On the Central Limit Theorem for Fréchet Means: Theory and Applications

Stephan F. Huckemann

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

Sept. 3, 2019
Geometric Statistics
Aug. 30 – Sept. 5, Toulouse

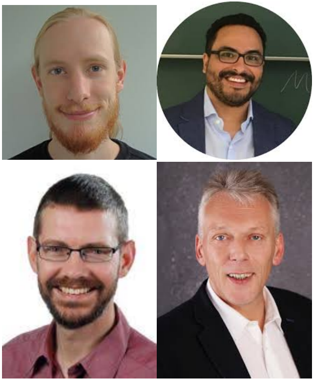
supported by the

Niedersachsen Vorab of the Volkswagen Foundation

Outline

People Having Contributed to this Talk

- Benjamin Eltzner (Univ. of Göttingen)
- Fernando Galaz-García (Univ. of Karlsruhe)
- Thomas Hotz (Univ. of Ilmenau)
- Wilderich Tuschmann (Univ. of Karlsruhe)



Motivation

- We have data $X_1, \ldots, X_n$ on manifolds or stratified spaces.

Motivation

- We have data $X_1, \ldots, X_n$ on manifolds or stratified spaces.
- We want to do inference: statistical testing,

Motivation

- We have data $X_1, \ldots, X_n$ on manifolds or stratified spaces.
- We want to do inference: statistical testing,
- controlling the error of the first kind,

$$\mathbb{P}\{\text{ accept } H_0 | H_0 \text{ is true}\} \geq 1 - \alpha\,,$$

Motivation

- We have data $X_1, \dots, X_n$ on manifolds or stratified spaces.
- We want to do inference: statistical testing,
- controlling the error of the first kind,

$$\mathbb{P}\{\text{ accept } H_0 | H_0 \text{ is true}\} \geq 1 - \alpha\,,$$

- asymptotically exact as $n \to \infty$,

Motivation

- We have data $X_1, \ldots, X_n$ on manifolds or stratified spaces.
- We want to do inference: statistical testing,
- controlling the error of the first kind,

$$\mathbb{P}\{\text{ accept } H_0 | H_0 \text{ is true}\} \geq 1 - \alpha\,,$$

- asymptotically exact as $n \to \infty$,
- or even exact for finite n ($\rightsquigarrow$ Thomas' talk at 15:30).

Motivation

- We have data $X_1, \ldots, X_n$ on manifolds or stratified spaces.
- We want to do inference: statistical testing,
- controlling the error of the first kind,

$$\mathbb{P}\{\text{ accept } H_0 | H_0 \text{ is true}\} \geq 1 - \alpha\,,$$

- asymptotically exact as $n \to \infty$,
- or even exact for finite n ($\rightsquigarrow$ Thomas' talk at 15:30).

Motivation

- We have data $X_1, \ldots, X_n$ on manifolds or stratified spaces.
- We want to do inference: statistical testing,
- controlling the error of the first kind,

$$\mathbb{P}\{\text{ accept } H_0 | H_0 \text{ is true}\} \geq 1 - \alpha\,,$$

- asymptotically exact as $n \to \infty$,
- or even exact for finite n ($\rightsquigarrow$ Thomas' talk at 15:30).

Here we do nonparametric asymptotics.

Euclidean Analog

Let i.i.d. $X, X_1, X_2, \ldots \in \mathbb{R}^D$ and $\bar{X}_n = \frac{X_1+\ldots+X_n}{n}$

Theorem (The Strong Law)

If $\mathbb{E}[X]$ exists then for $n \to \infty$

$$\bar{X}_n \to \mathbb{E}[X] \text{ a.s.}$$

Theorem (The Central Limit Theorem)

If $\mathbb{E}[\|X\|^2] < \infty$ then for $n \to \infty$

$$\sqrt{n}\left(\bar{X}_n - \mathbb{E}[X]\right) \stackrel{\mathcal{D}}{\to} \mathcal{N}(0, \mathrm{cov}[X])$$

Euclidean Analog

Let i.i.d. $X, X_1, X_2, \ldots \in \mathbb{R}^D$ and $\bar{X}_n = \frac{X_1+\ldots+X_n}{n}$

Theorem (The Strong Law)

If $\mathbb{E}[X]$ exists then for $n \to \infty$

$$\bar{X}_n \to \mathbb{E}[X] \text{ a.s.}$$

Theorem (The Central Limit Theorem)

If $\mathbb{E}[\|X\|^2] < \infty$ then for $n \to \infty$

$$\sqrt{n}\left(\bar{X}_n - \mathbb{E}[X]\right) \stackrel{\mathcal{D}}{\to} \mathcal{N}(0, \mathrm{cov}[X])$$

Test statistic for $\mathbb{E}[X]$: $\mathrm{cov}[X]^{-1/2}\sqrt{n}\left(\bar{X}_n - \mathbb{E}[X]\right) \stackrel{\mathcal{D}}{\to} \mathcal{N}(0, I)$

Euclidean Analog

Let i.i.d. $X, X_1, X_2, \ldots \in \mathbb{R}^D$ and $\bar{X}_n = \frac{X_1+\ldots+X_n}{n}$

Theorem (The Strong Law)

If $\mathbb{E}[X]$ exists then for $n \to \infty$

$$\bar{X}_n \to \mathbb{E}[X] \text{ a.s.}$$

Theorem (The Central Limit Theorem)

If $\mathbb{E}[\|X\|^2] < \infty$ then for $n \to \infty$

$$\sqrt{n}\left(\bar{X}_n - \mathbb{E}[X]\right) \stackrel{\mathcal{D}}{\to} \mathcal{N}(0, \mathrm{cov}[X])$$

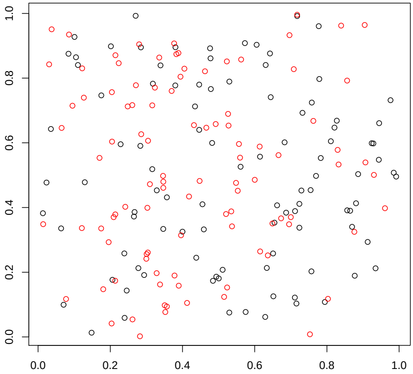
Test statistic for $\mathbb{E}[X]$: $\mathrm{cov}[X]^{-1/2}\sqrt{n}\left(\bar{X}_n - \mathbb{E}[X]\right) \stackrel{\mathcal{D}}{\to} \mathcal{N}(0, I)$

plugging in $\Sigma_n^X = \frac{1}{n-1}\sum_{j=1}^n (X_j - \bar{X}_n)(X_j - \bar{X}_n)^T$ for $\mathrm{cov}[X]$.

Test for Equality of Means

Two groups of random variables

$$X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} X \in \mathbb{R}^D \quad Y_1, \ldots, Y_m \overset{\text{i.i.d.}}{\sim} Y \in \mathbb{R}^D$$



Test H_0: $\mathbb{E}[X] = \mathbb{E}[Y]$

Hotelling Test for Equality of Means

- Under H_0 and either $\mathrm{cov}[X] = \mathrm{cov}[Y]$ or $n/m \to 1$,

$$T^2 := \frac{n+m-2}{\frac{1}{n}+\frac{1}{m}}(\bar{X}_n - \bar{Y}_m)^T (n\Sigma_n^X + m\Sigma_m^Y)^{-1}(\bar{X}_n - \bar{Y}_m)$$

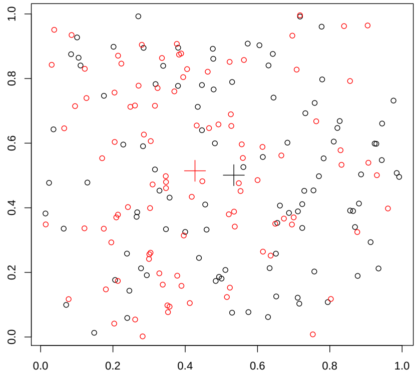
$\stackrel{\mathcal{D}}{\to}$ explicitly known limit $(n, m \to \infty,\ 0 < \lim n/m < \infty)$

Hotelling Test for Equality of Means

- Under H_0 and either $\mathrm{cov}[X] = \mathrm{cov}[Y]$ or $n/m \to 1$,

$$T^2 := \frac{n+m-2}{\frac{1}{n}+\frac{1}{m}}(\bar{X}_n - \bar{Y}_m)^T(n\Sigma_n^X + m\Sigma_m^Y)^{-1}(\bar{X}_n - \bar{Y}_m)$$

$\stackrel{\mathcal{D}}{\to}$ explicitly known limit ($n, m \to \infty$, $0 < \lim n/m < \infty$)

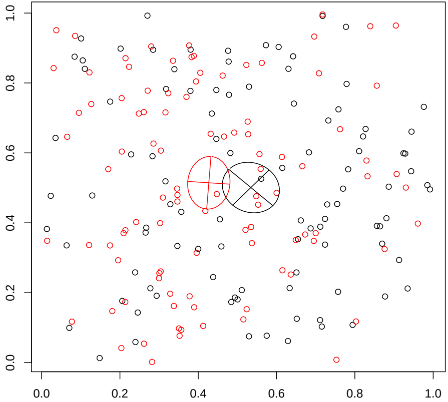


Reject H_0 with significance ($\alpha = 0.05$), not highly ($\alpha = 0.01$).

Principal Component Analysis (PCA)

Spectral decomposition $\mathrm{cov}[X] = \Gamma\Lambda\Gamma^T$.

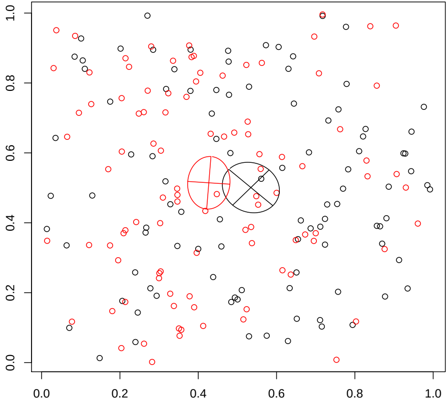
- With eigenvectors $\Gamma = (\gamma_1, \ldots, \gamma_m) \in SO(m)$ to



Principal Component Analysis (PCA)

Spectral decomposition $\operatorname{cov}[X] = \Gamma\Lambda\Gamma^T$.

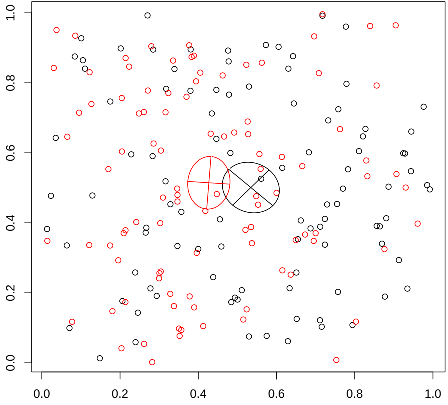
- With eigenvectors $\Gamma = (\gamma_1, \ldots, \gamma_m) \in SO(m)$ to
- eigenvalues $\lambda_1 \geq \ldots \geq \lambda_m \geq 0$, $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_m)$



Principal Component Analysis (PCA)

Spectral decomposition $\mathrm{cov}[X] = \Gamma\Lambda\Gamma^T$.

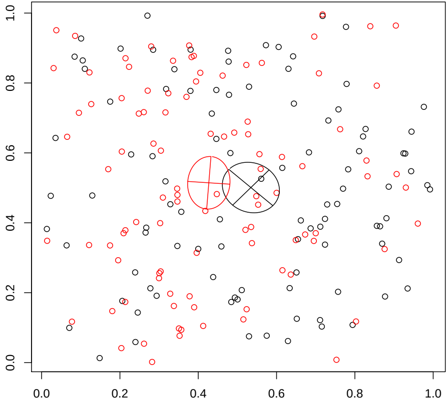
- With eigenvectors $\Gamma = (\gamma_1, \ldots, \gamma_m) \in SO(m)$ to
- eigenvalues $\lambda_1 \geq \ldots \geq \lambda_m \geq 0$, $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_m)$
- giving main modes of variation $\to$ dimension reduction.



Principal Component Analysis (PCA)

Spectral decomposition $\mathrm{cov}[X] = \Gamma\Lambda\Gamma^T$.

- With eigenvectors $\Gamma = (\gamma_1, \dots, \gamma_m) \in SO(m)$ to
- eigenvalues $\lambda_1 \geq \dots \geq \lambda_m \geq 0$, $\Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_m)$
- giving main modes of variation $\to$ dimension reduction.
- Test for PCs γ_j ? Note, $\gamma_j \in \mathbb{S}^{m-1}$. Actually in $\mathbb{R}P^{m-1}$.



