

Nonparametric Bayes Inference on Manifolds with Applications to Hypersphere and Shape Analysis

Abhishek Bhattacharya
Department of Statistics, Duke University
Joint work with Prof. D.Dunson

June 2010

Contents

- 1 Background & Motivation
- 2 Nonparametric Density Estimation on Manifolds
 - Properties
- 3 Applications to Landmark-Based Planar Shapes
- 4 NP Bayes Classification
 - Classification on d dimensional Sphere
- 5 NP Bayes Testing
 - Example 1
 - Example 2

- To develop very general class of nonparametric Bayes models & inferential procedures
- Not just limited to Euclidean spaces
- Particular emphasis on Riemannian manifolds such as shape spaces and hyperspheres

- To develop very general class of nonparametric Bayes models & inferential procedures
- Not just limited to Euclidean spaces
- Particular emphasis on Riemannian manifolds such as shape spaces and hyperspheres
- Nonparametric Density estimation on manifolds

- To develop very general class of nonparametric Bayes models & inferential procedures
- Not just limited to Euclidean spaces
- Particular emphasis on Riemannian manifolds such as shape spaces and hyperspheres
- Nonparametric Density estimation on manifolds
- Classification with features lying on the manifold &

- To develop very general class of nonparametric Bayes models & inferential procedures
- Not just limited to Euclidean spaces
- Particular emphasis on Riemannian manifolds such as shape spaces and hyperspheres
- Nonparametric Density estimation on manifolds
- Classification with features lying on the manifold &
- Hypothesis testing

Motivating Application 1 - Morphometrics

- Interest in studying shape of an organism & using shape for classification

Motivating Application 1 - Morphometrics

- Interest in studying shape of an organism & using shape for classification
- Example: Data on images of gorilla skulls and their gender
- Since different images obtained under different orientations, scale etc, it is important to be invariant to translations, rotations & scaling

Motivating Application 1 - Morphometrics

- Interest in studying shape of an organism & using shape for classification
- Example: Data on images of gorilla skulls and their gender
- Since different images obtained under different orientations, scale etc, it is important to be invariant to translations, rotations & scaling
- How to nonparametric estimate a shape density across organisms?

Motivating Application 1 - Morphometrics

- Interest in studying shape of an organism & using shape for classification
- Example: Data on images of gorilla skulls and their gender
- Since different images obtained under different orientations, scale etc, it is important to be invariant to translations, rotations & scaling
- How to nonparametric estimate a shape density across organisms?
- How to test for differences in shape & obtain a good method for classification based on shape?

Motivating Application 2 - Spatial Statistics

- Data correspond to locations on a sphere and other characteristics at each location

Motivating Application 2 - Spatial Statistics

- Data correspond to locations on a sphere and other characteristics at each location
- Example: Volcano locations across the globe with varying types of volcanos

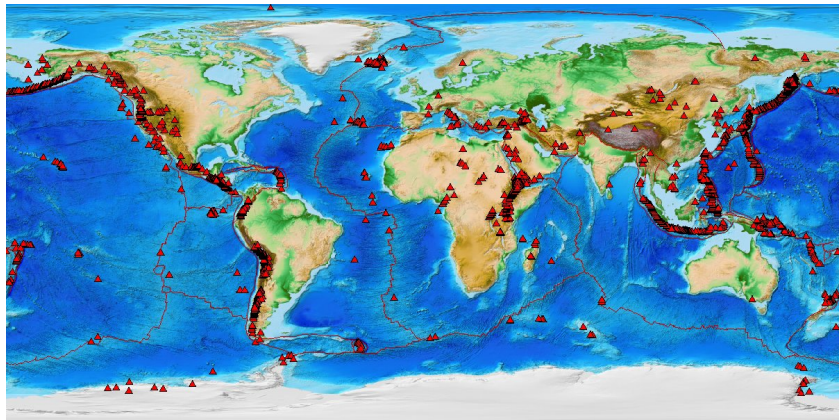
Motivating Application 2 - Spatial Statistics

- Data correspond to locations on a sphere and other characteristics at each location
- Example: Volcano locations across the globe with varying types of volcanos
- Important to not treat the locations as Euclidean - may be o.k. approximation locally but distorts the geometry & can lead to relatively poor results

Motivating Application 2 - Spatial Statistics

- Data correspond to locations on a sphere and other characteristics at each location
- Example: Volcano locations across the globe with varying types of volcanos
- Important to not treat the locations as Euclidean - may be o.k. approximation locally but distorts the geometry & can lead to relatively poor results
- Interest in testing whether type of volcano varies with location

Plot of NOAA Volcano Location Data



Direction to Take

- Parametric models - clearly too inflexible

Direction to Take

- Parametric models - clearly too inflexible
- Nonparametric Bayes - current models focus on Euclidean spaces

