Huckemann

Euclidear

BP/BL-CLT

(A2): Cut Loc

(710). Lilip. 1

(AU). Officary

Generalizatio

Application

Outlook

Reference

References

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



BP/BL-CLT

BP/BL-GLI

(A5): Emp. F

(A6): Smoor

PCA/

Applications

Outlook

References

Reference

- 1 Euclidean Statistics to be Generalized
- 2 The BP/BL-CLT (2005/2017)
- 3 Condition (A2) Dissected: The Cut Locus
- 4 Condition (A5) Dissected: Empirical Processes
- **6** Condition (A6) Dissected: Smeariness
- 6 Generalized Fréchet Means
- **7** PCA, Their Bootstrap Inference and Applications
- 8 Wrap Up and Outlook

Huckemann

Euclidean

(A5): Emp. P

(A6): Smeary

PCA/

Applications

Outlook

Reference

References

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)









Huckemann

Euclidean

RP/RL-CLT

(AO), Cut I aa

(A5): Emp. P

(A6): Smeary

Conorolization

PCA/

Application

Outlook

Reference

References

Motivation

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

Huckemann

Euclidean

BP/BL-CLT

(A2): Cut Loc

(A5): Emp. P

(A6): Smeary

Generalization

PCA/ Application

Outlook

Reference

Reference

Motivation

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

(A2): Cut Loc

(A5): Emp. P

(A6): Smeary

Generalization

PCA/ Applications

. . .

1 telefelles

References

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,

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

(A6): Smeary

Generalization

PCA/ Applications

Outlook

Reference

References

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 $n \to \infty$,

(A6): Smeary

DCA/

Applications

Outlook

Reference

References

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 → ∞,
- or even exact for finite $n (\sim Thomas')$ talk at 15:30).

(A6): Smeary

DCA/

Applications

Outlook

Reference

References

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 → ∞,
- or even exact for finite $n (\sim Thomas')$ talk at 15:30).

(Ab): Silleary

PCA/

Applications

Outlook

Reference

References

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 $n (\sim Thomas')$ talk at 15:30).

Here we do nonparametric asymptotics.

(40). 0.....

(710). Omeary

PCA/

Application

Outloo

Reference

References

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 $n \to \infty$

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

Theorem (The Central Limit Theorem)

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

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

(AC), Cmaam

PCA/

Application

Outloor

Reference

References

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 $n \to \infty$

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

Theorem (The Central Limit Theorem)

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

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

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

References

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 $n \to \infty$

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

Theorem (The Central Limit Theorem)

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

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

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

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

Euclidean

DD/DL CLT

BP/BL-CLI

(Ab): Emp. F

(A6): Smeary

Generalizatio

Application

Outlook

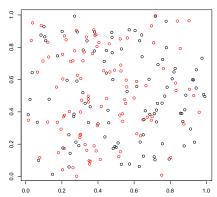
Reference

References

Test for Equality of Means

Two groups of random variables

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



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

Huckemann

Euclidean

RD/RL CLT

(A2): Cut Loc

(A5): Emp. Pi

(A6): Smeary

Conoralization

Gorioranzanorn

Application

Outlook

Reference

References

Hotelling Test for Equality of Means

• Under H_0 and either cov[X] = cov[Y] or $n/m \rightarrow 1$,

$$T^{2} := \frac{n + m - 2}{\frac{1}{n} + \frac{1}{m}} (\bar{X}_{n} - \bar{Y}_{m})^{T} (n\Sigma_{n}^{X} + m\Sigma_{m}^{Y})^{-1} (\bar{X}_{n} - \bar{Y}_{m})$$

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

Huckemann

Euclidean

DD/DL CLT

(A2): Cut Loc

(A5): Emp. Pr

(A6): Smeary

Concretion

PCA/

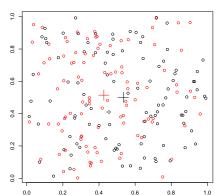
. .

Hotelling Test for Equality of Means

• Under H_0 and either cov[X] = cov[Y] or $n/m \to 1$,

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

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



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

Huckemann

Euclidean

DD/DL CLT

BP/BL-CLI

(A5): 5:::: 5

(A3). Lilip. 1

(Ab). Silleary

Generalization

PCA/

Outlook

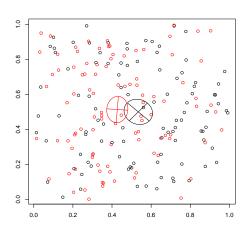
Poforonce

References

Principal Component Analysis (PCA)

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

• With eigenvectors $\Gamma = (\gamma_1, \dots, \gamma_m) \in SO(m)$ to



Huckemann

Euclidean

BP/BL-CLT

(A5): Emp. E

(A3). Lilip. 1

(Ab). Silleary

Generalizatio

PCA/

Outlook

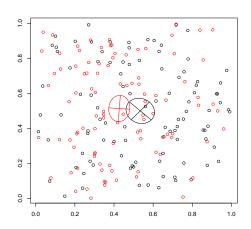
Poforonce

References

Principal Component Analysis (PCA)

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

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



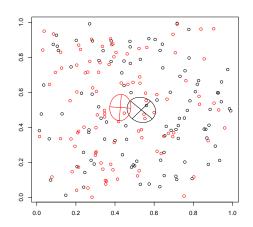
Huckemann

Euclidean

Principal Component Analysis (PCA)

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

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



Huckemann

Euclidean

BP/BL-CLT

BP/BL-CL

S

A2): Gut Lo

(A5): Emp. F

(A6): Smeary

Generalization

PCA/

Application

Outlook

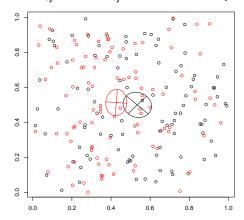
Reference

References

Principal Component Analysis (PCA)

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

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



Huckemann

Fuclidea

BP/BL-CLT

(40) 0 11

(Δ5): Emp. P

(A6): Sr

Conordization

Controlanzation

Application

Outlook

Reference

References

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) .

Huckemann

Fuclidea

BP/BL-CLT

....

(A5): Emp. P

(Ab): Smeary

Generalization

PCA/ Application

.

- -----

. .0.0.0.0.00

References

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

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

Huckemann

Euclidear

BP/BL-CLT

(A2): Gut Loc

(A5): Emp. P

(A6): Smeary

PCA/

Application

Outlook

Reference

References

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),

Huckemann

Euclidear

BP/BL-CLT

(/IL). Out Loc

(710). Linp. 1

(Ao). Silleary

PCA/

Application

Outlook

nelelelices

References

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

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

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

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

CLT for Fréchet Means Huckemann

BP/BL-CLT

$$\mathbf{X} \mapsto \mathbf{p}$$

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

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

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

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

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

 $x \mapsto \rho(X, \phi(x))^2$ a. s. $\in \mathcal{C}^2(V)$, (A3) $\mu_n \stackrel{\mathbb{P}}{\to} \mu$ for a measurable selection of sample means

Huckemann

Euclidear

BP/BL-CLT

BP/BL-CLI

(AE), Emm. D.

(40) 0

(A6): Smear

Generalization

PCA/ Application

Application

Outlook

nelelelice

References

The Bhattacharya and Patrangenaru (2005) CLT More assumptions:

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

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

BP/BL-CLI

(A5): Emp. Pi

(A6): Smeary

(Ab). Silleary

PCA/ Application

Application

Outloor

References

References

The Bhattacharya and Patrangenaru (2005) CLT More assumptions:

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

 $\exists H := \mathbb{E} \left[H(X, \phi^{-1}(\mu)) \right], H(X, X) = \operatorname{Hess}|_X \rho^2(X, \phi(X))$
(we cannot do without, e.g. valid on compact M)

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

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

BP/BL-CLT

(A5): Emp. Pi

(A6): Smeary

(710). Omoary

PCA/ Applications

Application

Outloor

References

References

The Bhattacharya and Patrangenaru (2005) CLT More assumptions:

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

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

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

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

(A6) H is not singular.

(A6): Smeary

Generalizati

PCA/ Applications

Outloo

References

References

The Bhattacharya and Patrangenaru (2005) CLT More assumptions:

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

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

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

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

(A6) H is not singular.

References

The Bhattacharya and Patrangenaru (2005) CLT More assumptions:

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

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

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

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

(A6) H is not singular.

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

Under Assumptions (A1) — (A6):

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

Huckemann

Euclidear

BP/BL-CLT

(A2): Cut Locu

(A5): Emp. P

(A6): Sme

Generalization

PCA/

Application

Outlook

Reference

References

Sketch of Proof

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

Huckemann

Euclidean

BP/BL-CLT

(A2): Cut Loc

(A5): Emp. P

(A6): Smeary

Generalization

PCA/ Application

, ipplioatio

Reference

Reference

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): $x_n \stackrel{\text{a.s.}}{\rightarrow} 0$.