The Bhattacharya and Patrangenaru (2005) CLT

Data $X_1, \ldots, X_n \stackrel{\text{i.i.d.}}{\sim} X$ on a Riemannian D-manifold (M, ρ).

The Bhattacharya and Patrangenaru (2005) CLT

Data $X_1, \ldots, X_n \stackrel{\text{i.i.d.}}{\sim} X$ on a Riemannian D-manifold (M, ρ).
Fréchet functions

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

The Bhattacharya and Patrangenaru (2005) CLT

Data $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} X$ on a Riemannian D-manifold (M, ρ).

Fréchet functions

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

Assumptions:

(A1) unique Fréchet mean $\mu \in \operatorname{argmin}_{p\in M} F(p)$ (difficult: Karcher (1977); Kendall (1990); Le (1998); Groisser (2005); Afsari (2011), not covered here),

The Bhattacharya and Patrangenaru (2005) CLT

Data $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} X$ on a Riemannian D-manifold (M, ρ).

Fréchet functions

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

Assumptions:

(A1) unique Fréchet mean $\mu \in \operatorname{argmin}_{p \in M} F(p)$ (difficult: Karcher (1977); Kendall (1990); Le (1998); Groisser (2005); Afsari (2011), not covered here),

(A2) in a local chart (U, ϕ), $\mu \in U \subseteq M$, $\phi^{-1}(U) = V \subseteq \mathbb{R}^D$,

$$x \mapsto \rho\big(X, \phi(x)\big)^2 \text{ a.s.} \in \mathcal{C}^2(V)\,,$$

The Bhattacharya and Patrangenaru (2005) CLT

Data $X_1, \ldots, X_n \stackrel{\text{i.i.d.}}{\sim} X$ on a Riemannian D-manifold (M, ρ).
Fréchet functions

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

Assumptions:

(A1) unique Fréchet mean $\mu \in \operatorname{argmin}_{p \in M} F(p)$ (difficult: Karcher (1977); Kendall (1990); Le (1998); Groisser (2005); Afsari (2011), not covered here),

(A2) in a local chart (U, ϕ), $\mu \in U \subseteq M$, $\phi^{-1}(U) = V \subseteq \mathbb{R}^D$,

$$x \mapsto \rho\big(X, \phi(x)\big)^2 \text{ a. s.} \in \mathcal{C}^2(V)\,,$$

(A3) $\mu_n \stackrel{\mathbb{P}}{\to} \mu$ for a measurable selection of sample means

$$\mu_n \in \operatorname*{argmin}_{p \in M} F_n(p)$$

(guaranteed by Ziezold (1977); Bhattacharya and Patrangenaru (2003) under very general conditions).

The Bhattacharya and Patrangenaru (2005) CLT

More assumptions:

(A4) $\exists G := \operatorname{cov}\left[\operatorname{grad}|_{x=\phi^{-1}(\mu)}\rho^2\big(X,\phi(x)\big)\right]$,

$\exists H := \mathbb{E}\left[H\big(X,\phi^{-1}(\mu)\big)\right]$, $H(X,x) = \operatorname{Hess}|_x\rho^2\big(X,\phi(x)\big)$

(we cannot do without, e.g. valid on compact M)

The Bhattacharya and Patrangenaru (2005) CLT

More assumptions:

(A4) $\exists G := \mathrm{cov}\left[\mathrm{grad}|_{x=\phi^{-1}(\mu)}\rho^2\big(X,\phi(x)\big)\right]$,

$\exists H := \mathbb{E}\left[H\big(X,\phi^{-1}(\mu)\big)\right]$, $H(X,x) = \mathrm{Hess}|_x\rho^2\big(X,\phi(x)\big)$

(we cannot do without, e.g. valid on compact M)

(A5) as $\epsilon \to 0$,

$$\mathbb{E}\left[\sup_{x=\phi^{-1}(\mu),\|x-x'\|<\epsilon}\big|H(X,x) - H(X,x')\big|\right] \to 0$$

The Bhattacharya and Patrangenaru (2005) CLT

More assumptions:

(A4) $\exists G := \operatorname{cov}\left[\operatorname{grad}|_{x=\phi^{-1}(\mu)}\rho^2\big(X,\phi(x)\big)\right]$,

$\exists H := \mathbb{E}\left[H\big(X,\phi^{-1}(\mu)\big)\right]$, $H(X,x) = \operatorname{Hess}|_x\rho^2\big(X,\phi(x)\big)$

(we cannot do without, e.g. valid on compact M)

(A5) as $\epsilon \to 0$,

$$\mathbb{E}\left[\sup_{x=\phi^{-1}(\mu),\|x-x'\|<\epsilon}\big|H(X,x) - H(X,x')\big|\right] \to 0$$

(A6) H is not singular.

The Bhattacharya and Patrangenaru (2005) CLT

More assumptions:

(A4) $\exists G := \mathrm{cov}\left[\mathrm{grad}|_{x=\phi^{-1}(\mu)}\rho^2\big(X,\phi(x)\big)\right]$,

$\exists H := \mathbb{E}\left[H\big(X,\phi^{-1}(\mu)\big)\right]$, $H(X,x) = \mathrm{Hess}|_x\rho^2\big(X,\phi(x)\big)$

(we cannot do without, e.g. valid on compact M)

(A5) as $\epsilon \to 0$,

$$\mathbb{E}\left[\sup_{x=\phi^{-1}(\mu),\|x-x'\|<\epsilon}\big|H(X,x)-H(X,x')\big|\right] \to 0$$

(A6) H is not singular.

The Bhattacharya and Patrangenaru (2005) CLT

More assumptions:

(A4) $\exists G := \mathrm{cov}\left[\mathrm{grad}|_{x=\phi^{-1}(\mu)}\rho^2\big(X,\phi(x)\big)\right]$,
$\exists H := \mathbb{E}\left[H\big(X,\phi^{-1}(\mu)\big)\right]$, $H(X,x) = \mathrm{Hess}|_x\rho^2\big(X,\phi(x)\big)$
(we cannot do without, e.g. valid on compact M)

(A5) as $\epsilon \to 0$,

$$\mathbb{E}\left[\sup_{x=\phi^{-1}(\mu),\|x-x'\|<\epsilon}\Big|H(X,x) - H(X,x')\Big|\right] \to 0$$

(A6) H is not singular.

Theorem (Bhattacharya and Patrangenaru (2005); Bhattacharya and Lin (2017))

Under Assumptions (A1) — (A6):

$$\sqrt{n}\left(\phi^{-1}(\mu_n) - \phi^{-1}(\mu)\right) \stackrel{\mathcal{D}}{\to} \mathcal{N}\left(0, H^{-1}GH^{-1}\right) .$$

Sketch of Proof

- W.l.o.g $\phi^{-1}(\mu) = 0$, $\phi^{-1}(\mu_n) = x_n$.

Sketch of Proof

- W.l.o.g $\phi^{-1}(\mu) = 0$, $\phi^{-1}(\mu_n) = x_n$.
- SLLN by Ziezold (1977); Bhattacharya and Patrangenaru (2003): $x_n \stackrel{a.s.}{\to} 0$.

Sketch of Proof

- W.l.o.g $\phi^{-1}(\mu) = 0$, $\phi^{-1}(\mu_n) = x_n$.
- SLLN by Ziezold (1977); Bhattacharya and Patrangenaru (2003): $x_n \stackrel{a.s.}{\to} 0$.
- Fréchet functions:

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

Sketch of Proof

- W.l.o.g $\phi^{-1}(\mu) = 0$, $\phi^{-1}(\mu_n) = x_n$.
- SLLN by Ziezold (1977); Bhattacharya and Patrangenaru (2003): $x_n \stackrel{\text{a.s.}}{\to} 0$.
- Fréchet functions:

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

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

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

(A2) $\Rightarrow$ holds also a.s. for random $x_0 = x_n$

Sketch of Proof

- W.l.o.g $\phi^{-1}(\mu) = 0$, $\phi^{-1}(\mu_n) = x_n$.
- SLLN by Ziezold (1977); Bhattacharya and Patrangenaru (2003): $x_n \stackrel{\text{a.s.}}{\to} 0$.
- Fréchet functions:

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

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

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

(A2) $\Rightarrow$ holds also a.s. for random $x_0 = x_n$

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

$$\mathrm{Hess}|_{x=\widetilde{x}} F_n(x) \stackrel{\mathbb{P}}{\to} \mathbb{E}\left[\mathrm{Hess}|_{x=0}\rho(X, x)^2\right] = H\,,$$

Sketch of Proof

- W.l.o.g $\phi^{-1}(\mu) = 0$, $\phi^{-1}(\mu_n) = x_n$.
- SLLN by Ziezold (1977); Bhattacharya and Patrangenaru (2003): $x_n \stackrel{a.s.}{\to} 0$.
- Fréchet functions:

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

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

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

  (A2) $\Rightarrow$ holds also a.s. for random $x_0 = x_n$
- generalized weak law ($n \to \infty$ and $x_0 \to 0$)

$$\mathrm{Hess}|_{x=\widetilde{x}} F_n(x) \stackrel{\mathbb{P}}{\to} \mathbb{E}\left[\mathrm{Hess}|_{x=0}\rho(X,x)^2\right] = H\,,$$

  (A5) $\Rightarrow$ holds also for random $x_0 = x_n$, and
  (A6) $\Rightarrow$ $\mathbb{E}\left[\mathrm{Hess}|_{x=0}\rho(X,x)^2\right] > 0$

Sketch of Proof

- W.l.o.g $\phi^{-1}(\mu) = 0$, $\phi^{-1}(\mu_n) = x_n$.
- SLLN by Ziezold (1977); Bhattacharya and Patrangenaru (2003): $x_n \stackrel{\text{a.s.}}{\to} 0$.
- Fréchet functions:

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

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

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

  (A2) $\Rightarrow$ holds also a.s. for random $x_0 = x_n$
- generalized weak law ($n \to \infty$ and $x_0 \to 0$)

$$\mathrm{Hess}|_{x=\widetilde{x}} F_n(x) \stackrel{\mathbb{P}}{\to} \mathbb{E}\left[\mathrm{Hess}|_{x=0}\rho(X,x)^2\right] = H\,,$$

  (A5) $\Rightarrow$ holds also for random $x_0 = x_n$, and
  (A6) $\Rightarrow \mathbb{E}\left[\mathrm{Hess}|_{x=0}\rho(X,x)^2\right] > 0$
  $\Rightarrow$ BP-CLT.

(A2) Dissected: The Cut Locus

Corollary (2.3 from Bhattacharya and Lin (2017))

Instead of

(A2) *in a local chart* (U, ϕ), $\mu \in U \subseteq M$, $\phi^{-1}(U) = V \subseteq \mathbb{R}^D$,

$$x \mapsto \rho\big(X, \phi(x)\big)^2 \text{ is a.s.} \in \mathcal{C}^2(V)$$

it suffices to require

(C) *there is a neighborhood* $W \subseteq M$ *of the cut locus* $\mathrm{Cut}(\mu)$ *of* μ *such that* $\mathbb{P}\{X \in W\} = 0$.

(A2) Dissected: The Cut Locus

Corollary (2.3 from Bhattacharya and Lin (2017))

Instead of

(A2) *in a local chart* (U, ϕ), $\mu \in U \subseteq M$, $\phi^{-1}(U) = V \subseteq \mathbb{R}^D$,

$$x \mapsto \rho\big(X, \phi(x)\big)^2 \text{ is a.s.} \in \mathcal{C}^2(V)$$

it suffices to require

(C) *there is a neighborhood* $W \subseteq M$ *of the cut locus* $\mathrm{Cut}(\mu)$ *of* μ *such that* $\mathbb{P}\{X \in W\} = 0$.

This is problematic, because

Example (Eltzner et al. (2019))

On the flat cylinder $M = \mathbb{S}^1 \times \mathbb{R}$ *there is a r.v.* X *that satisfies (C) but not (A2).*

(A2) Dissected: The Cut Locus

Corollary (2.3 from Bhattacharya and Lin (2017))

Instead of

(A2) *in a local chart* (U,ϕ), $\mu \in U \subseteq M$, $\phi^{-1}(U) = V \subseteq \mathbb{R}^D$,

$$x \mapsto \rho\big(X, \phi(x)\big)^2 \textit{ is } \mathrm{a.s.} \in \mathcal{C}^2(V)$$

it suffices to require

(C) *there is a neighborhood* $W \subseteq M$ *of the cut locus* $\mathrm{Cut}(\mu)$ *of* μ *such that* $\mathbb{P}\{X \in W\} = 0$.

