Bayesian Clustering of Curves with Applications to Juvenile and Hominin Cochlear Shapes

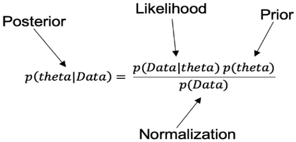
Anis Fradi, Chafik Samir, and José Braga

MLOMA December 21, 2023

Clermont, France

Bayesian inference: Some applications

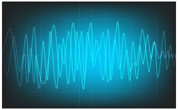
Bayesian inference: derives the posterior probability from a "prior" information and a "likelihood" function from the observed data.



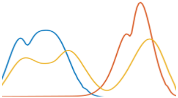
Models:

Applications:

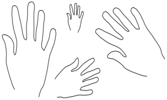
Vectors



Signals



Functions



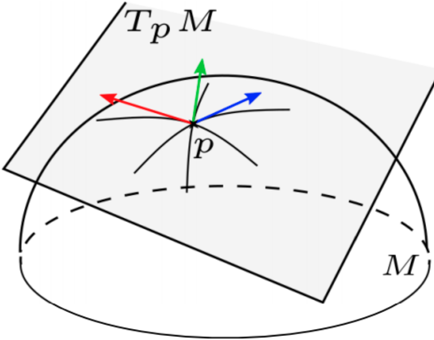
Shapes

Contents

Introduction: Riemannian manifold

Definition

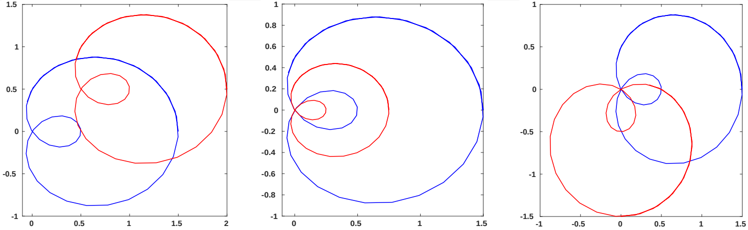
In differential geometry, a **Riemannian manifold** (M, g) is a nonlinear space M equipped with a positive-definite inner product g on the tangent space T_pM at each point p.



Introduction: Shapes of curves

Definition

A **shape** is the geometrical information that remains when affine transformations (translation, scaling and rotation) are filtered out.



translation scaling rotation

- Shape analysis of curves deals with the Generalized Procrustes Analysis (**GPA**) algorithm where the curve is represented with n landmarks in $\mathbb{R}^d$ $\Longrightarrow$ finite vectors on the sphere.
- Shape analysis of curves deals with the tangent PCA (**TPCA**) $\Longrightarrow$ finite vectors projected into the tangent space of the sphere + PCA.

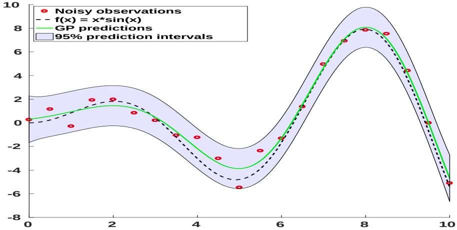
Introduction: Gaussian process (GP)

- A covariance is a bi-variate function $c : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$; $(x, x') \mapsto c(x, x')$ that characterizes the dependence between two random variables.

- A function $f : \mathbb{R} \to \mathbb{R}$ is modeled with a centered **GP**, denoted $f \sim \mathcal{GP}(0, c)$, if $\mathbf{f} = (f(x_1), ..., f(x_N))^T \sim \mathcal{N}(0, \mathbf{C})$ for any $N \geq 1$.

- **C** refers to the covariance matrix constructed from the covariance function c such that: $\mathbf{C}_{ij} = c(x_i, x_j) = cov(f(x_i), f(x_j))$.

- An example of a GP regression model:



Problem formulation

- Observations:
  $\beta_1, \ldots, \beta_N$ where $\beta_i : I = [0,1] \to \mathbb{R}^d$; $d \geq 1$.
- Goal:
  Assign each curve to one of K clusters with $K << N$.
- Problem:
  In practice, we can not observe all β_i.
- Notations:
  - Let $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)$ be a discretization of I.
  - We only observe a discretization of β_i, i.e., $\beta_i \circ F(\boldsymbol{\xi})$ where F is a reparametrization, identified with a **cumulative distribution function** (CDF) defined on $I = [0,1]$, belonging to

$$\mathcal{F} = \left\{ F : I \to I \mid F(0) = 0,\ F(1) = 1,\ \text{and } \dot{F} \text{ is nonnegative} \right\}$$

- References:
  Kendall (1984) ; Srivastava et al. (2011).

