# Limit Theorems for PCA-Like Analysis on Noneuclidean Spaces

Stephan F. Huckemann and Benjamin Eltzner

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

April 2, 2019

Workshop I: Geometric Processing at UCLA

supported by the

Niedersachsen Vorab of the Volkswagen Foundation,

and the DFG SFB 755 + GK 2088

# Review of Classical Statistics

Tasks:

- understanding (e.g. dimension reduction)
- discrimination
- classification
- prediction

# Review of Classical Statistics

Tasks:

- understanding (e.g. dimension reduction)
- discrimination
- classification
- prediction = fortune telling

# Review of Classical Statistics

Tasks:

- understanding (e.g. dimension reduction)
- discrimination
- classification
- prediction = fortune telling

Tools:

- machine learning (ML)

# Review of Classical Statistics

Tasks:

- understanding (e.g. dimension reduction)
- discrimination
- classification
- prediction = fortune telling

Tools:

- machine learning (ML)
- – works miracously if enough training data available

# Review of Classical Statistics

Tasks:

- understanding (e.g. dimension reduction)
- discrimination
- classification
- prediction = fortune telling

Tools:

- machine learning (ML)
- – works miracously if enough training data available
- else
  - model building
  - estimation
  - testing

# Review of Classical Statistics

Tasks:

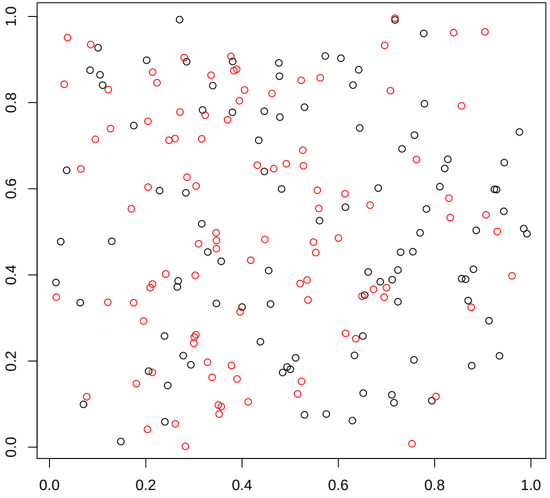
- understanding (e.g. dimension reduction)
- discrimination
- classification
- prediction = fortune telling

Tools:

- machine learning (ML)
- – works miracously if enough training data available
- else (ML also requires some basic)
  - model building
  - estimation
  - testing

# Some Classical Statistics

Two patterns of $n = 100$ random points each. What is their law? Same?



# Statistical Tests

- Basic model: two groups of random variables

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

# Statistical Tests

- Basic model: two groups of random variables

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

- with means

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \bar{Y}_m = \frac{1}{m}\sum_{j=1}^{m} Y_j$$

and empirical covariances

$$\Sigma_n^X = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)(X_i - \bar{X}_n)^T, \quad \Sigma_m^Y = \frac{1}{m}\sum_{i=1}^{m}(Y_j - \bar{Y}_m)(Y_j - \bar{Y}_m)^T$$

# Statistical Tests

- Basic model: two groups of random variables

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

- with means $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \bar{Y}_m = \frac{1}{m}\sum_{j=1}^{m} Y_j$

  and empirical covariances

$$\Sigma_n^X = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)(X_i - \bar{X}_n)^T, \quad \Sigma_m^Y = \frac{1}{m}\sum_{i=1}^{m}(Y_j - \bar{Y}_m)(Y_j - \bar{Y}_m)^T$$

- CLT, under additional model assumptions: population cov$[X]$ and cov$[Y]$ exist,

$$\sqrt{n}(\bar{X}_n - \mathbb{E}[X]) \overset{\mathcal{D}}{\to} \mathcal{N}(0, \mathrm{cov}[X])$$
$$\sqrt{m}(\bar{Y}_m - \mathbb{E}[Y]) \overset{\mathcal{D}}{\to} \mathcal{N}(0, \mathrm{cov}[Y])$$

# Statistical Tests

- Basic model: two groups of random variables

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

- with means $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \bar{Y}_m = \frac{1}{m}\sum_{j=1}^{m} Y_j$

  and empirical covariances

$$\Sigma_n^X = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)(X_i - \bar{X}_n)^T, \quad \Sigma_m^Y = \frac{1}{m}\sum_{i=1}^{m}(Y_j - \bar{Y}_m)(Y_j - \bar{Y}_m)^T$$

- CLT, under additional model assumptions:
  population cov$[X]$ and cov$[Y]$ exist and are of full rank,

$$\sqrt{n}\sqrt{\Sigma_n^X}^{-1}(\bar{X}_n - \mathbb{E}[X]) \overset{\mathcal{D}}{\to} \mathcal{N}(0, I_D)$$

$$\sqrt{m}\sqrt{\Sigma_m^Y}^{-1}(\bar{Y}_m - \mathbb{E}[Y]) \overset{\mathcal{D}}{\to} \mathcal{N}(0, I_D)$$

# Hypothesis Test

- E.g.: $H_0 : \mathbb{E}[X] = \mathbb{E}[Y]$ against $H_1 : \mathbb{E}[X] \neq \mathbb{E}[Y]$

# Hypothesis Test

- E.g.: $H_0 : \mathbb{E}[X] = \mathbb{E}[Y]$ against $H_1 : \mathbb{E}[X] \neq \mathbb{E}[Y]$
- Ideally, find r.v. $Z$ with computable law $\mathbb{P}^Z = \mathbb{P} \circ Z^{-1}$ such that $T(X_1, \ldots, X_n, Y_1, \ldots, Y_m) \sim Z$ under $H_0$.