This is problematic, because

Example (Eltzner et al. (2019))

On the flat cylinder $M = \mathbb{S}^1 \times \mathbb{R}$ *there is a r.v.* X *that satisfies (C) but not (A2).*

Theorem (Le and Barden (2014))

$\mathbb{P}\{X \in \mathrm{Cut}(\mu)\} = 0$.

Stability of the Cut Locus

Let M be a complete, connected Riemannian D-manifold. We say that (the cut loci of) M is (are)

topologically stable if $\forall\, p \in M$, neighborhoods W of $\mathrm{Cut}(p)$, $\exists \delta = \delta_{W,p}$ such that $\mathrm{Cut}\big(B(p,\delta)\big) \subseteq W$;

geometrically stable if $\forall\, p \in M$, $\epsilon > 0$, $\exists \delta = \delta_{\epsilon,p}$ such that $\mathrm{Cut}\big(B(p,\delta)\big) \subseteq B\big(\mathrm{Cut}(p),\epsilon\big)$.

Stability of the Cut Locus

Let M be a complete, connected Riemannian D-manifold. We say that (the cut loci of) M is (are)

> **topologically stable** if $\forall\, p \in M$, neighborhoods W of $\mathrm{Cut}(p)$, $\exists \delta = \delta_{W,p}$ such that $\mathrm{Cut}\big(B(p,\delta)\big) \subseteq W$;
> **geometrically stable** if $\forall\, p \in M$, $\epsilon > 0$, $\exists \delta = \delta_{\epsilon,p}$ such that $\mathrm{Cut}\big(B(p,\delta)\big) \subseteq B\big(\mathrm{Cut}(p),\epsilon\big)$.

Theorem (Eltzner et al. (2019))

1. *M topologically stable $\Rightarrow$ M geometrically stable;*

Stability of the Cut Locus

Let M be a complete, connected Riemannian D-manifold. We say that (the cut loci of) M is (are)

topologically stable if $\forall\, p \in M$, neighborhoods W of $\mathrm{Cut}(p)$, $\exists \delta = \delta_{W,p}$ such that $\mathrm{Cut}\big(B(p,\delta)\big) \subseteq W$;

geometrically stable if $\forall\, p \in M$, $\epsilon > 0$, $\exists \delta = \delta_{\epsilon,p}$ such that $\mathrm{Cut}\big(B(p,\delta)\big) \subseteq B\big(\mathrm{Cut}(p),\epsilon\big)$.

Theorem (Eltzner et al. (2019))

1. *M topologically stable $\Rightarrow$ M geometrically stable;*
2. *M compact $\Rightarrow$ M topologically stable;*

Stability of the Cut Locus

Let M be a complete, connected Riemannian D-manifold. We say that (the cut loci of) M is (are)

> topologically stable if $\forall\, p \in M$, neighborhoods W of $\mathrm{Cut}(p)$, $\exists \delta = \delta_{W,p}$ such that $\mathrm{Cut}\big(B(p,\delta)\big) \subseteq W$;
> geometrically stable if $\forall\, p \in M$, $\epsilon > 0$, $\exists \delta = \delta_{\epsilon,p}$ such that $\mathrm{Cut}\big(B(p,\delta)\big) \subseteq B\big(\mathrm{Cut}(p),\epsilon\big)$.

Theorem (Eltzner et al. (2019))

1. *M topologically stable $\Rightarrow$ M geometrically stable;*
2. *M compact $\Rightarrow$ M topologically stable;*
3. *M topologically stable and (C) $\Rightarrow$ (A2);*

Stability of the Cut Locus

Let M be a complete, connected Riemannian D-manifold. We say that (the cut loci of) M is (are)

> topologically stable if $\forall\, p \in M$, neighborhoods W of $\mathrm{Cut}(p)$, $\exists \delta = \delta_{W,p}$ such that $\mathrm{Cut}\big(B(p,\delta)\big) \subseteq W$;
>
> geometrically stable if $\forall\, p \in M$, $\epsilon > 0$, $\exists \delta = \delta_{\epsilon,p}$ such that $\mathrm{Cut}\big(B(p,\delta)\big) \subseteq B\big(\mathrm{Cut}(p),\epsilon\big)$.

Theorem (Eltzner et al. (2019))

1. *M topologically stable $\Rightarrow$ M geometrically stable;*
2. *M compact $\Rightarrow$ M topologically stable;*
3. *M topologically stable and (C) $\Rightarrow$ (A2);*
4. *M topologically stable $\Rightarrow$ Bhattacharya and Lin (2017, Cor. 2.3) holds.*

Stability of the Cut Locus

Let M be a complete, connected Riemannian D-manifold. We say that (the cut loci of) M is (are)

topologically stable if $\forall\, p \in M$, neighborhoods W of $\mathrm{Cut}(p)$, $\exists \delta = \delta_{W,p}$ such that $\mathrm{Cut}\big(B(p,\delta)\big) \subseteq W$;

geometrically stable if $\forall\, p \in M$, $\epsilon > 0$, $\exists \delta = \delta_{\epsilon,p}$ such that $\mathrm{Cut}\big(B(p,\delta)\big) \subseteq B\big(\mathrm{Cut}(p),\epsilon\big)$.

Theorem (Eltzner et al. (2019))

1. *M topologically stable $\Rightarrow$ M geometrically stable;*
2. *M compact $\Rightarrow$ M topologically stable;*
3. *M topologically stable and (C) $\Rightarrow$ (A2);*
4. *M topologically stable $\Rightarrow$ Bhattacharya and Lin (2017, Cor. 2.3) holds.*

Stability of the Cut Locus

Let M be a complete, connected Riemannian D-manifold. We say that (the cut loci of) M is (are)

> topologically stable if $\forall\, p \in M$, neighborhoods W of $\mathrm{Cut}(p)$, $\exists \delta = \delta_{W,p}$ such that $\mathrm{Cut}\big(B(p,\delta)\big) \subseteq W$;
>
> geometrically stable if $\forall\, p \in M$, $\epsilon > 0$, $\exists \delta = \delta_{\epsilon,p}$ such that $\mathrm{Cut}\big(B(p,\delta)\big) \subseteq B\big(\mathrm{Cut}(p),\epsilon\big)$.

Theorem (Eltzner et al. (2019))

1. *M topologically stable $\Rightarrow$ M geometrically stable;*
2. *M compact $\Rightarrow$ M topologically stable;*
3. *M topologically stable and (C) $\Rightarrow$ (A2);*
4. *M topologically stable $\Rightarrow$ Bhattacharya and Lin (2017, Cor. 2.3) holds.*

Example (Eltzner et al. (2019))

1. The flat cylinder $M = \mathbb{S}^1 \times \mathbb{R}$ is metrically stable;
2. The Beltrami trumpet (pseudosphere) is not metrically stable.

What Else Can Go Wrong?

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. $x \geq 0$

What Else Can Go Wrong?

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

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

$$2nF_n(x) = \sum_{X_j \in [x-\pi, \pi]} (X_j - x)^2 + \sum_{X_j < x-\pi} (X_j + 2\pi - x)^2$$

What Else Can Go Wrong?

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. $x \geq 0$

$$2nF_n(x) = \sum_{X_j \in [x-\pi, \pi]} (X_j - x)^2 + \sum_{X_j < x-\pi} (X_j + 2\pi - x)^2$$

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

What Else Can Go Wrong?

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

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

$$2nF_n(x) = \sum_{X_j \in [x-\pi, \pi]} (X_j - x)^2 + \sum_{X_j < x-\pi} (X_j + 2\pi - x)^2$$

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

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

What Else Can Go Wrong?

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

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

$$2nF_n(x) = \sum_{X_j \in [x-\pi, \pi]} (X_j - x)^2 + \sum_{X_j < x-\pi} (X_j + 2\pi - x)^2$$

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

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

$f(-\pi) > 0$ possible, $\lightning$(A5)

What Else Can Go Wrong?

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

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

$$2nF_n(x) = \sum_{X_j \in [x-\pi, \pi]} (X_j - x)^2 + \sum_{X_j < x-\pi} (X_j + 2\pi - x)^2$$

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

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

$f(-\pi) > 0$ possible, $\lightning$(A5)

Even $f(-\pi) = \frac{1}{2\pi}$ possible, $\lightning$(A6)

A More General CLT

- With unique (A1) population mean $\mu = \phi(0)$, measurable selection $\mu_n = \phi(x_n)$ of sample means,

A More General CLT

- With unique (A1) population mean $\mu = \phi(0)$, measurable selection $\mu_n = \phi(x_n)$ of sample means,
- Taylor with $2 \leq r$, $R \in SO(m)$ and $T_1, \ldots, T_m \neq 0$,

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

A More General CLT

- With unique (A1) population mean $\mu = \phi(0)$, measurable selection $\mu_n = \phi(x_n)$ of sample means,
- Taylor with $2 \leq r$, $R \in SO(m)$ and $T_1, \ldots, T_m \neq 0$,

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

- Donsker cond.: $\exists\ \dot\rho_0(X) := \mathrm{grad}_x \rho(X, \phi(x))^2|_{x=0}$ a.s. with $\exists\,\mathrm{cov}[\dot\rho_0(X)]$, m'ble function $\dot\rho : M \to \mathbb{R}$ such that $\mathbb{E}[\dot\rho(X)^2] < \infty$ and $\forall x_1, x_2 \in V$,

$$|\rho(X, \phi(x_1))^2 - \rho(X, \phi(x_2))^2| \leq \dot\rho(X)\|x_1 - x_2\|\ \mathrm{a.\,s.},$$

A More General CLT

- With unique (A1) population mean $\mu = \phi(0)$, measurable selection $\mu_n = \phi(x_n)$ of sample means,
- Taylor with $2 \leq r$, $R \in SO(m)$ and $T_1, \ldots, T_m \neq 0$,

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

- Donsker cond.: $\exists\ \dot\rho_0(X) := \mathrm{grad}_x \rho(X, \phi(x))^2|_{x=0}$ a.s. with $\exists\,\mathrm{cov}[\dot\rho_0(X)]$, m'ble function $\dot\rho : M \to \mathbb{R}$ such that $\mathbb{E}[\dot\rho(X)^2] < \infty$ and $\forall x_1, x_2 \in V$,

$$|\rho(X, \phi(x_1))^2 - \rho(X, \phi(x_2))^2| \leq \dot\rho(X)\|x_1 - x_2\|\ \mathrm{a.\,s.},$$

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

A More General CLT

- With unique (A1) population mean $\mu = \phi(0)$, measurable selection $\mu_n = \phi(x_n)$ of sample means,
- Taylor with $2 \leq r$, $R \in SO(m)$ and $T_1, \ldots, T_m \neq 0$,

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

- Donsker cond.: $\exists\ \dot\rho_0(X) := \mathrm{grad}_x \rho(X, \phi(x))^2|_{x=0}$ a.s. with $\exists\,\mathrm{cov}[\dot\rho_0(X)]$, m'ble function $\dot\rho : M \to \mathbb{R}$ such that $\mathbb{E}[\dot\rho(X)^2] < \infty$ and $\forall x_1, x_2 \in V$,

$$|\rho(X, \phi(x_1))^2 - \rho(X, \phi(x_2))^2| \leq \dot\rho(X)\|x_1 - x_2\|\ \mathrm{a.\,s.},$$

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

A More General CLT

- With unique (A1) population mean $\mu = \phi(0)$, measurable selection $\mu_n = \phi(x_n)$ of sample means,
- Taylor with $2 \leq r$, $R \in SO(m)$ and $T_1, \dots, T_m \neq 0$,

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

- Donsker cond.: $\exists\ \dot\rho_0(X) := \mathrm{grad}_x \rho(X, \phi(x))^2|_{x=0}$ a.s. with $\exists\,\mathrm{cov}[\dot\rho_0(X)]$, m'ble function $\dot\rho : M \to \mathbb{R}$ such that $\mathbb{E}[\dot\rho(X)^2] < \infty$ and $\forall x_1, x_2 \in V$,

$$|\rho(X, \phi(x_1))^2 - \rho(X, \phi(x_2))^2| \leq \dot\rho(X)\|x_1 - x_2\|\ \mathrm{a.\,s.},$$

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

Theorem (Eltzner and H. 2018)

$\sqrt{n}\, Rx_n |Rx_n|^{r-2} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma)$ (power component-wise), suitable $\Sigma > 0$.