Examples of F and β: $d = 3$ and $n = 100$

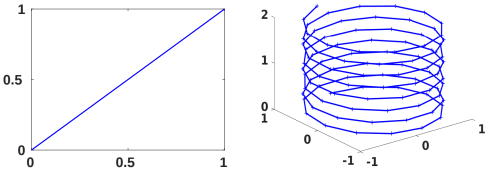


Figure: F is the **uniform** CDF on I: $F(\xi) = \xi$ and $\beta \circ F(\boldsymbol{\xi})$.

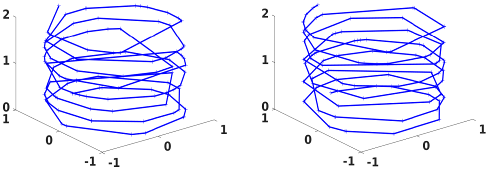


Figure: The same curve β with two **different** CDFs F_1 and F_2. The difference in locations is $||\beta \circ F_1 - \beta \circ F_2|| \neq 0$.

Main idea

- <u>Issues:</u>
  1. $\mathcal{F}$ is a **group** of diffeomorphisms without any geometric structure.
  2. Minimizing a cost function on $F \in \mathcal{F}$ is complicated and intractable.

- <u>Solutions:</u>
  1. $\mathcal{F}$ is isometrically mapped to the Hilbert **upper-hemisphere**

$$\mathcal{H} = \left\{ \psi \equiv \sqrt{\dot{F}} \mid \psi \text{ is nonnegative, and } ||\psi||_{\mathbb{L}^2} = \left(\int_I \psi(t)^2 dt \right)^{1/2} = 1 \right\}$$

  2. $(\mathcal{H}, <.,.>_{\mathbb{L}^2})$ is a complete **Riemannian manifold**.
  3. Given $\psi \in \mathcal{H}$ and $g \in T_\psi(\mathcal{H})$ the **geodesic path** with an initial position ψ and a direction g at any time instant t satisfies

$$\psi(t) = \cos\left(t||g||_{\mathbb{L}^2}\right)\psi + \sin\left(t||g||_{\mathbb{L}^2}\right)\frac{g}{||g||_{\mathbb{L}^2}}$$

CDF expansion

- If $\psi \sim \mathcal{GP}(0, c) \implies$ Its **Karhunen-Loève** expansion is

$$\psi(t) = \sum_{l=1}^{\infty} a_l \phi_l(t), \text{ with } a_l \overset{\text{ind}}{\sim} \mathcal{N}(0, \lambda_l) \text{ and } (\phi_l)_l \text{ is a } \mathbb{L}^2 \text{ basis}$$

$\hookrightarrow (\lambda_l)_l$ and $(\phi_l)_l$ refers to **eigen-values** and **eigen-functions** of c.

- Identifying ψ and F with their **truncated versions** at order m

$$\psi_m(t) = \sum_{l=1}^{m} a_l \phi_l(t) \quad \text{and} \quad F_m(\xi) = \int_0^{\xi} \psi_m^2(t) dt$$

Proposition

F_m is a CDF if and only if $A = \left(a_1, \ldots, a_m\right)^T \in \mathcal{S}^{m-1}$ where

$$\mathcal{S}^{m-1} = \left\{ A \in \mathbb{R}^m \mid ||A||_2 = \left(\sum_{l=1}^{m} a_l^2 \right)^{1/2} = 1 \right\}$$

Square-root velocity function (SRVF)

- Problems:
  1. The $\mathbb{L}^2$ metric is not a good choice to quantify the dissimilarity between curves $\Longrightarrow$ The **elastic** metric.
  2. The implementation of the elastic metric is hard in practice.