Euclidean

BP/BL-CLT

(--)-----

(A3). Lilip. 1

(A6): Smeary

Generalization

PCA/ Application

Outlook

Reference

Reference

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): $x_n \stackrel{\text{a.s.}}{\rightarrow} 0$.
- Fréchet functions:

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

Euclidean

BP/BL-CLT

(AZ). Out Loc

(AC) C

(AU). Silleary

PCA/

Applications

Outlook

References

References

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): $x_n \stackrel{\text{a.s.}}{\rightarrow} 0$.
 - Fréchet functions:

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

• Taylor expansion (with suitable \tilde{x} between 0 and x_0),

$$\sqrt{n} \operatorname{grad}|_{x=x_0} F_n(x) = \sqrt{n} \operatorname{grad}|_{x=0} F_n(x) + \operatorname{Hess}|_{x=\widetilde{x}} F_n(x) \sqrt{n} x_0$$
,
(A2) \Rightarrow holds also a.s. for random $x_0 = x_n$

BP/BL-CLT

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): $x_n \stackrel{\text{a.s.}}{\rightarrow} 0$.
- Fréchet functions:

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

• Taylor expansion (with suitable \tilde{x} between 0 and x_0),

$$\sqrt{n} \operatorname{grad}|_{x=x_0} F_n(x) = \sqrt{n} \operatorname{grad}|_{x=0} F_n(x) + \operatorname{Hess}|_{x=\widetilde{x}} F_n(x) \sqrt{n} x_0$$
,
(A2) \Rightarrow holds also a.s. for random $x_0 = x_n$

• generalized weak law $(n \to \infty \text{ and } x_0 \to 0)$

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

Euclidean

BP/BL-CLT

(--)-----

(AC) C

(AU). Officary

PCA/

0 11 1

D - (-

References

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): $x_n \stackrel{\text{a.s.}}{\rightarrow} 0$.
- Fréchet functions:

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

• Taylor expansion (with suitable \tilde{x} between 0 and x_0),

$$\sqrt{n} \operatorname{grad}|_{x=x_0} F_n(x) = \sqrt{n} \operatorname{grad}|_{x=0} F_n(x) + \operatorname{Hess}|_{x=\widetilde{x}} F_n(x) \sqrt{n} x_0$$
,
(A2) \Rightarrow holds also a.s. for random $x_0 = x_n$

• generalized weak law $(n \to \infty \text{ and } x_0 \to 0)$

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

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

CLT for Fréchet Means Huckemann

Euclidean

BP/BL-CLT

A5): Emp. Pr

Generalization

Applications
Outlook

Refere

References

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 \tilde{x} between 0 and x_0),

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

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

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

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

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

Huckemann

Euclidean

BP/BL-CLT

(A2): Cut Locus

(A5): Emp. P

(A6): Smeary

Generalization

PCA/ Application:

Outloo

Reference:

References

(A2) Dissected: The Cut Locus

Corollary (2.3 from Bhattacharya and Lin (2017))

Instead of

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

$$\mathbf{X}\mapsto
hoig(\mathbf{X},\phi(\mathbf{X})ig)^2$$
 is a.s. $\in\mathcal{C}^2(\mathbf{V})$

it suffices to require

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

Huckemann

Euclidean

BP/BL-CLT

(A2): Cut Locus

(A5): Emp. P

(Ab): Smeary

PCA/

Applications

Outloo

References

References

(A2) Dissected: The Cut Locus

Corollary (2.3 from Bhattacharya and Lin (2017))

Instead of

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

$$\mathbf{X}\mapsto
hoig(\mathbf{X},\phi(\mathbf{X})ig)^2$$
 is a.s. $\in \mathcal{C}^2(\mathbf{V})$

it suffices to require

(C) there is a neighborhood $W \subseteq M$ of the cut locus $Cut(\mu)$ 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).

onoman

RP/RL-CLT

(A2): Cut Locus

(A6): Smeary

PCA/ Applications

References

References

(A2) Dissected: The Cut Locus

Corollary (2.3 from Bhattacharya and Lin (2017))

Instead of

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

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

it suffices to require

(C) there is a neighborhood $W \subseteq M$ of the cut locus $Cut(\mu)$ 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))

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

Huckemann

Euclidean

BP/BL-CLT

(A2): Cut Locus

(A5): Emp. P

(A6): Smeary

Generalization

PCA/ Application

Application

Reference

References

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

Huckemann

Euclidean

BP/BL-CLT

(A2): Cut Locus

.

Au). Officary

PCA/

Applications

D (

References

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

Theorem (Eltzner et al. (2019))

1 *M* topologically stable ⇒ *M* geometrically stable;

Huckemann

Euclidean

BP/BL-CLT

(A2): Cut Locus

(A5): Emp. P

(A6): Smeary

Generalizati

Applications

Outlook

References

References

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

- M topologically stable ⇒ M geometrically stable;
- ② M compact ⇒ M topologically stable;

Huckemann

Euclidean

BP/BL-CLT

(A2): Cut Locus

PCA/ Application

Outlook

Reference

References

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

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

Huckemann

Euclidean

BP/BL-CLT

(A2): Cut Locus

(A5): Emp. F

(AO). Silleary

PCA/

Application

nelelelice:

References

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

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

Huckemann

Euclidean

BP/BL-CLT

(A2): Cut Locus

(A5): Emp. F

(AO). Silleary

PCA/

Application

nelelelice:

References

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

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

Huckemann

Euclidean

BP/BL-CLT

(A2): Cut Locus

(A6): Smeary

(* 10)1 0111011,

PCA/ Applications

Outlook

References

References

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

Theorem (Eltzner et al. (2019))

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

Huckemann

Euclidear

BP/BL-CLT

(A2): Cut Locus

(A5): Emp. Pr

(A6): Smeary

Generalization

PCA/ Application

0.41--1.

Reference

References

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. $x \geq 0$

Huckemann

Euclidear

BP/BL-CLT

(A2): Cut Locus

(A5): Emp. Pi

(A6): Smeary

Generalizat

PCA/ Application

Outlook

Reference

References

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. $x \ge 0$

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

Huckemann

Euclidear

BP/BL-CLT

(A2): Cut Locus

(A5): Emp. Pi

(A6): Smeary

PCA/

Application

Outlook

References

References

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. $x \ge 0$

$$2nF_n(x) = \sum_{X_j \in [x-\pi,\pi]} (X_j - x)^2 + \sum_{X_j < x-\pi} (X_j + 2\pi - x)^2$$
$$= \sum_{i=1}^n (X_j - x)^2 + 4\pi \sum_{X_i < x-\pi} (X_j - x + \pi)$$

Huckemann

Euclidear

BP/BL-CLT

(A2): Cut Locus

(A5): Emp. Pi

(Ab): Smeary

PCA/

Application

Outlook

Reference

References

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. $x \ge 0$

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

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

(A5): Emp. P

(A6): Smeary

_

PCA/

Application

Outlook

Reference

References

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. $x \ge 0$

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

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

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

(A6): Smeary

Generalization

PCA/ Applications

Outlook

References

References

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. $x \ge 0$

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

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

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

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

Huckemann

Euclidear

BP/BL-CLT

(A2): Cut Loci

(A5): Emp. Pr.

(A6): Smear

Generalizations

PCA/

Application

Outlook

. 1010101100

References

A More General CLT

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

Huckemann

Euclidean

BP/BL-CLT

(A2): Cut Loc

(A5): Emp. Pr.

(A6): Smeary

Conorolization

Application

Outlook

Reference

References

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 $T_1, \ldots, T_m \ne 0$,

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

Huckemann

Euclidean

BP/BL-CLT

(A2): Cut Lo

(A5): Emp. Pr.

(A6): Smeary

Generalization

PCA/

Application

D (

110101011001

References

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 $T_1, \ldots, T_m \ne 0$,

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

• Donsker cond.: $\exists \ \dot{\rho}_0(X) := \operatorname{grad}_x \rho(X, \phi(x))^2|_{x=0}$ a.s. with $\exists \operatorname{cov}[\dot{\rho}_0(X)]$, m'ble function $\dot{\rho} : M \to \mathbb{R}$ such that $\mathbb{E}[\dot{\rho}(X)^2] < \infty$ and $\forall x_1, x_2 \in V$,

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

Huckemann

Euclidean

BP/BL-CLT

(A2): Cut Loc

(A5): Emp. Pr.

(A6): Smeary

Generalization

DCA/

Application

Outlook

Reference

References

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 $T_1, \ldots, T_m \ne 0$,

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

• Donsker cond.: $\exists \ \dot{\rho}_0(X) := \operatorname{grad}_x \rho(X, \phi(x))^2|_{x=0}$ a.s. with $\exists \operatorname{cov}[\dot{\rho}_0(X)]$, m'ble function $\dot{\rho} : M \to \mathbb{R}$ such that $\mathbb{E}[\dot{\rho}(X)^2] < \infty$ and $\forall x_1, x_2 \in V$,

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

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

Huckemann

Euclidean

BP/BL-CLT

(A2): Cut Loc

(A5): Emp. Pr.

