PROCEEDINGS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 145, Number 1, January 2017, Pages 413–428
http://dx.doi.org/10.1090/proc/13216
Article electronically published on July 26, 2016

OMNIBUS CLTS FOR FRÉCHET MEANS AND NONPARAMETRIC INFERENCE ON NON-EUCLIDEAN SPACES

RABI BHATTACHARYA AND LIZHEN LIN

(Communicated by Mark M. Meerschaert)

ABSTRACT. Two central limit theorems for sample Fréchet means are derived, both significant for nonparametric inference on non-Euclidean spaces. The first theorem encompasses and improves upon most earlier CLTs on Fréchet means and broadens the scope of the methodology beyond manifolds to diverse new non-Euclidean data, including those on certain stratified spaces which are important in the study of phylogenetic trees. It does not require that the underlying distribution Q have a density and applies to both intrinsic and extrinsic analysis. The second theorem focuses on intrinsic means on Riemannian manifolds of dimensions $d > 2$ and breaks new ground by providing a broad CLT without any of the earlier restrictive support assumptions. It makes the statistically reasonable assumption of a somewhat smooth density of Q. The excluded case of dimension $d = 2$ proves to be an enigma, although the first theorem does provide a CLT in this case as well under a support restriction. The second theorem immediately applies to spheres S^d, $d > 2$, which are also of considerable importance in applications to axial spaces and to landmarks-based image analysis, as these spaces are quotients of spheres under a Lie group $\mathcal{G}$ of isometries of S^d.

1. Introduction

The present article focuses on the nonparametric, or model independent, statistical analysis of manifold-valued and other non-Euclidean data that arise in many areas of science and technology. The basic idea is to use means for comparisons among distributions, as one does with Euclidean data. On a metric space (S, ρ) there is a notion of the mean μ of a distribution Q, perhaps first formulated in detail in [22], as the minimizer of the expected squared distance from a point,

$$\mu = \arg\min_p \int \rho^2(p, q) Q(dq), \tag{1.1}$$

assuming the integral is finite (for some p) and the minimizer is unique, in which case one says that the *Fréchet mean of* Q *exists*. This μ is called the *Fréchet mean of* Q. In general, the set of minimizers is called the *Fréchet mean set* of Q, denoted C_Q. It turns out that uniqueness is crucial for making comparisons among distributions. Usually the minimizer is unique under relatively minor restrictions if the distance ρ

Received by the editors February 28, 2015 and, in revised form, November 12, 2015 and March 27, 2016.
2010 *Mathematics Subject Classification.* Primary 60F05, 62E20, 60E05, 62G20.
Key words and phrases. Inference on manifolds, Fréchet means, Omnibus central limit theorem, stratified spaces.

is the Euclidean distance inherited by the embedding J of a d-dimensional manifold M in a Euclidean space E^N, such that $J(M)$ is closed. Indeed, under the relabeling of M by $J(M)$, the Fréchet mean set in this case is given by

$$\arg\min_{p\in J(M)} \|p - m(Q\circ J^{-1})\|^2, \tag{1.2}$$

where $\|x\|$ is the Euclidean norm on E^N and $m(Q\circ J^{-1})$ is the usual Euclidean mean of the induced distribution $Q\circ J^{-1}$ on E^N. Thus the minimizer is unique if and only if the projection of the Euclidean mean on the image $J(M)$ of M is unique, in which case it is called an *extrinsic mean.* On the other hand, if ρ_g is the geodesic distance on a Riemannian manifold M with metric tensor g having positive sectional curvature (in some region of M), then conditions for uniqueness are known only for Q with support in a relatively small geodesic ball [1,30,31], which is too restrictive an assumption from the point of view of statistical applications. If the Fréchet mean exists under ρ_g it is called the *intrinsic mean.* A complete characterization of uniqueness of (1.1) for $\rho=\rho_g$ on the circle S^1 for probabilities Q with a continuous density ([12], [10]) indicates that the intrinsic mean exists broadly, without any support restrictions, if Q has a smooth density.

An important question that arises in the use of Fréchet means in nonparametric statistics is the choice of the distance ρ on M. There are in general uncountably many embeddings J and metric tensors g on a manifold M. For intrinsic analysis there are often natural choices for the metric tensor g. A good choice for extrinsic analysis is to find an embedding J: $M\to E^N$ with $J(M)$ closed, which is *equivariant* under a large Lie group $\mathcal{G}$ of actions on M. This means that there is a homomorphism $g\to\Phi_g$ on $\mathcal{G}$ into the general linear group $GL(N,\mathbb{R})$ such that $J\circ g=\Phi_g\circ J$ $\forall g\in\mathcal{G}$. Such embeddings and extrinsic means under them have been derived for Kendall type shape spaces in [14], [15], [4], [3], [19], and [8]. In most data examples that have been analyzed, using a natural metric tensor g and an equivariant J under a large group $\mathcal{G}$, the sample intrinsic and extrinsic means are virtually indistinguishable and the inferences based on the two different methodologies yield almost identical results [10]. This provides an affirmation of good choices of distances. It also strongly suggests that the intrinsic mean is unique in many, perhaps most, statistical applications.