Direction to Take

- Parametric models - clearly too inflexible
- Nonparametric Bayes - current models focus on Euclidean spaces
- Very little done in non-Euclidean spaces

Kernel Mixtures on Compact Metric Spaces

- M = compact metric space, with X a random variable on M

Kernel Mixtures on Compact Metric Spaces

- M = compact metric space, with X a random variable on M
- Assume that the distribution of X has a density wrt base measure V on M

Kernel Mixtures on Compact Metric Spaces

- M = compact metric space, with X a random variable on M
- Assume that the distribution of X has a density wrt base measure V on M
- Let $K(m; \mu, \kappa)$ denote a probability kernel on M with location $\mu \in M$ and inverse-scale/precision $\kappa \in \mathbb{R}^+$

Kernel Mixtures on Compact Metric Spaces

- M = compact metric space, with X a random variable on M
- Assume that the distribution of X has a density wrt base measure V on M
- Let $K(m; \mu, \kappa)$ denote a probability kernel on M with location $\mu \in M$ and inverse-scale/precision $\kappa \in \mathbb{R}^+$
- We focus on the kernel mixture model with

$$f(m; P, \kappa) = \int_M K(m; \mu, \kappa) P(d\mu), \quad (P, \kappa) \sim \Pi_1$$

- This induces a prior Π on space of all densities $\mathcal{D}(M)$

Some Questions

- How do we know that a particular prior choice is “good”?

Some Questions

- How do we know that a particular prior choice is “good”?
- How do we modify such kernel mixture models to obtain “good” methods for classification & testing?

Some Questions

- How do we know that a particular prior choice is “good”?
- How do we modify such kernel mixture models to obtain “good” methods for classification & testing?
- How to check if the estimates are consistent?

Some Questions

- How do we know that a particular prior choice is “good”?
- How do we modify such kernel mixture models to obtain “good” methods for classification & testing?
- How to check if the estimates are consistent?
- There are some standard choices for Π_1 - e.g., Dirichlet process (DP) prior on P leads to DP mixture model for density

Some Questions

- How do we know that a particular prior choice is “good”?
- How do we modify such kernel mixture models to obtain “good” methods for classification & testing?
- How to check if the estimates are consistent?
- There are some standard choices for Π_1 - e.g., Dirichlet process (DP) prior on \mathcal{P} leads to DP mixture model for density
- Given a random sample, we can devise algorithm to draw from the posterior of f and estimate it

Some Questions

- How do we know that a particular prior choice is “good”?
- How do we modify such kernel mixture models to obtain “good” methods for classification & testing?
- How to check if the estimates are consistent?
- There are some standard choices for Π_1 - e.g., Dirichlet process (DP) prior on \mathcal{P} leads to DP mixture model for density
- Given a random sample, we can devise algorithm to draw from the posterior of f and estimate it

Large Support & Consistency

- In Bayesian inference “nonparametric” implies large support

Large Support & Consistency

- In Bayesian inference “nonparametric” implies large support
- Regardless of *true* data-generating model, the prior can generate densities arbitrary close to the truth

Large Support & Consistency

- In Bayesian inference “nonparametric” implies large support
- Regardless of *true* data-generating model, the prior can generate densities arbitrary close to the truth

Large Support & Consistency

- In Bayesian inference “nonparametric” implies large support
- Regardless of *true* data-generating model, the prior can generate densities arbitrary close to the truth
- Under some conditions, implies posterior consistency - the posterior gives most of its mass in arbitrarily small neighborhood of the truth

Assumptions for Large Support & Weak Posterior Consistency

Let P_t denote the true distribution and f_t be its density. Assume

A1 K is continuous on $M \times M \times \mathbb{R}^+$

A2 For any cont. ϕ

$$\lim_{\kappa \rightarrow \infty} \sup_{x \in M} \left| \phi(x) - \int_M K(x; \mu, \kappa) \phi(\mu) V(d\mu) \right| = 0$$

A3 For any $\kappa_0 > 0$, $\exists \kappa \geq \kappa_0$ s.t. $(P_t, \kappa) \in \text{supp}(\Pi_1)$
(weak support)

A4 f_t is continuous

A5 f_t is strictly positive

Theorem (Bhattacharya & Dunson, 2010a)

*Under assumptions **A1-A4**, given any $\epsilon > 0$,*

$$\Pi\left(\left\{f : \sup_{x \in M} |f(x) - f_t(x)| < \epsilon\right\}\right) > 0.$$

and Kullback-Leibler (KL) condition satisfied

Using Schwartz theorem (Schwartz, 1965) **weak posterior consistency** follows

Theorem (Bhattacharya & Dunson, 2010a)