- Solutions:
  1. A curve β can be represented by its **SRVF** (or q-function)

$$
\begin{aligned}
q : I &\rightarrow \mathbb{R}^d \\
\xi &\mapsto q(\xi) = \begin{cases} \dfrac{\dot{\beta}(\xi)}{\sqrt{||\dot{\beta}(\xi)||_2}} & \text{if } \dot{\beta}(\xi) \neq 0. \\ 0 & \text{otherwise.} \end{cases}
\end{aligned}
$$

  2. $\beta \circ F$ is then represented by

$$q^*(\xi) = \sqrt{\dot{F}(\xi)}\, q(F(\xi))$$

Advantages of SRVF

- The elastic metric defined on the shape space of curves β reduces to a $\mathbb{L}^2$ metric on the space of SRVFs q.

- **Invariance** to: translation, scaling, rotation and reparametrization since $||q_1^* - q_2^*|| = ||q_1 - q_2||$.

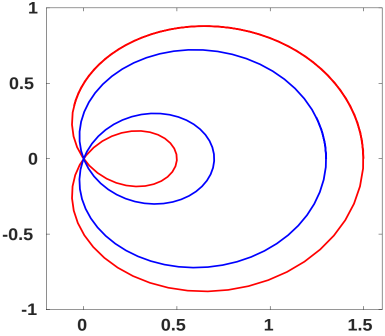
- Given a random sample $q_1, \ldots, q_N$, their **Fréchet mean** $\tilde{q}(\xi)$ minimizing the Fréchet variance

$$\mathbb{V}(q) = \frac{1}{N} \sum_{i=1}^{N} \inf_{F_i \in \mathcal{F}} ||q - q_i^*||^2$$

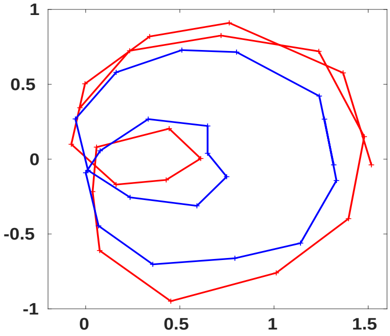
results to be the **Euclidean mean**, i.e., $\tilde{q}(\xi) = \frac{1}{N} \sum_{i=1}^{N} q_i^*(\xi)$.

Example of true and observed curves ($N = 2$, $K = 2$, $\sigma^2 = 0.1$)

True curve to be estimated for k-th cluster: $\tilde{q}^k(\boldsymbol{\xi})$, $k = 1, ..., K$



Observed curve: $q_i^*(\boldsymbol{\xi}) | C_i = k \sim \mathcal{N}(\tilde{q}^k(\boldsymbol{\xi}), \sigma^2 \mathcal{I})$, $i = 1, ..., N$



Bayesian clustering with GMM: K clusters

Finding the **optimal** truncated CDF F_m^k depending on A^k for k-th cluster.

Assumptions

- Let $\pi_k = \mathbb{P}(C_i = k)$ with $k = 1, \dots, K$.
- Let $q_i^*(\boldsymbol{\xi}) | C_i = k \sim \mathcal{N}(\tilde{q}^k(\boldsymbol{\xi}), \sigma^2 \mathcal{I})$.

Bayesian inference on coefficients A^k

- Likelihood:
  $\mathbb{P}(q_1, \dots, q_N | A^1, \dots, A^K, \pi_1, \dots, \pi_K, \tilde{q}^{1,m}(\boldsymbol{\xi}), \dots, \tilde{q}^{K,m}(\boldsymbol{\xi}), \sigma^2) \propto$
  $\prod_{i=1}^N \Big(\sum_{k=1}^K \pi_k \exp\big(-\frac{1}{2\sigma^2} ||q_i^*(\boldsymbol{\xi}) - \tilde{q}^{k,m}(\boldsymbol{\xi})||_2^2 \big) \Big)$
- Prior: $\mathbb{P}(A^k) \propto \exp\big(-\frac{1}{2} \sum_{l=1}^m \frac{{a_l^k}^2}{\lambda_l} \big) \times \delta_{\{A^k \in \mathcal{S}^{m-1}\}}$
- Log-posterior:
  $\log \mathbb{P}(A^1, \dots, A^K | q_1, \dots, q_N, \pi_1, \dots, \pi_K, \tilde{q}^{1,m}(\boldsymbol{\xi}), \dots, \tilde{q}^{K,m}(\boldsymbol{\xi}), \sigma^2) \propto$
  $\sum_{i=1}^N \log \Big(\sum_{k=1}^K \pi_k \exp\big(-\frac{1}{2\sigma^2} ||q_i^*(\boldsymbol{\xi}) - \tilde{q}^{k,m}(\boldsymbol{\xi})||_2^2 \big) \Big) - \frac{1}{2} \sum_{k=1}^K \sum_{l=1}^m \frac{{a_l^k}^2}{\lambda_l}$

