Analysis of Non-Euclidean Data: Use of Differential Geometry in Statistics

Rabi Bhattacharya, The University of Arizona, Tucson, AZ

[Research supported by NSF grant DMS1406872]

June, 2016

Based on joint work with A. Bhattacharya (BB), Lizhen Lin and V. Patrangenaru (BP)
CONTENTS

1. Introduction


3. Applications and Examples.
   (a). $S^d$ (Sphere) Paleomagnetism.
   (b). Kendall’s Planar Shape Space $\Sigma^k_2$-Two-Sample Tests for (1) Schizophrenia and (2) Male & Female Gorilla Skulls.
   (c). 3D Shape Space $R\Sigma^k_3$-Match Pair Test for Glaucoma.
   (d). The space $\text{Sym}^+(p)$ of positive definite matrices.
   (e). Stratified Spaces (1) $\Sigma^k_m$ ($m > 2$), (2) Open Book.

GEOMETRY. A Manifold $M$ of dimension $d$ – a metric space with each point having a neighborhood diffeomorphic to an open set in $\mathbb{R}^d$; these maps on intersecting neighborhoods are smoothly connected.

- **EXAMPLE 1.** Sphere $S^d = \{x \in \mathbb{R}^{d+1} : \|x\| = 1\}$ ($d \geq 1$). (Covered by two stereographic maps)
- Extrinsic, or chord, distance $d(p, q) = \|p - q\|$ (Euclidean distance inherited from an embedding $J : M \to \mathbb{R}^N$). On $S^d$, $J$ is the inclusion map: $S^d \to \mathbb{R}^{d+1}$. 

Rabi Bhattacharya, The University of Arizona, Tucson, AZ
Geometry

- A tangent vector \( \nu \) at \( p \) is the derivative \( dc(t)/dt \) at \( t = 0 \) of a smooth curve \( c(t) \), \( 0 \leq t \leq a \), with \( c(0) = p \), on \( M \) (computed in local coordinates in a nbd. of \( p \)). The set of tangent vectors at \( p \) is a \( d \)-dimensional vector space \( T_p(M) \).

\[ T_p(S^d) = \{ \nu \in \mathbb{R}^{d+1} : \nu \text{ orthogonal to } p \} \]
Geodesic, or intrinsic, distance $\rho_g(p, q)$: Arc length minimizing distance along smooth curves [depends on a metric tensor $g$ providing inner products smoothly on the tangent spaces of $M$]. Arc length of a curve $c(t)$, $0 \leq t \leq a$, from $p$ to $q$ is $\int_{[0,a]} |dc(t)/dt| \, dt$. On $S^d$, $\rho_g(p, q) = \text{arc length along the big circle joining } p \text{ and } q$. 

Rabi Bhattacharya, The University of Arizona, Tucson, AZ

Analysis of Non-Euclidean Data: Use of Differential Geometry in
Geodesics on $M$ – (locally) minimize geodesic distances between points. A geodesic from $p$ is entirely determined by (the initial point $p$ and) a tangent vector $v$ at $p$. On $S^d$ geodesics are the big circles.
**Geometry**

- **Geodesics** on $M$ – (locally) minimize geodesic distances between points. A geodesic from $p$ is entirely determined by (the initial point $p$ and) a tangent vector $v$ at $p$. On $S^d$ geodesics are the big circles.

- **Cut point** of $p$ is the point along a geodesic from $p$ beyond which the geodesic arc length is not distance minimizing. Cut locus of $p$ is the collection of all cut points of $p$. On $S^d$ the cut point of $p$ (along each geodesic) is $-p$ (antipodal point), so the cut locus of $p$ is $\{-p\}$. 
Example 2.

**EXAMPLE 2.** \( M = S^d / G \) – the space of orbits of \( S^d \) under a (Lie) group of isometries \( G \) of \( S^d \).

For \( p \in S^d \), \([p] = \{ hp : h \in G \} \) is the orbit of \( p \), and \( M = \{ [p] : p \in S^d \} \).

Intrinsic distance \( \rho_g([p],[q]) = \inf \{ \rho_g(hp, h'q) : h, h' \in G \} \).
Example 2(a). Axial Space $\mathbb{R}P^d$

- Axial Space $\mathbb{R}P^d = \{\text{Set of all lines passing through the origin in } \mathbb{R}^{d+1}\}$, also identified as $\{\text{the set of pairs of points } (p, -p) : p \in S^d\}$, and as $S^d/G$, where $G = \{h, \text{Id}\}$, with $hp = -p$.

- Cut locus of a point can be identified with $\mathbb{R}P^{d-1}$. 
Example 2 (b): Kendall’s planar shape space $\Sigma_2^k \ (k > 2)$.

- A $k$-ad is a set of $k$ points $\{(x_1, y_1), \ldots, (x_k, y_k)\}$ in $\mathbb{R}^2$, not all the same.
- $\Sigma_2^k$ is the set of all $k$-ads modulo translation scaling and rotation in the plane.
Example 2 (b): Kendall’s planar shape space $\Sigma^k_2 \ (k > 2)$.