Our focus in this article is to provide the asymptotic distribution theory which is the basis of nonparametric inference based on Fréchet means. The omnibus CLT Theorem 2.2 implies earlier results on CLT's and, in particular, extends them to certain stratified spaces. Unfortunately, for the intrinsic CLT a support condition is still needed for the theorem to apply. In Section 3 we remove these support conditions for CLT's on S^d, $d>2$, assuming statistically reasonable smooth densities. The implications of these results for axial spaces and Kendall's shape spaces, etc., are indicated.

Finally, it is important to distinguish the intrinsic mean on a Riemannian manifold (M,g) from the *Karcher mean* of Q which minimizes the Fréchet function restricted to an open set S containing the support of Q.

2. An omnibus CLT for the Fréchet mean

Let (S, ρ) be a metric space and Q a probability measure on its Borel σ-field. Define the *Fréchet function* of Q as

$$F(p) = \int \rho^2(p, q) Q(dq) \ (p \in S). \tag{2.1}$$

Assume that F is finite on S and has a unique minimizer $\mu = \arg\min_p F(p)$. Then μ is called the *Fréchet mean* of Q (with respect to the distance ρ). Under broad conditions, the Fréchet sample mean μ_n of the empirical distribution $Q_n = \dfrac{1}{n}\sum_{j=1}^n \delta_{Y_j}$ based on independent S-valued random variables Y_j $(j = 1, \dots, n)$ with common distribution Q is a consistent estimator of μ. That is, $\mu_n \to \mu$ almost surely, as $n \to \infty$. Here μ_n may be taken to be any measurable selection from the (random) set of minimizers of the Fréchet function of Q_n, namely, $F_n(p) = \dfrac{1}{n}\sum_{j=1}^n \rho^2(p, Y_j)$ (see [44], [14], [15] and [10]).

We make the following assumptions.

- (A1) (Uniqueness of μ) The Fréchet mean μ of Q is unique.
- (A2) $\mu \in G$, where G is a measurable subset of S, and there is a homeomorphism $\phi : G \to U$, where U is an open subset of $\mathbb{R}^s$ for some $s \geq 1$ and G is given its relative topology on S. Also,

$$x \mapsto h(x; q) := \rho^2(\phi^{-1}(x), q) \tag{2.2}$$

  is twice continuously differentiable on U, for every q outside a Q-null set.
- (A3) $P(\mu_n \in G) \to 1$ as $n \to \infty$.
- (A4) Let $D_r h(x; q) = \partial h(x; q)/\partial x_r$, $D_{r,r'} = D_r D_{r'}$, $1 \leq r, r' \leq s$. Then

$$E|D_r h(\phi(\mu); Y_1)|^2 < \infty, \ E|D_{r,r'} h(\phi(\mu); Y_1)| < \infty \text{ for } r, r' = 1, \dots, s. \tag{2.3}$$

- (A5) (Locally uniform L^1-smoothness of the Hessian) 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

$$E|u_{r,r'}(\epsilon; Y_1)| \to 0 \text{ as } \epsilon \to 0 \text{ for all } 1 \leq r, r' \leq s. \tag{2.4}$$

- (A6) (Nonsingularity of the Hessian) The matrix $\Lambda = [E D_{r,r'} h(\phi(\mu); Y_1)]_{r,r'=1,\dots,s}$ is nonsingular.

Remark 2.1. Observe that $Eh(x, Y_1) = F(\phi^{-1}(x)) = ED_r h(x, Y_1) = D_r F(\phi^{-1}(x))$, $1 \leq r \leq s$, $x \in U$. Also, $ED_r h(\phi(\mu), Y_1) = D_r F(\phi^{-1}(x)) \mid_{x=\phi(\mu)} = 0$, $1 \leq r \leq s$, since $F(\phi^{-1}(x))$ attains a minimum at $x = \phi(\mu)$.

Theorem 2.2. *Under assumptions* (A1)-(A6),

$$n^{1/2}[\phi(\mu_n) - \phi(\mu)] \xrightarrow{\mathcal{L}} N(0, \Lambda^{-1} C \Lambda^{-1}), \ \textit{as } n \to \infty, \tag{2.5}$$

where C is the covariance matrix of $\{D_r h(\phi(\mu); Y_1), r = 1, \dots, s\}$.