Spherical Hamiltonian Monte Carlo (HMC) sampling: 10^4 iterations

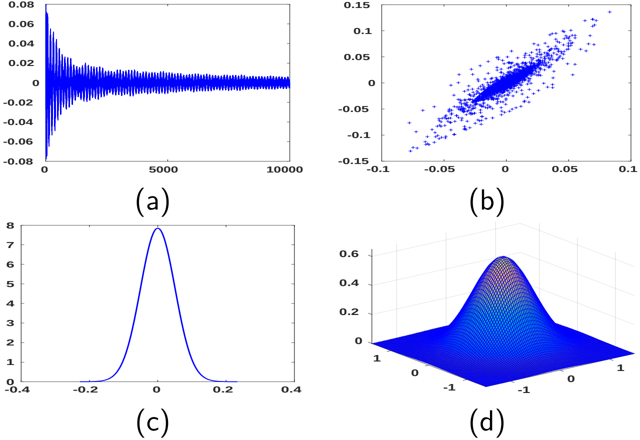


Figure: The **Markov chain** trajectory of: a_1^1 (a) and (a_1^1, a_2^1) (b). The nonparametric **density estimation** of: a_1^1 (c) and (a_1^1, a_2^1) (d).

Experimental results: Methods

- Our method: spherical HMC sampling for A^k with an extra MCMC sampling for π_k, $\tilde{q}^{k,m}(\boldsymbol{\xi})$, and σ^2.

- Probability that i-th curve belongs to k-th cluster

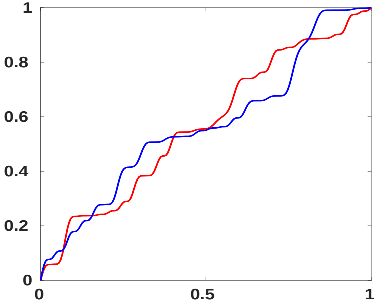
$$\mathbb{P}(C_i = k|q_i) = \frac{\pi_k \exp\left(-\frac{1}{2\sigma^2}||q_i^*(\boldsymbol{\xi}) - \tilde{q}^{k,m}(\boldsymbol{\xi})||_2^2\right)}{\sum_{k=1}^{K} \pi_k \exp\left(-\frac{1}{2\sigma^2}||q_i^*(\boldsymbol{\xi}) - \tilde{q}^{k,m}(\boldsymbol{\xi})||_2^2\right)}$$

- Comparison:
  1. GPA-kmeans and GPA-kmedoids, when applying the GPA to $\beta_i(\boldsymbol{\xi})$ $\Longrightarrow$ vectors belonging to $\mathcal{S}^{nd-d-1-\frac{1}{2}d(d-1)}$.
  2. TPCA-kmeans and TPCA-GMM, when applying the PCA to shape vectors projected onto the tangent space of the sphere $\Longrightarrow$ vectors belonging to $\mathbb{R}^2$ or $\mathbb{R}^3$.

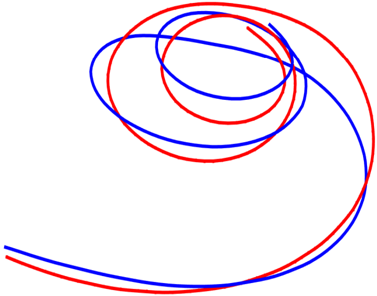
Experimental results: First dataset

- Dataset: 94 cochlea for juvenile.
- Dimension: $n \times d = 200 \times 3$.
- Cluster 1: girls & Cluster 2: boys.

Cluster of girls (blue) and Cluster of boys (red).