# Hypothesis Test

- E.g.: $H_0 : \mathbb{E}[X] = \mathbb{E}[Y]$ against $H_1 : \mathbb{E}[X] \neq \mathbb{E}[Y]$
- Ideally, find r.v. $Z$ with computable law $\mathbb{P}^Z = \mathbb{P} \circ Z^{-1}$ such that $T(X_1, \ldots, X_n, Y_1, \ldots, Y_m) \sim Z$ under $H_0$.
- Level $0 < \alpha < 1$, typically $\alpha = 0.05$ (significant).

# Hypothesis Test

- E.g.: $H_0 : \mathbb{E}[X] = \mathbb{E}[Y]$ against $H_1 : \mathbb{E}[X] \neq \mathbb{E}[Y]$
- Ideally, find r.v. $Z$ with computable law $\mathbb{P}^Z = \mathbb{P} \circ Z^{-1}$ such that $T(X_1, \ldots, X_n, Y_1, \ldots, Y_m) \sim Z$ under $H_0$.
- Level $0 < \alpha < 1$, typically $\alpha = 0.05$ (significant).
- Confidence region $K_\alpha \subseteq \mathbb{R}^k$ such that

$$\mathbb{P}\{Z \in K_\alpha\} = 1 - \alpha\,.$$

# Hypothesis Test

- E.g.: $H_0 : \mathbb{E}[X] = \mathbb{E}[Y]$ against $H_1 : \mathbb{E}[X] \neq \mathbb{E}[Y]$
- Ideally, find r.v. $Z$ with computable law $\mathbb{P}^Z = \mathbb{P} \circ Z^{-1}$ such that $T(X_1, \ldots, X_n, Y_1, \ldots, Y_m) \sim Z$ under $H_0$.
- Level $0 < \alpha < 1$, typically $\alpha = 0.05$ (significant).
- Confidence region $K_\alpha \subseteq \mathbb{R}^k$ such that

$$\mathbb{P}\{Z \in K_\alpha\} = 1 - \alpha\,.$$

- Reject $H_0$ if $T \notin K_\alpha$. Don't determine $K_\alpha$ before computing $T$!

# Hypothesis Test

- E.g.: $H_0 : \mathbb{E}[X] = \mathbb{E}[Y]$ against $H_1 : \mathbb{E}[X] \neq \mathbb{E}[Y]$
- Ideally, find r.v. $Z$ with computable law $\mathbb{P}^Z = \mathbb{P} \circ Z^{-1}$ such that $T(X_1, \dots, X_n, Y_1, \dots, Y_m) \sim Z$ under $H_0$.
- Level $0 < \alpha < 1$, typically $\alpha = 0.05$ (significant).
- Confidence region $K_\alpha \subseteq \mathbb{R}^k$ such that

$$\mathbb{P}\{Z \in K_\alpha\} = 1 - \alpha\,.$$

- Reject $H_0$ if $T \notin K_\alpha$. Don't determine $K_\alpha$ before computing $T$!
- Sometimes only find $T \xrightarrow{\mathcal{D}} Z$ as $n, m \to \infty$ (test holds asymptotically the level).

# Hypothesis Test

- E.g.: $H_0 : \mathbb{E}[X] = \mathbb{E}[Y]$ against $H_1 : \mathbb{E}[X] \neq \mathbb{E}[Y]$
- Ideally, find r.v. $Z$ with computable law $\mathbb{P}^Z = \mathbb{P} \circ Z^{-1}$ such that $T(X_1, \ldots, X_n, Y_1, \ldots, Y_m) \sim Z$ under $H_0$.
- Level $0 < \alpha < 1$, typically $\alpha = 0.05$ (significant).
- Confidence region $K_\alpha \subseteq \mathbb{R}^k$ such that

$$\mathbb{P}\{Z \in K_\alpha\} = 1 - \alpha\,.$$

- Reject $H_0$ if $T \notin K_\alpha$. Don't determine $K_\alpha$ before computing $T$!
- Sometimes only find $T \stackrel{\mathcal{D}}{\to} Z$ as $n, m \to \infty$ (test holds asymptotically the level).
- $Z_n \stackrel{\mathcal{D}}{\to} Z \Leftrightarrow \int f \, d\,\mathbb{P}^{Z_n} \to \int f \, d\,\mathbb{P}^Z \;\; \forall f \in C_b(\mathbb{R}^k \to \mathbb{R})$.

# Hotelling Test for Equality of Means

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

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

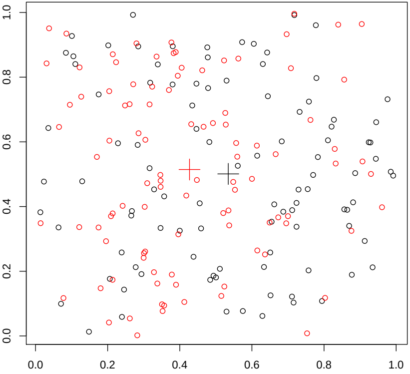
$\stackrel{\mathcal{D}}{\to}$ explicitly known limit ($n, m \to \infty$, $0 < \lim n/m < \infty$)

# Hotelling Test for Equality of Means

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

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

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

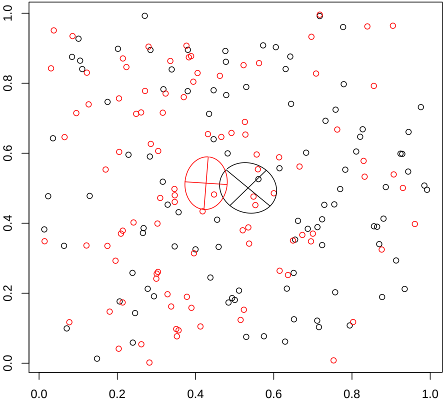



Reject $H_0$ with significance ($\alpha = 0.05$), not highly ($\alpha = 0.01$).