- A $k$-ad is a set of $k$ points $\{(x_1, y_1), \ldots, (x_k, y_k)\}$ in $\mathbb{R}^2$, not all the same.
- $\Sigma^k_2$ is the set of all $k$-ads modulo translation scaling and rotation in the plane.
- That is, first subtract from each $(x_i, y_i)$ the mean of the $k$ points; then divide the centered vector by its Euclidean norm to get the pre-shape sphere identified as $S^{2k-3}$. Then let $\Sigma^k_2 = S^{2k-3}/G$, where $G = SO(2)$, the space of rotations in the plane (a Lie group of isometries of dimension 1).
Example 2 (b): Kendall’s shape space $\Sigma_2^k$.

- Hence $\Sigma_2^k$ has dimension $2k - 4$ and is called Kendall’s space of planar shapes.

- When the $k$-ads are represented as points on the complex plane, and are centered, then it lies on a space isomorphic to $\mathbb{C}^{k-1}$, and scaling and a rotation of a point $p$ in $\mathbb{C}^{k-1}$ can be represented as $\{\lambda p : \lambda \in \mathbb{C}\}$, i.e., a complex line passing through the origin and $p$. The space of all such points is the complex projective space $\mathbb{C}P^{k-2}$. Cut Locus of a point can be identified with $\mathbb{C}P^{k-3}$, that is, with $\Sigma_2^{k-1}$. 

Rabi Bhattacharya, The University of Arizona, Tucson, AZ
Fréchet function of a probability distribution $Q$ is

$$F(p) = \int \rho^2(p, x) Q(dx), \ p \in M.$$
Fréchet function of a probability distribution $Q$ is

$$F(p) = \int \rho^2(p, x) Q(dx), \ p \in M.$$

Fréchet mean set is the set of minimizers of $F$. A unique minimizer is called the Fréchet mean of $Q$, say $\mu$. Sample Fréchet mean $\mu_n$ is a measurable selection from the mean set of the empirical $Q_n$ based on i.i.d. $X_1, \cdots, X_n \sim Q$. 
Proposition 1. (Ziezold, 1977; BP, 2003) Let $F$ be finite. (i) then the Fréchet mean set is nonempty compact. (ii) in case of a unique minimum. $\mu_n \to \mu$ (with probability one).

Remark 1.

The *extrinsic mean* based on $\rho$ inherited from Euclidean space $E^N$ via an embedding $J$

$$J : M \to E^N$$

is given by $\mu = J^{-1}(P_J(M)\mu_J(Q))$, if the projection $P_{J(M)}$ of the Euclidean mean $\mu_J$ of $Q \circ J^{-1}$ on $J(M)$ is unique.
FRÉCHET MEANS ON METRIC SPACES

- **Proposition 1.**(Ziezold,1977; BP,2003) Let $F$ be finite. (i) then the Fréchet mean set is nonempty compact. (ii) in case of a unique minimum. $\mu_n \to \mu$ (with probability one).

- **Remark 1.**
  - The *extrinsic mean* based on $\rho$ inherited from Euclidean space $E^N$ via an embedding $J$

$$J : M \to E^N$$

is given by $\mu = J^{-1}(P_{J(M)}\mu_J(Q))$, if the projection $P_{J(M)}$ of the Euclidean mean $\mu_J$ of $Q \circ J^{-1}$ on $J(M)$ is unique.

- If $M$ is Riemannian and $\rho$ is the geodesic distance, then the Fréchet minimizer is called the *intrinsic mean* (if unique) (an open problem).
Remark 2. The embedding $J$ considered here is equivariant under the action of a large Lie group $G$:

$\exists$ a group homomorphism $\Phi : G \rightarrow GL(N, E^N)$, $g \rightarrow \Phi(g)$ such that

$$\Phi(g)(J(x)) = J(gx), \quad \forall g \in G, \ x \in M.$$
AN OMNIBUS CENTRAL LIMIT THEOREM
(Bhattacharya and Lin (2016))

We make the following assumptions.

(A1) The Fréchet mean $\mu$ of $Q$ is unique.

(A2) $\mu \in G$, $G \subset M$, $\exists$ a homeomorphism $\phi : G \rightarrow U$, open $\subset \mathbb{R}^s$ ($s \geq 1$) and

$$ x \mapsto h(x; q) := \rho^2(\phi^{-1}(x), q) $$

is $C^2$ on $U$, for every $q$ outside a $Q$-null set.

(A3) $P(\mu_n$ belongs to $G) \rightarrow 1$ as $n \rightarrow \infty$. 
(A4) Let $D_r h(x; q) = \partial h(x; q) / \partial x_r$, $r = 1, \ldots, s$. Then

$$E|D_r h(\phi(\mu); Y_1)|^2 < \infty, \ E|D_{r, r'} h(\phi(\mu); Y_1)| < \infty \ r, r' = 1, \ldots, s.$$  \hspace{1cm} (2)

(A5) Let $u_{r, r'}(\epsilon; q) = \sup \{|D_{r, r'} h(\theta; q) - D_{r, r'} h(\phi(\mu); q)| : |\theta - \phi(\mu)| < \epsilon\}$. Then

$$E|u_{r, r'}(\epsilon; Y_1)| \to 0 \ \text{as} \ \epsilon \to 0 \ \text{for all} \ 1 \leq r, r' \leq s.$$ \hspace{1cm} (3)

(A6) The matrix $\Lambda = [ED_{r, r'} h(\phi(\mu); Y_1)]_{r, r'=1,\ldots,s}$ is nonsingular.
Theorem 2.1 (Bhattacharya and Lin (2016))

Under assumptions (A1)-(A6),

\[ n^{1/2} \left[ \phi(\mu_n) - \phi(\mu) \right] \xrightarrow{L} N(0, \Lambda^{-1} C \Lambda^{-1}), \text{ as } n \to \infty, \quad (4) \]

where \( C \) is the covariance matrix of \( \{D_r h(\phi(\mu); Y_1), r = 1, \ldots, s\} \).
A CENTRAL LIMIT THEOREM FOR FRÉCHET MEAN

Remark 3. For the intrinsic mean, Theorem 2.1 holds only if $Q$ assigns probability zero to a neighborhood of the cut locus of $\mu$. 