Proof. The function $x \to F_n(\phi^{-1}x) = \dfrac{1}{n}\sum_{j=1}^n h(x, Y_j)$ on U attains a minimum at $\phi(\mu_n) \in U$ for all sufficiently large n (almost surely). For all such n one therefore has the first order condition

$$\nabla F_n(\phi^{-1}\nu_n) = \frac{1}{n}\sum_{j=1}^n \nabla h(\nu_n, Y_j) = 0, \tag{2.6}$$

where $\nu = \phi(\mu)$, $\nu_n = \phi(\mu_n)$ (column vectors in U). Here ∇ is the gradient $(D_1, \ldots, D_r)$. A Taylor expansion yields

$$0 = \frac{1}{n}\sum_{j=1}^{n} \nabla\, h(\nu_n, Y_j) = \frac{1}{n}\sum_{j=1}^{n} \nabla\, h(\nu, Y_j) + \Lambda_n(\nu_n - \nu) \tag{2.7}$$

where Λ_n is the $s \times s$ matrix given by

$$\Lambda_n = \frac{1}{n}\sum_{j=1}^{n}[D_{r,r'} h(\theta_{n,r,r'}, Y_j)]_{r,r'=1,\ldots,s}, \tag{2.8}$$

and $\theta_{n,r,r'}$ lies on the line segment joining ν_n and ν. We will show that

$$\Lambda_n \to \Lambda \text{ in probability}, \quad \text{as } n \to \infty. \tag{2.9}$$

Fix $r, r' \in \{1, \ldots, s\}$. For $\delta > 0$, write $Eu_{r,r'}(\delta, Y_1) = \gamma(\delta)$. There exists $n = n(\delta)$ such that $P(|\nu_n - \nu| > \delta) < \delta$ for $n > n(\delta)$. Now

$$\begin{aligned} E\big|[\frac{1}{n}\sum_{j=1}^{n} D_{r,r'} h(\nu_n, Y_j) - \frac{1}{n}\sum_{j=1}^{n} D_{r,r'} h(\nu, Y_j)] \cdot 1_{[|\nu_n - \nu| \le \delta]}\big| &\le E\frac{1}{n}\sum_{j=1}^{n} u_{r,r'}(\delta, Y_j) \\ &= Eu_{r,r'}(\delta, Y_1) = \gamma(\delta) \to 0 \end{aligned}$$

as $\delta \to 0$. Hence, by Chebyshev's inequality for first moments, for $n > n(\delta)$ one has for every $\epsilon > 0$,

$$P(\big|\frac{1}{n}\sum_{j=1}^{n} D_{r,r'} h(\nu_n, Y_j) - \frac{1}{n}\sum_{j=1}^{n} D_{r,r'} h(\nu, Y_j)\big| > \epsilon) \le \delta + \gamma(\delta)/\epsilon \to 0 \ \text{ as } \delta \to 0. \tag{2.10}$$

This shows that

$$\big[\frac{1}{n}\sum_{j=1}^{n} D_{r,r'} h(\nu_n, Y_j) - \frac{1}{n}\sum_{j=1}^{n} D_{r,r'} h(\nu, Y_j)\big] \to 0; \text{ in probability as } n \to \infty. \tag{2.11}$$

Next, by the strong law of large numbers,

$$\frac{1}{n}\sum_{j=1}^{n} D_{r,r'} h(\nu, Y_j) \to ED_{r,r'} h(\nu, Y_1) \ \text{ almost surely, } \text{ as } n \to \infty. \tag{2.12}$$

Since (2.10) – (2.12) hold for all r,r', (2.9) follows. The set of symmetric $s \times s$ positive definite matrices is open in the set of all $s \times s$ symmetric matrices, so that (2.9) implies that Λ_n is nonsingular with probability going to 1 and $\Lambda_n^{-1} \to \Lambda^{-1}$ in probability, as $n \to \infty$. Note that $E\nabla h(\nu, Y_1) = 0$ (see Remark 2.1). Therefore, using (A4), by the classical CLT and Slutsky's Lemma, (2.7) leads to

$$\sqrt{n}(\nu_n - \nu) = \Lambda_n^{-1}[-(1/\sqrt{n})\frac{1}{n}\sum_{j=1}^{n} \nabla\, h(\nu, Y_j)] \xrightarrow{\mathcal{L}} N(0, \Lambda^{-1} C \Lambda^{-1}), \tag{2.13}$$

as $n \to \infty$. □

A preliminary version of Theorem 2.2 was presented in [11].