# Principal Component Analysis (PCA)

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

- With eigenvectors $\Gamma = (\gamma_1, \ldots, \gamma_m) \in SO(m)$ to
- eigenvalues $\lambda_1 \geq \ldots \geq \lambda_m \geq 0$, $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_m)$
- giving main modes of variation
- If $X$ is Gaussian, $\Gamma^T X$ has independent entries.
- Test for PCs $\gamma_j$ ? Note, $\gamma_j \in \mathbb{S}^{m-1}$. Actually in $\mathbb{R}P^{m-1}$.



# Asymptotics of Euclidean PCA

$$\gamma_1 \in \underset{\gamma \in \mathbb{R}P^{m-1}}{\operatorname{argmax}} \; \gamma^T \operatorname{cov}[X]\gamma\,, \qquad \max = \lambda_1$$

$$\gamma_2 \in \underset{\gamma_1 \perp \gamma \in \mathbb{R}P^{m-1}}{\operatorname{argmax}} \; \gamma^T \operatorname{cov}[X]\gamma\,, \qquad \max = \lambda_2$$

$$\hat{\gamma}_1 \in \underset{\gamma \in \mathbb{R}P^{m-1}}{\operatorname{argmax}} \; \gamma \Sigma_n^X \gamma \qquad \max = \hat{\lambda}_1$$

$$\vdots$$

**Theorem (Anderson (1963), never (?) reproduced)**

*If $X$ Gaussian,* $\sqrt{n}(\hat{\gamma}_1 - \gamma_1) \to \mathcal{N}\left(0, \sum_{k=2}^{m} \frac{\lambda_1 \lambda_k}{(\lambda_1 - \lambda_k)^2} \gamma_k \gamma_k^T\right)$ .

# Asymptotics of Euclidean PCA

$$\gamma_1 \in \underset{\gamma \in \mathbb{R}P^{m-1}}{\operatorname{argmax}} \ \gamma^T \operatorname{cov}[X]\gamma\,, \qquad \max = \lambda_1$$

$$\gamma_2 \in \underset{\gamma_1 \perp \gamma \in \mathbb{R}P^{m-1}}{\operatorname{argmax}} \ \gamma^T \operatorname{cov}[X]\gamma\,, \qquad \max = \lambda_2$$

$$\hat\gamma_1 \in \underset{\gamma \in \mathbb{R}P^{m-1}}{\operatorname{argmax}} \ \gamma \Sigma_n^X \gamma \qquad \max = \hat\lambda_1$$

$$\vdots$$

**Theorem (Anderson (1963), never (?) reproduced)**

*If X Gaussian,* $\sqrt{n}(\hat\gamma_1 - \gamma_1) \to \mathcal{N}\left(0, \sum_{k=2}^m \frac{\lambda_1\lambda_k}{(\lambda_1-\lambda_k)^2}\gamma_k\gamma_k^T\right)$ .

- Alternate proof by Watson (1983); Ruymgaart and Yang (1997) using perturbation theory
- $\exists$ more involed nonnormal versions
- all unaware of noneuclidean stats.

# The Beginnings of Noneuclidean Statistics

- Kendall's shape spaces
- $k$ landmarks matrices $X_1, \ldots, X_n \in \mathbb{R}^{m \times k}$
- modulo translation, rotation and scaling
- Gower (1975) minimizes the Procrustes sum of squares

$$\underset{\substack{\lambda_i > 0,\, g_i \in SO(m), \\ a_i \in \mathbb{R}^m,\, 1 \le i \le n}}{\operatorname{argmin}} \sum_{1 \le i,j \le n} \left\| \lambda_i g_i (X_i - a_i \mathbf{1}_{m\times m}^T) - \lambda_j g_j (X_j - a_j \mathbf{1}_{m\times m}^T) \right\|^2$$

# The Beginnings of Noneuclidean Statistics

- Kendall's shape spaces
- $k$ landmarks matrices $X_1, \ldots, X_n \in \mathbb{R}^{m \times k}$
- modulo translation, rotation and scaling
- Gower (1975) minimizes the Procrustes sum of squares

$$\underset{\substack{\lambda_i > 0, g_i \in SO(m), \\ a_i \in \mathbb{R}^m, 1 \le i \le n}}{\operatorname{argmin}} \sum_{1 \le i,j \le n} \left\| \lambda_i g_i (X_i - a_i \mathbf{1}_{m \times m}^T) - \lambda_j g_j (X_j - a_j \mathbf{1}_{m \times m}^T) \right\|^2$$

- under the constraining condition that

$$\underbrace{\left\| \sum_{i=1}^n \lambda_i g_i (X_i - a_i \mathbf{1}_{m \times m}^T) \right\|}_{\rightsquigarrow \hat{\mu} = \text{Procrustes mean}} = 1$$

- PCA of the residuals $Y_i - \operatorname{trace}(Y_i^T \hat{\mu}) \hat{\mu}$ for optimal $Y_j = \lambda_i g_i (X_i - a_i \mathbf{1}_{m \times m}^T)$ is Procrustes analysis.
- Unaware of noneuclidean stats.