Rabi Bhattacharya, The University of Arizona, Tucson, AZ
Theorem 2.2 (Bhattacharya and Lin (2016))

Let $C(B)$ denote the set of cut loci of points $p \in B$. Also, let $\phi$ be the log map, or $\text{Exp}^{-1}$. Suppose that $Q$ has an intrinsic mean $\mu$, and that $Q$ is absolutely continuous in a neighborhood $W$ of the cut locus of $\mu$ with a continuous density there with respect to the volume measure. Assume also that

(i) $Q(C(B(\mu; \epsilon))) = O(\epsilon^{d-c}), \epsilon \to 0$, for some $c$, $0 \leq c < d$;

(ii) on some neighborhood $V$ of $\nu = \phi(\mu) = 0$ the function $\theta \to F(\phi^{-1}(\theta))$ is twice continuously differentiable with a nonsingular Hessian $\Lambda(\theta)$, and

(iii) (A4) holds with $\phi(\mu)$ replaced by $\theta$, $\forall \theta \in V$.

Then, if $d > c + 2$, one has the CLT (4) for sample mean $\mu_n$. 
A CLT for Intrinsic Means on $S^d$

Corollary 2.3

Let $M = S^d$, $d > 2$. If $Q$ has a $C^2$ density and has a unique intrinsic mean then the CLT holds for the sample intrinsic mean.
2(a). Example 1 \((S^d)\).

- Let \(X_1, \ldots, X_n\) be i.i.d on \(S^d\). The von Mises-Fisher distribution on the sphere \(S^d\) has the following density (w.r.t. the uniform distribution on \(S^d\)).

\[
f(x; \mu, \tau) = C_d(\tau) \exp\{\tau < x, \mu >\}, x \in S^d \quad (\mu \in S^d, \tau \geq 0).
\]

(5)

Intrinsic & extrinsic means are both \(\mu\). The MLE of \(\mu\) is the sample extrinsic mean \(\mu_{n,E}\).
APPLICATIONS of $S^d$

Application 1 (Paleomagnetism). In a seminal paper, Fisher (1953) estimated mean directions of the magnetic pole for two sets of data—from a recent period and from a geologically different period in the past. Using the model (5), he constructed confidence regions for the two mean directions, and provided convincing evidence that the polarities had nearly reversed. We compare Fisher’s confidence regions for the extrinsic mean for two sets of data.
Figure: Confidence regions for the direction of earth’s magnetic poles, using Fisher’s method (red) and the nonparametric extrinsic method (blue), in Fisher’s first example.
**Introduction**

Fréchet Mean on Metric Spaces

**Examples and Applications**

Nonparametric Bayes Theory on Manifolds.

---

**Figure:** Confidence regions for the direction of earth’s magnetic poles, using Fisher’s method (red) and the nonparametric extrinsic method (blue), based on the Jurassic period data of Irving (1963).

In both cases, Fishers confidence regions are about 10% larger (in area) than those given by the nonparametric method.
2(b). KENDALL’S SHAPE SPACES $\Sigma^k_m$.

Each observation $x = (x_1, \ldots, x_k)$ of $k > m$ points in $m$-dimension (not all the same)-$k$ locations on an $m$-dimensional object. $k$-ads are equivalent mod $G$: a group $G$ of transformations.
2(b). **KENDALL’S SHAPE SPACES** $\Sigma^k_m$.

- Each observation $x = (x_1, \ldots, x_k)$ of $k > m$ points in $m$-dimension (not all the same)-$k$ locations on an $m$-dimensional object. $k$-ads are equivalent mod $G$: a group $G$ of transformations.

- $G$ is generated by translations, scaling (to unit size), rotations.
2(b). KENDALL’S SHAPE SPACES $\Sigma^k_m$.

- Each observation $x = (x_1, \ldots, x_k)$ of $k > m$ points in $m$-dimension (not all the same)-$k$ locations on an $m$-dimensional object. $k$-ads are equivalent mod $G$: a group $G$ of transformations.
- $G$ is generated by translations, scaling (to unit size), rotations.
- **Preshape**

  $u = (x_1 - \langle x \rangle, \ldots, x_k - \langle x \rangle)/\|x - \langle x \rangle\|$

  $u \in S^{m(k-1)-1}$, the *preshape sphere*.
- **Shape** of $k$-ad $\sigma(x) \in S^{m(k-1)-1}/SO(m) = \Sigma^k_m$. 