Under assumptions **A1-A4**, given any $\epsilon > 0$,

$$\Pi\left(\left\{f : \sup_{x \in M} |f(x) - f_t(x)| < \epsilon\right\}\right) > 0.$$

and Kullback-Leibler (KL) condition satisfied

Using Schwartz theorem (Schwartz, 1965) **weak posterior consistency** follows

$$\Pi(U|X_1, \dots, X_n) \longrightarrow 1 \text{ a.s. as } n \rightarrow \infty$$

U being a weak open neighborhood of f_t

Assumptions for Strong Posterior Consistency

A6 $\Pi_1(\mathcal{M}(M) \times (n^{(1/a)}, \infty)) < \exp(-n\beta)$ for some $a > a_1 a_3$ and $\beta > 0$, where

Assumptions for Strong Posterior Consistency

A6 $\Pi_1(\mathcal{M}(M) \times (n^{(1/a)}, \infty)) < \exp(-n\beta)$ for some $a > a_1 a_3$ and $\beta > 0$, where

A7 $\exists \mathcal{K}_1, a_1, A_1 > 0$ such that for all $\mathcal{K} \geq \mathcal{K}_1, \mu, \nu \in M$,

$$\sup_{m \in M, \kappa \in [0, \mathcal{K}]} |K(m; \mu, \kappa) - K(m; \nu, \kappa)| \leq A_1 \mathcal{K}^{a_1} \rho(\mu, \nu).$$

A8 $\exists a_3 > 0$ s.t given any $\epsilon > 0$, M can be covered by finitely many ϵ -diameter balls of number of the order ϵ^{-a_3} .

Strong Posterior Consistency

Theorem (Bhattacharya & Dunson, 2010b)

*Under assumptions **A1-A8**, strong posterior consistency holds, i.e. the posterior probability of any total variation neighborhood of f_t converges to 1 almost surely.*

Introduction to Planar Shape Space

- k landmark points or k -ad picked from an object in 2D - similarity shape removes effects of translation, rotation & scaling.

Introduction to Planar Shape Space

- k landmark points or k -ad picked from an object in 2D - similarity shape removes effects of translation, rotation & scaling.
- Denote k -ad by complex k vector $z = (z_1, \dots, z_k)' \in \mathbb{C}^k$ - remove translation by subtracting centroid, $z_c = z - \bar{z}$.

Introduction to Planar Shape Space

- k landmark points or k -ad picked from an object in 2D - similarity shape removes effects of translation, rotation & scaling.
- Denote k -ad by complex k vector $z = (z_1, \dots, z_k)' \in \mathbb{C}^k$ - remove translation by subtracting centroid, $z_c = z - \bar{z}$.
- Remove scaling by normalizing coordinates of z_c to obtain point w on complex unit sphere - referred to as preshape

Introduction to Planar Shape Space

- k landmark points or k -ad picked from an object in 2D - similarity shape removes effects of translation, rotation & scaling.
- Denote k -ad by complex k vector $z = (z_1, \dots, z_k)' \in \mathbb{C}^k$ - remove translation by subtracting centroid, $z_c = z - \bar{z}$.
- Remove scaling by normalizing coordinates of z_c to obtain point w on complex unit sphere - referred to as preshape
- Similarity shape of z is orbit of w under all rotations in 2D - the space of all such orbits is the **planar shape space** Σ_2^k

Example: Shapes of Gorilla Skulls

- 8 landmarks chosen on the midline plane of the 2D image of 29 male and 30 female gorilla skulls (*Dryden and Mardia, 98*).

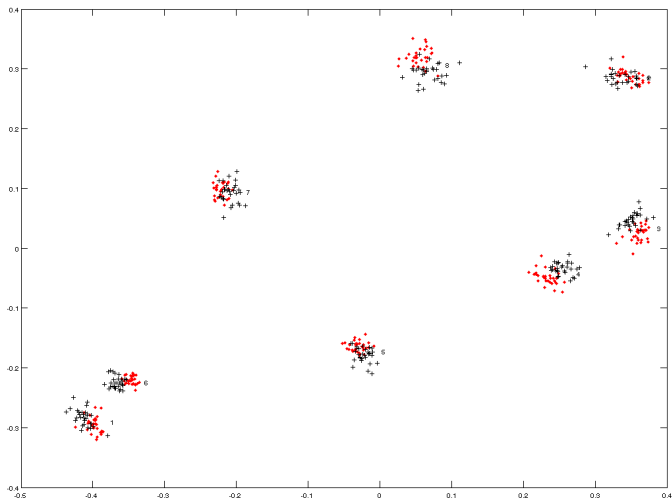
Example: Shapes of Gorilla Skulls

- 8 landmarks chosen on the midline plane of the 2D image of 29 male and 30 female gorilla skulls (*Dryden and Mardia, 98*).
- Goal: Study the shapes of the skulls and use that to detect difference in shapes between the sexes and predict the sex.