# Fréchet (1948) Means

- Random data $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} X \in Q$ on a metric space $(Q, d)$
- $\mu \in \operatorname{argmin}_{q \in Q} \mathbb{E}[d(X, q)^2] =$ population Fréchet mean
- $\hat{\mu} \in \operatorname{argmin}_{q \in Q} \sum_{j=1}^{b} d(X_j, q)^2 =$ sample Fréchet mean

# Fréchet (1948) Means

- Random data $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} X \in Q$ on a metric space $(Q, d)$
- $\mu \in \text{argmin}_{q \in Q} \mathbb{E}[d(X, q)^2]$ = population Fréchet mean
- $\hat{\mu} \in \text{argmin}_{q \in Q} \sum_{j=1}^{b} d(X_j, q)^2$ = sample Fréchet mean
- $d$ can be
  - intrinsic distance
  - extrinsic distance
  - just a loss function, e.g. residual distance in the tangent space of $q$
- Is the Procrustes mean a Fréchet mean?

# Fréchet (1948) Means

- Random data $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} X \in Q$ on a metric space $(Q, d)$
- $\mu \in \text{argmin}_{q \in Q} \mathbb{E}[d(X, q)^2]$ = population Fréchet mean
- $\hat{\mu} \in \text{argmin}_{q \in Q} \sum_{j=1}^{b} d(X_j, q)^2$ = sample Fréchet mean
- $d$ can be
    - intrinsic distance
    - extrinsic distance
    - just a loss function, e.g. residual distance in the tangent space of $q$
- Is the Procrustes mean a Fréchet mean?
- Yes, $d$ is the quotient on a sphere modulo rotations of the residual distance loss

# Fréchet (1948) Means

- Random data $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} X \in Q$ on a metric space $(Q, d)$
- $\mu \in \text{argmin}_{q \in Q} \mathbb{E}[d(X, q)^2]$ = population Fréchet mean
- $\hat{\mu} \in \text{argmin}_{q \in Q} \sum_{j=1}^{b} d(X_j, q)^2$ = sample Fréchet mean
- $d$ can be
    - intrinsic distance
    - extrinsic distance
    - just a loss function, e.g. residual distance in the tangent space of $q$
- Is the Procrustes mean a Fréchet mean?
- Yes, $d$ is the quotient on a sphere modulo rotations of the residual distance loss
- Procrustes analysis (GPA) = PCA in the tangent space of the Procrustes mean.

# Fréchet (1948) Means

- Random data $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} X \in Q$ on a metric space $(Q, d)$
- $\mu \in \operatorname{argmin}_{q \in Q} \mathbb{E}[d(X, q)^2]$ = population Fréchet mean
- $\hat{\mu} \in \operatorname{argmin}_{q \in Q} \sum_{j=1}^{b} d(X_j, q)^2$ = sample Fréchet mean
- $d$ can be
  - intrinsic distance
  - extrinsic distance
  - just a loss function, e.g. residual distance in the tangent space of $q$
- Is the Procrustes mean a Fréchet mean?
- Yes, $d$ is the quotient on a sphere modulo rotations of the residual distance loss
- Procrustes analysis (GPA) = PCA in the tangent space of the Procrustes mean.
- Principal geodesic analysis (PGA) by Fletcher and Joshi (2004) = PCA in the tangent space of an intrinsic mean.

# Asymptotics of Fréchet Means

Under uniqueness of $\mu$

- $\hat{\mu} \to \mu$ a.s. by Ziezold (1977) under finite expectation, $d$ quasimetric and $Q$ separable
- $\sqrt{n}\,(\phi(\hat{\mu}) - \phi(\mu)) \stackrel{\mathcal{D}}{\to} \mathcal{N}(0, \Sigma)$ by Bhattacharya and Patrangenaru (2005) if $Q$ is a manifold, $\phi$ is a local chart near $\mu$ and more regularity assumptions
- same for Procrustes means by S.H. (2011a)

# Asymptotics of Fréchet Means

Under uniqueness of $\mu$

- $\hat{\mu} \to \mu$ a.s. by Ziezold (1977) under finite expectation, $d$ quasimetric and $Q$ separable
- $\sqrt{n}\,(\phi(\hat{\mu}) - \phi(\mu)) \stackrel{\mathcal{D}}{\to} \mathcal{N}(0, \Sigma)$ by Bhattacharya and Patrangenaru (2005) if $Q$ is a manifold, $\phi$ is a local chart near $\mu$ and more regularity assumptions
- same for Procrustes means by S.H. (2011a)
- How to set up a two-sample test?
- Need an estimator of $\mathrm{var}[\phi(\hat{\mu})]$?

