pi \cong \pi$, w.l.o.g. $x \geq 0$

$$\begin{aligned} 2nF_n(x) &= \sum_{x-\pi \leq X_j} (X_j - x)^2 + \sum_{X_j < x-\pi} (X_j + 2\pi - x)^2 \\ &= \sum_{j=1}^{n} (X_j - x)^2 + 4\pi \sum_{X_j < x-\pi} (X_j - x + \pi) \end{aligned}$$

$\mathrm{Hess}|_x F_n(x) = 1$ a.s., $\mathrm{Hess}|_{x=0} F(x) = 1 - 2\pi f(-\pi)$.
$f(-\pi) > 0$ possible! Even $f(-\pi) = \frac{1}{2\pi}$ possible!

A Smeary CLT

- With smooth chart ϕ of P near unique population mean $p^* = \phi(0)$, ρ^2 smooth, Fréchet functions F, F_n,

A Smeary CLT

- With smooth chart ϕ of P near unique population mean $p^* = \phi(0)$, ρ^2 smooth, Fréchet functions F, F_n,
- Taylor with $2 \leq r$, $R \in SO(m)$ and $T_1, \dots, T_m \neq 0$,

$$F(x) = F(0) + \sum_{j=1}^{m} T_j |(Rx)_j|^r + o(\|x\|^r)\,,$$

A Smeary CLT

- With smooth chart ϕ of P near unique population mean $p^* = \phi(0)$, ρ^2 smooth, Fréchet functions F, F_n,
- Taylor with $2 \leq r$, $R \in SO(m)$ and $T_1, \dots, T_m \neq 0$,

$$F(x) = F(0) + \sum_{j=1}^{m} T_j |(Rx)_j|^r + o(\|x\|^r)\,,$$

- Donsker cond.: $\exists\ \dot\rho_0(X) := \mathrm{grad}_x \rho(X,x)|_{x=0}$ a.s., m'ble function $\dot\rho : Q \to \mathbb{R}$ such that $\mathbb{E}[\dot\rho(X)^2] < \infty$ and

$$|\rho(X,x_1) - \rho(X,x_2)| \leq \dot\rho(X)\|x_1 - x_2\| \,\mathrm{a.\,s.}\, \forall x_1, x_2 \in U,$$

A Smeary CLT

- With smooth chart ϕ of P near unique population mean $p^* = \phi(0)$, ρ^2 smooth, Fréchet functions F, F_n,
- Taylor with $2 \leq r$, $R \in SO(m)$ and $T_1, \dots, T_m \neq 0$,

$$F(x) = F(0) + \sum_{j=1}^{m} T_j |(Rx)_j|^r + o(\|x\|^r),$$

- Donsker cond.: $\exists\ \dot\rho_0(X) := \mathrm{grad}_x \rho(X, x)|_{x=0}$ a.s., m'ble function $\dot\rho : Q \to \mathbb{R}$ such that $\mathbb{E}[\dot\rho(X)^2] < \infty$ and

$$|\rho(X, x_1) - \rho(X, x_2)| \leq \dot\rho(X)\|x_1 - x_2\| \,\mathrm{a.\,s.}\ \forall x_1, x_2 \in U,$$

- if $\hat p_n \in E_n$ m'ble, use some van der Vaart (2000),

A Smeary CLT

- With smooth chart ϕ of P near unique population mean $p^* = \phi(0)$, ρ^2 smooth, Fréchet functions F, F_n,
- Taylor with $2 \leq r$, $R \in SO(m)$ and $T_1, \dots, T_m \neq 0$,

$$F(x) = F(0) + \sum_{j=1}^{m} T_j |(Rx)_j|^r + o(\|x\|^r)\,,$$

- Donsker cond.: $\exists\ \dot\rho_0(X) := \mathrm{grad}_x \rho(X, x)|_{x=0}$ a.s., m'ble function $\dot\rho : Q \to \mathbb{R}$ such that $\mathbb{E}[\dot\rho(X)^2] < \infty$ and

$$|\rho(X, x_1) - \rho(X, x_2)| \leq \dot\rho(X)\|x_1 - x_2\| \,\mathrm{a.\,s.}\, \forall x_1, x_2 \in U,$$