Example: Shapes of Gorilla Skulls

- 8 landmarks chosen on the midline plane of the 2D image of 29 male and 30 female gorilla skulls (*Dryden and Mardia, 98*).
- Goal: Study the shapes of the skulls and use that to detect difference in shapes between the sexes and predict the sex.
- Two mutually independent iid sample of planar shapes of sizes 29 and 30 on Σ_2^k , $k = 8$.

Gorilla Skull Preshapes: Females (red), Males (+)



Mean shape

- Σ_2^k is a compact Riemannian manifold

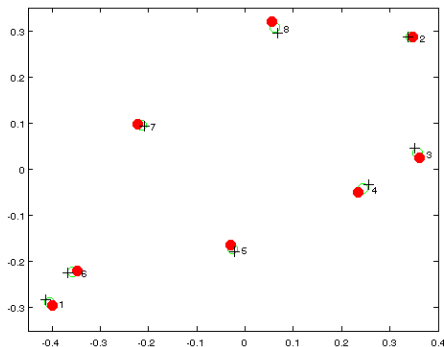
Mean shape

- Σ_2^k is a compact Riemannian manifold
- geodesic distance can be defined between two shapes, an embedding into Euclidean space induces the “extrinsic distance”

Mean shape

- Σ_2^k is a compact Riemannian manifold
- geodesic distance can be defined between two shapes, an embedding into Euclidean space induces the “extrinsic distance”
- Using an integrated distance square loss function, can define a notion of mean of a probability - **intrinsic** corresponding to geodesic distance or **extrinsic** using the extrinsic distance (*Bhattacharya, R. & Patrangenaru, 2003*, *Bhattacharya, A. & Bhattacharya, R., 2008*)

Gorilla Skulls: Extrinsic Mean Shapes Plot



SAMPLE EX. MEANS FOR FEMALES (r.), MALES (+) ALONG
WITH THE POOLED SAMPLE EX. MEAN (go)

Density modelling on Σ_2^k

- Simple parametric kernel corresponds to the complex Watson distribution (Dryden & Mardia, 98)

$$CW(m; \mu, \kappa) = c^{-1}(\kappa) \exp(\kappa |z^* v|^2),$$

z, v preshapes of $m, \mu \in \Sigma_2^k$, $*$ denotes complex conjugate transpose

Density modelling on Σ_2^k

- Simple parametric kernel corresponds to the complex Watson distribution (Dryden & Mardia, 98)

$$CW(m; \mu, \kappa) = c^{-1}(\kappa) \exp(\kappa |z^* v|^2),$$

z, v preshapes of $m, \mu \in \Sigma_2^k$, $*$ denotes complex conjugate transpose

- We use the kernel location mixture model

$$f(m; P, \kappa) = \int CW(m; \mu, \kappa) P(d\mu), \quad m \in \Sigma_2^k,$$

Density modelling on Σ_2^k

- Simple parametric kernel corresponds to the complex Watson distribution (Dryden & Mardia, 98)

$$CW(m; \mu, \kappa) = c^{-1}(\kappa) \exp(\kappa |z^* v|^2),$$

z, v preshapes of $m, \mu \in \Sigma_2^k$, $*$ denotes complex conjugate transpose

- We use the kernel location mixture model

$$f(m; P, \kappa) = \int CW(m; \mu, \kappa) P(d\mu), \quad m \in \Sigma_2^k,$$

- For prior Π_1 , we let $P \sim DP(\omega_0 P_0)$, with $P_0 = CW(\mu_0, \sigma_0)$ corresponding to a complex Watson base &

Density modelling on Σ_2^k

- Simple parametric kernel corresponds to the complex Watson distribution (Dryden & Mardia, 98)

$$CW(m; \mu, \kappa) = c^{-1}(\kappa) \exp(\kappa |z^* v|^2),$$

z, v preshapes of $m, \mu \in \Sigma_2^k$, $*$ denotes complex conjugate transpose

- We use the kernel location mixture model

$$f(m; P, \kappa) = \int CW(m; \mu, \kappa) P(d\mu), \quad m \in \Sigma_2^k,$$

- For prior Π_1 , we let $P \sim DP(\omega_0 P_0)$, with $P_0 = CW(\mu_0, \sigma_0)$ corresponding to a complex Watson base &
- $\kappa \sim \text{Gam}(a, b)$.