Rabi Bhattacharya, The University of Arizona, Tucson, AZ

Analysis of Non-Euclidean Data: Use of Differential Geometry in
Case $m = 2$. Planar shapes. $M = \Sigma^k_2$.

$$\sigma(x) = \sigma(u) \equiv [u] = \{ e^{i\theta} u : -\pi < \theta \leq \pi \}.$$ 

$M \simeq S^{2k-3}/SO(2) \simeq \mathbb{C}P^{k-2}$ (Complex projective space).
Case $m = 2$. Planar shapes. $M = \Sigma^k_2$.

$$\sigma(x) = \sigma(u) \equiv [u] = \{e^{i\theta}u : -\pi < \theta \leq \pi\}.$$ 

$M \simeq S^{2k-3}/SO(2) \simeq \mathbb{C}P^{k-2}$ (Complex projective space)

**Extrinsic mean $\mu_E$:** Embedding:

$$J : \sigma(x) \mapsto uu^* \in S(k, \mathbb{C})(k \times k \text{ Hermitian matrices})$$
KENDALL’S SHAPE SPACES

- **Proposition 2.** (BP(2003)) \( \mu_E \) exists iff the largest eigenvalue \( \lambda \) of \( E(uu^*) \) is simple. \[ J(\mu_E) = \mu_0 \mu_0^*, \] \( \mu_0 \) unit eigenvector for \( \lambda \).

- Case \( m > 2 \). \( \Sigma_m^k \) has singularities. Action of \( SO(m) \) is not free on \( M \).
Remark 4 (Riemannian Submersions). Other than regular submanifolds of $E^D$, such as $S^d$, which inherit the Euclidean metric tensor, most of the manifolds of interest here are of the form of $M = N/G$, where $N$ is a Riemannian manifold and $G$ is a Lie group of isometries on $N$. $M$ is the space of orbits $O_x = \{gx, g \in G\}$ ($x \in N$). The tangent space $T_p(M)$ at $p = O_x \in M$ is the horizontal subspace of $T_x(N)$, orthogonal to the direction along the orbit. $T_p(M)$ inherits the metric tensor from $T_x(N)$ on this subspace.

Example: $N = S^{2(k-1)-1} = S^{2k-3}$. $G = SO(2)$, $M = \Sigma^k_2$. 
EXAMPLES & APPLICATIONS FOR $\Sigma^k_2$

Two-sample problem: discrimination between two shape distributions.


$k = 13$ landmarks were recorded on the midsagittal slice of the brain scan of each of $n_1 = 14$ schizophrenic patients and $n_2 = 14$ normal patients (Bookstein (1991)). Shape space $\Sigma^{13}_2$. 
Figure: (a) and (b) show 13 landmarks for 14 normal and 14 schizophrenic children respectively along with the respective mean shapes. * correspond to the mean shapes’ landmarks.
Figure: The sample extrinsic means for the 2 groups along with the pooled sample mean, corresponding to Figure 3.

\[ p\text{-value: nonparametric tests (intrinsic & extrinsic) } 4 \times 10^{-11} \]

Goodalls parametric test 0.01 Hotellings $T^2$ test 0.66
k = 8 landmarks, \(n_1 = 29\) male skulls, \(n_2 = 30\) female skulls (BP (2005), BB(2008), (2012), Dryden & Mardia (1998)). Shape space \(\Sigma^8_2\).
Figure: (a) and (b) show 8 landmarks from skulls of 30 female and 29 male gorillas respectively along with the respective sample mean shapes. * correspond to the mean shapes’ landmarks.

p-value: Nonparametric tests (intrinsic & extrinsic) $< 10^{-16}$
Parametric test (Hotellings t2 test, boxs m-test) 0.0001
2 (c). 3D Shape Space $RΣ^k_3$-Match Pair Test for Glaucoma

- Assume affine span of each k-ad is $\mathbb{R}^m$, with preshape $u = (u_1, \ldots, u_k) \in S^{m(k-1)-1}$. 

$\lambda_1 \geq \ldots \geq \lambda_k$ be the eigenvalues of $E(u_i \cdot u_j)$, with corresponding orthonormal eigenvectors $U_1, \ldots, U_k$. Then (i) $\mu_E$ exists iff $\lambda_m > \lambda_{m+1}$ and then (ii) $J(\mu_E) = (v_1, \ldots, v_m)(v_1, \ldots, v_m)^t$, where $v_j = (\lambda_j - \bar{\lambda} + 1/m)U_j$, with $\bar{\lambda} = (\lambda_1 + \ldots + \lambda_m)/m$. 

Rabi Bhattacharya, The University of Arizona, Tucson, AZ
2 (c). 3D Shape Space $RΣ^k_3$-Match Pair Test for Glaucoma

- Assume affine span of each k-ad is $\mathbb{R}^m$, with preshape $u = (u_1, \ldots, u_k) \in S^{m(k-1)-1}$.