(A6): Smeary

Generalization

DCA/

Application

Outlook

Reference

References

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 $T_1, \ldots, T_m \ne 0$,

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

• Donsker cond.: $\exists \ \dot{\rho}_0(X) := \operatorname{grad}_x \rho(X, \phi(x))^2|_{x=0}$ a.s. with $\exists \operatorname{cov}[\dot{\rho}_0(X)]$, m'ble function $\dot{\rho} : M \to \mathbb{R}$ such that $\mathbb{E}[\dot{\rho}(X)^2] < \infty$ and $\forall x_1, x_2 \in V$,

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

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

Huckemann

Euclidean

BP/BL-CLT

(AZ). Gut Loc

(A5): Emp. Pr.

(Ab): Smeary

Don

Application

References

References

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 $T_1, \ldots, T_m \ne 0$,

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

• Donsker cond.: $\exists \ \dot{\rho}_0(X) := \operatorname{grad}_x \rho(X, \phi(x))^2|_{x=0}$ a.s. with $\exists \operatorname{cov}[\dot{\rho}_0(X)]$, m'ble function $\dot{\rho} : M \to \mathbb{R}$ such that $\mathbb{E}[\dot{\rho}(X)^2] < \infty$ and $\forall x_1, x_2 \in V$,

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

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

Theorem (Eltzner and H. 2018) $\sqrt{n} Rx_n |Rx_n|^{r-2} \stackrel{\mathcal{D}}{\to} \mathcal{N}(0, \Sigma)$ (power component-wise), suitable $\Sigma > 0$.

Huckemann

Euclidean

BP/BL-CLT

(A5): Emp. Pr.

(A6): Smeary

(710): Omodiy

PCA/

Applications

Deference

References

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 $T_1, \ldots, T_m \ne 0$,

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

• Donsker cond.: $\exists \ \dot{\rho}_0(X) := \operatorname{grad}_x \rho(X, \phi(x))^2|_{x=0}$ a.s. with $\exists \operatorname{cov}[\dot{\rho}_0(X)]$, m'ble function $\dot{\rho} : M \to \mathbb{R}$ such that $\mathbb{E}[\dot{\rho}(X)^2] < \infty$ and $\forall x_1, x_2 \in V$,

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

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

Theorem (Eltzner and H. 2018) $\sqrt{n} Rx_n |Rx_n|^{r-2} \stackrel{\mathcal{D}}{\to} \mathcal{N}(0, \Sigma)$ (power component-wise), suitable $\Sigma > 0$. x_n has rate $n^{-\frac{1}{2(r-1)}}$, is r-2-smearv.

Huckemann

Fuclidea

RP/RL-CLT

(A0), Cut I aa

(A5): Emp. Pi

(A6): Smeary

Canavalization

PCA/

Application

Outlook

Reference

References

Smeariness: The Beast is Real

∃ arbitrary smeariness on S¹ (Hotz and H., 2015);

Huckemann

Fuclidea

RP/RL-CLT

(A2): Cut Loc

(A5): Emp. P

(A6): Smeary

(* 10)1 01110411)

PCA/ Applications

Application

Outlook

releterice

References

- ∃ arbitrary smeariness on S¹ (Hotz and H., 2015);
- $\exists r-2=2$ smeariness on \mathbb{S}^m for all $m\in\mathbb{N}$ (Eltzner and H., 2018);

Huckemann

Fuclidear

BP/BL-CLT

(A2): Cut Loc

(A5): Emp. P

(A6): Smeary

_

PCA/

Application

Outlook

1 telefelles

References

- ∃ arbitrary smeariness on S¹ (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);

Huckemann

Fuclidea

BP/BL-CLT

(A2): Cut Loc

(A5): Emp. P

(A6): Smeary

Generalizatio

PCA/ Application

Outland

i telefelice:

References

- ∃ arbitrary smeariness on S¹ (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 (!);

Huckemann

Fuclidear

RP/RL-CLT

(A2): Cut Loc

(A5): Emp. P

(A6): Smeary

PCA/

Application

Outlook

References

References

- ∃ arbitrary smeariness on S¹ (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.

Huckemann

Euclidean

DD/DL CLT

(A5). ⊑IIIp. F

(A6): Smeary

Generalization

PCA/

Applicatio

Outlook

i telefelice

References

Finite Sample Smeariness

Table 1.5 Orientations of 76 turtles after laying eggs (Gould's data cited by

Direction (in degrees) clockwise from north									
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				

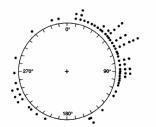
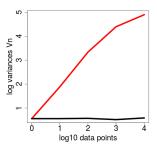


Figure 1.5 Circular plot of the turtle data of Table 1.5.

from Mardia and Jupp (2000).



Bootstrapped variance black = Euclidean in

$$[-\pi,\pi]\subset\mathbb{R},$$

red = circular $\sim n^{2/3}$?

Huckemann

Cuelidade

DD/DL CLT

BP/BL-CLT

(--)- ----

(A5): Emp. P

(A6): Smeary

Generalization

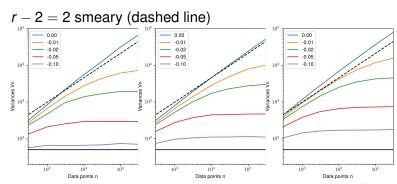
PCA/

. . .

Deference

Poforonco

Two-Smeariness (Eltzner and H. 2018)



On a sphere
$$\mathbb{S}^m$$
 with dimension (all derivatives $O(m^{-1/2})$)
 $m=2$ $m=10$ $m=100$

Huckemann

Fuclidear

BP/BL-CLT

(A2): Cut Lo

(A5): Emp. P

(A6): Smear

Generalizations

PCA/ Application

Outlook

Reference

References

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

Huckemann

Fuclidear

BP/BL-CLT

(A5): Emp. P

(A6): Smeary

Generalizations

PCA/ Application

Deference

Deference

Separating Data from Descriptor Space

- 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. Γ(Q) = space of geodesics on Q

Huckemann

Euclidea

BP/BL-CLT

(A2): Cut Lo

(A5): Emp. Pi

(A6): Smeary

Generalizations

PCA/

, ippiioatio

Outloor

i telefelice

Reference

Separating Data from Descriptor Space

- 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. Γ(Q) = space of geodesics on Q
- $\rho: Q \times P \to [0, \infty)$ continuous = link function

Huckemann

Euclidea

BP/BL-CLT

(A5): Emp. D

(A5): Emp. Pr

Generalizations

PCA/ Applications

Outlook

Reference

Reference

Separating Data from Descriptor Space

- 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
- $\gamma \in \operatorname{argmin}_{p \in P} \mathbb{E}[\rho(X, p)^2] = \operatorname{generalized}$ population Fréchet mean

Huckemann

Fuclidea

BP/BL-CLT

(A5): Emp. P

(A5): Emp. Pt

Generalizations

PCA/ Application

Outlook

Reference

References

Separating Data from Descriptor Space

- 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
- $\gamma \in \operatorname{argmin}_{p \in P} \mathbb{E}[\rho(X, p)^2] = \operatorname{generalized}$ population Fréchet mean
- $\hat{\gamma} \in \operatorname{argmin}_{p \in P} \sum_{j=1}^{b} \rho(X_j, p)^2 = \operatorname{generalized sample}$ Fréchet mean

Huckemann

Fuclidea

BP/BL-CLT

(A5): Emp. P

(A5): Emp. Pi

Generalizations

PCA/ Application

Outlook

Reference

References

Separating Data from Descriptor Space

- 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
- $\gamma \in \operatorname{argmin}_{p \in P} \mathbb{E}[\rho(X, p)^2] = \operatorname{generalized}$ population Fréchet mean
- $\hat{\gamma} \in \operatorname{argmin}_{p \in P} \sum_{j=1}^{b} \rho(X_j, p)^2 = \operatorname{generalized sample}$ Fréchet mean

Huckemann

Fuclidea

BP/BL-CLT

(AZ). Gut Loc

(A5): Emp. Pi

Generalizations

PCA/

Applicatio

Outloor

Reference

References

Separating Data from Descriptor Space

- 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
- $\gamma \in \operatorname{argmin}_{p \in P} \mathbb{E}[\rho(X, p)^2] = \operatorname{generalized}$ population Fréchet mean
- $\hat{\gamma} \in \operatorname{argmin}_{p \in P} \sum_{j=1}^{b} \rho(X_j, p)^2 = \operatorname{generalized sample}$ Fréchet mean
- If γ is unique,