Theoretical Properties & Computation for DPMs of Complex Watsons

- The kernel and prior can be shown to satisfy the sufficient conditions in our theorems to ensure L_∞ and KL support (*Bhattacharya & Dunson, 2010a*)

Theoretical Properties & Computation for DPMs of Complex Watsons

- The kernel and prior can be shown to satisfy the sufficient conditions in our theorems to ensure L_∞ and KL support (*Bhattacharya & Dunson, 2010a*)
- This directly implies W.P.C. at all continuous, positive true densities

Theoretical Properties & Computation for DPMs of Complex Watsons

- The kernel and prior can be shown to satisfy the sufficient conditions in our theorems to ensure L_∞ and KL support (*Bhattacharya & Dunson, 2010a*)
- This directly implies W.P.C. at all continuous, positive true densities
- S.P.C. follows if the scale parameter for the gamma prior depends on n in an appropriate manner (*Bhattacharya & Dunson, 2010b*)

Theoretical Properties & Computation for DPMs of Complex Watsons

- The kernel and prior can be shown to satisfy the sufficient conditions in our theorems to ensure L_∞ and KL support (*Bhattacharya & Dunson, 2010a*)
- This directly implies W.P.C. at all continuous, positive true densities
- S.P.C. follows if the scale parameter for the gamma prior depends on n in an appropriate manner (*Bhattacharya & Dunson, 2010b*)
- A Markov chain can be build for posterior computaton

NP Bayes Classification through Joint Modeling

- Let Y be a categorical variable taking values in $\mathbb{Y} = \{1, 2, \dots, L\}$. eg. Gorilla gender, presence/absence of disease, animal/plant species, etc

NP Bayes Classification through Joint Modeling

- Let Y be a categorical variable taking values in $\mathbb{Y} = \{1, 2, \dots, L\}$. eg. Gorilla gender, presence/absence of disease, animal/plant species, etc
- Goal: nonparametric estimation of classification function $\Pr(Y = l | X = x)$ with X predictors lying on M

NP Bayes Classification through Joint Modeling

- Let Y be a categorical variable taking values in $\mathbb{Y} = \{1, 2, \dots, L\}$. eg. Gorilla gender, presence/absence of disease, animal/plant species, etc
- Goal: nonparametric estimation of classification function $\Pr(Y = l | X = x)$ with X predictors lying on M
- Idea: Nonparametrically model the joint of Y, X to induce a prior on the classification function

Joint Kernel Mixture Model

- Model the joint distribution of (X, Y) via

$$f(x, y; P, \kappa) = \int_{\mathbb{X} \times \mathcal{S}_{L-1}} \nu_y K(x; \mu, \kappa) P(d\mu d\nu), \quad (x, y) \in M \times \mathbb{Y},$$

Joint Kernel Mixture Model

- Model the joint distribution of (X, Y) via

$$f(x, y; P, \kappa) = \int_{\mathbb{X} \times \mathcal{S}_{L-1}} \nu_y K(x; \mu, \kappa) P(d\mu d\nu), \quad (x, y) \in M \times \mathbb{Y},$$

- $\nu = (\nu_1, \dots, \nu_L)' \in \mathcal{S}_{L-1}$ is a probability vector on the simplex $\mathcal{S}_{L-1} = \{\nu \in [0, 1]^L : \sum \nu_l = 1\}$

Joint Kernel Mixture Model

- Model the joint distribution of (X, Y) via

$$f(x, y; P, \kappa) = \int_{\mathbb{X} \times \mathcal{S}_{L-1}} \nu_y K(x; \mu, \kappa) P(d\mu d\nu), \quad (x, y) \in M \times \mathbb{Y},$$

- $\nu = (\nu_1, \dots, \nu_L)' \in \mathcal{S}_{L-1}$ is a probability vector on the simplex $\mathcal{S}_{L-1} = \{\nu \in [0, 1]^L : \sum \nu_l = 1\}$
- $K(\cdot; \mu, \kappa)$ is a kernel located at $\mu \in M$ with precision $\kappa \in \mathbb{R}^+$

Joint Kernel Mixture Model

- Model the joint distribution of (X, Y) via

$$f(x, y; P, \kappa) = \int_{\mathbb{X} \times \mathcal{S}_{L-1}} \nu_y K(x; \mu, \kappa) P(d\mu d\nu), \quad (x, y) \in M \times \mathbb{Y},$$

- $\nu = (\nu_1, \dots, \nu_L)' \in \mathcal{S}_{L-1}$ is a probability vector on the simplex $\mathcal{S}_{L-1} = \{\nu \in [0, 1]^L : \sum \nu_l = 1\}$
- $K(\cdot; \mu, \kappa)$ is a kernel located at $\mu \in M$ with precision $\kappa \in \mathbb{R}^+$
- $P \in \mathcal{M}(M \times \mathcal{S}_{L-1})$ is a mixing measure

Joint Kernel Mixture Model

- Model the joint distribution of (X, Y) via

$$f(x, y; P, \kappa) = \int_{\mathbb{X} \times \mathcal{S}_{L-1}} \nu_y K(x; \mu, \kappa) P(d\mu d\nu), \quad (x, y) \in M \times \mathbb{Y},$$