Corollary 2.3 (CLT for intrinsic means-I). *Let (M, g) be a d-dimensional complete Riemannian manifold with metric tensor g and geodesic distance ρ_g. Suppose Q is a probability measure on M with intrinsic mean μ_I and that Q assigns zero mass to a neighborhood, however small, of the* cut locus *of μ_I. Let $\phi = \exp \mu_I^{-1}$ be the* *inverse exponential, or* log-*, function at* μ_I *defined on a neighborhood* G *of* $\mu = \mu_I$ *onto its image* U *in the tangent space* $T_{\mu_I}(M)$*. Assume that the assumptions* (A4)-(A6) *hold. Then, with* $s = d$*, the CLT* (2.5) *holds for the intrinsic sample mean* $\mu_n = \mu_{n,I}$*, say.*

Remark 2.4. Corollary 2.3 improves the CLT for the intrinsic mean due to [15], and also Theorem 2.3 and Theorem 5.3 in [10].

For the case of the *extrinsic mean*, let M be a d-dimensional differentiable manifold, and $J : M \to E^N$ an embedding of M into an N-dimensional Euclidean space. Assume that $J(M)$ is closed in E^N, which is always the case, in particular, if M is compact. The extrinsic distance $\rho_{E,J}$ on M is defined as $\rho_{E,J}(p,q) = \|J(p) - J(q)\|$ for $p, q \in M$, where $\|\cdot\|$ denotes the Euclidean norm of E^N. The image μ in $J(M)$ of the *extrinsic mean* $\mu_{E,J}$ is then given by $\mu = P(m)$, where m is the usual mean of $Q \circ J^{-1}$ thought of as a probability on the Euclidean space E^N, and P is the orthogonal projection defined on an N-dimensional neighborhood V of m into $J(M)$ minimizing the Euclidean distance between $p \in V$ and $J(M)$. If the projection P is unique on V, then the projection $\mu_n = P(m_n)$ of the Euclidean mean $m_n = \sum_{j=1}^n J(Y_j)/n$ on $J(M)$ is, with probability tending to one as $n \to \infty$, unique and lies in an open neighborhood G of $\mu = P(m)$ in $J(M)$. Theorem 2.2 immediately implies the following result of [14] (also see [10], Proposition 4.3).

Corollary 2.5 (CLT for extrinsic means on a manifold)**.** *Assume that* P *is uniquely defined in a neighborhood of the* N*-dimensional Euclidean mean* m *of* $Q \circ J^{-1}$*. Let* ϕ *be a diffeomorphism on a neighborhood* G *of* $\mu = P(m)$ *in* $J(M)$ *onto an open set* U *in* $\mathbb{R}^d$*. Assume* (A1), (A4)-(A6)*. Then, using the notation of* (2.5)*,*

$$\sqrt{n}\,[\phi(\mu_n) - \phi(\mu)] = \sqrt{n}\,[\phi(P(m_n)) - \phi(P(m))] \xrightarrow{\mathcal{L}} N(0, \Lambda^{-1} C \Lambda^{-1}),\ \text{as } n \to \infty.$$

Remark 2.6. In Corollary 2.5, one may, in particular, choose (U, ϕ) to be a coordinate neighborhood of $\mu = P(m)$ in $J(M)$. In [14], however, ϕ is chosen to be the linear orthogonal projection on G into the tangent space $T_\mu J(M)$.

Remark 2.7. In the case $S = M$ is a Riemannian manifold and $(G = M)$, the dispersion matrix in Theorem 2.2 (and Theorem 3.3 in the next section) is related to the sectional curvature of M. For M with constant curvature such as S^d one may express this matrix explicitly (see [9]). Recently, [32] has extended this result to the important case of planar shape space Σ_2^k and, more generally, to manifolds with constant holomorphic curvature.

We now turn to applications of Theorem 2.2 to the so-called *stratified spaces* S which are made up of several subspaces of different dimensions. In particular, we next consider an example where S is a *space of nonpositive curvature* (NPC), which is not in general a differentiable manifold, but has a metric with properties of a geodesic distance (namely, minimum length of curves between points) and which is also somewhat analogous to differentiable manifolds of nonpositive curvature. These spaces were originally studied by A. D. Alexandrov and developed further by Yu. G. Reshetnyak and M. Gromov (see [41] for a detailed treatment). Unlike differentiable manifolds of positive curvature where uniqueness of the intrinsic mean is known only under very restrictive conditions (see [30], [31] and [1]), on an NPC space the Fréchet mean is always unique if the Fréchet function (2.1) is finite [41].

We will consider a stratified NPC space S which is the union of a finite number of disjoint sets U_k, each of which in its relative topology in S is homeomorphic to an open subset of $\mathbb{R}^s$, including possibly the degenerate case $s = 0$, $\mathbb{R}^0$ being a singleton.

The results described below originated in a SAMSI working group (`http://www samsi.info/working-groups/data-analysis-sample-spaces-manifold-stratification`), and are further developed in [13], [26]. Also see [6], [36] and [28].