A More General CLT

- With unique (A1) population mean $\mu = \phi(0)$, measurable selection $\mu_n = \phi(x_n)$ of sample means,
- Taylor with $2 \le r$, $R \in SO(m)$ and $T_1, \dots, T_m \neq 0$,

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

- Donsker cond.: $\exists\ \dot\rho_0(X) := \mathrm{grad}_x \rho(X, \phi(x))^2|_{x=0}$ a.s. with $\exists\,\mathrm{cov}[\dot\rho_0(X)]$, m'ble function $\dot\rho : M \to \mathbb{R}$ such that $\mathbb{E}[\dot\rho(X)^2] < \infty$ and $\forall x_1, x_2 \in V$,

$$|\rho(X,\phi(x_1))^2 - \rho(X,\phi(x_2))^2| \le \dot\rho(X)\|x_1 - x_2\|\ \mathrm{a.\,s.},$$

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

Theorem (Eltzner and H. 2018)

$\sqrt{n}\, Rx_n |Rx_n|^{r-2} \xrightarrow{\mathcal{D}} \mathcal{N}(0,\Sigma)$ (power component-wise), suitable $\Sigma > 0$. x_n has rate $n^{-\frac{1}{2(r-1)}}$, is $r-2$-smeary.

Smeariness: The Beast is Real

- $\exists$ arbitrary smeariness on $\mathbb{S}^1$ (Hotz and H., 2015);

Smeariness: The Beast is Real

- $\exists$ arbitrary smeariness on $\mathbb{S}^1$ (Hotz and H., 2015);
- $\exists$ $r - 2 = 2$ smeariness on $\mathbb{S}^m$ for all $m \in \mathbb{N}$ (Eltzner and H., 2018);

Smeariness: The Beast is Real

- $\exists$ arbitrary smeariness on $\mathbb{S}^1$ (Hotz and H., 2015);
- $\exists$ $r - 2 = 2$ smeariness on $\mathbb{S}^m$ for all $m \in \mathbb{N}$ (Eltzner and H., 2018);
- $\exists$ $r - 2 = 2$ smeariness on $\mathbb{S}^m$ for all $m \geq 5$ with (C) (Eltzner, 2019);

Smeariness: The Beast is Real

- $\exists$ arbitrary smeariness on $\mathbb{S}^1$ (Hotz and H., 2015);
- $\exists$ $r - 2 = 2$ smeariness on $\mathbb{S}^m$ for all $m \in \mathbb{N}$ (Eltzner and H., 2018);
- $\exists$ $r - 2 = 2$ smeariness on $\mathbb{S}^m$ for all $m \geq 5$ with (C) (Eltzner, 2019);
- smeariness is measure dependent (!);

Smeariness: The Beast is Real

- $\exists$ arbitrary smeariness on $\mathbb{S}^1$ (Hotz and H., 2015);
- $\exists$ $r - 2 = 2$ smeariness on $\mathbb{S}^m$ for all $m \in \mathbb{N}$ (Eltzner and H., 2018);
- $\exists$ $r - 2 = 2$ smeariness on $\mathbb{S}^m$ for all $m \geq 5$ with (C) (Eltzner, 2019);
- smeariness is measure dependent (!);
- smeariness, although only for nullset of the parameter space influences finite sample rates nearby.

Finite Sample Smeariness

Table 1.5 Orientations of 76 turtles after laying eggs (Gould's data cited by Stephens, 1969e)

Direction (in degrees) clockwise from north									
8	9	13	13	14	18	22	27	30	34
38	38	40	44	45	47	48	48	48	48
50	53	56	57	58	58	61	63	64	64
64	65	65	68	70	73	78	78	78	83
83	88	88	88	90	92	92	93	95	96
98	100	103	106	113	118	138	153	153	155
204	215	223	226	237	238	243	244	250	251
257	268	285	319	343	350				

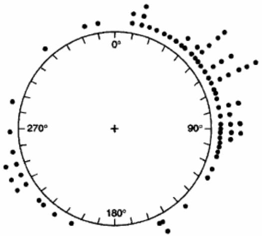
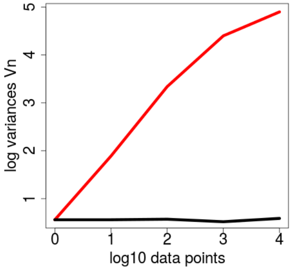


Figure 1.5 Circular plot of the turtle data of Table 1.5.

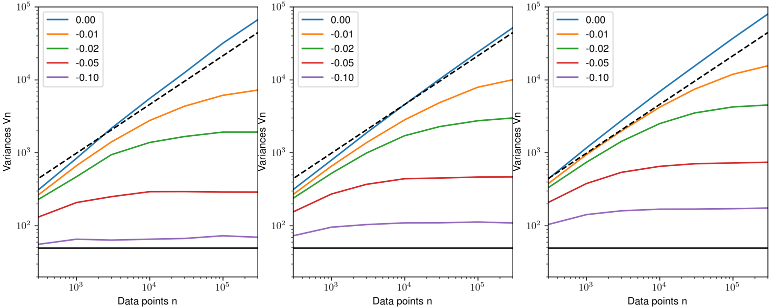
from Mardia and Jupp (2000).



Bootstrapped variance black = Euclidean in $[-\pi, \pi] \subset \mathbb{R}$, red = circular $\sim n^{2/3}$?

Two-Smeariness (Eltzner and H. 2018)

$r - 2 = 2$ smeary (dashed line)



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

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

Separating Data from Descriptor Space

Generalized Fréchet Means (S.H 2011a,b):

- Random $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} X \in Q$ on a data space Q

Separating Data from Descriptor Space

Generalized Fréchet Means (S.H 2011a,b):

- Random $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} X \in Q$ on a data space Q
- $P =$ descriptor space, e.g. $\Gamma(Q) =$ space of geodesics on Q

Separating Data from Descriptor Space

Generalized Fréchet Means (S.H 2011a,b):

- Random $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} X \in Q$ on a data space Q
- $P =$ descriptor space, e.g. $\Gamma(Q) =$ space of geodesics on Q
- $\rho : Q \times P \to [0, \infty)$ continuous = link function

Separating Data from Descriptor Space

Generalized Fréchet Means (S.H 2011a,b):

- Random $X_1, \ldots, X_n \stackrel{\text{i.i.d.}}{\sim} X \in Q$ on a data space Q
- $P =$ descriptor space, e.g. $\Gamma(Q) =$ space of geodesics on Q
- $\rho : Q \times P \to [0, \infty)$ continuous = link function
- $\gamma \in \operatorname{argmin}_{p \in P} \mathbb{E}[\rho(X, p)^2] =$ generalized population Fréchet mean

Separating Data from Descriptor Space

Generalized Fréchet Means (S.H 2011a,b):

- Random $X_1, \ldots, X_n \stackrel{\text{i.i.d.}}{\sim} X \in Q$ on a data space Q
- $P =$ descriptor space, e.g. $\Gamma(Q) =$ space of geodesics on Q
- $\rho : Q \times P \to [0, \infty)$ continuous = link function
- $\gamma \in \operatorname{argmin}_{p \in P} \mathbb{E}[\rho(X, p)^2] =$ generalized population Fréchet mean
- $\hat{\gamma} \in \operatorname{argmin}_{p \in P} \sum_{j=1}^{b} \rho(X_j, p)^2 =$ generalized sample Fréchet mean

Separating Data from Descriptor Space

Generalized Fréchet Means (S.H 2011a,b):

- Random $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} X \in Q$ on a data space Q
- $P =$ descriptor space, e.g. $\Gamma(Q) =$ space of geodesics on Q
- $\rho : Q \times P \to [0, \infty)$ continuous = link function
- $\gamma \in \operatorname{argmin}_{p \in P} \mathbb{E}[\rho(X, p)^2] =$ generalized population Fréchet mean
- $\hat{\gamma} \in \operatorname{argmin}_{p \in P} \sum_{j=1}^{b} \rho(X_j, p)^2 =$ generalized sample Fréchet mean

Separating Data from Descriptor Space

Generalized Fréchet Means (S.H 2011a,b):

- Random $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} X \in Q$ on a data space Q
- $P =$ descriptor space, e.g. $\Gamma(Q) =$ space of geodesics on Q
- $\rho : Q \times P \to [0, \infty)$ continuous = link function
- $\gamma \in \text{argmin}_{p \in P} \mathbb{E}[\rho(X, p)^2] =$ generalized population Fréchet mean
- $\hat{\gamma} \in \text{argmin}_{p \in P} \sum_{j=1}^{b} \rho(X_j, p)^2 =$ generalized sample Fréchet mean
- If γ is unique,

Separating Data from Descriptor Space

Generalized Fréchet Means (S.H 2011a,b):

- Random $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} X \in Q$ on a data space Q
- $P =$ descriptor space, e.g. $\Gamma(Q) =$ space of geodesics on Q
- $\rho : Q \times P \to [0, \infty)$ continuous = link function
- $\gamma \in \operatorname{argmin}_{p \in P} \mathbb{E}[\rho(X, p)^2] =$ generalized population Fréchet mean
- $\hat{\gamma} \in \operatorname{argmin}_{p \in P} \sum_{j=1}^{b} \rho(X_j, p)^2 =$ generalized sample Fréchet mean
- If γ is unique,
  - $\hat{\gamma} \to \gamma$ a.s. by S.H. (2011b) under weak regularity conditions

Separating Data from Descriptor Space

Generalized Fréchet Means (S.H 2011a,b):

- Random $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} X \in Q$ on a data space Q
- P = descriptor space, e.g. $\Gamma(Q)$ = space of geodesics on Q
- $\rho : Q \times P \to [0, \infty)$ continuous = link function
- $\gamma \in \operatorname{argmin}_{p \in P} \mathbb{E}[\rho(X, p)^2]$ = generalized population Fréchet mean
- $\hat{\gamma} \in \operatorname{argmin}_{p \in P} \sum_{j=1}^{b} \rho(X_j, p)^2$ = generalized sample Fréchet mean
- If γ is unique,
    - $\hat{\gamma} \to \gamma$ a.s. by S.H. (2011b) under weak regularity conditions
    - $\sqrt{n}\,(\phi(\hat{\gamma}) - \phi(\gamma)) \overset{\mathcal{D}}{\to} \mathcal{N}(0, \Sigma)$ by S.H. (2011a) if P is near γ a manifold with local chart ϕ, under regularity conditions adapted from (A1) – (A6).

Application: The CLT of Classical PCA

Original CLT proof by Anderson (1963) has not been reproduced (not even by himself). $\exists$ nonnormal perturbation theory proofs, e.g. Davis (1977); Watson (1983); Ruymgaart and Yang (1997).

Application: The CLT of Classical PCA

Original CLT proof by Anderson (1963) has not been reproduced (not even by himself). $\exists$ nonnormal perturbation theory proofs, e.g. Davis (1977); Watson (1983); Ruymgaart and Yang (1997).

$$Q = \mathbb{R}^m,\ P = G(m,k) \ni p = \mathrm{span}(\overbrace{v_{k+1},\ldots,v_m}^{=:W})^\perp:$$

Application: The CLT of Classical PCA

Original CLT proof by Anderson (1963) has not been reproduced (not even by himself). $\exists$ nonnormal perturbation theory proofs, e.g. Davis (1977); Watson (1983); Ruymgaart and Yang (1997).

$Q = \mathbb{R}^m$, $P = G(m,k) \ni p = \mathrm{span}(\overbrace{v_{k+1},\ldots,v_m}^{=:W})^\perp$:
$\mathrm{cov}[X] = V\Lambda V^T$, $\lambda_1 = \ldots = \lambda_k > \lambda_{k+1} \geq \ldots \geq \lambda_m > 0$;

Application: The CLT of Classical PCA

Original CLT proof by Anderson (1963) has not been reproduced (not even by himself). $\exists$ nonnormal perturbation theory proofs, e.g. Davis (1977); Watson (1983); Ruymgaart and Yang (1997).

$$Q = \mathbb{R}^m,\ P = G(m,k) \ni p = \mathrm{span}(\overbrace{v_{k+1},\ldots,v_m}^{=:W})^\perp:$$

$$\mathrm{cov}[X] = V\Lambda V^T,\ \lambda_1 = \ldots = \lambda_k > \lambda_{k+1} \geq \ldots \geq \lambda_m > 0;$$

$$\mathrm{cov}[X_1,\ldots,X_n] = \hat{V}\hat{\Lambda}\hat{V}^T,\ \hat{\lambda}_1 \geq \ldots \geq \hat{\lambda}_k \geq \hat{\lambda}_{k+1} \geq \ldots \hat{\lambda}_m \geq 0;$$

Application: The CLT of Classical PCA

Original CLT proof by Anderson (1963) has not been reproduced (not even by himself). $\exists$ nonnormal perturbation theory proofs, e.g. Davis (1977); Watson (1983); Ruymgaart and Yang (1997).