- $\nu = (\nu_1, \dots, \nu_L)' \in \mathcal{S}_{L-1}$ is a probability vector on the simplex $\mathcal{S}_{L-1} = \{\nu \in [0, 1]^L : \sum \nu_l = 1\}$
- $K(\cdot; \mu, \kappa)$ is a kernel located at $\mu \in M$ with precision $\kappa \in \mathbb{R}^+$
- $P \in \mathcal{M}(M \times \mathcal{S}_{L-1})$ is a mixing measure
- Given (μ, ν, κ) , $X \sim K(\cdot; \mu, \kappa)$ & Y independent $\text{Multi}(\nu)$

Joint Kernel Mixture Model

- Model the joint distribution of (X, Y) via

$$f(x, y; P, \kappa) = \int_{\mathbb{X} \times \mathcal{S}_{L-1}} \nu_y K(x; \mu, \kappa) P(d\mu d\nu), \quad (x, y) \in M \times \mathbb{Y},$$

- $\nu = (\nu_1, \dots, \nu_L)' \in \mathcal{S}_{L-1}$ is a probability vector on the simplex $\mathcal{S}_{L-1} = \{\nu \in [0, 1]^L : \sum \nu_l = 1\}$
- $K(\cdot; \mu, \kappa)$ is a kernel located at $\mu \in M$ with precision $\kappa \in \mathbb{R}^+$
- $P \in \mathcal{M}(M \times \mathcal{S}_{L-1})$ is a mixing measure
- Given (μ, ν, κ) , $X \sim K(\cdot; \mu, \kappa)$ & Y independent $\text{Multi}(\nu)$
- Association between X & Y through P

Prior Selection

- Set a prior Π_1 on (P, κ) such as $DP(\omega_0 P_0) \otimes \pi_1$ with P_0 and π_1 a dist. on $M \times S_{c-1}$ and \mathbb{R}^+ respectively

Prior Selection

- Set a prior Π_1 on (P, κ) such as $DP(\omega_0 P_0) \otimes \pi_1$ with P_0 and π_1 a dist. on $M \times S_{c-1}$ and \mathbb{R}^+ respectively
- This induces a prior Π on the joint density f of (X, Y) & hence on $\Pr(Y = j | X = x)$

Prior Selection

- Set a prior Π_1 on (P, κ) such as $DP(\omega_0 P_0) \otimes \pi_1$ with P_0 and π_1 a dist. on $M \times S_{c-1}$ and \mathbb{R}^+ respectively
- This induces a prior Π on the joint density f of (X, Y) & hence on $\Pr(Y = j | X = x)$
- We can use this for classification of new subjects based on features x_{n+1}

Prior Selection

- Set a prior Π_1 on (P, κ) such as $DP(\omega_0 P_0) \otimes \pi_1$ with P_0 and π_1 a dist. on $M \times S_{c-1}$ and \mathbb{R}^+ respectively
- This induces a prior Π on the joint density f of (X, Y) & hence on $\Pr(Y = j | X = x)$
- We can use this for classification of new subjects based on features x_{n+1}
- Under similar conditions to those discussed above, we obtain L_∞ , KL support & L_1 consistency for the classification function (*Bhattacharya & Dunson, 2010c*).

- Let M be the unit sphere $S^d = \{x \in \mathbb{R}^{d+1} : \sum_j x_j^2 = 1\}$.

- Let M be the unit sphere $S^d = \{x \in \mathbb{R}^{d+1} : \sum_j x_j^2 = 1\}$.
- Take kernel K to be vMF density (*von Mises, 1918*)

$$\text{vMF}(x; \mu, \kappa) = c^{-1}(\kappa) \exp(\kappa x' \mu)$$

- Let M be the unit sphere $S^d = \{x \in \mathbb{R}^{d+1} : \sum_j x_j^2 = 1\}$.
- Take kernel K to be vMF density (*von Mises, 1918*)

$$\text{vMF}(x; \mu, \kappa) = c^{-1}(\kappa) \exp(\kappa x' \mu)$$

- The model satisfies large support & consistency conditions (*Bhattacharya & Dunson, 2010c*)

- Let M be the unit sphere $S^d = \{x \in \mathbb{R}^{d+1} : \sum_j x_j^2 = 1\}$.
- Take kernel K to be vMF density (*von Mises, 1918*)

$$\text{vMF}(x; \mu, \kappa) = c^{-1}(\kappa) \exp(\kappa x' \mu)$$

- The model satisfies large support & consistency conditions (*Bhattacharya & Dunson, 2010c*)
- A simple Gibbs sampler can be implemented with conjugate sampling steps

- Let M be the unit sphere $S^d = \{x \in \mathbb{R}^{d+1} : \sum_j x_j^2 = 1\}$.
- Take kernel K to be vMF density (*von Mises, 1918*)

$$\text{vMF}(x; \mu, \kappa) = c^{-1}(\kappa) \exp(\kappa x' \mu)$$

- The model satisfies large support & consistency conditions (*Bhattacharya & Dunson, 2010c*)
- A simple Gibbs sampler can be implemented with conjugate sampling steps
- Better performance than discriminant analysis methods based on mixtures of Gaussians (*BD, 2010c*)