Let Q be a probability measure on S. We define the *Wasserstein distance* d_W on the space $\mathcal{P}(S)$ of probability measures on the Borel sigma-field of S as

$$d_W^2(Q_1, Q_2) = \inf\{E\rho^2(\boldsymbol{X}, \boldsymbol{Y}) : \mathcal{L}(\boldsymbol{X}) = Q_1, \mathcal{L}(\boldsymbol{Y}) = Q_2\}, \tag{2.14}$$

where $\mathcal{L}(\boldsymbol{Z})$ denotes the law, or distribution, of $\boldsymbol{Z}$. That is, the infimum on the right is over the set of all (joint) distributions of $(\boldsymbol{X}, \boldsymbol{Y})$ (in $\mathcal{P}(S \times S)$) with marginals Q_1 and Q_2. For considering finite Fréchet functions the appropriate space of probabilities that we consider below is $\{\widetilde{Q} \in \mathcal{P}(S)$: Fréchet function of $\widetilde{Q}$ is finite$\}$, endowed with the Wasserstein distance.

On a stratified space S, we say that the *Fréchet mean* μ of Q is *sticky* on U_k if there exists a Wasserstein neighborhood of Q such that for every $\widetilde{Q}$ in this neighborhood the Fréchet mean of $\widetilde{Q}$ lies in the same stratum U_k.

As an immediate consequence of Theorem 2.2, we get the following result.

Proposition 2.8. *Suppose the Fréchet mean μ of Q on a stratified NPC space S is sticky on a stratum U_k which is not degenerate. Then, with $G = U_k$, the CLT in Theorem* 2.2 *holds under the given assumptions* (2.2) *and* (A4)-(A6). *In the degenerate case, i.e., $U_k = \{\mu\}$, the sample Fréchet mean μ_n equals μ for all sufficiently large n, almost surely.*

Example 2.9 (Open book)**.** Let $S = (\bigcup_{k=1,\dots,K} H_k) \cup S_0$ where $H_k := \{k\} \times H$, $H = \mathbb{R}^D \times [0, \infty)$, $S_0 = \{0\} \times \mathbb{R}^D$, with the boundary point $(k; 0, x^1, \dots, x^D)$ of H_k identified with the point $(0, x^1, \dots, x^D)$ of S_0 for all k. That is, S is the union of K copies of the half space H glued together at the common border or *spine* $S_0 = \{0\} \times \mathbb{R}^D$. We express S as the disjoint union $S = (\bigcup_{k=1,\dots,K} S_k) \cup S_0$, where the k-th *leaf* is $S_k = \{(k; x^0, x^1, \dots, x^D)\}$ with $x^0 \in (0, \infty)$, $x^j \in \mathbb{R}$ for $j = 1, \dots, D$. For a point $\boldsymbol{x} = (x^0, x^1, \dots, x^D) \in H$ we define its *reflection across the spine* S_0 as $R\boldsymbol{x} = (-x^0, x^1, \dots, x^D)$. Using $\|\cdot\|$ for the Euclidean norm, the distance ρ on S is then defined by

(2.15)

$$\begin{aligned} &\rho((k; \boldsymbol{x}), (k; \boldsymbol{y})) = \|\boldsymbol{x} - \boldsymbol{y}\| \ \forall \boldsymbol{x}, \boldsymbol{y} \in H = \mathbb{R}^D \times [0, \infty), k = 1, \dots, K; \\ &\rho((k; \boldsymbol{x}), (k'; \boldsymbol{y})) = \|\boldsymbol{x} - R\boldsymbol{y}\| = \|R\boldsymbol{x} - \boldsymbol{y}\| \ \forall \boldsymbol{x}, \boldsymbol{y} \in H = \mathbb{R}^D \times [0, \infty), \text{ if } k \neq k'. \end{aligned}$$

Note that while the zero-th coordinate x^0 of $\boldsymbol{x}$ is nonnegative, that of $R\boldsymbol{x}$ is $-x^0$ and is negative or zero, so that if $k \neq k'$, then

(2.16)

$$\rho^2\left((k; x^0, x^1, \dots, x^D), (k'; y^0, y^1, \dots, y^D)\right) = (x^0 + y^0)^2 + \|(x^1, \dots, x^D) - (y^1, \dots, y^D)\|^2.$$

We now provide an exposition of a characterization of sticky Fréchet means on open books due to [26]: "Sticky central limit theorems on open books", with slightly different notation and terminology. Assume that $w_k = \mu(S_k) > 0$ for all