$Q = \mathbb{R}^m$, $P = G(m,k) \ni p = \mathrm{span}(\overbrace{v_{k+1},\ldots,v_m}^{=:W})^\perp$:
$\mathrm{cov}[X] = V\Lambda V^T$, $\lambda_1 = \ldots = \lambda_k > \lambda_{k+1} \geq \ldots \geq \lambda_m > 0$;
$\mathrm{cov}[X_1,\ldots,X_n] = \hat{V}\hat{\Lambda}\hat{V}^T$, $\hat{\lambda}_1 \geq \ldots \geq \hat{\lambda}_k \geq \hat{\lambda}_{k+1} \geq \ldots \hat{\lambda}_m \geq 0$;
$d(p,p')^2 = \min_{R\in O(m-k)} \|W - RW'\|^2$

Application: The CLT of Classical PCA

Original CLT proof by Anderson (1963) has not been reproduced (not even by himself). ∃ nonnormal perturbation theory proofs, e.g. Davis (1977); Watson (1983); Ruymgaart and Yang (1997).

$$Q = \mathbb{R}^m,\ P = G(m,k) \ni p = \mathrm{span}(\overbrace{v_{k+1},\ldots,v_m}^{=:W})^\perp:$$
$$\mathrm{cov}[X] = V\Lambda V^T,\ \lambda_1 = \ldots = \lambda_k > \lambda_{k+1} \geq \ldots \geq \lambda_m > 0;$$
$$\mathrm{cov}[X_1,\ldots,X_n] = \hat V\hat\Lambda\hat V^T,\ \hat\lambda_1 \geq \ldots \geq \hat\lambda_k \geq \hat\lambda_{k+1} \geq \ldots \hat\lambda_m \geq 0;$$
$$d(p,p')^2 = \min_{R\in O(m-k)} \|W - RW'\|^2$$
$$\rho(X,p)^2 = \|X - WW^TX\|^2 = \|X\|^2 - \mathrm{trace}(W^TXX^TW);$$

Application: The CLT of Classical PCA

Original CLT proof by Anderson (1963) has not been reproduced (not even by himself). ∃ nonnormal perturbation theory proofs, e.g. Davis (1977); Watson (1983); Ruymgaart and Yang (1997).

$$Q = \mathbb{R}^m,\ P = G(m,k) \ni p = \mathrm{span}(\overbrace{v_{k+1},\ldots,v_m}^{=:W})^\perp:$$

$$\mathrm{cov}[X] = V\Lambda V^T,\ \lambda_1 = \ldots = \lambda_k > \lambda_{k+1} \geq \ldots \geq \lambda_m > 0;$$

$$\mathrm{cov}[X_1,\ldots,X_n] = \hat{V}\hat{\Lambda}\hat{V}^T,\ \hat{\lambda}_1 \geq \ldots \geq \hat{\lambda}_k \geq \hat{\lambda}_{k+1} \geq \ldots \hat{\lambda}_m \geq 0;$$

$$d(p,p')^2 = \min_{R\in O(m-k)} \|W - RW'\|^2$$

$$\rho(X,p)^2 = \|X - WW^TX\|^2 = \|X\|^2 - \mathrm{trace}(W^TXX^TW);$$

$\Rightarrow$ (A1), Taylor with $r = 2$;

Application: The CLT of Classical PCA

Original CLT proof by Anderson (1963) has not been reproduced (not even by himself). $\exists$ nonnormal perturbation theory proofs, e.g. Davis (1977); Watson (1983); Ruymgaart and Yang (1997).

$$Q = \mathbb{R}^m,\ P = G(m,k) \ni p = \mathrm{span}(\overbrace{v_{k+1},\ldots,v_m}^{=:W})^\perp:$$
$$\mathrm{cov}[X] = V\Lambda V^T,\ \lambda_1 = \ldots = \lambda_k > \lambda_{k+1} \geq \ldots \geq \lambda_m > 0;$$
$$\mathrm{cov}[X_1,\ldots,X_n] = \hat{V}\hat{\Lambda}\hat{V}^T,\ \hat{\lambda}_1 \geq \ldots \geq \hat{\lambda}_k \geq \hat{\lambda}_{k+1} \geq \ldots \hat{\lambda}_m \geq 0;$$
$$d(p,p')^2 = \min_{R\in O(m-k)} \|W - RW'\|^2$$
$$\rho(X,p)^2 = \|X - WW^TX\|^2 = \|X\|^2 - \mathrm{trace}(W^TXX^TW);$$
$$\Rightarrow \text{(A1), Taylor with } r = 2;$$
$$\rho(X,p')^2 - \rho(X,p)^2 = \mathrm{trace}(W^TXX^TW) - \mathrm{trace}(W'^TXX^TW'),$$

Application: The CLT of Classical PCA

Original CLT proof by Anderson (1963) has not been reproduced (not even by himself). ∃ nonnormal perturbation theory proofs, e.g. Davis (1977); Watson (1983); Ruymgaart and Yang (1997).

$Q = \mathbb{R}^m$, $P = G(m,k) \ni p = \mathrm{span}(\overbrace{v_{k+1},\ldots,v_m}^{=:W})^\perp$:
$\mathrm{cov}[X] = V\Lambda V^T$, $\lambda_1 = \ldots = \lambda_k > \lambda_{k+1} \geq \ldots \geq \lambda_m > 0$;
$\mathrm{cov}[X_1,\ldots,X_n] = \hat{V}\hat{\Lambda}\hat{V}^T$, $\hat{\lambda}_1 \geq \ldots \geq \hat{\lambda}_k \geq \hat{\lambda}_{k+1} \geq \ldots \hat{\lambda}_m \geq 0$;
$d(p,p')^2 = \min_{R\in O(m-k)} \|W - RW'\|^2$
$\rho(X,p)^2 = \|X - WW^TX\|^2 = \|X\|^2 - \mathrm{trace}(W^TXX^TW)$;
$\Rightarrow$ (A1), Taylor with $r = 2$;
$\rho(X,p')^2 - \rho(X,p)^2 = \mathrm{trace}(W^TXX^TW) - \mathrm{trace}(W'^TXX^TW')$,
with $E[\|X\|^4] < \infty \Rightarrow$ Donsker;

Application: The CLT of Classical PCA

Original CLT proof by Anderson (1963) has not been reproduced (not even by himself). $\exists$ nonnormal perturbation theory proofs, e.g. Davis (1977); Watson (1983); Ruymgaart and Yang (1997).

$Q = \mathbb{R}^m$, $P = G(m,k) \ni p = \mathrm{span}(\overbrace{v_{k+1}, \dots, v_m}^{=:W})^\perp$:
$\mathrm{cov}[X] = V\Lambda V^T$, $\lambda_1 = \dots = \lambda_k > \lambda_{k+1} \geq \dots \geq \lambda_m > 0$;
$\mathrm{cov}[X_1, \dots, X_n] = \hat V \hat\Lambda \hat V^T$, $\hat\lambda_1 \geq \dots \geq \hat\lambda_k \geq \hat\lambda_{k+1} \geq \dots \hat\lambda_m \geq 0$;
$d(p,p')^2 = \min_{R \in O(m-k)} \|W - RW'\|^2$
$\rho(X,p)^2 = \|X - WW^TX\|^2 = \|X\|^2 - \mathrm{trace}(W^TXX^TW)$;
$\Rightarrow$ (A1), Taylor with $r = 2$;
$\rho(X,p')^2 - \rho(X,p)^2 = \mathrm{trace}(W^TXX^TW) - \mathrm{trace}(W'^TXX^TW')$,
with $E[\|X\|^4] < \infty \Rightarrow$ Donsker;
$\Rightarrow \sqrt{n}$ Gaussian CLT.

More Applications

- Geodesic PCA (GPCA) on Riemannian spaces by S.H et al. (2010):

More Applications

- Geodesic PCA (GPCA) on Riemannian spaces by S.H et al. (2010):
    - $P_1 = \Gamma(Q) =$ all geodesics on Q,
      $\rightsquigarrow \gamma_1$ and $\hat{\gamma}_1$ = 1st geodesic PCs

More Applications

- Geodesic PCA (GPCA) on Riemannian spaces by S.H et al. (2010):
  - $P_1 = \Gamma(Q) =$ all geodesics on Q,
    $\rightsquigarrow \gamma_1$ and $\hat{\gamma}_1$ = 1st geodesic PCs
  - $P_2 = \{p \in \Gamma(Q) : \gamma_1 \perp p, \gamma_1 \cap p \neq \emptyset\}$
    $\hat{P}_2 = \{p \in \Gamma(Q) : \hat{\gamma}_1 \perp p, \hat{\gamma}_1 \cap p \neq \emptyset\}$
    $\rightsquigarrow \gamma_2$ and $\hat{\gamma}_2$ = 2nd geodesic PCs

More Applications

- Geodesic PCA (GPCA) on Riemannian spaces by S.H et al. (2010):
  - $P_1 = \Gamma(Q) =$ all geodesics on Q,
    $\rightsquigarrow \gamma_1$ and $\hat{\gamma}_1$ = 1st geodesic PCs
  - $P_2 = \{p \in \Gamma(Q) : \gamma_1 \perp p, \gamma_1 \cap p \neq \emptyset\}$
    $\hat{P}_2 = \{p \in \Gamma(Q) : \hat{\gamma}_1 \perp p, \hat{\gamma}_1 \cap p \neq \emptyset\}$
    $\rightsquigarrow \gamma_2$ and $\hat{\gamma}_2$ = 2nd geodesic PCs
  - $\vdots$

More Applications

- Geodesic PCA (GPCA) on Riemannian spaces by S.H et al. (2010):
    - $P_1 = \Gamma(Q) =$ all geodesics on Q,
      $\rightsquigarrow \gamma_1$ and $\hat{\gamma}_1$ = 1st geodesic PCs
    - $P_2 = \{p \in \Gamma(Q) : \gamma_1 \perp p, \gamma_1 \cap p \neq \emptyset\}$
      $\hat{P}_2 = \{p \in \Gamma(Q) : \hat{\gamma}_1 \perp p, \hat{\gamma}_1 \cap p \neq \emptyset\}$
      $\rightsquigarrow \gamma_2$ and $\hat{\gamma}_2$ = 2nd geodesic PCs
    - $\vdots$

More Applications

- Geodesic PCA (GPCA) on Riemannian spaces by S.H et al. (2010):
  - $P_1 = \Gamma(Q) =$ all geodesics on Q,
    $\rightsquigarrow \gamma_1$ and $\hat{\gamma}_1$ = 1st geodesic PCs
  - $P_2 = \{p \in \Gamma(Q) : \gamma_1 \perp p, \gamma_1 \cap p \neq \emptyset\}$
    $\hat{P}_2 = \{p \in \Gamma(Q) : \hat{\gamma}_1 \perp p, \hat{\gamma}_1 \cap p \neq \emptyset\}$
    $\rightsquigarrow \gamma_2$ and $\hat{\gamma}_2$ = 2nd geodesic PCs
  - $\vdots$
- If M is a Riemannian manifold and G a Lie group acting properly and isometrically on G then the shape space $Q := M/G$ is a Riemann stratified space, so is $\Gamma(Q)$.

More Applications

- Geodesic PCA (GPCA) on Riemannian spaces by S.H et al. (2010):
  - $P_1 = \Gamma(Q) =$ all geodesics on Q,
    $\rightsquigarrow \gamma_1$ and $\hat{\gamma}_1$ = 1st geodesic PCs
  - $P_2 = \{p \in \Gamma(Q) : \gamma_1 \perp p, \gamma_1 \cap p \neq \emptyset\}$
    $\hat{P}_2 = \{p \in \Gamma(Q) : \hat{\gamma}_1 \perp p, \hat{\gamma}_1 \cap p \neq \emptyset\}$
    $\rightsquigarrow \gamma_2$ and $\hat{\gamma}_2$ = 2nd geodesic PCs
  - $\vdots$
- If M is a Riemannian manifold and G a Lie group acting properly and isometrically on G then the shape space $Q := M/G$ is a Riemann stratified space, so is $\Gamma(Q)$.
- A shape space has an open and dense top-dimensional manifold part Q^* (cf. Bredon (1972)).