# Asymptotics of Fréchet Means

Under uniqueness of $\mu$

- $\hat{\mu} \to \mu$ a.s. by Ziezold (1977) under finite expectation, $d$ quasimetric and $Q$ separable
- $\sqrt{n}\,(\phi(\hat{\mu}) - \phi(\mu)) \overset{\mathcal{D}}{\to} \mathcal{N}(0, \Sigma)$ by Bhattacharya and Patrangenaru (2005) if $Q$ is a manifold, $\phi$ is a local chart near $\mu$ and more regularity assumptions
- same for Procrustes means by S.H. (2011a)
- How to set up a two-sample test?
- Need an estimator of $\mathrm{var}[\phi(\hat{\mu})]$?
- Bootstrap $\hat{\mu}^*$ from the data

# Global Concepts of PCA on Noneuclidean Spaces

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

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

# Global Concepts of PCA on Noneuclidean Spaces

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

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

# Example

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

# Example

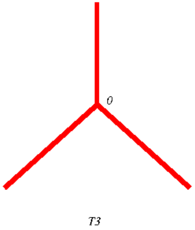
- Geodesic PCA (GPCA) on Riemannian spaces by S.H et al. (2010):
  - $P_1 = \Gamma(Q) =$ all geodesics on $Q$,
    $\rightsquigarrow \gamma_1$ and $\hat{\gamma}_1$ = 1st geodesic PCs
  - $P_2 = \{p \in \Gamma(Q) : \gamma_1 \perp p, \gamma_1 \cap p \neq \emptyset\}$
    $\hat{P}_2 = \{p \in \Gamma(Q)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

# Example

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

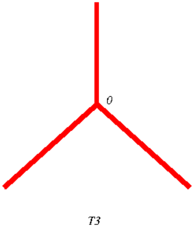
# Lack of Manifold Stability

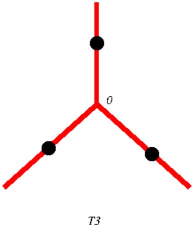
Consider $Q =$ three spider



# Lack of Manifold Stability

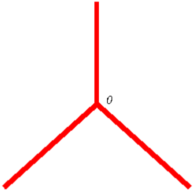
Consider $Q =$ three spider



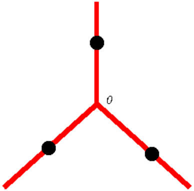



# Lack of Manifold Stability

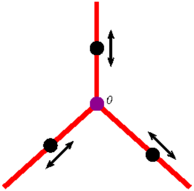
Consider $Q$ = three spider




*T3*



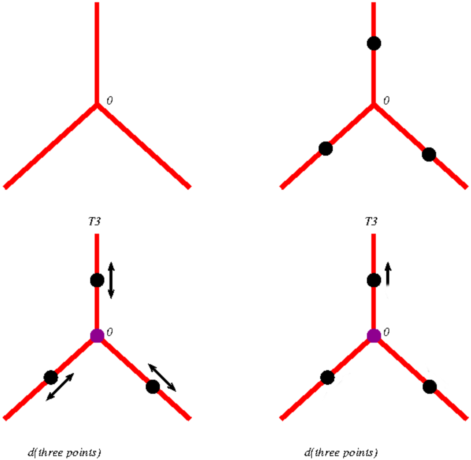

*T3*




*d(three points)*

# Lack of Manifold Stability

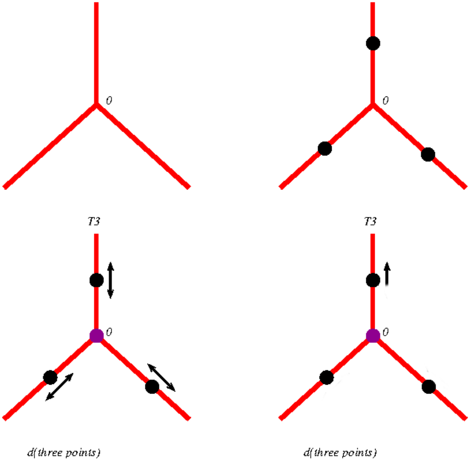
Consider $Q =$ three spider




Hotz et al. (2013) and S.H. et al (2015).

# Lack of Manifold Stability

Consider $Q =$ three spider




Hotz et al. (2013) and S.H. et al (2015).

Stickiness may kill all fluctuation.

# Nestedness and Equivalence

Euclidean PCA is

- nested, $\mu \in$ 1st PC
- minimizing residual variance over *k*-dim affine subspaces

$$\Leftrightarrow$$

maximizing projected variance over *k*-dim affine subspaces

# Nestedness and Equivalence

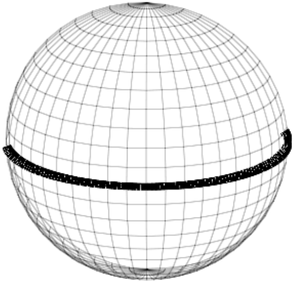
Euclidean PCA is

- nested, $\mu \in$ 1st PC
- minimizing residual variance over *k*-dim affine subspaces

$$\Leftrightarrow$$

maximizing projected variance over *k*-dim affine subspaces

How about GPCA?



# Nestedness and Equivalence

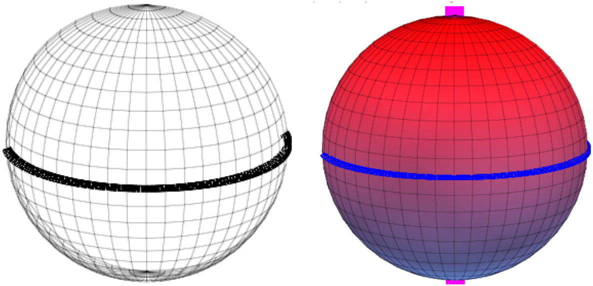
Euclidean PCA is

- nested, $\mu \in$ 1st PC
- minimizing residual variance over *k*-dim affine subspaces

$$\Leftrightarrow$$

maximizing projected variance over *k*-dim affine subspaces

How about GPCA?