- if $\hat p_n \in E_n$ m'ble, use some van der Vaart (2000),

A Smeary CLT

- With smooth chart ϕ of P near unique population mean $p^* = \phi(0)$, ρ^2 smooth, Fréchet functions F, F_n,
- Taylor with $2 \le r$, $R \in SO(m)$ and $T_1, \dots, T_m \neq 0$,

$$F(x) = F(0) + \sum_{j=1}^{m} T_j |(Rx)_j|^r + o(\|x\|^r)\,,$$

- Donsker cond.: $\exists\ \dot\rho_0(X) := \mathrm{grad}_x \rho(X, x)|_{x=0}$ a.s., m'ble function $\dot\rho : Q \to \mathbb{R}$ such that $\mathbb{E}[\dot\rho(X)^2] < \infty$ and

$$|\rho(X, x_1) - \rho(X, x_2)| \le \dot\rho(X)\|x_1 - x_2\| \,\mathrm{a.\,s.}\ \forall x_1, x_2 \in U,$$

- if $\hat p_n \in E_n$ m'ble, use some van der Vaart (2000),

Theorem (Eltzner and H. 2018)

$\sqrt{n}\,\phi(\hat p_n)^r \stackrel{\mathcal{D}}{\to} \mathcal{N}(0, \Sigma)$ *(power component-wise), suitable* $\Sigma > 0$.

A Smeary CLT

- With smooth chart ϕ of P near unique population mean $p^* = \phi(0)$, ρ^2 smooth, Fréchet functions F, F_n,
- Taylor with $2 \leq r$, $R \in SO(m)$ and $T_1, \dots, T_m \neq 0$,

$$F(x) = F(0) + \sum_{j=1}^{m} T_j |(Rx)_j|^r + o(\|x\|^r)\,,$$

- Donsker cond.: $\exists\ \dot\rho_0(X) := \mathrm{grad}_x \rho(X,x)|_{x=0}$ a.s., m'ble function $\dot\rho : Q \to \mathbb{R}$ such that $\mathbb{E}[\dot\rho(X)^2] < \infty$ and

$$|\rho(X,x_1) - \rho(X,x_2)| \leq \dot\rho(X)\|x_1 - x_2\| \,\mathrm{a.\,s.}\ \forall x_1, x_2 \in U,$$

- if $\hat p_n \in E_n$ m'ble, use some van der Vaart (2000),

Theorem (Eltzner and H. 2018)

$\sqrt{n}\,\phi(\hat p_n)^r \stackrel{\mathcal{D}}{\to} \mathcal{N}(0,\Sigma)$ (power component-wise), suitable $\Sigma > 0$. $\phi(\hat p_n)$ has rate $n^{-\frac{1}{2(r-1)}}$, is $r-2$-smeary.

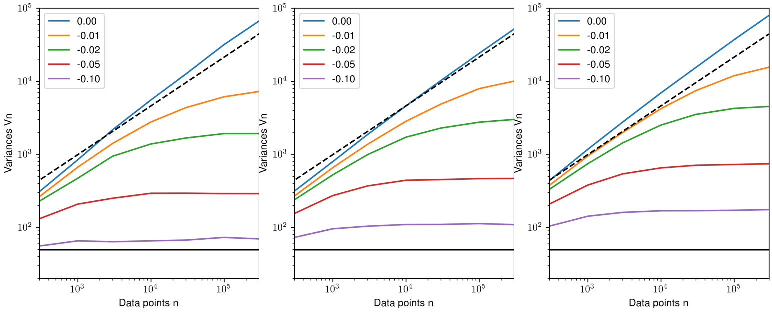
k-Smeariness

If

$$n^{\frac{1}{2(k+1)}}\left(\phi(\hat{p}_n)-\phi(p)\right)$$

has a non-trivial distribution as $n \to \infty$.

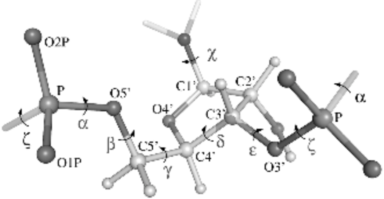
- $k = 2$ smeary (dashed line)

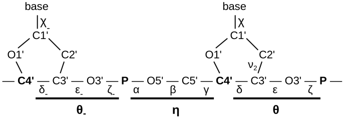


On a sphere $\mathbb{S}^m$ with dimension (all derivatives $O(m^{-1/2})$)