- Shape $\sigma(x) \in S^{m(k-1)-1}/O(m) = M$. Embedding

  \[ J : \sigma(x) \mapsto ((u_i \cdot u_j)) \quad (M \rightarrow S_0^+(k, R)) \]

(Bandulasiri and Patrangenaru (2005), Bandulasiri and BP (2009), Dryden, Kume, Le, Wood (2008))
2 (c). 3D Shape Space $R\Sigma_3^k$-Match Pair Test for Glaucoma

- Assume affine span of each k-ad is $\mathbb{R}^m$, with preshape $u = (u_1, \ldots, u_k) \in S^{m(k-1)-1}$.
- Shape $\sigma(x) \in S^{m(k-1)-1}/O(m) = M$. Embedding

$$J : \sigma(x) \mapsto ((u_i \cdot u_j)) \ (M \to S_{0+}(k, R))$$

(Bandulasiri and Patrangenaru (2005), Bandulasiri and BP (2009), Dryden, Kume, Le, Wood (2008))

- Proposition 3. (A. Bhattacharya (2008)). Let $\lambda_1 \geq \ldots \geq \lambda_k$ be the eigenvalues of $E((u_i \cdot u_j))$, with corresponding orthonormal eigenvectors $U_1, \ldots, U_k$. Then (i) $\mu_E$ exists iff $\lambda_m > \lambda_{m+1}$ and then (ii) $J(\mu_E) = (\nu_1, \ldots, \nu_m)(\nu_1, \ldots, \nu_m)^t$, where $\nu_j = (\lambda_j - \bar{\lambda} + 1/m)^{1/2}U_j$, with $\bar{\lambda} = (\lambda_1 + \ldots + \lambda_m)/m$. 
APPLICATION of \((RΣ_{3}^{k})\)

To detect any shape change of the inner eye due to glaucoma, 3D images of the optical nerve head (ONH) of both eyes of 12 mature rhesus monkeys were recorded. One of the eyes was subjected to increased intraocular pressure (IOP). \(k = 5\) landmarks of the inner eye were measured on each eye. For this match pair experiment, the manifold is \(RΣ_{3}^{k} \times RΣ_{3}^{k}\). The null hypothesis is that the (extrinsic) mean lies on the diagonal of this product manifold (BP(2005), BB(2009)). \(p\)-value of the nonparametric chisquare test is (BB(2009)) \(1.55 \times 10^{-5}\).
2 (d). Sym\(^+(p)\)–p × p Positive Definite Matrices

1. **Euclidean metric:** \( \|A\|^2 = \text{Trace}(A)^2 \). Sym\(^+(p)\) open convex subset of Sym\((p)\), and \( Q \) on Sym\(^+(p)\) has Euclidean mean \( \mu_E = \int AQ(dA) \).
2 (d). $\text{Sym}^+(p)$—$p \times p$ Positive Definite Matrices

1. **Euclidean metric**: $\|A\|^2 = \text{Trace}(A)^2$. $\text{Sym}^+(p)$ open convex subset of $\text{Sym}(p)$, and $Q$ on $\text{Sym}^+(p)$ has Euclidean mean $\mu_E = \int AQ(dA)$.

2. **log-Euclidean metric** (Arsigney et al. (2006)).
   
   $J \equiv \log : \text{Sym}^+(p) \rightarrow \text{Sym}(p)$ is the inverse of the exponential map $B \rightarrow e^B$, $\text{Sym}(p) \rightarrow \text{Sym}^+(p)$. (Diffeomorphism).
   
   $d_{LE}(A_1, A_2) = \|\log(A_1) - \log(A_2)\|$. 

   $\mu_{LE} = \exp(\int (\log(A))Q(dA))$. (**Extrinsic mean** under $J$).

   Also, **intrinsic mean** under bi-invariant metric of $\text{Sym}^+(p)$ as a Lie group: $A_1 \circ A_2 = \exp(\log(A_1) + \log(A_2))$ (zero-curvature).
2 (d). \( \text{Sym}^+(p) \) – \( p \times p \) Positive Definite Matrices

1. **Euclidean metric**: \( \| A \|^2 = \text{Trace}(A)^2 \). \( \text{Sym}^+(p) \) open convex subset of \( \text{Sym}(p) \), and \( Q \) on \( \text{Sym}^+(p) \) has Euclidean mean \( \mu_E = \int A Q(dA) \).

2. **log-Euclidean metric** (Arsigney et al. (2006)).
   
   \( J \equiv \log : \text{Sym}^+(p) \to \text{Sym}(p) \) is the inverse of the exponential map \( B \to e^B \), \( \text{Sym}(p) \to \text{Sym}^+(p) \). (Diffeomorphism).
   
   \( d_{LE}(A_1, A_2) = \| \log(A_1) - \log(A_2) \| \).
   
   \( \mu_{LE} = \exp(\int (\log(A)) Q(dA)) \). (Extrinsic mean under \( J \)).
   
   Also, **intrinsic mean** under bi-invariant metric of \( \text{Sym}^+(p) \) as a Lie group: \( A_1 \circ A_2 = \exp(\log(A_1) + \log(A_2)) \) (zero-curvature).

3. **Affine invariant metric**.
   
   \( d_{AI}^2(A_1, A_2) = \| \log(A_1^{-1/2} A_2 A_1^{-1/2}) \|^2 \).
   
   \( \langle B_1, B_2 \rangle = \text{Trace}(A^{-1} B_1 A^{-1} B_2) \) (Non-positive curvature).
2 \textit{(d). Sym}^+(p) – p \times p \textbf{ Positive Definite Matrices}

\textbf{APPLICATIONS (p=3).} DTI (Diffusion Tensor Imaging) provides measurements of the diffusion matrix of water molecules in tiny voxels in the white matter of the brain. \textit{Anisotropy} in the presence of the structural barriers of nerve fibers is reduced when a trauma occurs (Parkinsons, Alzheimers,...). Challenges to statistical inference. Also see Schwartzman (2014).
APPLICATIONS ($p = 3$)- HIV IMAGING DATA

- Diffusion-weighted images were acquired for each of 46 subjects with 28 HIV+ subjects and 18 healthy controls.
- In the previous DTI findings, the diffusion tensors in the splenium of the corpus callosum were found significantly different between the HIV+ and control group. We examine the finite sample performance of our method by using this fiber tract.
APPLICATIONS ($p = 3$)- HIV IMAGING DATA

- Diffusion-weighted images were acquired for each of 46 subjects with 28 HIV+ subjects and 18 healthy controls.