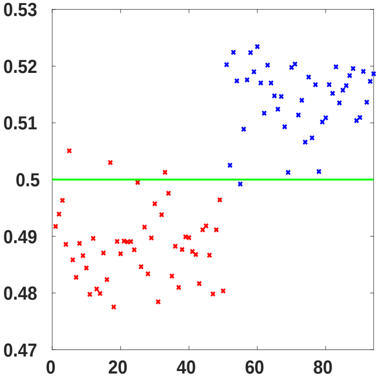
The optimal reparametrization



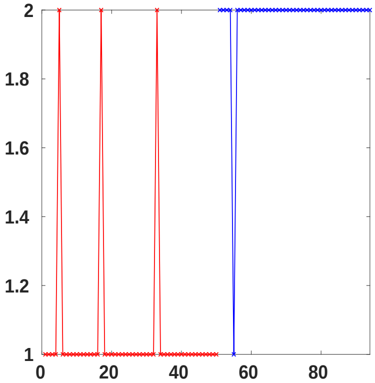
The Fréchet mean

Experimental results: Clustering

Cluster of girls (blue) and Cluster of boys (red).



The probability $\mathbb{P}(C_i = 1|q_i)$



The resulting cluster

Experimental results: Accuracy rates

Table: Mean clustering error (MCE), specificity (SP) and sensibility (SE) for juvenile cochlea.

Methods	MCE	SP	SE
TPCA-GMM	41.75%	58.07%	58.5%
TPCA-kmeans	41.54%	58.44%	58.5%
GPA-kmeans	25.11%	74.4%	75.45%
GPA-kmedoids	10.85%	89.8%	88.41%
Proposed	**4.26%**	**94%**	**97.73%**

Experimental results: Second dataset

- Dataset: 80 cochlea for hominin.
- Dimension: $n \times d = 200 \times 3$.
- Cluster 1: Modern humans (HSS) & Cluster 2: Paranthropus (PAR) & Cluster 3: Gorillas (GOR) & Cluster 4: Chimpanzees (PAN) & Cluster 5: Australopithecus (AUS).

Figure: The Fréchet mean of each cluster.

Experimental results: Accuracy rates

The probability: $\mathbb{P}(C_i = k | q_i)$.

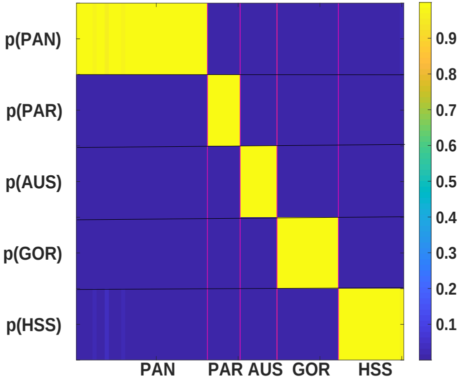
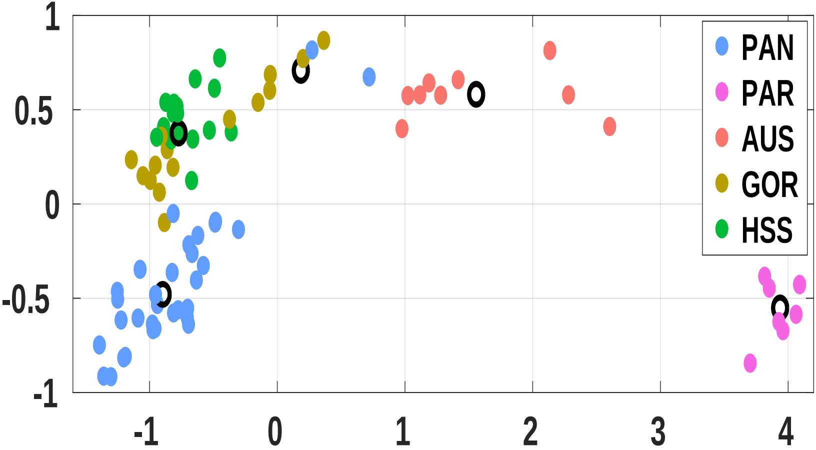


Table: Mean clustering error (MCE) for hominin cochlea.

Methods	MCE
TPCA-GMM	15%
TPCA-kmeans	20%
GPA-kmeans	12.5%
GPA-kmedoids	10%
Proposed	**0%**

Experimental results: Illustration

TPCA with PC1= 82.5% and PC2= 10.2% of variance.



Thank you for your attention !!!

CNRS PRIME research project.
More details about cochlear data collection and analysis:

jose.braga@univ-tlse3.fr

chafik.samir@uca.fr