Fuclidear

BP/BL-CLT

(A2): Cut Lo

(A5): Emp. P

Generalizations

Generalization

PCA/ Application

Outlool

Reference

References

Separating Data from Descriptor Space

- 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
- $\gamma \in \operatorname{argmin}_{p \in P} \mathbb{E}[\rho(X, p)^2] = \operatorname{generalized}$ population Fréchet mean
- $\hat{\gamma} \in \operatorname{argmin}_{p \in P} \sum_{j=1}^{b} \rho(X_j, p)^2 = \operatorname{generalized sample}$ Fréchet mean
- If γ is unique,
 - $\hat{\gamma} \rightarrow \gamma$ a.s. by S.H. (2011b) under weak regularity conditions

Reference

Separating Data from Descriptor Space

- 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
- $\gamma \in \operatorname{argmin}_{p \in P} \mathbb{E}[\rho(X, p)^2] = \operatorname{generalized}$ population Fréchet mean
- $\hat{\gamma} \in \operatorname{argmin}_{p \in P} \sum_{j=1}^{b} \rho(X_j, p)^2 = \operatorname{generalized sample}$ Fréchet mean
- If γ is unique,
 - $\hat{\gamma} \rightarrow \gamma$ a.s. by S.H. (2011b) under weak regularity conditions
 - $\sqrt{n}(\phi(\hat{\gamma}) \phi(\gamma)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma)$ by S.H. (2011a) if P is near γ a manifold with local chart ϕ , under regularity conditions adapted from (A1) (A6).

Huckemann

Fuclidear

BP/BL-CLT

(A2): Cut Loc

(A5): Emp. P

(A6): Smear

Gerreranzar

PCA/ Applications

O. 41- -1.

Reference

References

Application: The CLT of Classical PCA

Huckemann

Euclidea

BP/BL-CLT

(AC), Conseque

PCA/ Applications

Application

Reference

D (

Application: The CLT of Classical PCA

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

Huckemann

Fuclidea

BP/BL-CLT

(A5): Emp. P

(A6): Smeary

PCA/

Applications

Outlook

nelelelice

References

Application: The CLT of Classical PCA

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

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

Huckemann

Fuclidea

BP/BL-CLT

(A5): Emp. P

(A6): Smeary

PCA/

Applications

Outlook

nelelelices

References

Application: The CLT of Classical PCA

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

$$\operatorname{cov}[X] = V \wedge V^{T}, \lambda_{1} = \dots = \lambda_{k} > \lambda_{k+1} \geq \dots \geq \lambda_{m} > 0;$$

$$\operatorname{cov}[X_{1}, \dots, X_{n}] = \hat{V} \hat{\Lambda} \hat{V}^{T}, \hat{\lambda}_{1} \geq \dots \geq \hat{\lambda}_{k} \geq \hat{\lambda}_{k+1} \geq \dots \hat{\lambda}_{m} \geq 0;$$

Huckemann

Euclidear

BP/BL-CLT

(A5): Emp. P

(A6): Smeary

PCA/

Applications

Outlook

References

References

Application: The CLT of Classical PCA

$$Q = \mathbb{R}^m, P = G(m, k) \ni p = \operatorname{span}(\widehat{v_{k+1}, \dots, v_m})^{\perp}:$$

$$\operatorname{cov}[X] = V \wedge V^T, \ \lambda_1 = \dots = \lambda_k > \lambda_{k+1} \ge \dots \ge \lambda_m > 0;$$

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

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

Huckemann

Euclidean

BP/BL-CLT

(A5): Emp. P

(A6): Smeary

PCA/

Applications

Outlook

Reference

References

Application: The CLT of Classical PCA

$$Q = \mathbb{R}^m, P = G(m, k) \ni p = \operatorname{span}(\widehat{v_{k+1}, \dots, v_m})^{\perp}:$$

$$\operatorname{cov}[X] = V \wedge V^T, \ \lambda_1 = \dots = \lambda_k > \lambda_{k+1} \ge \dots \ge \lambda_m > 0;$$

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

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

$$\rho(X, p)^2 = \|X - WW^T X\|^2 = \|X\|^2 - \operatorname{trace}(W^T X X^T W);$$

Huckemann

Euclidear

BP/BL-CLT

(A5): Emp. P

(A6): Smeary

PCA/

Applications

Catioon

11010101100

References

Application: The CLT of Classical PCA

$$Q = \mathbb{R}^{m}, P = G(m, k) \ni p = \operatorname{span}(\widehat{v_{k+1}, \dots, v_{m}})^{\perp}:$$

$$\operatorname{cov}[X] = V \wedge V^{T}, \lambda_{1} = \dots = \lambda_{k} > \lambda_{k+1} \geq \dots \geq \lambda_{m} > 0;$$

$$\operatorname{cov}[X_{1}, \dots, X_{n}] = \hat{V} \hat{\Lambda} \hat{V}^{T}, \hat{\lambda}_{1} \geq \dots \geq \hat{\lambda}_{k} \geq \hat{\lambda}_{k+1} \geq \dots \hat{\lambda}_{m} \geq 0;$$

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

$$\rho(X, p)^{2} = \|X - WW^{T}X\|^{2} = \|X\|^{2} - \operatorname{trace}(W^{T}XX^{T}W);$$

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

Huckemann

Euclidear

BP/BL-CLT

(A5): Emp. Pi

(A6): Smeary

PCA/

Applications

Outlook

Reference

References

Application: The CLT of Classical PCA

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

$$\operatorname{cov}[X] = V \wedge V^{T}, \ \lambda_{1} = \dots = \lambda_{k} > \lambda_{k+1} \geq \dots \geq \lambda_{m} > 0;$$

$$\operatorname{cov}[X_{1}, \dots, X_{n}] = \widehat{V} \hat{\Lambda} \widehat{V}^{T}, \ \hat{\lambda}_{1} \geq \dots \geq \hat{\lambda}_{k} \geq \hat{\lambda}_{k+1} \geq \dots \hat{\lambda}_{m} \geq 0;$$

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

$$\rho(X, p)^{2} = \|X - WW^{T}X\|^{2} = \|X\|^{2} - \operatorname{trace}(W^{T}XX^{T}W);$$

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

$$\rho(X, p')^{2} - \rho(X, p)^{2} = \operatorname{trace}(W^{T}XX^{T}W) - \operatorname{trace}(W'^{T}XX^{T}W'),$$

Euclidean

BP/BL-CLT

(A5): Emp. Pi

(A6): Smeary

PCA/

Applications

Outlook

Reference

References

Application: The CLT of Classical PCA

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

$$\operatorname{cov}[X] = V \wedge V^T, \ \lambda_1 = \dots = \lambda_k > \lambda_{k+1} \ge \dots \ge \lambda_m > 0;$$

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

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

$$\rho(X, p)^2 = \|X - WW^T X\|^2 = \|X\|^2 - \operatorname{trace}(W^T X X^T W);$$

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

$$\rho(X, p')^2 - \rho(X, p)^2 = \operatorname{trace}(W^T X X^T W) - \operatorname{trace}(W'^T X X^T W'),$$
with $E[\|X\|^4] < \infty \Rightarrow \text{Donsker};$

Euclidean

BP/BL-CLT

(A5): Emp. Pi

(A6): Smeary

PCA/

Applications

Outlook

Reference

References

Application: The CLT of Classical PCA

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

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

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

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

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

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

$$\rho(X,p')^2 - \rho(X,p)^2 = \operatorname{trace}(W^TXX^TW) - \operatorname{trace}(W'^TXX^TW'),$$
with $E[\|X\|^4] < \infty \Rightarrow \operatorname{Donsker};$

$$\Rightarrow \sqrt{n} \text{ Gaussian CLT}.$$

Huckemann

Euclidean

BP/BL-CLT

(A2): Cut Loc

(A5): Emp. P

(A6): Smeary

Reneralizations

PCA/ Applications

пррпоцію

Reference

References

More Applications

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

Huckemann

Euclidean

BP/BL-CLT

(A2): Cut Loca

(A5): Emp. P

(A6): Smoor

(710). Omoary

adridianzation

PCA/ Applications

прриодио

Reference

References

- Geodesic PCA (GPCA) on Riemannian spaces by S.H et al. (2010):
 - P₁ = Γ(Q) = all geodesics on Q,
 → γ₁ and γ̂₁ = 1st geodesic PCs

Huckemann

PCA/ **Applications**

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

Huckemann

PCA/ **Applications**

- Geodesic PCA (GPCA) on Riemannian spaces by S.H. et al. (2010):
 - $P_1 = \Gamma(Q) = \text{all geodesics on } Q$, $\rightarrow \gamma_1$ 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

Huckemann

PCA/ **Applications**

- Geodesic PCA (GPCA) on Riemannian spaces by S.H. et al. (2010):
 - $P_1 = \Gamma(Q) = \text{all geodesics on } Q$, $\rightarrow \gamma_1$ 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

Huckemann

Euclidean

BP/BL-CLT

(A2): Cut Loc

(A5): Emp. F

(A6): Smear

Conorolizatio

PCA/

Applications

Outlook

Reference

References

- Geodesic PCA (GPCA) on Riemannian spaces by S.H et al. (2010):
 - P₁ = Γ(Q) = all geodesics on Q,
 → γ₁ and γ̂₁ = 1st geodesic PCs
 - $P_2 = \{ p \in \Gamma(Q) : \gamma_1 \perp p, \gamma_1 \cap p \neq \emptyset \}$ $\hat{P}_2 = \{ p \in \Gamma(Q) : \hat{\gamma}_1 \perp p, \hat{\gamma}_1 \cap p \neq \emptyset \}$ $\rightarrow \gamma_2$ and $\hat{\gamma}_2 = 2$ nd geodesic PCs
 - •
- 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 Γ(Q).