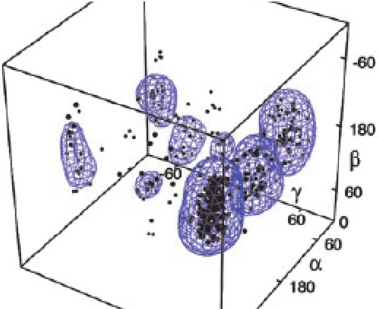
$m = 2$ $m = 10$ $m = 100$

Dimension Reduction in RNA Structure Analysis





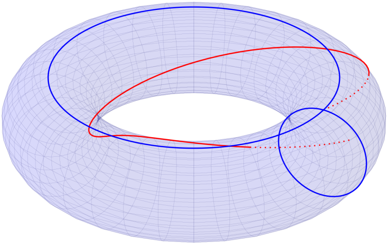
- 7 dihedral angles $\in (\mathbb{S}^1)^7$, 2 pseudotorsion angles $\in (\mathbb{S}^1)^2$,
- = shape, i.e. translational / rotational invariant

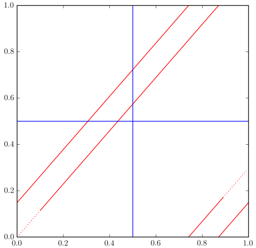


- Murray et al. (2003) using www.rscb.org:
- C2'-pucker RNA clusters in many 1D groups in heminucleotide angles.
- Can we verify (improve? understand?) by PCA?

PCA on a Torus $\mathbb{S}^1 \times \ldots \times \mathbb{S}^1$

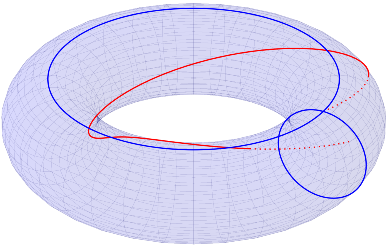
- Only very few geodesics are not winding around,
- an uncountable number of geodesics is dense and
- every data set can be perfectly approximated.
- Standard geometry of $(\mathbb{S}^1)^k$ is not statistically benign.

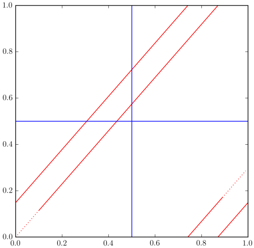




PCA on a Torus $\mathbb{S}^1 \times \ldots \times \mathbb{S}^1$

- Only very few geodesics are not winding around,
- an uncountable number of geodesics is dense and
- every data set can be perfectly approximated.
- Standard geometry of $(\mathbb{S}^1)^k$ is not statistically benign.





- Altis et al. (2008); Kent and Mardia (2009, 2015) allow only few geodesics.
- Tangent space PCA (Euclidean) for $(\mathbb{S}^1)^k$ "$\subset$" $\mathbb{R}^k$.
- Dihedral PCA Altis et al. (2008); Sargsyan et al. (2012) $(\mathbb{S}^1)^k \subset \mathbb{R}^{2k}$.

Euclidean vs. Spherical PCA

P_k = all "canonical" k-dim. subspaces in m-dim. Q.

$\dim(P_k)$

- $= \dim G(m,k) + \sharp$ translates
  $= (m-k)k + m - k = (m-k)(k+1)$ for $Q = \mathbb{R}^m$,
  canonically nested,

Euclidean vs. Spherical PCA

P_k = all "canonical" k-dim. subspaces in m-dim. Q.

$\dim(P_k)$

- $= \dim G(m,k) + \sharp$ translates
  $= (m-k)k + m - k = (m-k)(k+1)$ for $Q = \mathbb{R}^m$,
  canonically nested,
- $= \dim G(m+1,k+1) = (m-k)(k+1)$ for $Q = \mathbb{S}^m$,
  great subspheres, non-nested,

Euclidean vs. Spherical PCA

P_k = all “canonical” k-dim. subspaces in m-dim. Q.

$\dim(P_k)$

- $= \dim G(m,k) + \sharp$ translates
  $= (m-k)k + m - k = (m-k)(k+1)$ for $Q = \mathbb{R}^m$,
  canonically nested,
- $= \dim G(m+1,k+1) = (m-k)(k+1)$ for $Q = \mathbb{S}^m$,
  great subspheres, non-nested,
- $= \dim G(m+1,k+1) + (m-k) = (m-k)(k+2)$ for
  $Q = \mathbb{S}^m$, small subspheres, non-nested,
  statistically more benign than Euclidean PCA.