# Sequences of Nested Subspaces

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

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

# Sequences of Nested Subspaces

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 talk?)

- forward or backward nested or all at once

# Sequences of Nested Subspaces

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 talk?)

- forward or backward nested or all at once

How about asymptotics of such nested random subspaces?

# Backward Nested Families of Descriptors

$Q$ (topological, separable = ts): Data space

# Backward Nested Families of Descriptors

$Q$ (topological, separable = ts): Data space

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

# Backward Nested Families of Descriptors

$Q$ (topological, separable = ts): Data space

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

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

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

# Backward Nested Families of Descriptors

$Q$ (topological, separable = ts): Data space

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

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

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

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

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

# Backward Nested Families of Descriptors

$Q$ (topological, separable = ts): Data space

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

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

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

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

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

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

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

is BNFD from $P_m$ to $P_j$ from the space

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

# Backward Nested Families of Descriptors

$Q$ (topological, separable = ts): Data space

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

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

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

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

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

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

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

is BNFD from $P_m$ to $P_j$ from the space

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

with projection along each descriptor

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

# Backward Nested Families of Descriptors

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

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

# Backward Nested Fréchet Means

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

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

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

# Backward Nested Fréchet Means

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

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

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

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

# Strong Law

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

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

$$f_n \to f \ a.s.$$

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

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

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

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

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

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

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

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

Idea of proof:

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

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

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

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

Idea of proof:

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

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

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

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

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

Idea of proof:

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

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

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

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

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

Idea of proof:

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

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

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

# The Joint Central Limit Theorem

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

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

and typical regularity conditions, where

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

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

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

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

vanishes at $\eta = \eta'$.

# Factoring Charts

If the following diagram commutes we say the chart factors

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

# Factoring Charts

If the following diagram commutes we say the chart factors

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

Then

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

# Factoring Charts

If the following diagram commutes we say the chart factors

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

Then

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

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

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

and projection to the $\theta$ coordinate preserves Gaussianity.

# Two-Sample Descriptor Test

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

# Two-Sample Descriptor Test

| | | | | | |
|---|---|---|---|---|---|
| Data: | $\underbrace{X_1, \dots, X_n}$ | , | $\underbrace{Y_1, \dots, Y_m}$ | $\in$ | $Q$ |
| | $\downarrow$ | | $\downarrow$ | | |
| Descriptors: | $f^X, p^X$ | | $f^Y, p^Y$ | $\in$ | $T, P$ |
| | $\downarrow$ | $\phi$ | $\downarrow$ | | |
| Coordinates: | $Z^X$ | | $Z^Y$ | $\in$ | $\mathbb{R}^D$ |

Under $H_0 : X \sim Y$,

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

# Two-Sample Descriptor Test

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

Under $H_0 : X \sim Y$,

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

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

# Bootstrapping

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

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

# Bootstrapping

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

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

# Bootstrapping

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

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

# Bootstrapping

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

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

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

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

# Bootstrapping

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

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

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

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

# Bootstrapping

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

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

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

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

# Bootstrapping

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

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

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

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

# Bootstrapping

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

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

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

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

# Bootstrapping

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

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

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

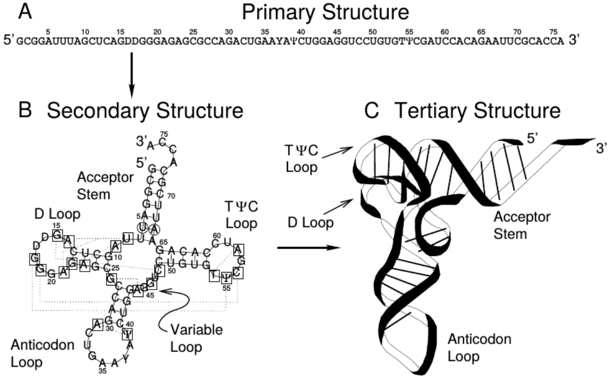
- $W^*_{1,b}, \dots, W^*_{n+m,b}$ from $X_1, \dots, X_n, Y_1, \dots, Y_m$
- set $X^*_{j,b} = W^*_{j,b}$ for $j = 1, \dots, n$
- set $Y^*_{j,b} = W^*_{j+n,b}$ for $j = 1, \dots, m$
- compute the empirical quantile $c^*_{1-\alpha}$ such that
- $\mathbb{P}\{(Z^{X^*} - Z^{Y^*})^T A^{-1} (Z^{X^*} - Z^{Y^*})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

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 $\alpha$.