$k = 1, \ldots, K$. Define the following *k-th folding map* f_k on S into $\mathbb{R}^{D+1}$ as

$$f_k((k:\boldsymbol{x})) = \boldsymbol{x}, f_k((k':\boldsymbol{x})) = R\boldsymbol{x} \text{ if } k' \neq k \ (k = 1, \ldots, K), \tag{2.17}$$

and denote by m_k the usual (one-dimensional) mean of the zero-th coordinate of f_k:

(2.18)

$$m_k = \int z^0 (Q \circ f_k^{-1})(d\boldsymbol{z}) = -(1/2) \left[\partial/\partial x^0 \int_{\mathbb{R}^{d+1}} \|\boldsymbol{z} - \boldsymbol{x}\|^2 (Q \circ f_k^{-1})(d\boldsymbol{z}) \right]_{x^0 = 0}.$$

Let $\widetilde{Q}$ be the distribution induced by Q on H under the projection π on S into H defined by $\pi(k;\boldsymbol{x}) = \boldsymbol{x}$ (and $\pi(\boldsymbol{x}) = \boldsymbol{x}$ on S_0). Let Q_k be the measure $\widetilde{Q}$ restricted to $\pi(S_k)$. Note that $Q_k = Q \circ f_k^{-1}$ restricted to S_k. Also, let Q_0 be the restriction of Q (or $\widetilde{Q}$) to S_0. In view of the additive nature of ρ^2, the minimization of the Fréchet function is achieved separately for the zero-th coordinate x^0 of $\boldsymbol{x}$ along with the leaf on which it lies and the remaining D coordinates $(x^1, \ldots, x^D)$. The last D coordinate of the Fréchet mean on S is simply the mean μ_{1D}, say, of $(x^1, \ldots, x^D)$ under $\widetilde{Q}$. The position of the Fréchet mean μ, or whether it is sticky on the spine S_0 or to some other stratum, is determined by m_k $(k = 1, \ldots, K)$. Since the integral on the right side of (2.18) is the Fréchet function of Q evaluated on the leaf S_k at the spine, it follows from (2.18) that *if $m_k > 0$, then, for a while, the Fréchet function is strictly decreasing on S_k along the zero-th coordinate as it moves away from the spine S_0.* On the other hand, if $m_k > 0$, then $m_{k'} < 0$ for all $k' \neq k$. For this note that $m_k = \int_H \boldsymbol{z}^0 Q_k(d\boldsymbol{z}) - \sum_{1 \leq k' \neq k} \int_H \boldsymbol{z}^0 Q_{k'}(d\boldsymbol{z})$. Comparing this with the corresponding expression for $m_{k'}$, we see that $m_{k'} \leq \int_H \boldsymbol{z}^0 Q_{k'}(d\boldsymbol{z}) - \int_H \boldsymbol{z}^0 Q_k(d\boldsymbol{z}) < 0$, since $m_k > 0$. Hence *the Fréchet function is strictly increasing on $S_{k'}$ for all $k' \neq k$ along the zero-th coordinate as it increases, i.e., as the point moves away from the spine S_0.* It follows that $\mu \in S_k$. Also, if $m_k > 0$, then there exists a neighborhood of Q in the Wasserstein distance on which $m_k > 0$. That is, if $m_k > 0$ for some k, then μ is sticky on the stratum S_k, and Theorem 2.2 applies with $s = D + 1 = d$. It is clear that the Fréchet mean in this case is $\mu = (k; m_k, \mu_{1D})$, and the asymptotic distribution of $\pi(\mu_n)$ is normal with mean (m_k, μ_{1D}) and covariance matrix $n^{-1}\Sigma$, where Σ is the $d \times d$ covariance matrix of $Q \circ f_k^{-1}$, which follows from the classical multivariate CLT for i.i.d. summands with common distribution $Q \circ f_k^{-1}$. The above argument also shows that if $m_k < 0$ for all $k = 1, \ldots, K$, then μ belongs to S_0, and it is *sticky on the spine* S_0, so that Theorem 2.2 applies with $s = D$. In this case $\mu = (0, \mu_{1D})$ and, with probability tending to one as $n \to \infty$, μ_n lies in S_0, with its zero-th coordinate as 0 and its remaining D coordinates comprising the mean of n i.i.d. vectors with the common distribution that of $(X_1, \ldots, X_D)$ under Q. Thus, again, by the classical multivariate CLT for i.i.d. summands, the asymptotic distribution of $\mu_n = \pi(\mu_n)$ on S_0 is normal $N((0, \mu_{1D}), n^{-1}\Sigma_0)$. Note that Σ_0 is the same as the $D \times D$ upper sub-matrix of Σ.