Euclidean vs. Spherical PCA

P_k = all "canonical" k-dim. subspaces in m-dim. Q.

$\dim(P_k)$

- $= \dim G(m,k) + \sharp$ translates
  $= (m-k)k + m - k = (m-k)(k+1)$ for $Q = \mathbb{R}^m$,
  canonically nested,
- $= \dim G(m+1,k+1) = (m-k)(k+1)$ for $Q = \mathbb{S}^m$,
  great subspheres, non-nested,
- $= \dim G(m+1,k+1) + (m-k) = (m-k)(k+2)$ for
  $Q = \mathbb{S}^m$, small subspheres, non-nested,
  statistically more benign than Euclidean PCA.

- make this nested $\to$ principal nested (great)subspheres
  (PN(G)S) by Jung et al. (2012).

Sausage Transformation

$(\mathbb{S}^1)^k \to \mathbb{S}^k$?

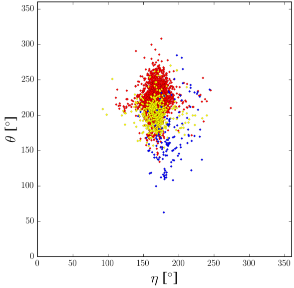
Data Driven Torus (T) PCA for $(\mathbb{S}^1)^k$

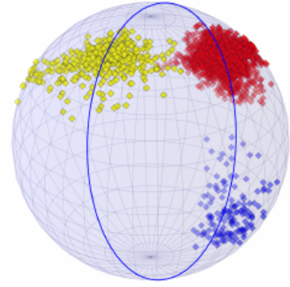
- Choose a codimension 2 subtorus furthest from data (opposite to mean, or largest gap) $\to \mathbb{S}^k / \sim$ glued along "that" $\mathbb{S}^{k-2}$,
- ideally, data near equatorial circle (EC) orthogonal (no deformation),
- center and number new angles by highest variance inside, or outside,

$$\sum_{l=1}^{k} d\psi_l^2 \to d\phi_1^2 + \sum_{l=2}^{k} \left(\prod_{j=1}^{l-1} \sin^2 \phi_j \right) d\phi_l^2 \,,$$

- halve all angles (but the last) – otherwise we obtain several copies of $\mathbb{S}^k / \sim$ glued together,
- do a variant of PNS (non-glued small subspheres, optimized by $\mathbb{S}^k / \sim$ distance).

Separation of Clusters by 7D Torus PCA

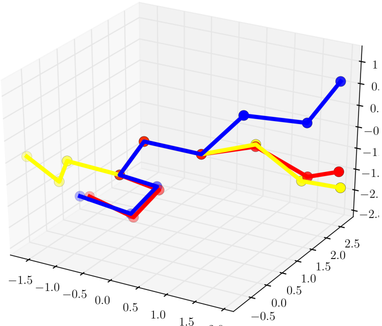




1: α-helix well known

2: helical-like less known

7: low-density new



Wrap UP: Challenges

- Stastically non-benign geometries:

Wrap UP: Challenges

- Stastically non-benign geometries:
    - Boldly change geometry,

Wrap UP: Challenges

- Stastically non-benign geometries:
  - Boldly change geometry,
  - works also for PCA on polyspheres: $\mathbb{S}^{k_1} \times \cdots \times \mathbb{S}^{k_r}$.

Wrap UP: Challenges

- Stastically non-benign geometries:
  - Boldly change geometry,
  - works also for PCA on polyspheres: $\mathbb{S}^{k_1} \times \cdots \times \mathbb{S}^{k_r}$.
- The classical BP-CLT misses the dirty CLTs

Wrap UP: Challenges

- Stastically non-benign geometries:
  - Boldly change geometry,
  - works also for PCA on polyspheres: $\mathbb{S}^{k_1} \times \cdots \times \mathbb{S}^{k_r}$.
- The classical BP-CLT misses the dirty CLTs
- Stickiness is a rather dead end for statistics on (phylogenetic) trees.

Wrap UP: Challenges

- Stastically non-benign geometries:
  - Boldly change geometry,
  - works also for PCA on polyspheres: $\mathbb{S}^{k_1} \times \cdots \times \mathbb{S}^{k_r}$.
- The classical BP-CLT misses the dirty CLTs
- Stickiness is a rather dead end for statistics on (phylogenetic) trees.
  - Challenge: Systematic treatment.