# Application: tRNA

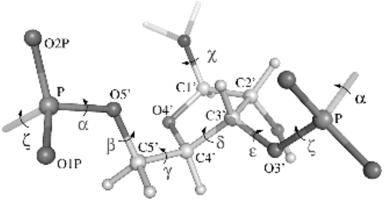


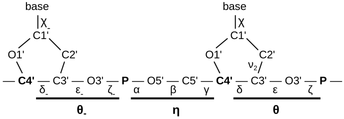

from Gardner (2003),

```
# STOCKHOLM 1.0
#=GF ID     trna
#=GF DE     S.cerevisiae tRNA-PHE75 data-set

yeast-PHE1    GCGGAUUUAGCUCAGUUGGGAGAGCGCCAGACUGAAGAUUUGGAGGUCCUGUGUUCGAUC
yeast-PHE2    GCGGAUUUAGCUCAGUUGGGAGAGCGCCAGACUGAAGAUCUGGAGGUCCUGUGUUCGAUC
e.coli-PHE    GCGGAUUUAGCUCAGUUGGGAGAGCGCCAGACUGAAUAUCUGGAGGUCCUGUGUUCGAUC
yeast-PHE3    GCGGACUUAGCUCAGUUGGGAGAGCGCCAGACUGAAGAUCUGGAGGUCCUGUGUUCGAUC
N.crassa-PHE  GCGGGUUUAGCUCAGUUGGGAGAGCGUCAGACUGAAGAUCUGAAGGUCGUGUGUUCGAUC
L.luteus-PHE  GCGGGGAUAGCUCAGUUGGGAGAGCGUCAGACUGAAGAUCUGAAGGUCCCGUGUUCGAUC
H.vulgarePHE  GCGGGGAUAGCUCAGUUGGGAGAGCGUCAGACUGAAGAUCUGAAGGUCGCGUGUUCGAUC
```

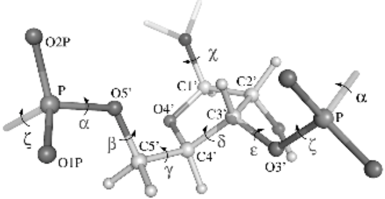
# Single Residue Analysis

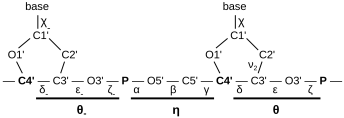


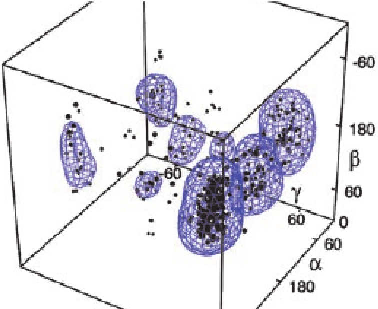

- 7 dihedral angles $\in (\mathbb{S}^1)^7$, 2 pseudotorsion angles $\in (\mathbb{S}^1)^2$,
- = shape, i.e. translational / rotational invariant.

# Single Residue Analysis



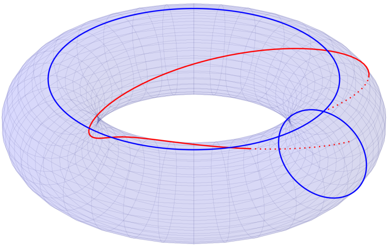

- 7 dihedral angles $\in (\mathbb{S}^1)^7$, 2 pseudotorsion angles $\in (\mathbb{S}^1)^2$,
- = shape, i.e. translational / rotational invariant.

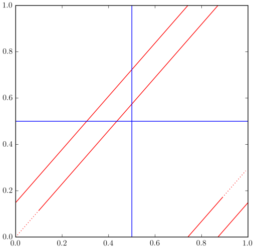



- Murray et al. (2003) using www.rscb.org:
- C2'-pucker RNA clusters in many 1D groups in heminucleotide angles.
- Can we verify (improve? understand?) by PCA?

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

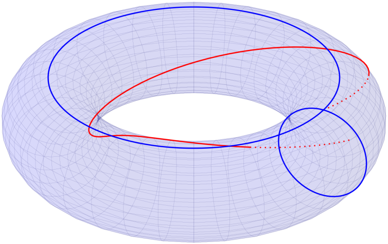
- Only very few geodesics are not winding around,
- an uncountable number of geodesics is dense and
- every data set can be perfectly approximated.
- Standard geometry of $(\mathbb{S}^1)^k$ is not statistically benign.

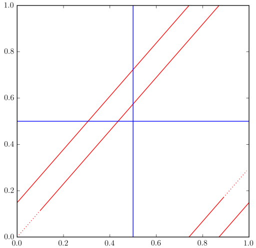




# PCA on a Torus $\mathbb{S}^1 \times \ldots \times \mathbb{S}^1$

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






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

# Euclidean vs. Spherical PCA

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

$\dim(P_k)$

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

# Euclidean vs. Spherical PCA

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

$\dim(P_k)$

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

# Euclidean vs. Spherical PCA

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

$\dim(P_k)$

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

# Sausage Transformation

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

# Sausage Transformation

- $(\mathbb{S}^1)^k \to \mathbb{S}^k$?
- Embed (non-isometrically) $\mathbb{S}^1 \times \mathbb{S}^1 = T = \left\{(x, y, z) \in \mathbb{R}^3 : \left(\sqrt{x^2 + y^2} - 2\right)^2 + z^2 = 1\right\}$,

# Sausage Transformation

- Embed (non-isometrically) $\mathbb{S}^1 \times \mathbb{S}^1 = T = \left\{(x, y, z) \in \mathbb{R}^3 : \left(\sqrt{x^2 + y^2} - 2\right)^2 + z^2 = 1\right\}$,
- cut open along the ring $(x - 2)^2 + z^2 = 1$,

# Sausage Transformation

- Embed (non-isometrically) $\mathbb{S}^1 \times \mathbb{S}^1 = T = \left\{(x, y, z) \in \mathbb{R}^3 : \left(\sqrt{x^2 + y^2} - 2\right)^2 + z^2 = 1\right\}$,
- cut open along the ring $(x - 2)^2 + z^2 = 1$,
- wrap up the two shores of the ring to become points that are identified,
- yields topology of a two-sphere with north and south pole identified: a sausage with its two ends connected.