(A2): Cut Loc

(A5): Emp. P

(A6): Smear

PCA/

Applications

Outlook

Reference

References

- Geodesic PCA (GPCA) on Riemannian spaces by S.H et al. (2010):
 - P₁ = Γ(Q) = all geodesics on Q,
 → γ₁ and γ̂₁ = 1st geodesic PCs
 - $P_2 = \{ p \in \Gamma(Q) : \gamma_1 \perp p, \gamma_1 \cap p \neq \emptyset \}$ • $\hat{P}_2 = \{ p \in \Gamma(Q) : \hat{\gamma}_1 \perp p, \hat{\gamma}_1 \cap p \neq \emptyset \}$ • γ_2 and $\hat{\gamma}_2$ = 2nd geodesic PCs
 - •
- 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 Γ(Q).
- A shape space has an open and dense top-dimensional manifold part Q* (cf. Bredon (1972)).

(A2): Cut Loc

(A5): Emp. P

(A6): Smear

PCA/

Applications

Outlook

Reference

References

- Geodesic PCA (GPCA) on Riemannian spaces by S.H et al. (2010):
 - P₁ = Γ(Q) = all geodesics on Q,
 → γ₁ and γ̂₁ = 1st geodesic PCs
 - $P_2 = \{ p \in \Gamma(Q) : \gamma_1 \perp p, \gamma_1 \cap p \neq \emptyset \}$ • $\hat{P}_2 = \{ p \in \Gamma(Q) : \hat{\gamma}_1 \perp p, \hat{\gamma}_1 \cap p \neq \emptyset \}$ • γ_2 and $\hat{\gamma}_2$ = 2nd geodesic PCs
 - •
- 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 Γ(Q).
- A shape space has an open and dense top-dimensional manifold part Q* (cf. Bredon (1972)).

(A2): Cut Lo

(A5): Emp. Pi

(A6): Smeary

Generalizatio

PCA/ Applications

Outloo

Reference

Reference

- Geodesic PCA (GPCA) on Riemannian spaces by S.H et al. (2010):
 - P₁ = Γ(Q) = all geodesics on Q,
 → γ₁ and γ̂₁ = 1st geodesic PCs
 - $P_2 = \{ p \in \Gamma(Q) : \gamma_1 \perp p, \gamma_1 \cap p \neq \emptyset \}$ • $\hat{P}_2 = \{ p \in \Gamma(Q) : \hat{\gamma}_1 \perp p, \hat{\gamma}_1 \cap p \neq \emptyset \}$ • γ_2 and $\hat{\gamma}_2$ = 2nd geodesic PCs
- 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 Γ(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.

Huckemann

Fuclidear

DD/DL OLT

BP/BL-CLT

(A5), 5, 5, 5, 7

(AO): O----

Conorolizatio

PCA/

Applications

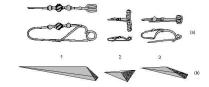
Outloo

Reference

References

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)).



Huckemann

Fuclidean

RD/RL CLT

BP/BL-CLI

(A5), 5, 5, 5, 7

(A5): Emp. Pr

Generalization

PCA/ Applications

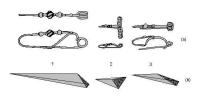
Outlook

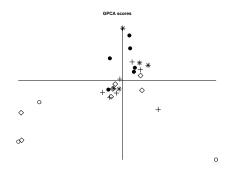
Reference

References

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.

Huckemann

Euclidean

BP/BL-CL

(A2): Cut Loc

(A5): Emp. P

(A6): Smeary

Generalization

PCA/ Applications

Outlook

neletetice

References

Two-Sample Descriptor Test

Data: X_1, \dots, X_n , $Y_1, \dots, Y_m \in Q$

Descriptors: p^{X} p^{Y} $\in P$

Coordinates: Z^X ϕ^{-1} \downarrow Z^Y $\in \mathbb{R}^D$

Huckemann

Euclidean

BP/BL-CLT

(A0), Cut I a

(A5): Emp. P

(710). Officially

Generalization

PCA/ Applications

Outlool

References

References

Two-Sample Descriptor Test

Data:
$$X_1, \dots, X_n \longrightarrow Y_1, \dots, Y_m \in Q$$

Descriptors: p^X $\downarrow \qquad \phi^{-1}$

Coordinates: Z^X ϕ^{-1} \downarrow Z^Y $\in \mathbb{R}$

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

$$\frac{nm}{n+m} (m+n-2) (Z^{X} - Z^{Y})^{T} \Big(n \widehat{\text{cov}}[Z_{1...n}^{X}] + m \widehat{\text{cov}}[Z_{1...m}^{Y}] \Big)^{-1} \cdot (Z^{X} - Z^{Y}) \sim \mathcal{T}^{2}(k, n+m-2)$$

Means Huckemann PCA/ **Applications**

CLT for Fréchet

Two-Sample Descriptor Test

 X_1,\ldots,X_n , Y_1,\ldots,Y_m Descriptors:

Coordinates:

Data:

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

 $\frac{nm}{n+m}(m+n-2)(Z^{X}-Z^{Y})^{T}\Big(n\widehat{\text{cov}}[Z_{1...n}^{X}]+m\widehat{\text{cov}}[Z_{1...m}^{Y}]\Big)^{-1}$ $(Z^X - Z^Y) \sim \mathcal{T}^2(k, n+m-2)$

But how to access $\widehat{\text{cov}}[Z_1^X]_n$ and $\widehat{\text{cov}}[Z_1^Y]_m$?

Huckemann

Euclidean

BF/BL-GL

(A5): Emp. D

(A6): Smearv

PCA/

Applications

0 11 1

Reference

References

Bootstrapping

For b = 1, ..., B, resample:

•
$$X_{1,b}^*, \dots, X_{n,b}^*$$
 from X_1, \dots, X_n gives $\widehat{\text{cov}}[Z_{1\dots n}^X]$

Huckemann

Euclidear

BP/BL-CL

(AE), E---- D

(710). Officary

PCA/

Applications

Outlook

- - - - -

For b = 1, ..., B, resample:

• $X_{1,b}^*, \dots, X_{n,b}^*$ from X_1, \dots, X_n gives $\widehat{\text{cov}}[Z_{1...n}^X]$

Bootstrapping

• $Y_{1,b}^*, \dots, Y_{m,b}^*$ from Y_1, \dots, Y_m gives $\widehat{\text{cov}}[Z_{1...m}^Y]$

Huckemann

PCA/

Applications

Bootstrapping

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

- $X_{1,b}^*, \ldots, X_{n,b}^*$ from X_1, \ldots, X_n gives $\widehat{\text{cov}}[Z_1^X]$
- $Y_{1,h}^*, \ldots, Y_{m,h}^*$ from Y_1, \ldots, Y_m gives $\widehat{\text{cov}}[Z_1^Y]_m$
- set $A = n\widehat{\text{cov}}[Z_1^X] + m\widehat{\text{cov}}[Z_1^Y]$

Lucilueari

BP/BL-CL1

(Δ5): Emp. P

(710). Lilip. 1

(710). Officary

PCA/

Applications

Outlook

Reference

Reference

Bootstrapping

For b = 1, ..., B, resample:

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

Again, for b = 1, ..., B', resample:

• $W_{1,b}^*, \dots, W_{n+m,b}^*$ from $X_1, \dots, X_n, Y_1, \dots, Y_m$

Lacilacan

BP/BL-CL1

(A5): Emp. P

(A6): Smearv

(710). Officially

PCA/

Applications

11010101100

References

Bootstrapping

For b = 1, ..., B, resample:

•
$$X_{1,b}^*, \dots, X_{n,b}^*$$
 from X_1, \dots, X_n gives $\widehat{\text{cov}}[Z_{1\dots n}^X]$

•
$$Y_{1,b}^*, \ldots, Y_{m,b}^*$$
 from Y_1, \ldots, Y_m gives $\widehat{\text{cov}}[Z_{1...m}^Y]$

• set
$$A = n\widehat{\operatorname{cov}}[Z_{1...n}^X] + m\widehat{\operatorname{cov}}[Z_{1...m}^Y]$$

•
$$W_{1,b}^*, \dots, W_{n+m,b}^*$$
 from $X_1, \dots, X_n, Y_1, \dots, Y_m$

• set
$$X_{j,b}^* = W_{j,b}^*$$
 for $j = 1, ..., n$

BP/BL-CL

(A5): Emp. P

(A6): Smeary

Coporolizatio

PCA/ Applications

Application

Reference

D - f - ...