Wrap UP: Challenges

- Stastically non-benign geometries:
  - Boldly change geometry,
  - works also for PCA on polyspheres: $\mathbb{S}^{k_1} \times \cdots \times \mathbb{S}^{k_r}$.
- The classical BP-CLT misses the dirty CLTs
- Stickiness is a rather dead end for statistics on (phylogenetic) trees.
  - Challenge: Systematic treatment.
  - Try out different tropical geometry (Maclagan and Sturmfels (2015); Yoshida et al. (2017))?

Wrap UP: Challenges

- Stastically non-benign geometries:
  - Boldly change geometry,
  - works also for PCA on polyspheres: $\mathbb{S}^{k_1} \times \cdots \times \mathbb{S}^{k_r}$.
- The classical BP-CLT misses the dirty CLTs
- Stickiness is a rather dead end for statistics on (phylogenetic) trees.
  - Challenge: Systematic treatment.
  - Try out different tropical geometry (Maclagan and Sturmfels (2015); Yoshida et al. (2017))?
- Smeariness may give misleading asymptotics in high diemsion low sample size (HDLS).

Wrap UP: Challenges

- Stastically non-benign geometries:
  - Boldly change geometry,
  - works also for PCA on polyspheres: $\mathbb{S}^{k_1} \times \cdots \times \mathbb{S}^{k_r}$.
- The classical BP-CLT misses the dirty CLTs
- Stickiness is a rather dead end for statistics on (phylogenetic) trees.
  - Challenge: Systematic treatment.
  - Try out different tropical geometry (Maclagan and Sturmfels (2015); Yoshida et al. (2017))?
- Smeariness may give misleading asymptotics in high diemsion low sample size (HDLS).
  - Challenge: systematic treatment?

References

Altis, A., M. Otten, P. H. Nguyen, H. Rainer, and G. Stock (2008). Construction of the free energy landscape of biomolecules via dihedral angle principal component analysis. *The Journal of Chemical Physics 128*(24).

Anderson, T. (1963). Asymptotic theory for principal component analysis. *Ann. Math. Statist. 34*(1), 122–148.

Bhattacharya, R. N. and V. Patrangenaru (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds I. *The Annals of Statistics 31*(1), 1–29.

Bhattacharya, R. N. and V. Patrangenaru (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds II. *The Annals of Statistics 33*(3), 1225–1259.

Billera, L., S. Holmes, and K. Vogtmann (2001). Geometry of the space of phylogenetic trees. *Advances in Applied Mathematics 27*(4), 733–767.

Eltzner, B., S. F. Huckemann, and K. V. Mardia (2015). Deformed torus PCA with applications to RNA structure. arXiv:1511.04993.

Fletcher, P. T. and S. C. Joshi (2004). Principal geodesic analysis on symmetric spaces: Statistics of diffusion tensors. *ECCV Workshops CVAMIA and MMBIA*, 87–98.

Gower, J. C. (1975). Generalized Procrustes analysis. *Psychometrika 40*, 33–51.

Haeckel, E. (1879). *The Evolution of Man, vol 2*. ICON Group International.

Hendriks, H. and Z. Landsman (1996). Asymptotic behaviour of sample mean location for manifolds. *Statistics & Probability Letters 26*, 169–178.

Hotz, T. and S. Huckemann (2015). Intrinsic means on the circle: Uniqueness, locus and asymptotics. *Annals of the Institute of Statistical Mathematics 67*(1), 177–193.

Hotz, T., S. Huckemann, H. Le, J. S. Marron, J. Mattingly, E. Miller, J. Nolen, M. Owen, V. Patrangenaru, and S. Skwerer (2013). Sticky central limit theorems on open books. *Annals of Applied Probability 23*(6), 2238–2258.

Huckemann, S. (2011a). Inference on 3D Procrustes means: Tree boles growth, rank-deficient diffusion tensors and perturbation models. *Scandinavian Journal of Statistics 38*(3), 424–446.

Huckemann, S. (2011b). Intrinsic inference on the mean geodesic of planar shapes and tree discrimination by leaf growth. *The Annals of Statistics 39*(2), 1098–1124.

Huckemann, S., T. Hotz, and A. Munk (2010). Intrinsic shape analysis: Geodesic principal component analysis for Riemannian manifolds modulo Lie group actions (with discussion). *Statistica Sinica 20*(1), 1–100.