More Applications

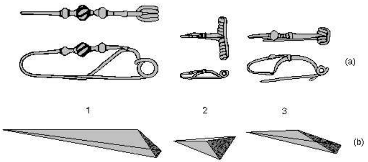
- Geodesic PCA (GPCA) on Riemannian spaces by S.H et al. (2010):
  - $P_1 = \Gamma(Q) =$ all geodesics on Q, $\rightsquigarrow \gamma_1$ and $\hat{\gamma}_1$ = 1st geodesic PCs
  - $P_2 = \{p \in \Gamma(Q) : \gamma_1 \perp p, \gamma_1 \cap p \neq \emptyset\}$ $\hat{P}_2 = \{p \in \Gamma(Q) : \hat{\gamma}_1 \perp p, \hat{\gamma}_1 \cap p \neq \emptyset\}$ $\rightsquigarrow \gamma_2$ and $\hat{\gamma}_2$ = 2nd geodesic PCs
  - $\vdots$
- If M is a Riemannian manifold and G a Lie group acting properly and isometrically on G then the shape space $Q := M/G$ is a Riemann stratified space, so is $\Gamma(Q)$.
- A shape space has an open and dense top-dimensional manifold part Q^* (cf. Bredon (1972)).

More Applications

- Geodesic PCA (GPCA) on Riemannian spaces by S.H et al. (2010):
  - $P_1 = \Gamma(Q) =$ all geodesics on Q,
    $\rightsquigarrow \gamma_1$ and $\hat{\gamma}_1$ = 1st geodesic PCs
  - $P_2 = \{p \in \Gamma(Q) : \gamma_1 \perp p, \gamma_1 \cap p \neq \emptyset\}$
    $\hat{P}_2 = \{p \in \Gamma(Q) : \hat{\gamma}_1 \perp p, \hat{\gamma}_1 \cap p \neq \emptyset\}$
    $\rightsquigarrow \gamma_2$ and $\hat{\gamma}_2$ = 2nd geodesic PCs
  - $\vdots$
- If M is a Riemannian manifold and G a Lie group acting properly and isometrically on G then the shape space $Q := M/G$ is a Riemann stratified space, so is $\Gamma(Q)$.
- A shape space has an open and dense top-dimensional manifold part Q^* (cf. Bredon (1972)).
- Manifold stability for intrinsic means (singularities are repulsive for means) not for Procrustes means (!), cf. S.H. (2012). Open for GPCs.

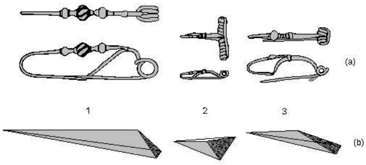
Euclidean visualization of scores, o.g. projection onto GPCs (H. et al, 2010)

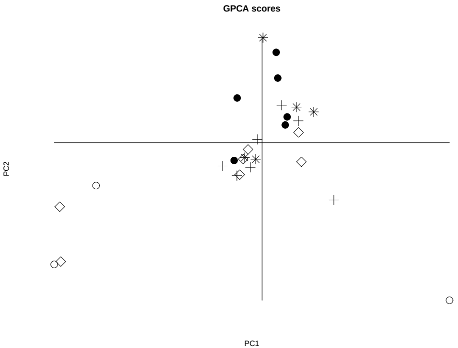
28 tetrahedral iron-age fibulae from a grave site in Münsingen, Switzerland (Hodson et al. (1966) and Small (1996)).



Euclidean visualization of scores, o.g. projection onto GPCs (H. et al, 2010)

28 tetrahedral iron-age fibulae from a grave site in Münsingen, Switzerland (Hodson et al. (1966) and Small (1996)).





Groups from old to young: filled circles, stars, crosses, diamonds and circles.

PC2: Shape change; PC1: Stronger effect, diversification.

Two-Sample Descriptor Test

$$
\begin{array}{llcccll}
\text{Data:} & \underbrace{X_1,\ldots,X_n} & , & \underbrace{Y_1,\ldots,Y_m} & \in & Q \\
& \downarrow & & \downarrow & & \\
\text{Descriptors:} & p^X & & p^Y & \in & P \\
& \downarrow & \phi^{-1} & \downarrow & & \\
\text{Coordinates:} & Z^X & & Z^Y & \in & \mathbb{R}^D
\end{array}
$$

Two-Sample Descriptor Test

$$\begin{array}{lcccccc}
\text{Data:} & \underbrace{X_1,\dots,X_n} & , & \underbrace{Y_1,\dots,Y_m} & \in & Q \\
 & \downarrow & & \downarrow & & \\
\text{Descriptors:} & p^X & & p^Y & \in & P \\
 & \downarrow & \phi^{-1} & \downarrow & & \\
\text{Coordinates:} & Z^X & & Z^Y & \in & \mathbb{R}^D
\end{array}$$

Under $H_0 : \mu^X = \mu^Y$,

$$\frac{nm}{n+m}(m+n-2)(Z^X - Z^Y)^T\left(n\,\widehat{\mathrm{cov}}[Z^X_{1\dots n}] + m\,\widehat{\mathrm{cov}}[Z^Y_{1\dots m}]\right)^{-1} \cdot (Z^X - Z^Y) \sim \mathcal{T}^2(k, n+m-2)$$

Two-Sample Descriptor Test

$$\begin{array}{llccccc} \text{Data:} & \underbrace{X_1,\ldots,X_n} & , & \underbrace{Y_1,\ldots,Y_m} & \in & Q \\ & \downarrow & & \downarrow & & \\ \text{Descriptors:} & p^X & & p^Y & \in & P \\ & \downarrow & \phi^{-1} & \downarrow & & \\ \text{Coordinates:} & Z^X & & Z^Y & \in & \mathbb{R}^D \end{array}$$

Under $H_0 : \mu^X = \mu^Y$,

$$\frac{nm}{n+m}(m+n-2)(Z^X - Z^Y)^T \left(n\,\widehat{\mathrm{cov}}[Z^X_{1\ldots n}] + m\,\widehat{\mathrm{cov}}[Z^Y_{1\ldots m}]\right)^{-1} \cdot (Z^X - Z^Y) \sim \mathcal{T}^2(k, n+m-2)$$

But how to access $\widehat{\mathrm{cov}}[Z^X_{1\ldots n}]$ and $\widehat{\mathrm{cov}}[Z^Y_{1\ldots m}]$?

Bootstrapping

For $b = 1, \ldots, B$, resample:

- $X^*_{1,b}, \ldots, X^*_{n,b}$ from $X_1, \ldots, X_n$ gives $\widehat{\text{cov}}[Z^X_{1\ldots n}]$

Bootstrapping

For $b = 1, \ldots, B$, resample:

- $X^*_{1,b}, \ldots, X^*_{n,b}$ from $X_1, \ldots, X_n$ gives $\widehat{\text{cov}}[Z^X_{1\ldots n}]$
- $Y^*_{1,b}, \ldots, Y^*_{m,b}$ from $Y_1, \ldots, Y_m$ gives $\widehat{\text{cov}}[Z^Y_{1\ldots m}]$

Bootstrapping

For $b = 1, \ldots, B$, resample:

- $X^*_{1,b}, \ldots, X^*_{n,b}$ from $X_1, \ldots, X_n$ gives $\widehat{\text{cov}}[Z^X_{1\ldots n}]$
- $Y^*_{1,b}, \ldots, Y^*_{m,b}$ from $Y_1, \ldots, Y_m$ gives $\widehat{\text{cov}}[Z^Y_{1\ldots m}]$
- set $A = n\,\widehat{\text{cov}}[Z^X_{1\ldots n}] + m\,\widehat{\text{cov}}[Z^Y_{1\ldots m}]$

Bootstrapping

For $b = 1, \ldots, B$, resample:

- $X^*_{1,b}, \ldots, X^*_{n,b}$ from $X_1, \ldots, X_n$ gives $\widehat{\text{cov}}[Z^X_{1\ldots n}]$
- $Y^*_{1,b}, \ldots, Y^*_{m,b}$ from $Y_1, \ldots, Y_m$ gives $\widehat{\text{cov}}[Z^Y_{1\ldots m}]$
- set $A = n\,\widehat{\text{cov}}[Z^X_{1\ldots n}] + m\,\widehat{\text{cov}}[Z^Y_{1\ldots m}]$

Again, for $b = 1, \ldots, B'$, resample:

- $W^*_{1,b}, \ldots, W^*_{n+m,b}$ from $X_1, \ldots, X_n, Y_1, \ldots, Y_m$

Bootstrapping

For $b = 1, \dots, B$, resample:

- $X^*_{1,b}, \dots, X^*_{n,b}$ from $X_1, \dots, X_n$ gives $\widehat{\operatorname{cov}}[Z^X_{1\dots n}]$
- $Y^*_{1,b}, \dots, Y^*_{m,b}$ from $Y_1, \dots, Y_m$ gives $\widehat{\operatorname{cov}}[Z^Y_{1\dots m}]$
- set $A = n\,\widehat{\operatorname{cov}}[Z^X_{1\dots n}] + m\,\widehat{\operatorname{cov}}[Z^Y_{1\dots m}]$

Again, for $b = 1, \dots, B'$, resample:

- $W^*_{1,b}, \dots, W^*_{n+m,b}$ from $X_1, \dots, X_n, Y_1, \dots, Y_m$
- set $X^*_{j,b} = W^*_{j,b}$ for $j = 1, \dots, n$

Bootstrapping

For $b = 1, \ldots, B$, resample:

- $X^*_{1,b}, \ldots, X^*_{n,b}$ from $X_1, \ldots, X_n$ gives $\widehat{\text{cov}}[Z^X_{1\ldots n}]$
- $Y^*_{1,b}, \ldots, Y^*_{m,b}$ from $Y_1, \ldots, Y_m$ gives $\widehat{\text{cov}}[Z^Y_{1\ldots m}]$
- set $A = n\,\widehat{\text{cov}}[Z^X_{1\ldots n}] + m\,\widehat{\text{cov}}[Z^Y_{1\ldots m}]$

Again, for $b = 1, \ldots, B'$, resample:

- $W^*_{1,b}, \ldots, W^*_{n+m,b}$ from $X_1, \ldots, X_n, Y_1, \ldots, Y_m$
- set $X^*_{j,b} = W^*_{j,b}$ for $j = 1, \ldots, n$
- set $Y^*_{j,b} = W^*_{j+n,b}$ for $j = 1, \ldots, m$

Bootstrapping

For $b = 1, \ldots, B$, resample:

- $X^*_{1,b}, \ldots, X^*_{n,b}$ from $X_1, \ldots, X_n$ gives $\widehat{\text{cov}}[Z^X_{1\ldots n}]$
- $Y^*_{1,b}, \ldots, Y^*_{m,b}$ from $Y_1, \ldots, Y_m$ gives $\widehat{\text{cov}}[Z^Y_{1\ldots m}]$
- set $A = n\,\widehat{\text{cov}}[Z^X_{1\ldots n}] + m\,\widehat{\text{cov}}[Z^Y_{1\ldots m}]$

Again, for $b = 1, \ldots, B'$, resample:

- $W^*_{1,b}, \ldots, W^*_{n+m,b}$ from $X_1, \ldots, X_n, Y_1, \ldots, Y_m$
- set $X^*_{j,b} = W^*_{j,b}$ for $j = 1, \ldots, n$
- set $Y^*_{j,b} = W^*_{j+n,b}$ for $j = 1, \ldots, m$
- compute the empirical quantile $c^*_{1-\alpha}$ such that

Bootstrapping

For $b = 1, \dots, B$, resample:

- $X^*_{1,b}, \dots, X^*_{n,b}$ from $X_1, \dots, X_n$ gives $\widehat{\mathrm{cov}}[Z^X_{1\dots n}]$
- $Y^*_{1,b}, \dots, Y^*_{m,b}$ from $Y_1, \dots, Y_m$ gives $\widehat{\mathrm{cov}}[Z^Y_{1\dots m}]$
- set $A = n\,\widehat{\mathrm{cov}}[Z^X_{1\dots n}] + m\,\widehat{\mathrm{cov}}[Z^Y_{1\dots m}]$

Again, for $b = 1, \dots, B'$, resample:

- $W^*_{1,b}, \dots, W^*_{n+m,b}$ from $X_1, \dots, X_n, Y_1, \dots, Y_m$
- set $X^*_{j,b} = W^*_{j,b}$ for $j = 1, \dots, n$
- set $Y^*_{j,b} = W^*_{j+n,b}$ for $j = 1, \dots, m$
- compute the empirical quantile $c^*_{1-\alpha}$ such that
- $\mathbb{P}\left\{(Z^{X^*} - Z^{Y^*})^T A^{-1} (Z^{X^*} - Z^{Y^*}) \leq c^*_{1-\alpha} \right.$ $\left. | X_1, \dots, X_n, Y_1, \dots, Y_m \right\} \geq 1 - \alpha$

Bootstrapping

For $b = 1, \dots, B$, resample:

- $X^*_{1,b}, \dots, X^*_{n,b}$ from $X_1, \dots, X_n$ gives $\widehat{\operatorname{cov}}[Z^X_{1\dots n}]$
- $Y^*_{1,b}, \dots, Y^*_{m,b}$ from $Y_1, \dots, Y_m$ gives $\widehat{\operatorname{cov}}[Z^Y_{1\dots m}]$
- set $A = n\,\widehat{\operatorname{cov}}[Z^X_{1\dots n}] + m\,\widehat{\operatorname{cov}}[Z^Y_{1\dots m}]$

Again, for $b = 1, \dots, B'$, resample:

- $W^*_{1,b}, \dots, W^*_{n+m,b}$ from $X_1, \dots, X_n, Y_1, \dots, Y_m$
- set $X^*_{j,b} = W^*_{j,b}$ for $j = 1, \dots, n$
- set $Y^*_{j,b} = W^*_{j+n,b}$ for $j = 1, \dots, m$
- compute the empirical quantile $c^*_{1-\alpha}$ such that
- $\mathbb{P}\left\{(Z^{X^*} - Z^{Y^*})^T A^{-1} (Z^{X^*} - Z^{Y^*}) \leq c^*_{1-\alpha} \,\middle|\, X_1, \dots, X_n, Y_1, \dots, Y_m\right\} \geq 1 - \alpha$

Then, the test

reject H_0 if $(Z^X - Z^Y)^T A^{-1} (Z^X - Z^Y) > c^*_{1-\alpha}$

has the asymptotic level α.