- In the previous DTI findings, the diffusion tensors in the **spleenium of the corpus callosum** were found significantly different between the HIV+ and control group. We examine the finite sample performance of our method by using this fiber tract.

- Diffusion tensor were constructed for 75 voxels along the fiber.

- In order to detect meaningful group differences, registration is crucial. The 46 HIV DTI data used in our studies, including the splenium tracts and diffusion tensors on them, were registered in the same atlas space.
APPLICATIONS \((p = 3)\)- HIV IMAGING DATA

- We first carry out the two-sample testing (voxel-wise) using a testing statistics based on the usual Euclidean distance.

\[
(\bar{X} - \bar{Y})\Sigma^{-1}(\bar{X} - \bar{Y})^T
\]

where \(\bar{X}\) and \(\bar{Y}\) are the sample mean vector of dimension 6 of \(X\) and \(Y\) respectively, \(\Sigma = (1/n_1\Sigma_X + 1/n_2\Sigma_Y)\).

- The testing statistics has a asymptotic chisquare distribution \(\chi^2(6)\).

- Next is a plot of the p-values along the fiber tracks.

- We can apply the *Benjamin-Yekutieli procedure* to control false discovery rate.
APPLICATIONS ($p = 3$)- HIV IMAGING DATA

p–values along the fibers using Euclidean distance

p–values

Arc length
FALSE DISCOVERY RATE. BENJAMIN-HOCHBERG PROCEDURE

- Set $\alpha = 0.05$. Apply *Benjamini-Hochberg procedure* to the tests.
- Reject only the $k$ null hypothesis with the smallest $p$-values, where $k = \max\{i : p(i) \leq \frac{1}{m} \alpha\}$. 

In our example we first order the 75 $p$-values corresponding to the tests carried out at all the locations. The ordered $p$-values are compared with the vector $\{0.05/75, 0.1/75, \ldots, 0.05\}$, which gives the result $k = 58$. Therefore we reject the 58 null hypotheses corresponding to the first 58 ordered $p$-values. The false discovery rate is smaller than $\frac{m_0}{m} \alpha \leq \alpha$. 

Rabi Bhattacharya, The University of Arizona, Tucson, AZ
FALSE DISCOVERY RATE. BENJAMIN-HOCHBERG PROCEDURE

- Set $\alpha = 0.05$. Apply *Benjamini-Hochberg procedure* to the tests.
- Reject only the $k$ null hypothesis with the smallest $p$-values, where $k = \max\{i : p(i) \leq \frac{1}{m}\alpha\}$.
- In our example we first order the 75 $p$-values corresponding to the tests carried out at all the locations.
- The ordered $p$-values are compared with the vector $\{0.05/75, 0.1/75, \ldots, 0.05\}$, which gives the result $k = 58$. 
FALSE DISCOVERY RATE. BENJAMIN-HOCHBERG PROCEDURE

- Set $\alpha = 0.05$. Apply *Benjamini-Hochberg procedure* to the tests.
- Reject only the $k$ null hypothesis with the smallest $p$-values, where $k = \max\{i : p(i) \leq \frac{1}{m} \alpha\}$.
- In our example we first order the 75 $p$-values corresponding to the tests carried out at all the locations.
- The ordered $p$-values are compared with the vector \{0.05/75, 0.1/75, \ldots, 0.05\}, which gives the result $k = 58$.
- Therefore we reject the 58 null hypotheses corresponding to the first 58 ordered $p$-values.
- The false discovery rate is smaller than $m_0/m \alpha \leq \alpha$. 

Rabi Bhattacharya, The University of Arizona, Tucson, AZ
Second, we carry out two-sample testings based on the log-Euclidean distance of the DTI matrices. The matrix log of each raw diffusion matrix is first calculated. The testing statistics is based on the Euclidean distance of the 6 distinct values of the log matrices.

Next is a plot of the $p$-values along the fiber tracks.

To control false discovery rate, we also carry out the Benjamini-Hochberg procedure. We reject the first 48 tests based on the order $p$-values.
APPLICATIONS (p=3)- HIV IMAGING DATA

p-values along the fibers using log-Euclidean distance

Arc length

p-values

Rabi Bhattacharya, The University of Arizona, Tucson, AZ
Plot of the p-values

p-values along the fibers using Euclidean and log–Euclidean distance

- log–Euclidean
- 0.05
- Euclidean

Arc length

Rabi Bhattacharya, The University of Arizona, Tucson, AZ

Analysis of Non-Euclidean Data: Use of Differential Geometry in
2(e). **Stratified Spaces (1)** $\Sigma_m^k (m > 2)$

- *Stratified spaces* $S$ are made up of several subspaces of different dimensions.
- A familiar example is $\Sigma_m^k$ with $m > 2$. After translation and scaling the $k$-ads lie in (and fill out) a preshape sphere $S^{mk-m-1}$. The shape space is then viewed as $\Sigma_m^k = S^{mk-m-1}/\text{SO}(m)$.
2(e). **Stratified Spaces (1) $\Sigma^k_m (m > 2)$**