Bootstrapping

For b = 1, ..., B, resample:

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

- $W_{1,b}^*, \dots, W_{n+m,b}^*$ from $X_1, \dots, X_n, Y_1, \dots, Y_m$
- set $X_{j,b}^* = W_{j,b}^*$ for j = 1, ..., n
- set $Y_{j,b}^* = W_{j+n,b}^*$ for j = 1, ..., m

Euclidean

BP/BL-CL

(A5): Emp. P

(A6): Smeary

PCA/ Applications

Outlook

References

Bootstrapping

For b = 1, ..., B, resample:

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

- $W_{1,b}^*, \dots, W_{n+m,b}^*$ from $X_1, \dots, X_n, Y_1, \dots, Y_m$
- set $X_{i,b}^* = W_{i,b}^*$ for j = 1, ..., n
- set $Y_{j,b}^* = W_{j+n,b}^*$ for j = 1, ..., m
- ullet compute the empirical quantile c_{1-lpha}^* such that

BP/BL-CL

(A5): Emp. P

(A6): Smeary

GOHOTAIIZA

PCA/ Applications

Outlook

Reference

References

Bootstrapping

For b = 1, ..., B, resample:

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

- $W_{1,b}^*, \ldots, W_{n+m,b}^*$ from $X_1, \ldots, X_n, Y_1, \ldots, Y_m$
- set $X_{i,b}^* = W_{i,b}^*$ for j = 1, ..., n
- set $Y_{i,b}^* = W_{i+n,b}^*$ for j = 1, ..., m
- compute the empirical quantile $c_{1-\alpha}^*$ such that
- $\mathbb{P}\left\{(Z^{X^*}-Z^{Y^*})^TA^{-1}(Z^{X^*}-Z^{Y^*}) \leq c_{1-\alpha}^* | X_1,\ldots,X_n,Y_1,\ldots,Y_m\right\} \geq 1-\alpha$

Euclidean

BP/BL-CL

(AE), E---- D

(AO). Officary

PCA/

Applications

Outloor

Reference

References

Bootstrapping

For b = 1, ..., B, resample:

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

Again, for b = 1, ..., B', resample:

- $W_{1,h}^*, \ldots, W_{n+m,h}^*$ from $X_1, \ldots, X_n, Y_1, \ldots, Y_m$
- set $X_{i,b}^* = W_{i,b}^*$ for j = 1, ..., n
- set $Y_{i,b}^* = W_{i+n,b}^*$ for j = 1, ..., m
- ullet compute the empirical quantile c_{1-lpha}^* such that

•
$$\mathbb{P}\left\{(Z^{X^*}-Z^{Y^*})^TA^{-1}(Z^{X^*}-Z^{Y^*}) \leq c_{1-\alpha}^* | X_1,\ldots,X_n,Y_1,\ldots,Y_m\right\} \geq 1-\alpha$$

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 .

(A5): Emp. P

(A6): Smeary

PCA/ Applications

Outlook

Reference

References

Improved Power

Recall, for b = 1, ..., 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 $j = 1, ..., n$

• set
$$Y_{j,b}^* = W_{j+n,b}^*$$
 for $j = 1, ..., m$

ullet compute the empirical quantile c_{1-lpha}^* 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 $(Z^X - Z^Y)^T A^{-1} (Z^X - Z^Y) > c_{1-\alpha}^*$

Lacilacan

BP/BL-CL1

(A5): Emp. P

(A3). Lilip. 1

(10)1 01110411

PCA/

Applications

Outlook

Reference

References

Improved Power

Recall, for b = 1, ..., B', resample:

- $W_{1,b}^*, \ldots, W_{n+m,b}^*$ from $X_1, \ldots, X_n, Y_1, \ldots, Y_m$
- set $X_{i,b}^* = W_{i,b}^*$ for j = 1, ..., n
- set $Y_{j,b}^* = W_{j+n,b}^*$ for j = 1, ..., m
- ullet compute the empirical quantile c_{1-lpha}^* 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, \ldots, X_n and $Y^{*,b}$ from Y_1, \ldots, Y_m

Euclidean

BP/BL-CL1

(A5): Emp. P

(A6): Smeary

0 " "

PCA/

Applications

Outlook

Reference

References

Improved Power

Recall, for b = 1, ..., B', resample:

- $W_{1,b}^*, \ldots, W_{n+m,b}^*$ from $X_1, \ldots, X_n, Y_1, \ldots, Y_m$
- set $X_{i,b}^* = W_{i,b}^*$ for j = 1, ..., n
- set $Y_{i,b}^* = W_{i+n,b}^*$ for j = 1, ..., 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 $(Z^X - Z^Y)^T A^{-1} (Z^X - Z^Y) > c_{1-\alpha}^*$

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

Luciiuean

BP/BL-CL1

(Δ5): Emp. P

(A6): Smoory

0------

PCA/

Applications

Outlook

Reference

References

Improved Power

Recall, for b = 1, ..., B', resample:

- $W_{1,b}^*, \ldots, W_{n+m,b}^*$ from $X_1, \ldots, X_n, Y_1, \ldots, Y_m$
- set $X_{i,b}^* = W_{i,b}^*$ for j = 1, ..., n
- set $Y_{j,b}^* = W_{j+n,b}^*$ for j = 1, ..., 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 $(Z^X - Z^Y)^T A^{-1} (Z^X - Z^Y) > c_{1-\alpha}^*$

- $X^{*,b}$ from X_1, \ldots, X_n and $Y^{*,b}$ from Y_1, \ldots, Y_m
- set $d^{X^*} = Z^{X^*} Z^{\mu^X}$, $d^{Y^*} = Z^{Y^*} Z^{\mu^Y}$
- ullet compute the empirical quantile c_{1-lpha}^* such that

Euclidean

BP/BL-CLT

(A5): Emp. P

(A3). Lilip. I

(10)1 01110411

PCA/

Applications

Outlook

Reference

References

Improved Power

Recall, for b = 1, ..., 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, ..., n
- set $Y_{j,b}^* = W_{j+n,b}^*$ for j = 1, ..., m
- ullet compute the empirical quantile c_{1-lpha}^* 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 $(Z^X - Z^Y)^T A^{-1} (Z^X - Z^Y) > c_{1-\alpha}^*$

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

Euclidean

BP/BL-CLT

(A5): Emp. P

(A3). Lilip. I

(10)1 01110411

PCA/

Applications

Outlook

Reference

References

Improved Power

Recall, for b = 1, ..., 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, ..., n
- set $Y_{j,b}^* = W_{j+n,b}^*$ for j = 1, ..., m
- ullet compute the empirical quantile c_{1-lpha}^* 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 $(Z^X - Z^Y)^T A^{-1} (Z^X - Z^Y) > c_{1-\alpha}^*$

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

Huckemann

Euclidean

BP/BL-CLT

(Δ5): Emp. Pi

(A3). LIIIp. 11

(710). Omoar

PCA/ Applications

Outlook

References

Improved Power

Recall, for b = 1, ..., B', resample:

- $W_{1,b}^*, \ldots, W_{n+m,b}^*$ from $X_1, \ldots, X_n, Y_1, \ldots, Y_m$
- set $X_{i,b}^* = W_{i,b}^*$ for j = 1, ..., n
- set $Y_{i,b}^* = W_{i+n,b}^*$ for j = 1, ..., 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 $d^{X^*} = Z^{X^*} Z^{\mu^X}$, $d^{Y^*} = Z^{Y^*} Z^{\mu^Y}$
- compute the empirical quantile $c_{1-\alpha}^*$ such that
- $\mathbb{P}\left\{ (d^{X^*} d^{Y^*})^T A^{-1} (d^{X^*} d^{Y^*}) \le c_{1-\alpha}^* \right\} \ge 1 \alpha$

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 .

Huckemann

Euclidear

BP/BL-CLT

(A2): Cut Loc

(A5): Emp. P

(A6): Smeary

Generalization

PCA/ Applications

Application

Reference

References

Sequences of Nested Subspaces

Note:

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

(A6): Smeary

_ _ . .

PCA/ Applications

Outlook

Reference

References

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.