Improved Power

Recall, for $b = 1, \dots, B'$, resample:

- $W^*_{1,b}, \dots, W^*_{n+m,b}$ from $X_1, \dots, X_n$, $Y_1, \dots, Y_m$
- set $X^*_{j,b} = W^*_{j,b}$ for $j = 1, \dots, n$
- set $Y^*_{j,b} = W^*_{j+n,b}$ for $j = 1, \dots, m$
- compute the empirical quantile $c^*_{1-\alpha}$ such that
- $\mathbb{P}\left\{(Z^{X^*} - Z^{Y^*})^T A^{-1} (Z^{X^*} - Z^{Y^*}) \leq c^*_{1-\alpha}\right\} \geq 1 - \alpha$

Then, *reject H_0 if $(Z^X - Z^Y)^T A^{-1} (Z^X - Z^Y) > c^*_{1-\alpha}$*

Improved Power

Recall, for $b = 1, \dots, B'$, resample:

- $W^*_{1,b}, \dots, W^*_{n+m,b}$ from $X_1, \dots, X_n$, $Y_1, \dots, Y_m$
- set $X^*_{j,b} = W^*_{j,b}$ for $j = 1, \dots, n$
- set $Y^*_{j,b} = W^*_{j+n,b}$ for $j = 1, \dots, m$
- compute the empirical quantile $c^*_{1-\alpha}$ such that
- $\mathbb{P}\left\{(Z^{X^*} - Z^{Y^*})^T A^{-1} (Z^{X^*} - Z^{Y^*}) \le c^*_{1-\alpha}\right\} \ge 1 - \alpha$

Then, *reject H_0 if $(Z^X - Z^Y)^T A^{-1} (Z^X - Z^Y) > c^*_{1-\alpha}$*

To improve the power, resample

- $X^{*,b}$ from $X_1, \dots, X_n$ and $Y^{*,b}$ from $Y_1, \dots, Y_m$

Improved Power

Recall, for $b = 1, \ldots, B'$, resample:

- $W^*_{1,b}, \ldots, W^*_{n+m,b}$ from $X_1, \ldots, X_n$, $Y_1, \ldots, Y_m$
- set $X^*_{j,b} = W^*_{j,b}$ for $j = 1, \ldots, n$
- set $Y^*_{j,b} = W^*_{j+n,b}$ for $j = 1, \ldots, m$
- compute the empirical quantile $c^*_{1-\alpha}$ such that
- $\mathbb{P}\left\{(Z^{X^*} - Z^{Y^*})^T A^{-1}(Z^{X^*} - Z^{Y^*}) \leq c^*_{1-\alpha}\right\} \geq 1 - \alpha$

Then, *reject H_0 if $(Z^X - Z^Y)^T A^{-1}(Z^X - Z^Y) > c^*_{1-\alpha}$*

To improve the power, resample

- $X^{*,b}$ from $X_1, \ldots, X_n$ and $Y^{*,b}$ from $Y_1, \ldots, Y_m$
- set $d^{X^*} = Z^{X^*} - Z^{\mu^X}$, $d^{Y^*} = Z^{Y^*} - Z^{\mu^Y}$

Improved Power

Recall, for $b = 1, \dots, B'$, resample:

- $W^*_{1,b}, \dots, W^*_{n+m,b}$ from $X_1, \dots, X_n$, $Y_1, \dots, Y_m$
- set $X^*_{j,b} = W^*_{j,b}$ for $j = 1, \dots, n$
- set $Y^*_{j,b} = W^*_{j+n,b}$ for $j = 1, \dots, m$
- compute the empirical quantile $c^*_{1-\alpha}$ such that
- $\mathbb{P}\left\{(Z^{X^*} - Z^{Y^*})^T A^{-1} (Z^{X^*} - Z^{Y^*}) \leq c^*_{1-\alpha}\right\} \geq 1 - \alpha$

Then, *reject H_0 if $(Z^X - Z^Y)^T A^{-1}(Z^X - Z^Y) > c^*_{1-\alpha}$*

To improve the power, resample

- $X^{*,b}$ from $X_1, \dots, X_n$ and $Y^{*,b}$ from $Y_1, \dots, Y_m$
- set $d^{X^*} = Z^{X^*} - Z^{\mu^X}$, $d^{Y^*} = Z^{Y^*} - Z^{\mu^Y}$
- compute the empirical quantile $c^*_{1-\alpha}$ such that

Improved Power

Recall, for $b = 1, \dots, B'$, resample:

- $W^*_{1,b}, \dots, W^*_{n+m,b}$ from $X_1, \dots, X_n$, $Y_1, \dots, Y_m$
- set $X^*_{j,b} = W^*_{j,b}$ for $j = 1, \dots, n$
- set $Y^*_{j,b} = W^*_{j+n,b}$ for $j = 1, \dots, m$
- compute the empirical quantile $c^*_{1-\alpha}$ such that
- $\mathbb{P}\left\{(Z^{X^*} - Z^{Y^*})^T A^{-1}(Z^{X^*} - Z^{Y^*}) \leq c^*_{1-\alpha}\right\} \geq 1 - \alpha$

Then, *reject H_0 if $(Z^X - Z^Y)^T A^{-1}(Z^X - Z^Y) > c^*_{1-\alpha}$*

To improve the power, resample

- $X^{*,b}$ from $X_1, \dots, X_n$ and $Y^{*,b}$ from $Y_1, \dots, Y_m$
- set $d^{X^*} = Z^{X^*} - Z^{\mu^X}$, $d^{Y^*} = Z^{Y^*} - Z^{\mu^Y}$
- compute the empirical quantile $c^*_{1-\alpha}$ such that
- $\mathbb{P}\left\{(d^{X^*} - d^{Y^*})^T A^{-1}(d^{X^*} - d^{Y^*}) \leq c^*_{1-\alpha}\right\} \geq 1 - \alpha$

Improved Power

Recall, for $b = 1, \dots, B'$, resample:

- $W^*_{1,b}, \dots, W^*_{n+m,b}$ from $X_1, \dots, X_n, Y_1, \dots, Y_m$
- set $X^*_{j,b} = W^*_{j,b}$ for $j = 1, \dots, n$
- set $Y^*_{j,b} = W^*_{j+n,b}$ for $j = 1, \dots, m$
- compute the empirical quantile $c^*_{1-\alpha}$ such that
- $\mathbb{P}\left\{(Z^{X^*} - Z^{Y^*})^T A^{-1} (Z^{X^*} - Z^{Y^*}) \le c^*_{1-\alpha}\right\} \ge 1 - \alpha$

Then, *reject H_0 if $(Z^X - Z^Y)^T A^{-1} (Z^X - Z^Y) > c^*_{1-\alpha}$*

To improve the power, resample

- $X^{*,b}$ from $X_1, \dots, X_n$ and $Y^{*,b}$ from $Y_1, \dots, Y_m$
- set $d^{X^*} = Z^{X^*} - Z^{\mu^X}$, $d^{Y^*} = Z^{Y^*} - Z^{\mu^Y}$
- compute the empirical quantile $c^*_{1-\alpha}$ such that
- $\mathbb{P}\left\{(d^{X^*} - d^{Y^*})^T A^{-1} (d^{X^*} - d^{Y^*}) \le c^*_{1-\alpha}\right\} \ge 1 - \alpha$

Improved Power

Recall, for $b = 1, \dots, B'$, resample:

- $W^*_{1,b}, \dots, W^*_{n+m,b}$ from $X_1, \dots, X_n$, $Y_1, \dots, Y_m$
- set $X^*_{j,b} = W^*_{j,b}$ for $j = 1, \dots, n$
- set $Y^*_{j,b} = W^*_{j+n,b}$ for $j = 1, \dots, m$
- compute the empirical quantile $c^*_{1-\alpha}$ such that
- $\mathbb{P}\left\{(Z^{X^*} - Z^{Y^*})^T A^{-1} (Z^{X^*} - Z^{Y^*}) \leq c^*_{1-\alpha}\right\} \geq 1 - \alpha$

Then, *reject H_0 if $(Z^X - Z^Y)^T A^{-1} (Z^X - Z^Y) > c^*_{1-\alpha}$*

To improve the power, resample

- $X^{*,b}$ from $X_1, \dots, X_n$ and $Y^{*,b}$ from $Y_1, \dots, Y_m$
- set $d^{X^*} = Z^{X^*} - Z^{\mu^X}$, $d^{Y^*} = Z^{Y^*} - Z^{\mu^Y}$
- compute the empirical quantile $c^*_{1-\alpha}$ such that
- $\mathbb{P}\left\{(d^{X^*} - d^{Y^*})^T A^{-1} (d^{X^*} - d^{Y^*}) \leq c^*_{1-\alpha}\right\} \geq 1 - \alpha$

Then, *rejecting H_0 if $(Z^X - Z^Y)^T A^{-1} (Z^X - Z^Y) > c^*_{1-\alpha}$* has the asymptotic level α and we have simulated “close” to H_0.

Sequences of Nested Subspaces

Note:

- Euclidean PCA ist canonically nested.
- non-Euclidean PCA is not.

Sequences of Nested Subspaces

Note:

- Euclidean PCA ist canonically nested.
- non-Euclidean PCA is not.

For data on a sphere $Q = \mathbb{S}^m$, Jung et al. (2012) define principal nested spheres (PNS) by residual variance minimization

- $\mathbb{S}^m \supset \hat{\mathbb{S}}^{m-1} \supset \ldots \supset \hat{\mathbb{S}}^1 \supset \{\hat{\mu}\}$ (great spheres).
- or even small spheres,
- backward nested.

Sequences of Nested Subspaces

Note:

- Euclidean PCA ist canonically nested.
- non-Euclidean PCA is not.

For data on a sphere $Q = \mathbb{S}^m$, Jung et al. (2012) define principal nested spheres (PNS) by residual variance minimization

- $\mathbb{S}^m \supset \hat{\mathbb{S}}^{m-1} \supset \ldots \supset \hat{\mathbb{S}}^1 \supset \{\hat{\mu}\}$ (great spheres).
- or even small spheres,
- backward nested.

For more general spaces, Pennec (2018) defines barycentric subspaces (next two days)

- forward or backward nested or all at once.

Sequences of Nested Subspaces

Note:

- Euclidean PCA ist canonically nested.
- non-Euclidean PCA is not.

For data on a sphere $Q = \mathbb{S}^m$, Jung et al. (2012) define principal nested spheres (PNS) by residual variance minimization

- $\mathbb{S}^m \supset \hat{\mathbb{S}}^{m-1} \supset \ldots \supset \hat{\mathbb{S}}^1 \supset \{\hat{\mu}\}$ (great spheres).
- or even small spheres,
- backward nested.

For more general spaces, Pennec (2018) defines barycentric subspaces (next two days)

- forward or backward nested or all at once.

How about asymptotics of such nested random subspaces?