- *Stratified spaces* $S$ are made up of several subspaces of different dimensions.
- A familiar example is $\Sigma^k_m$ with $m > 2$. After translation and scaling the $k$-ads lie in (and fill out) a preshape sphere $S^{mk-m-1}$. The shape space is then viewed as $\Sigma^k_m = S^{mk-m-1}/SO(m)$.
- For simplicity, consider $m = 3$. One may split $\Sigma^k_3$ into two strata. The larger stratum $S_1$ corresponds to shapes of non-collinear $k$-ads.
2(e). Stratified Spaces (1) $\Sigma^k_m \ (m > 2)$

- *Stratified spaces* $S$ are made up of several subspaces of different dimensions.
- A familiar example is $\Sigma^k_m$ with $m > 2$. After translation and scaling the $k$-ads lie in (and fill out) a preshape sphere $S^{mk-m-1}$. The shape space is then viewed as $\Sigma^k_m = S^{mk-m-1}/SO(m)$.
- For simplicity, consider $m = 3$. One may split $\Sigma^k_3$ into two strata. The larger stratum $S_1$ corresponds to shapes of non-collinear $k$-ads.
- $S_1$ is a manifold of dimension $3k - 4 - 3 = 3k - 7$. The manifold is not complete in the geodesic distance.
2(e). Stratified Spaces (1) $\Sigma^k_m (m > 2)$

- *Stratified spaces* $S$ are made up of several subspaces of different dimensions.
- A familiar example is $\Sigma^k_m$ with $m > 2$. After translation and scaling the $k$-ads lie in (and fill out) a preshape sphere $S^{mk-m-1}$. The shape space is then viewed as $\Sigma^k_m = S^{mk-m-1}/\text{SO}(m)$.
- For simplicity, consider $m = 3$. One may split $\Sigma^k_3$ into two strata. The larger stratum $S_1$ corresponds to shapes of non-collinear $k$-ads.
- $S_1$ is a manifold of dimension $3k - 4 - 3 = 3k - 7$. The manifold is not complete in the geodesic distance.
- The other stratum $S_0$ comprises shapes of $k$-ads each $k$-ad being a set of $k$ collinear points in $\mathbb{R}^3$. Each orbit has dimension 3. The stratum $S_0$ may be given the structure of a differentiable manifold of dimension $k - 2$. 

Rabi Bhattacharya, The University of Arizona, Tucson, AZ

Analysis of Non-Euclidean Data: Use of Differential Geometry in
2(e). Stratified Spaces (2) Open Book (Hotz et al. (2013))

An *open book* \( O \) is the disjoint union of \( K \) *open leaves* \( L_j^+(H, j), 1 \leq j \leq K \), joined at the *spine* \( L_0 \) as the common boundary. Here \( H = (0, \infty) \times \mathbb{R}^d \) and \( L_0 = 0 \times \mathbb{R}^d \).
2(e). Stratified Spaces (2) Open Book (Hotz et al. (2013))

- An open book $\mathcal{O}$ is the disjoint union of $K$ open leaves $L_j^+(H,j)$, $1 \leq j \leq K$, joined at the spine $L_0$ as the common boundary. Here $H = (0, \infty) \times \mathbb{R}^d$ and $L_0 = 0 \times \mathbb{R}^d$.

- The distance $\rho$ on $L_j^+$, or $L_0$, is the usual Euclidean distance on $\mathbb{R}^{d+1}$, but for $j \neq k$, $\rho((x,j),(y,k)) = |x - Ry|$, where $R_y$ is the reflection, $R_y = (-y^{(0)}, y^{(1)}, \ldots, y^{(d)}) \forall y = (y^{(0)}, y^{(1)}, \ldots, y^{(d)}) \in [0, \infty) \times \mathbb{R}^d$.

- The open book is a geodesic space with non-positive curvature in the sense of A.D. Alexandrov and therefore $Q$ has a unique Fréchet mean.
2(e). Stratified Spaces (2) Open Book (Hotz et al. (2013))

Consider the map $F_j : \mathcal{O} \rightarrow \mathbb{R}^{d+1}$, $F_j((x, j)) = x$, $F_j((x, k)) = Rx$ if $k \neq j$. Write $m_j = \int x(0) (Q \circ F_j^{-1})(dx)$. Under the assumption $Q(L_j + j) > 0$ for $1 \leq j \leq K$, either (1) $m_j \geq 0$ for some $j$, and $m_k < 0 \forall k \neq j$, or (2) $m_j < 0 \forall j$. In case (2), the Fréchet mean is sticky, that is, with probability one, $\mu_N \in L_0$ for all sufficiently large $N$. Also, the classical CLT holds on the $d$-dimensional space $L_0$. 
2(e). Stratified Spaces (2) Open Book (Hotz et al. (2013))

- Consider the map $F_j : \mathcal{O} \to \mathbb{R}^{d+1}$, $F_j((x,j)) = x$, $F_j((x,k)) = Rx$ if $k \neq j$. Write $m_j = \int x^0(Q \circ F_j^{-1})(dx)$.
- Under the assumption $Q(L_j^+) > 0$ for $1 \leq j \leq K$, either (1) $m_j \geq 0$ for some $j$, and $m_k < 0 \ \forall \ k \neq j$, or (2) $m_j < 0 \ \forall \ j$. 
2(e). Stratified Spaces (2) Open Book (Hotz et al. (2013))

- Consider the map $F_j : \mathcal{O} \rightarrow \mathbb{R}^{d+1}$, $F_j((x, j)) = x$, $F_j((x, k)) = Rx$ if $k \neq j$. Write $m_j = \int x^{(0)}(Q \circ F_j^{-1})(dx)$.