(A6): Smeary

PCA/

Applications

Outlook

Reference

References

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.

Huckemann

Euclidean

BP/BL-CLT

BP/BL-CLI

(A5): Fmp F

(A6): Smoory

Conorolizatio

PCA/

Applications

Outlook

Reference

Reference

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?

Huckemann

Fuclidear

RP/RL-CL

(A2). Cut Loc

(A5): Emp. P

(A6): Sm

Generalization

PCA/ Applications

Applicatio

Reference

References

Backward Nested Families of Descriptors

Q (topological, separable = ts): Data space

Huckemann

Euclidear

BP/BL-CL

(, 12). Gut 200

(A5): Emp. Pi

(A6): Smeary

Generalization

PCA/ Applications

Applicatio

Reference

References

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, $P_m = \{Q\}$;

Huckemann

Euclidean

BP/BL-CL

(A5): Emp. Pr

(A6): Smeary

Generalization

PCA/

Applications

Outlook

Reference

References

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, ..., m) is itself a topological space giving rise to a topological space $\emptyset \neq S_p \subseteq P_{j-1}$ with

$$ho_p: p imes \mathcal{S}_p o [0,\infty)$$
 , continuous ;

Huckemann

Euclidean

BP/BL-CLT

(A5): Emp. Pr

(A6): Smeary

Generalization

PCA/ Applications

Outlook

Reference

References

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, ..., m) is itself a topological space giving rise to a topological space $\emptyset \neq S_p \subseteq P_{j-1}$ with

$$ho_{p}: p imes \mathcal{S}_{p}
ightarrow [0, \infty)$$
 , continuous ;

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

 $\pi_{p,s}: p o s$, measurable .

Euclidean

BP/BL-CL1

(A5): Emp. Pr

(A6): Smeary

Generalizations

PCA/ Applications

Outlook

References

References

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, ..., m) is itself a topological space giving rise to a topological space $\emptyset \neq S_p \subseteq P_{j-1}$ with

$$ho_{p}: p imes \mathcal{S}_{p} o [0, \infty)$$
 , continuous ;

(iii) $\forall \ p \in P_j \ (j=1,\ldots,m) \ \text{and} \ s \in S_p \ \exists \ \text{``projection''}$

$$\pi_{p,s}: p o s$$
 , measurable .

For $j \in \{1, ..., m\}$,

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

is BNFD from P_m to P_j from the space

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

Euclidean

BP/BL-CI

(A5): Emp. F

(A6): Smeary

PCA/ Applications

Outlook

References

References

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, ..., m) is itself a topological space giving rise to a topological space $\emptyset \neq S_p \subseteq P_{j-1}$ with

$$ho_{m{p}}: m{p} imes m{S_p}
ightarrow [\mathtt{0}, \infty)$$
 , continuous ;

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

$$\pi_{oldsymbol{
ho},oldsymbol{s}}:oldsymbol{
ho} ooldsymbol{s}$$
 , measurable .

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

is BNFD from P_m to P_i from the space

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

with projection along each descriptor

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

(A5): Emp. Pi

(A6): Smeary

Conorolization

Generalizatio

PCA/ Applications

Application

Outlook

Reference

References

Backward Nested Families of Descriptors

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

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

Outlook

Reference

References

Backward Nested Fréchet Means

Random elements $X_1, ..., X_n \stackrel{\text{i.i.d.}}{\sim} X$ 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, ..., 1,

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

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

Huckemann

Fuclidear

BP/BL-CLT

(A2): Gut Lo

(A5): Emp. P

0 " "

PCA/

Applications

Outlool

Reference

References

Backward Nested Fréchet Means

Random elements $X_1, ..., X_n \stackrel{\text{i.i.d.}}{\sim} X$ 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, ..., 1,

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

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

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

PCA/ **Applications**

Strong Law

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

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

$$f_n \rightarrow f$$
 a.s.

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

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

Huckemann

Euclidear

RP/RL-CL

(A2): Cut Loc

(A5): Emp. Pi

(A6): Smear

Generalization

PCA/ Applications

Applicatio

Reference

References

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

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

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

Huckemann

Fuclidear

BP/BL-CL

(A2): Cut Loc

(A5): Emp. P

(A6): Smeary

Caparalization

Gorioranzano

PCA/ Applications

Outlook

Reference

References

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

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

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

Idea of proof:

$$0 = \operatorname{grad}_{\eta} \sum_{k=1}^{n} \tau^{j}(\eta_{n}, X_{k}) + \sum_{l=j+1}^{m} \lambda_{n}^{l} \operatorname{grad}_{\eta} \sum_{k=1}^{n} \tau^{l}(\eta_{n}, X_{k})$$

CLT for Fréchet Means Huckemann

PCA/ **Applications**

Idea of proof:



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

$$\displaystyle \inf_{\psi^{-1}}$$

$$\begin{array}{c} 1 \\ -1 \\ \rightarrow \end{array} f^{j}$$

$$\stackrel{-1}{\rightarrow} f^j$$

$$f^{j-1}$$

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

 $\cdot (n'-n_n)$

 $0 = \operatorname{grad}_{\eta} \sum_{k=1}^{m} \tau^{j}(\eta_{n}, X_{k}) + \sum_{l=i+1}^{m} \lambda_{n}^{l} \operatorname{grad}_{\eta} \sum_{k=1}^{m} \tau^{l}(\eta_{n}, X_{k})$

 $= \operatorname{grad}_{\eta} \sum_{k=1}^{n} \tau^{j}(\eta', X_{k}) + \sum_{l=j+1}^{m} \lambda_{n}^{l} \operatorname{grad}_{\eta} \sum_{k=1}^{n} \tau^{l}(\eta', X_{k})$

 $+ \left(\operatorname{Hess}_{\eta} \sum_{k=1}^{n} \tau^{j}(\widetilde{\eta}_{n}, X_{k}) + \sum_{l=i+1}^{m} \lambda_{n}^{l} \operatorname{Hess}_{\eta} \sum_{k=1}^{n} \tau^{l}(\widetilde{\eta}_{n}, X_{k}) \right)$

$$\rightarrow$$

$$\rightarrow$$
 \mathcal{N}

$$\mathcal{N}(0,$$

$$\mathcal{N}(\mathbf{0}, B_{\psi})$$

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

$$\rightarrow \mathcal{N}(\mathbf{0}, \mathbf{D}_{i})$$

CLT for Fréchet Means Huckemann

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

 $+ \left(\operatorname{Hess}_{\eta} \sum_{k=1}^{n} \tau^{j}(\widetilde{\eta}_{n}, X_{k}) + \sum_{l=j+1}^{m} \lambda_{n}^{l} \operatorname{Hess}_{\eta} \sum_{k=1}^{n} \tau^{l}(\widetilde{\eta}_{n}, X_{k}) \right)$

PCA/ **Applications**

Idea of proof:

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

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

 $0 = \operatorname{grad}_{\eta} \sum_{k=1}^{m} \tau^{j}(\eta_{n}, X_{k}) + \sum_{l=i+1}^{m} \lambda_{n}^{l} \operatorname{grad}_{\eta} \sum_{k=1}^{m} \tau^{l}(\eta_{n}, X_{k})$

 $\cdot (n'-n_n)$

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

 $= \operatorname{grad}_{\eta} \sum_{k=1}^{n} \tau^{j}(\eta', X_{k}) + \sum_{l=i+1}^{m} \lambda_{n}^{l} \operatorname{grad}_{\eta} \sum_{k=1}^{n} \tau^{l}(\eta', X_{k})$

Fréchet Means Huckemann

CLT for

The Joint CLT [S.H. and Eltzner (2018)] With local chart $\eta \stackrel{\psi^{-1}}{\mapsto} f^{j-1} \mapsto \rho_{\mathcal{D}^j}(\pi_{f^j} \circ X, \mathcal{D}^{j-1})^2 := \tau^j(\eta, X)$:

PCA/ **Applications**

$$\operatorname{grad}_{\eta}$$

Idea of proof:

$$\sum_{j=1}^{n} \tau^{j} (t)$$

$$\sum_{j=1}^{n} \tau^{j}(\eta_{n}, \lambda_{n})$$

$$0 = \operatorname{grad}_{\eta} \sum_{k=1}^{n} \tau^{j}(\eta_{n}, X_{k}) + \sum_{l=i+1}^{m} \lambda_{n}^{l} \operatorname{grad}_{\eta} \sum_{k=1}^{n} \tau^{l}(\eta_{n}, X_{k})$$

$$+\sum_{l=j+1}^{m}\lambda$$

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

$$\sum_{l=j+1}^{m} \lambda_n^l$$

$$\lambda_n^I$$
 gra

$$\operatorname{rad}_{\eta} \sum_{i=1}^{n}$$

$$\sum_{k=1}^{n} \tau^{l}$$

$$\sum_{i=1}^{n} \tau^{I}(\eta_{n}, \cdot)$$

$$\sum_{n=1}^{\infty} \tau^{I}(\eta_{n}, \Sigma)$$

$$\frac{1}{1}$$

$$0 = \frac{1}{\sqrt{n}} \operatorname{grad}_{\eta} \sum_{k=1}^{n} \tau^{j}(\eta', X_{k}) + \frac{1}{\sqrt{n}} \sum_{l=j+1}^{m} \lambda_{n}^{l} \operatorname{grad}_{\eta} \sum_{k=1}^{n} \tau^{l}(\eta', X_{k})$$

$$= \frac{1}{\sqrt{n}}\operatorname{grad}_{\eta} \sum_{k=1}^{r} \tau^{j}(\eta, X_{k}) + \frac{1}{\sqrt{n}} \sum_{l=j+1}^{r} \lambda_{n} \operatorname{grad}_{\eta} \sum_{k=1}^{r} \tau^{l}(\eta, X_{k}) + \frac{1}{n} \left(\operatorname{Hess}_{\eta} \sum_{k=1}^{n} \tau^{j}(\widetilde{\eta}_{n}, X_{k}) + \sum_{l=j+1}^{m} \lambda_{n}^{l} \operatorname{Hess}_{\eta} \sum_{k=1}^{n} \tau^{l}(\widetilde{\eta}_{n}, X_{k}) \right)$$

$$\int_{\tau^{l}(\widetilde{\eta}_{n},X_{k})}^{t=1}$$

$$\lim_{l=j+1} |j+1| \qquad k=1$$

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

 $\sqrt{n}(n'-n_n)$

Huckemann

Euclidean

BP/BL-CL

(AE), E. D.

(A6): Smearv

Conorolizatio

PCA/

Applications

References

References

The Joint Central Limit Theorem

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

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

and typical regularity conditions, where

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

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

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

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

vanishes at $\eta = \eta'$.

Huckemann

Euclidean

BP/BL-CL

(A5), 5,...

(710): Link: 1

Generalization

PCA/ Applications

Outlook

Reference

Reference

Factoring Charts

If the following diagram commutes we say the chart factors

$$T_{m,j-1}$$
 \ni f^{j-1} $=$ (f^{j}, p^{j-1}) $\stackrel{\psi}{\rightarrow}$ η $=$ (θ, ξ) $\downarrow \pi^{\mathbb{R}^{\dim(\theta)}}$ P_{j-1} \ni p^{j-1} $\stackrel{\phi}{\rightarrow}$ θ

BP/BL-CL

(AE), E.... D

(A5): Emp. P

Generalization

PCA/ Applications

.

D (

Factoring Charts

If the following diagram commutes we say the chart factors

$$T_{m,j-1}$$
 \ni f^{j-1} $=$ (f^{j},p^{j-1}) $\stackrel{\psi}{\rightarrow}$ η $=$ (θ,ξ) $\downarrow \pi^{\mathbb{R}^{\dim(\theta)}}$ P_{j-1} \ni p^{j-1} $\stackrel{\phi}{\rightarrow}$ θ

Then

$$\eta = (\theta, \xi) \stackrel{\psi^{-1}}{\mapsto} f^{j-1} \quad \mapsto \quad \rho_{p^{j}}(\pi_{f^{j}} \circ X, p^{j-1})^{2} \\
= \quad \rho_{\pi^{P_{j}} \circ \psi_{2}^{-1}(\xi)} \left(\pi_{\psi_{2}^{-1}(\xi)} \circ X, \psi_{1}^{-1}(\theta)\right)^{2} \\
=: \quad \tau^{j}(\theta, \xi, X),$$

CLT for Fréchet Means Huckemann

PCA/ **Applications**

Factoring Charts

If the following diagram commutes we say the chart factors

$$T_{m,j-1} \ni f^{j-1} = (f^j, p^{j-1}) \stackrel{\psi}{ o} \eta = (\theta, \xi) \ \downarrow \pi^{P_{j-1}} \qquad \qquad \downarrow \pi^{\mathbb{R}^{\dim(\theta)}}$$
 $P_{j-1} \ni \qquad \qquad p^{j-1} \stackrel{\phi}{ o} \qquad \qquad \theta$

Then

$$\eta = (\theta, \xi) \stackrel{\psi^{-1}}{\mapsto} f^{j-1} \quad \mapsto \quad \rho_{p^{j}}(\pi_{f^{j}} \circ X, p^{j-1})^{2} \\
= \quad \rho_{\pi^{p_{j}} \circ \psi_{2}^{-1}(\xi)} \left(\pi_{\psi_{2}^{-1}(\xi)} \circ X, \psi_{1}^{-1}(\theta)\right)^{2} \\
= : \quad \tau^{j}(\theta, \xi, X).$$

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

$$\sqrt{n}H_{\psi}(\eta_n-\eta')=\sqrt{n}H_{\psi}\left(egin{array}{c} heta_n- heta' \ au_n- au' \end{array}
ight) \;\;
ightarrow \;\; \mathcal{N}(0,B_{\psi})$$

and projection to the θ coordinate preserves Gaussianity.

Huckemann

Euclidean

BP/BL-CLT

(AE): Emp. E

(A6): Smeary

PCA/

Applications

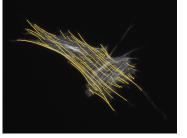
Outlool

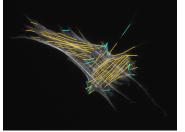
Reference

References

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:

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

Huckemann

market and

DD/DL CL

(AO) - O - + 1 -

(A5): Emp. P

(A6): Smeary

Conoralizatio

PCA/

Applications

Outlook

References

Applying the Bootstrap Two-Sample Test

١.					
		nested mean		jointly great circle and nested mean	
	Time	\leq 1 kPa	≥ 10 kPa	\leq 1 kPa	\geq 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

Euclidean

BP/BL-CLT

(A0), Cut I a

(A5): Emp. P

(A6): Smoory

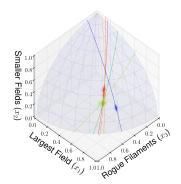
PCA/ Applications

Outlook

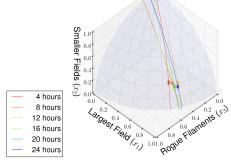
Reference

Poforonoos

Visualization



Left: ≤ 1 kPa.



Right: ≥ 10 kPa

Huckemann

Euclidean

BP/BL-CLT

.

(A5): Emp. P

(A6): Smeary

Generalization

PCA/

Application

Outlook

Reference

References

Wrap up and Outlook

Wrap up:

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

Huckemann

Euclidean

BP/BL-CLT

(A5): Emp. D

(A6): Smeary

GOTTOT GITE GETTOT

PCA/ Application

Outlook

Reference

Reference

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);

Huckemann

Euclidean

BP/BL-CLT

(A5): Emp. P.

(A6): Smeary

Generalization

PCA/ Application

Application

Outlook

1 1010101100

Reference

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:

Huckemann

Euclidean

BP/BL-CLT

(A5): Emp. P

(A6): Smeary

PCA/ Applications

πρριισατισι

Outlook

nelelelices

References

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.

Huckemann

Euclidean

BP/BL-CLT

(A5): Emp. P

(A6): Smeary

PCA/ Applications

πρριισατισι

Outlook

nelelelices

References

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.

Huckemann

Euclidean

BP/BL-CLT

(A5): Emp. P

(A6): Smeary

Gerreranzar

PCA/ Applications

Outlook

References

References

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?

Huckemann

Euclidean

BP/BL-CLI

(A5): Emp. P

(A6): Smeary

Generalizat

Application:

Outlook

References

References

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.

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

Huckemann

Euclidean

BP/BL-CLI

(A5): Emp. P

(A6): Smeary

DCA/

Applications

Outlook

References

References

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.

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

Huckemann

Euclidean

(AO): O::

(A5): Emp. P

(A6): Smeary

DO 4 /

Applications

Outlook

References

References

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.

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

Huckemann

Euclidean

DF/DL-GLI

(A5): Emp. P

(Ab): Smeary

DO 4 /

Applications

Outlook

References

Reference

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.

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

(A6): Smeary

PCA/

Applications

Outlook

Reference

Reference

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.

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

Huckemann

Euclidean

BP/BL-CLT

(A5): Fmp P

(A6): Smearv

Conorolination

PCA/

Outlook

0 41.0011

Reference

Reference

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.

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

Huckemann

Euclidean

BP/BL-CLT

(A5): Emp. P

(Ab): Smeary

PCA/

) ! Dutlook

References

Poforonoos

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. Stochastic. 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 Š. 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

Huckemann

Euclidear

BP/BL-CL7

() ----

(A5): Emp. P

(AU). Officary

PCA/

Applications

Reference

References

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.