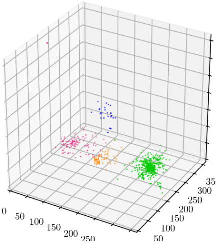
# Data Driven Torus (T) PCA for $(\mathbb{S}^1)^k$

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

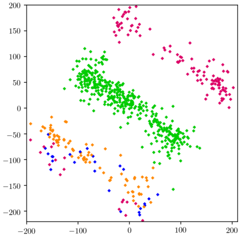
$$\sum_{l=1}^{k} d\psi_l^2 \to d\phi_1^2 + \sum_{l=2}^{k} \left( \prod_{j=1}^{l-1} \sin^2 \phi_j \right) d\phi_l^2 \,,$$

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

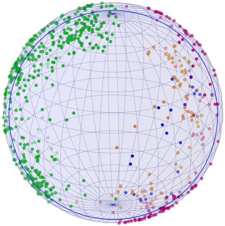
# The $C2'$-Endo Sugar Pucker Revisited



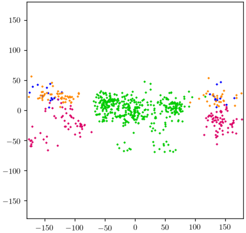

3 heminucleotide angles




TS-PCA of 7 angles



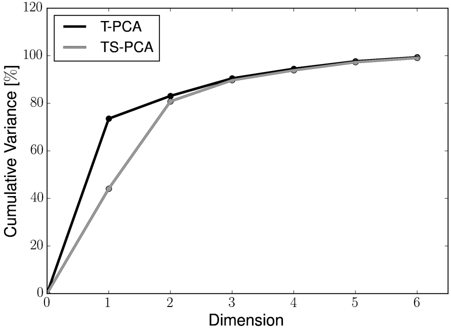
T-PCA of 7 angles




planar view

# T-PCA is one dimension ahead of TS-PCA

Indeed, for the *C*2′-endo sugar pucker:




Note that the 2D T-PCA component is almost a great sphere
⇒ no subsphere advantage in 2D.

cf. Eltzner et al. (2018).

# Outlook: Smeariness

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

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

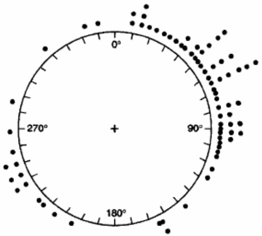



**Figure 1.5** Circular plot of the turtle data of Table 1.5.

from Mardia and Jupp (2000).

# Outlook: Smeariness

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

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

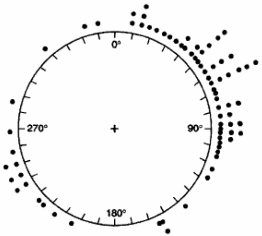
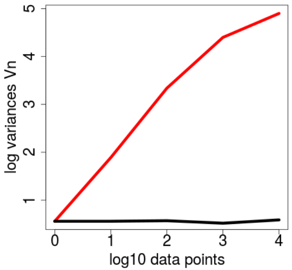



**Figure 1.5** Circular plot of the turtle data of Table 1.5.

from Mardia and Jupp (2000).

Recall

$n\mathrm{var}[\overline{X}_n^*] \to \mathrm{var}[X]$ a.s.




Bootstrapped variance black = Euclidean in $[-\pi, \pi] \subset \mathbb{R}$,

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

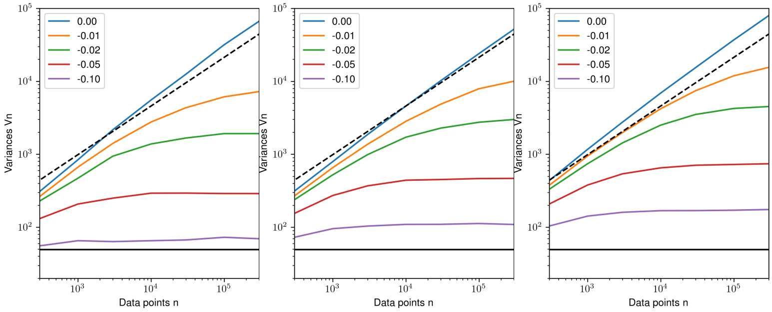
# $k$-Smeariness (Eltzner and S.H. 2019)

If (cf. also S.H and Hotz 2015)

$$n^{\frac{1}{2(k+1)}}\Big(\phi(\mu_n) - \phi(\mu)\Big)$$

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

- $k = 2$ smeary (dashed line)




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

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

# References

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

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

Bhattacharya, R. N. and V. Patrangenaru (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.

Eltzner, B., S. Huckemann, and K. V. Mardia (2018). Torus principal component analysis with applications to rna structure. *Ann. Appl. Statist. 12*(2), 1332–1359.

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

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

Fréchet, M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. *Annales de l'Institut de Henri Poincaré 10*(4), 215–310.

Gardner, P. P. (2003). *Simulating the RNA-world and computational ribonomics: a thesis presented for the degree of Doctor of Philosophy in Biomathematics at Massey University, Palmerston North, New Zealand*. Ph. D. thesis, Massey University.

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

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

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

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

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

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