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 $p \in P_j$ ($j = 1, \dots, m$) 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"

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

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 $j \in \{1,\dots,m\}$,

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

is BNFD from P_m to P_j from the space

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

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 $j \in \{1,\ldots,m\}$,

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

is BNFD from P_m to P_j from the space

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

with projection along each descriptor

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

Backward Nested Families of Descriptors

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

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

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 $j = m, \dots, 1$,

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

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

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 $j = m, \ldots, 1$,

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

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

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

Strong Law

Theorem (S.H. and Eltzner (2018))

If the BN population means $f = (p^m, \dots, p^j)$ are unique and $f_n = (p_n^m, \dots, p_n^j)$ 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 [S.H. and Eltzner (2018)]

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 [S.H. and Eltzner (2018)]

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 [S.H. and Eltzner (2018)]

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:

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

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

The Joint CLT [S.H. and Eltzner (2018)]

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:

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

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

The Joint CLT [S.H. and Eltzner (2018)]

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:

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

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

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 $\lambda_{j+1}, \dots \lambda_m \in \mathbb{R}$ 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

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

Factoring Charts

If the following diagram commutes we say the chart factors

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

Then

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

Factoring Charts

If the following diagram commutes we say the chart factors

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

Then

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

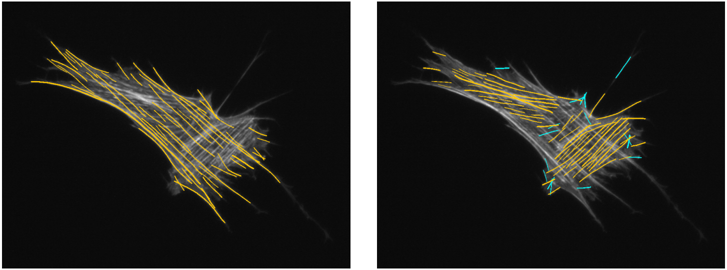
Taylor expansion at $\eta' = (\theta',\xi')$ gives a joint Gaussian CLT,

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

and projection to the θ coordinate preserves Gaussianity.

Application: Stem Cell Diversification (H. and Eltzner, 2018)

Actin-myosin structure of an adult stem cell after 16 hours.



Left: $m_1 =$ main orienation field filament pixels.
Right: $m_2 =$ smaller orienation field filament pixels,
Cyan: $m_3 =$ “rogue” filament pixels.

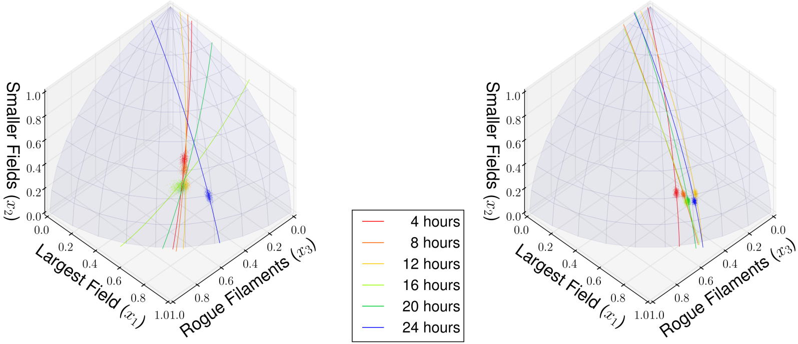
Composite data $m = m_1 + m_2 + m_3$ mapped to a sphere:

$$\left(\sqrt{\frac{m_1}{m}}, \sqrt{\frac{m_2}{m}}, \sqrt{\frac{m_3}{m}}\right)$$

Applying the Bootstrap Two-Sample Test

	nested mean		jointly great circle and nested mean	
Time	≤ 1 kPa	≥ 10 kPa	≤ 1 kPa	≥ 10 kPa
4h–8h	0.120	$< 10^{-3}$	0.308	$< 10^{-3}$
8h–12h	$< 10^{-3}$	$< 10^{-3}$	0.024	$< 10^{-3}$
12h–16h	0.126	$< 10^{-3}$	0.008	$< 10^{-3}$
16h–20h	0.468	0.626	0.494	0.462
20h–24h	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	0.014

Visualization



Left: $\leq$ 1 kPa.

Right: $\geq$ 10 kPa

Wrap up and Outlook

Wrap up:

- from the BP/BL-CLT requiring (A1), (A2), (A5), (A6) we have gone to

Wrap up and Outlook

Wrap up:

- from the BP/BL-CLT requiring (A1), (A2), (A5), (A6) we have gone to
- a more general CLT requiring only (A1), (Taylor), (Donsker);

Wrap up and Outlook

Wrap up:

- from the BP/BL-CLT requiring (A1), (A2), (A5), (A6) we have gone to
- a more general CLT requiring only (A1), (Taylor), (Donsker);
- smeariness may be a finite sample issue:

Wrap up and Outlook

Wrap up:

- from the BP/BL-CLT requiring (A1), (A2), (A5), (A6) we have gone to
- a more general CLT requiring only (A1), (Taylor), (Donsker);
- smeariness may be a finite sample issue:
- we have dealt with generalized, nested and flag Fréchet means and their bootstrap inference.

Wrap up and Outlook

Wrap up:

- from the BP/BL-CLT requiring (A1), (A2), (A5), (A6) we have gone to
- a more general CLT requiring only (A1), (Taylor), (Donsker);
- smeariness may be a finite sample issue:
- we have dealt with generalized, nested and flag Fréchet means and their bootstrap inference.

Wrap up and Outlook

Wrap up:

- from the BP/BL-CLT requiring (A1), (A2), (A5), (A6) we have gone to
- a more general CLT requiring only (A1), (Taylor), (Donsker);
- smeariness may be a finite sample issue:
- we have dealt with generalized, nested and flag Fréchet means and their bootstrap inference.

Open challenges:

- (A1): uniqueness of Fréchet means?

Wrap up and Outlook

Wrap up:

- from the BP/BL-CLT requiring (A1), (A2), (A5), (A6) we have gone to
- a more general CLT requiring only (A1), (Taylor), (Donsker);
- smeariness may be a finite sample issue:
- we have dealt with generalized, nested and flag Fréchet means and their bootstrap inference.

Open challenges:

- (A1): uniqueness of Fréchet means?
- Validity of the Taylor expansion of the Fréchet function?

Wrap up and Outlook

Wrap up:

- from the BP/BL-CLT requiring (A1), (A2), (A5), (A6) we have gone to
- a more general CLT requiring only (A1), (Taylor), (Donsker);
- smeariness may be a finite sample issue:
- we have dealt with generalized, nested and flag Fréchet means and their bootstrap inference.

Open challenges:

- (A1): uniqueness of Fréchet means?
- Validity of the Taylor expansion of the Fréchet function?
- Manifold stability for GPCs etc. e.g. on shape spaces?

Wrap up and Outlook

Wrap up:

- from the BP/BL-CLT requiring (A1), (A2), (A5), (A6) we have gone to
- a more general CLT requiring only (A1), (Taylor), (Donsker);
- smeariness may be a finite sample issue:
- we have dealt with generalized, nested and flag Fréchet means and their bootstrap inference.

Open challenges:

- (A1): uniqueness of Fréchet means?
- Validity of the Taylor expansion of the Fréchet function?
- Manifold stability for GPCs etc. e.g. on shape spaces?
- $\exists$ arbitrary smeariness on (non?)compact spaces?

Wrap up and Outlook

Wrap up:

- from the BP/BL-CLT requiring (A1), (A2), (A5), (A6) we have gone to
- a more general CLT requiring only (A1), (Taylor), (Donsker);
- smeariness may be a finite sample issue:
- we have dealt with generalized, nested and flag Fréchet means and their bootstrap inference.

Open challenges:

- (A1): uniqueness of Fréchet means?
- Validity of the Taylor expansion of the Fréchet function?
- Manifold stability for GPCs etc. e.g. on shape spaces?
- $\exists$ arbitrary smeariness on (non?)compact spaces?
- N.B: $\exists$ stickiness on all nonmanifold stratified spaces?

Wrap up and Outlook

Wrap up:

- from the BP/BL-CLT requiring (A1), (A2), (A5), (A6) we have gone to
- a more general CLT requiring only (A1), (Taylor), (Donsker);
- smeariness may be a finite sample issue:
- we have dealt with generalized, nested and flag Fréchet means and their bootstrap inference.

Open challenges:

- (A1): uniqueness of Fréchet means?
- Validity of the Taylor expansion of the Fréchet function?
- Manifold stability for GPCs etc. e.g. on shape spaces?
- $\exists$ arbitrary smeariness on (non?)compact spaces?
- N.B: $\exists$ stickiness on all nonmanifold stratified spaces?
- $\exists$ antismeariness (crispness?) $n^\gamma x_n = O_p(1)$ with $\gamma > 1/2$?

Wrap up and Outlook

Wrap up:

- from the BP/BL-CLT requiring (A1), (A2), (A5), (A6) we have gone to
- a more general CLT requiring only (A1), (Taylor), (Donsker);
- smeariness may be a finite sample issue:
- we have dealt with generalized, nested and flag Fréchet means and their bootstrap inference.

Open challenges:

- (A1): uniqueness of Fréchet means?
- Validity of the Taylor expansion of the Fréchet function?
- Manifold stability for GPCs etc. e.g. on shape spaces?
- $\exists$ arbitrary smeariness on (non?)compact spaces?
- N.B: $\exists$ stickiness on all nonmanifold stratified spaces?
- $\exists$ antismeariness (crispness?) $n^\gamma x_n = O_p(1)$ with $\gamma > 1/2$?
- $\vdots$

References

Afsari, B. (2011). Riemannian L^p center of mass: existence, uniqueness, and convexity. *Proceedings of the American Mathematical Society 139*, 655–773.

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

Bhattacharya, R. and L. Lin (2017). Omnibus CLTs for Fréchet means and nonparametric inference on non-Euclidean spaces. *Proceedings of the American Mathematical Society 145*(1), 413–428.

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.

Bredon, G. E. (1972). *Introduction to Compact Transformation Groups*, Volume 46 of *Pure and Applied Mathematics*. New York: Academic Press.

Davis, A. W. (1977). Asymptotic theory for principal component analysis: non-normal case. *Australian Journal of Statistics 19*, 206–212.

Eltzner, B. (2019). Measure dependent asymptotic rate of the mean: Geometrical and topological smeariness. *arXiv preprint arXiv:1908.04233*.

Eltzner, B., F. Galaz-García, S. F. Huckemann, and W. Tuschmann (2019). Stability of the cut locus and a central limit theorem for Fréchet means of Riemannian manifolds. *arXiv*.

Eltzner, B. and S. F. Huckemann (2018). A smeary central limit theorem for manifolds with application to high dimensional spheres. accepted (AOS), arXiv:1801.06581.

Groisser, D. (2005). On the convergence of some Procrustean averaging algorithms. *Stochastics: Internatl. J. Probab. Stochstic. Processes 77*(1), 51–60.

Hodson, F. R., P. H. Sneath, and J. E. Doran (1966). Some experiments in the numerical analysis of archeological data. *Biometrika 53*, 411–324.

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.

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. (2012). On the meaning of mean shape: Manifold stability, locus and the two sample test. *Annals of the Institute of Statistical Mathematics 64*(6), 1227–1259.

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. F. and B. Eltzner (2018). Backward nested descriptors asymptotics with inference on stem cell

References

Huckemann, S. F. and B. Eltzner (2018). Backward nested descriptors asymptotics with inference on stem cell differentiation. *The Annals of Statistics* (5), 1994 – 2019.

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

Karcher, H. (1977). Riemannian center of mass and mollifier smoothing. *Communications on Pure and Applied Mathematics XXX*, 509–541.

Kendall, W. S. (1990). Probability, convexity, and harmonic maps with small image I: Uniqueness and fine existence. *Proceedings of the London Mathematical Society 61*, 371–406.

Le, H. (1998). On the consistency of Procrustean mean shapes. *Advances of Applied Probability (SGSA) 30*(1), 53–63.

Le, H. and D. Barden (2014). On the measure of the cut locus of a Fréchet mean. *Bulletin of the London Mathematical Society 46*(4), 698–708.

Mardia, K. V. and P. E. Jupp (2000). *Directional Statistics*. New York: Wiley.

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.

Pennec, X. (2018). Barycentric subspace analysis on manifolds. *The Annals of Statistics 46*(6A), 2711–2746.

Ruymgaart, F. H. and S. Yang (1997). Some applications of Watson's perturbation approach to random matrices. *Journal of Multivariate Analysis 60*(1), 48–60.

Small, C. G. (1996). *The Statistical Theory of Shape*. New York: Springer-Verlag.

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

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

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.