- Under the assumption $Q(L_j^+) > 0$ for $1 \leq j \leq K$, either (1) $m_j \geq 0$ for some $j$, and $m_k < 0 \ \forall \ k \neq j$, or (2) $m_j < 0 \ \forall \ j$.

- In case (2), the Fréchet mean is sticky, that is, with probability one, $\mu_N \in L_0$ for all sufficiently large $N$. Also, the classical CLT holds on the $d$-dimensional space $L_0$. 
2(e). Stratified Spaces (2) Open Book

- Recall case (1) $m_j \geq 0$ for some $j$, and $m_k < 0 \forall k \neq j$.
- If in case (1), $m_j > 0$, then $\mu$ lies in the open leaf $L_j^+$, as do $\mu_N$ for all sufficient large $N$; hence the classical $(d + 1)$-dimensional CLT holds.
2(e). Stratified Spaces (2) Open Book

- Recall case (1) $m_j \geq 0$ for some $j$, and $m_k < 0 \quad \forall \; k \neq j$.
- If in case (1), $m_j > 0$, then $\mu$ lies in the open leaf $L^+_j$, as do $\mu_N$ for all sufficient large $N$; hence the classical $(d + 1)$-dimensional CLT holds.
- If, however, $m_j = 0$, $\mu \in L_0$; but $\mu_N \in L_j$ if (the empirical) $m_{j,n} > 0$ and $\mu_N \in L_0$ if $m_{j,n} \leq 0$; hence the asymptotic distribution centered at $\mu$ is the distribution of $((X^{(0)}_+, X^{(1)}, \ldots, X^{(d)}), j)$ on $L^{(+)}_j \cup L_0$, where $(X^{(0)}, X^{(1)}, \ldots, X^{(d)})$ has the Gaussian distribution stated under the preceding case $m_j > 0$, and $X^{(0)}_+ = \max\{X^{(0)}, 0\}$.
2(e). Stratified Spaces (2) Open Book

Remark 5. The study of this and some toy models of phylogenetic trees has been motivated in part by the pioneering work of Susan Holmes and her collaborators.
3 (a). NONPARAMETRIC BAYES ON MANIFOLDS-DENSITY ESTIMATION


- Density of $Q$ with a standard measure on $M$, represented as a mixture $P(d\theta)$ of a parametric family of densities $K(x; \theta)$ ($\theta \in \Theta$)

$$f(x; P) = \int_{\Theta} K(x; \theta) P(d\theta).$$

- $P$ is a probability measure on $\Theta$, is often imposed with a Dirichlet process prior (Ferguson (1974)).

$Q$ with a standard measure on $M$, represented as a mixture $P(d\theta)$ of a parametric family of densities $K(x; \theta)$ ($\theta \in \Theta$)
3 (a). NONPARAMETRIC BAYES ON MANIFOLDS-DENSITY ESTIMATION

- Density of $Q$ with a standard measure on $M$, represented as a mixture $P(d\theta)$ of a parametric family of densities $K(x; \theta)$ ($\theta \in \Theta$)

$$f(x; P) = \int_{\Theta} K(x; \theta) P(d\theta).$$

- $P$ is a probability measure on $\Theta$, is often imposed with a Dirichlet process prior (Ferguson (1974)).

- Sethuraman’s *stick-breaking* representation $\sum w_j \delta_{Y_j}$ of prior with $w_1 = u_1$, $w_j = u_j(1 - u_1) \cdots (1 - u_{j-1})$ ($j > 1$). Here $u_j$ are i.i.d Beta$(1, \alpha(\Theta))$, where $\alpha$ is the base measure on $\Theta$, $Y_j$ are i.i.d $\sim G = \alpha/\alpha(\Theta)$. Draws from posterior by MCMC.
3 (a). NONPARAMETRIC BAYES ON MANIFOLDS-DENSITY ESTIMATION

**Example.** A density on $\Sigma^k_2$ is estimated by the kernel method (KD) (Pelletier (2005)), NP Bayes and MLE. Simulation study yielded the following estimate of the mean $L^1$ distance of these methods: NP(0.44), KD (0.75), MLE (1.03).
3 (b). NONPARAMETRIC BAYES ON MANIFOLDS-CLASSIFICATIONS

Classifications. $\Sigma^8_2$ (Gorilla Skulls). $n_1 = 30$, $n_2 = 29$. 25 randomly chosen from each group as the training samples. The remaining 9 were classified.
Figure: Estimated shape densities of gorillas: Female(solid), Male(dot). Estimate(r), 95% C.R.(b,g).

Densities evaluated at a dense grid of points drawn from the unit speed geodesic starting at female extrinsic mean in direction of male extrinsic mean.
REFERENCES

REFERENCES

REFERENCES

- Irving, E. (1963), Wiley.
REFERENCES