Null and Alternative Hypothesis

- Instead of classifying a new feature, goal is to test whether the distribution of the features differs across the classes

Null and Alternative Hypothesis

- Instead of classifying a new feature, goal is to test whether the distribution of the features differs across the classes
- Test for independence between X and Y

Null and Alternative Hypothesis

- Instead of classifying a new feature, goal is to test whether the distribution of the features differs across the classes
- Test for independence between X and Y
- Alternative hypothesis H_1 corresponds to all joint densities $f(x, y)$ excluding

$$H_0 : f(x, y) = f(x)f(y) \quad \forall (x, y) \in M \times \mathbb{Y}$$

Null and Alternative Hypothesis

- Instead of classifying a new feature, goal is to test whether the distribution of the features differs across the classes
- Test for independence between X and Y
- Alternative hypothesis H_1 corresponds to all joint densities $f(x, y)$ excluding

$$H_0 : f(x, y) = f(x)f(y) \quad \forall (x, y) \in M \times \mathbb{Y}$$

- Model under H_1 is $f(x, y; P, \kappa)$. Set prior Π_1 on parameters (P, κ) - done as earlier for classification

Prior for Null Model

- Under H_0 use the same specification but replace $P(d\mu d\nu)$ with $P_1(d\mu)P_2(d\nu)$

Prior for Null Model

- Under H_0 use the same specification but replace $P(d\mu d\nu)$ with $P_1(d\mu)P_2(d\nu)$
- Distributions of predictor kernel locations & response probability vectors independent

Prior for Null Model

- Under H_0 use the same specification but replace $P(d\mu d\nu)$ with $P_1(d\mu)P_2(d\nu)$
- Distributions of predictor kernel locations & response probability vectors independent
- It follows that X, Y independent given parameters P_1, P_2, κ

Prior for Null Model

- Under H_0 use the same specification but replace $P(d\mu d\nu)$ with $P_1(d\mu)P_2(d\nu)$
- Distributions of predictor kernel locations & response probability vectors independent
- It follows that X, Y independent given parameters P_1, P_2, κ
- Set prior Π_0 on (P_1, P_2, κ)

Bayes Factor

- The Bayes-factor BF in favor of H_1 over H_0 , is then the ratio of the posterior probabilities of the two models =

$$\frac{\int \prod_{i=1}^n f(x_i, y_i; P, \kappa) \Pi_1(dP d\kappa)}{\int \prod_{i=1}^n f_X(x_i; P_1, \kappa) f_Y(y_i; P_2) \Pi_0(dP_1 dP_2 d\kappa)}$$

Bayes Factor

- The Bayes-factor BF in favor of H_1 over H_0 , is then the ratio of the posterior probabilities of the two models =

$$\frac{\int \prod_{i=1}^n f(x_i, y_i; P, \kappa) \Pi_1(dP d\kappa)}{\int \prod_{i=1}^n f_X(x_i; P_1, \kappa) f_Y(y_i; P_2) \Pi_0(dP_1 dP_2 d\kappa)}$$

- For DPM priors, by introducing a latent variable z which is the indicator of accepting H_1 , we devise a simple algorithm for computing BF (*Bhattacharya & Dunson, 2010c*)

Bayes Factor

- The Bayes-factor BF in favor of H_1 over H_0 , is then the ratio of the posterior probabilities of the two models =

$$\frac{\int \prod_{i=1}^n f(x_i, y_i; P, \kappa) \Pi_1(dP d\kappa)}{\int \prod_{i=1}^n f_X(x_i; P_1, \kappa) f_Y(y_i; P_2) \Pi_0(dP_1 dP_2 d\kappa)}$$

- For DPM priors, by introducing a latent variable z which is the indicator of accepting H_1 , we devise a simple algorithm for computing BF (*Bhattacharya & Dunson, 2010c*)
- The BF is consistent under H_1 : $BF \rightarrow \infty$ a.s. if H_1 is true (*BD, 2010c*)

Gorilla Skull Application

- Goal to test if the skull shape distribution differs accross gender and predict gender based on skull shape

Gorilla Skull Application

- Goal to test if the skull shape distribution differs accross gender and predict gender based on skull shape
- Introduce gender label Y and test for independence between X and Y

Gorilla Skull Application

- Goal to test if the skull shape distribution differs accross gender and predict gender based on skull shape
- Introduce gender label Y and test for independence between X and Y
- $BF(H_1 : H_0) > 10^{16}$.
Conclusion: Strong evidence in favor of H_1 hence shape dists for two sexes different.