Huckemann, S., J. C. Mattingly, E. Miller, and J. Nolen (2015). Sticky central limit theorems at isolated hyperbolic planar singularities. *Electronic Journal of Probability 20*(78), 1–34.

Huckemann, S. F. and B. Eltzner (2017). Backward nested descriptors asymptotics with inference on stem cell differentiation. *AOS*. accepted, arXiv preprint arXiv:1609.00814.

Jung, S., I. L. Dryden, and J. S. Marron (2012). Analysis of principal nested spheres. *Biometrika 99*(3), 551–568.

References

Huckemann, S. (2011b). Intrinsic inference on the mean geodesic of planar shapes and tree discrimination by leaf growth. *The Annals of Statistics 39*(2), 1098–1124.

Huckemann, S., T. Hotz, and A. Munk (2010). Intrinsic shape analysis: Geodesic principal component analysis for Riemannian manifolds modulo Lie group actions (with discussion). *Statistica Sinica 20*(1), 1–100.

Huckemann, S., J. C. Mattingly, E. Miller, and J. Nolen (2015). Sticky central limit theorems at isolated hyperbolic planar singularities. *Electronic Journal of Probability 20*(78), 1–34.

Huckemann, S. F. and B. Eltzner (2017). Backward nested descriptors asymptotics with inference on stem cell differentiation. *AOS*. accepted, arXiv preprint arXiv:1609.00814.

Jung, S., I. L. Dryden, and J. S. Marron (2012). Analysis of principal nested spheres. *Biometrika 99*(3), 551–568.

Jupp, P. E. (1988). Residuals for directional data. *Journal of Applied Statistics 15*(2), 137–147.

Kent, J. T. and K. V. Mardia (2009). Principal component analysis for the wrapped normal torus model. *Proceedings of the Leeds Annual Statistical Research (LASR) Workshop 2009*.

Kent, J. T. and K. V. Mardia (2015). The winding number for circular data. *Proceedings of the Leeds Annual Statistical Research (LASR) Workshop 2015*.

Kobayashi, S. and K. Nomizu (1969). *Foundations of Differential Geometry*, Volume II. Chichester: Wiley.

Maclagan, D. and B. Sturmfels (2015). *Introduction to tropical geometry*, Volume 161. American mathematical society Providence, RI.

McKilliam, R. G., B. G. Quinn, and I. V. L. Clarkson (2012). Direction estimation by minimum squared arc length. *IEEE Transactions on Signal Processing 60*(5), 2115–2124.

Murray, L. J. W., W. B. I. Arendall, D. C. Richardson, and J. S. Richardson (2003). RNA backbone is rotameric. *Proc. Natl Acad. Sci. USA 100*(24), 13904–13909.

Nye, T. M., X. Tang, G. Weyenberg, and Y. Yoshida (2016). Principal component analysis and the locus of the frechet mean in the space of phylogenetic trees. *arXiv preprint arXiv:1609.03045*.

Panaretos, V. M., T. Pham, and Z. Yao (2014). Principal flows. *Journal of the American Statistical Association 109*(505), 424–436.

Pennec, X. (2017). Barycentric subspace analysis on manifolds. *The Annals of Statistics*. accepted, arXiv:1607.02833.

Sargsyan, K., J. Wright, and C. Lim (2012). GeoPCA: a new tool for multivariate analysis of dihedral angles based on principal component geodesics. *Nucleic Acids Research 40*(3), e25.

Sommer, S. (2016). Anisotropically weighted and nonholonomically constrained evolutions on manifolds. *Entropy 18*(12), 425.

Sturm, K. (2003). Probability measures on metric spaces of nonpositive curvature. *Contemporary mathematics 338*, 357–390.

References

Watson, G. (1983). *Statistics on Spheres*. University of Arkansas Lecture Notes in the Mathematical Sciences, Vol. 6. New York: Wiley.

Yoshida, Y., L. Zhang, and X. Zhang (2017). Tropical principal component analysis and its application to phylogenetics.

Ziezold, H. (1977). Expected figures and a strong law of large numbers for random elements in quasi-metric spaces. *Transaction of the 7th Prague Conference on Information Theory, Statistical Decision Function and Random Processes A*, 591–602.

Ziezold, H. (1994). Mean figures and mean shapes applied to biological figure and shape distributions in the plane. *Biometrical Journal* (36), 491–510.

van der Vaart, A. (2000). *Asymptotic statistics*. Cambridge Univ. Press.

Thank you!