To complete the picture consider the case $m_k = 0$ for some k. Then once again $m_{k'} < 0$ for all $k' \neq k$, and the minimum of the Fréchet function occurs on $S_0 \cup S_k = \bar{S}_k$. Let $m_{k,n}$ be the sample mean of the zero-th coordinate under $Q \circ f_k^{-1}$. Since the set $\{Q' : m_{k'} < 0 \text{ for all } k' \neq k\}$ is open in the Wasserstein distance (in the set of probabilities $\{Q' :$ Fréchet function of Q' is finite$\}$), if $m_{k,n} \leq 0$, then the sample Fréchet mean μ_n lies in S_0. If $m_{k,n} > 0$, then μ_n lies in S_k. Since $E(m_{k,n}) = m_k = 0$, it follows by the classical CLT that the asymptotic distribution of μ_n is, with probability $\frac{1}{2}$, $N((0, \mu_{1D}), n^{-1}\Sigma_0)$ on S_0 and, with probability $\frac{1}{2}$, it has the asymptotic distribution on S_k of its numerical coordinates as the conditional distribution of $(X^0, X^1, \ldots, X^D)$, given $X^0 > 0$, where $(X^0, X^1, \ldots, X^D)$ has the distribution $N((0, \mu_{1D}), n^{-1}\Sigma)$.

We refer to other examples of stratified spaces such as considered in [21] and [5], where also Theorem 2.2 applies. These may be thought of as toy models for the study of phylogenetic trees pioneered by S. Holmes and her collaborators (see, e.g., [16], [24]).

3. A CLT for the intrinsic mean

We begin with the circle S^1. Under the assumption of a continuous density f of Q on S^1, a necessary and sufficient condition for the existence of a unique minimizer of the intrinsic Fréchet function on the circle S^1 was given in the manuscript [12], showing, in particular, the twice continuous differentiability of the intrinsic Fréchet function. It is further shown there that the Fréchet function is convex at $p \in S^1$ if $f(-p) < 1/2\pi$, concave if $f(-p) > 1/2\pi$. This work is mentioned in [25], p. 182, and also appears in [10], pp. 73-75, 31-33. Under a continuity assumption, a direct proof of the CLT of the Fréchet mean is given in [34], and extended further in [25] when the continuity assumption does not hold.

Proposition 3.1. *On S^d the Fréchet function is twice continuously differentiable if Q has a twice continuously differentiable density f.*

Proof. For this one expresses the Fréchet function as $F(p) = \int_{D_\pi} \|v\|^2 f(\exp_p v) m(dv)$ with a natural identification with the disc $D_\pi = \{v : 0 \leq \|v\| < \pi\}$ $(\subset \mathbb{R}^d)$ of the image of $S^d \backslash \{-p\}$ in $T_p S^d$ under the map $\log_p$, and $m(dv)$ denoting the measure induced on $T_p S^d$ from the volume measure on S^d by the map $\log_p$, thought of as a measure on D_π by corresponding identifications for all p. □

Remark 3.2. Since the squared intrinsic distance $\rho_g^2(p, q)$ is smooth in p for q outside any neighborhood of $\{-p\}$, it is probably enough to assume that f has continuous derivatives of order one, or even that f is continuous. Also, we expect Proposition 3.1 and its proof to carry over to more general Riemannian manifolds such as those which are *homogeneous* ([17], p. 154).

On a general complete connected d-dimensional Riemannian manifold (M, g), the *cut point* of a point p along a geodesic $\gamma(t)$, $t \geq 0$ $(\gamma(0) = p)$, is $\gamma(t_0)$, where $t_0 = \sup\{t \geq 0 : \gamma(u), 0 \leq u \leq t$, is the unique distance minimizing segment of γ between p and $\gamma(t)\}$. The set of all cut points of p along geodesics is called the *cut locus* of p and is denoted $C(p)$ ([17], p. 207). Suppose the intrinsic mean μ_I of a probability measure Q on M exists. Take $\mu = \mu_I$, $\phi(p) = \log_\mu(p)$ defined on $M \backslash C(\mu)$. Then $\phi^{-1}(x) = \exp_\mu(x)$ and $x \to h(x, q)$ is twice continuously differentiable on $J((M \backslash C(\mu)) \backslash C(q))$. Observe that $p \in C(q)$ if and only if $q \in C(p)$ ([17], p. 271). By a slight abuse of notation, we will denote by $C(U)$ the set of cut loci of all points in a set $U \subset M$. Let $B(\mu; \epsilon)$ denote the geodesic ball with center μ and radius ϵ. Then $\phi(B(\mu; \epsilon))$ is the ball in $T_\mu M$ with center $\nu = \phi(\mu) = 0$ and radius ϵ. We then have the following result.