Gorilla Skull Application

- Goal to test if the skull shape distribution differs accross gender and predict gender based on skull shape
- Introduce gender label Y and test for independence between X and Y
- $BF(H_1 : H_0) > 10^{16}$.
Conclusion: Strong evidence in favor of H_1 hence shape dists for two sexes different.
- Nonparametric frequentist test comparing extrinsic means (*Bhattacharya & Bhattacharya, 08*), p-value $< 10^{-16}$

Gorilla Skull Application

- Goal to test if the skull shape distribution differs across gender and predict gender based on skull shape
- Introduce gender label Y and test for independence between X and Y
- $BF(H_1 : H_0) > 10^{16}$.
Conclusion: Strong evidence in favor of H_1 hence shape dists for two sexes different.
- Nonparametric frequentist test comparing extrinsic means (*Bhattacharya & Bhattacharya, 08*), p-value $< 10^{-16}$
- In simulations, Bayes approach does better - frequentist test only detects changes in mean

Gorilla Skull Application

- Goal to test if the skull shape distribution differs accross gender and predict gender based on skull shape
- Introduce gender label Y and test for independence between X and Y
- $BF(H_1 : H_0) > 10^{16}$.
Conclusion: Strong evidence in favor of H_1 hence shape dists for two sexes different.
- Nonparametric frequentist test comparing extrinsic means (*Bhattacharya & Bhattacharya, 08*), p-value $< 10^{-16}$
- In simulations, Bayes approach does better - frequentist test only detects changes in mean
- Good out of sample classification performance

Volcano location data

- Information on locations and types of 1545 volcanoes across the globe from NOAA National Geophysical Data Center Volcano Location Database

Volcano location data

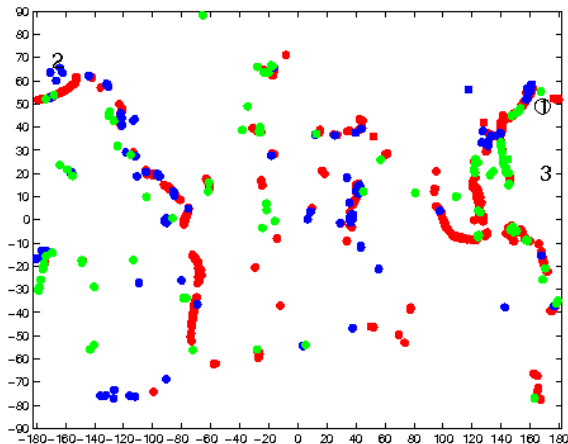
- Information on locations and types of 1545 volcanoes across the globe from NOAA National Geophysical Data Center Volcano Location Database
- Goal to test if there is any association between the location and type of the volcano

Volcano location data

- Information on locations and types of 1545 volcanoes across the globe from NOAA National Geophysical Data Center Volcano Location Database
- Goal to test if there is any association between the location and type of the volcano
- Considered 3 most common types - Strato, Shield and Submarine volcanoes, comprising 999 many

Volcano location data

- Information on locations and types of 1545 volcanoes across the globe from NOAA National Geophysical Data Center Volcano Location Database
- Goal to test if there is any association between the location and type of the volcano
- Considered 3 most common types - Strato, Shield and Submarine volcanoes, comprising 999 many
- Bivariate sample (X, Y) , $X \in \mathcal{S}^2$, $Y \in \{1, 2, 3\}$



Volcano

Locations(long-lat) for Strato(r), Shield(b), Submarine(g). Their sample ex. means 1,2,3. Full sample mean(o)

Findings

- Posterior probability for H_1 is ≈ 1

Findings

- Posterior probability for H_1 is ≈ 1
- Implies strong evidence of variation in the spatial distribution of volcanos according to type

Findings

- Posterior probability for H_1 is ≈ 1
- Implies strong evidence of variation in the spatial distribution of volcanos according to type
- For comparison, perform the frequentist np test for comparing extrinsic means corresponding to different types, p-value = 3.6×10^{-7}

- Considered a broad class of kernel mixture models for density estimation, classification & testing in general spaces

- Considered a broad class of kernel mixture models for density estimation, classification & testing in general spaces
- Theory to verify that a prior leads to large support & consistency

- Considered a broad class of kernel mixture models for density estimation, classification & testing in general spaces
- Theory to verify that a prior leads to large support & consistency
- Applied to hyperspheres & shapes - new computational methods also developed

- Considered a broad class of kernel mixture models for density estimation, classification & testing in general spaces
- Theory to verify that a prior leads to large support & consistency
- Applied to hyperspheres & shapes - new computational methods also developed
- Ongoing: Bayesian dimension reduction on manifolds & methods for manifold learning
- Work to appear at Biometrika & available as 3 discussion papers at Duke

THANK YOU

THANK YOU
THANKS TO GOD