Theorem 3.3 (CLT for intrinsic means-II)**.** *Suppose that Q has an intrinsic mean μ and that Q is absolutely continuous in a neighborhood W of the cut locus of μ 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 $\theta \to F\left(\phi^{-1}(\theta)\right)$ is twice continuously differentiable with a nonsingular Hessian $\Lambda(\theta)$; and (iii)* (A4) *holds with $\phi(\mu)$ replaced by θ, $\forall \theta \in V$. Then, if $d > c + 2$, one has the CLT* (2.5) *for the sample intrinsic mean μ_n.*

Proof. Without loss of generality we take the neighborhood V of $\nu = 0$ sufficiently small such that $C(\phi^{-1}(V)) \subset W$. Then $Z_n(\theta) := n^{-1} \sum_{1 \leq j \leq n} \operatorname{grad} h(\theta, Y_j)$ is well defined for $Y_j \notin C(\phi^{-1}\theta)$, $j = 1, \ldots, n$, that is, with probability one, provided $\theta \in V$, since $Q(C(\phi^{-1}\theta)) = 0$. By the classical CLT, $Z_n(0) := n^{-1} \sum_{1 \leq j \leq n} \operatorname{grad} h(0, Y_j)$ is of the order $O_p(n^{-1/2})$. Let B_n be the ball in $T_\mu M$ with center $\nu = \phi(\mu) = 0$ and radius $n^{-1/2} \log n$. By hypothesis, the probability that $Y_j \in C(\phi^{-1}(B_n))$ is $O((n^{-1/2} \log n)^{d-c})$. For $\phi^{-1}(B_n)$ is the geodesic ball $B(\mu; n^{-1/2} \log n)$; hence the probability that the set $\{Y_j : j = 1, \ldots, n\}$ intersects $C(\phi^{-1}(B_n))$ is $O(n(n^{-1/2} \log n)^{d-c}) = o(1)$ if $d > c + 2$. Hence with probability converging to 1, one may use a Taylor expansion of $Z_n(\theta)$ in B_n,

$$Z_n(\theta) = Z_n(\nu) + \Lambda_n(\theta)(\theta - \nu), \quad (\theta \in B_n), \ (\nu = 0), \tag{3.1}$$

where $\Lambda_n(\theta)$ is the $d \times d$ matrix whose (r, r') element is

$$n^{-1} \sum_{1 \leq j \leq n} D_{r,r'} h(\theta(n; r, r', Y_j), Y_j)$$

with $\theta(n; r, r', Y_j)$ lying on the line segment joining θ and $\nu = 0$. By hypothesis (ii), with probability converging to one as $n \to \infty$, $\Lambda_n(\theta)$ is nonsingular for all large n $(\theta \in B_n)$ since its difference (in norm) from the Hessian $\Lambda(\theta)$ goes to zero as $n \to \infty$, by the strong law of large numbers. Now, with probability going to 1, the function $\theta \to H_n(\theta) = 0 - \Lambda_n(\theta)^{-1} Z_n(\nu)$ maps $\bar{B}_n$ into itself, where $\bar{B}_n$ is the closure of B_n. For this argument recall that $Z_n(0) = O_p(n^{-1/2})$ by the classical CLT. By the *Brouwer fixed point theorem* ([35]), $H_n(\theta)$ has a fixed point. Letting ν_n denote a measurable selection from the set of fixed points in $\bar{B}_n$, it follows that, with probability going to 1, ν_n converges to ν and satisfies the first order equation (2.7). Hence one may take ν_n as the sample intrinsic mean (note that the Fréchet function is strictly convex in a neighborhood of ν). The CLT now follows as in the last line of the proof of Theorem 2.2. □

Remark 3.4. For $d \leq c + 2$ the condition (i) in Theorem 3.3 does not imply that the probability the set $\{Y_1, \ldots, Y_n\}$ intersects $C(\phi^{-1}(B_n))$ goes to zero. Intuitively one may think that the cut locus of the image under ϕ^{-1} of a small neighborhood of the *random line* joining ν_n and 0 intersecting $\{Y_1, \ldots, Y_n\}$ is negligible, but we do not know how to justify this intuition or that it is even true.

Corollary 3.5. *Suppose Q on $M = S^d$ $(d > 2)$ has an intrinsic mean μ and is absolutely continuous on a neighborhood W of $C(\mu)$ with a continuous density on W. Suppose that the hypotheses (ii), (iii) of Theorem* 3.3 *hold. Then the CLT for the sample intrinsic mean holds.*

Proof. It is enough to note that the hypothesis (i) in Theorem 3.3 holds. Note that in the present case $C(\mu) = \{-\mu\}$ and $C(\phi^{-1}(B_n))$ is the set $-\phi^{-1}(B_n) = \{-B(\mu; n^{-1/2}\log n)\} = B(-\mu; n^{-1/2}\log n)$. The probability that $\{Y_1, \ldots, Y_n\}$ intersects this last set is $O(n(n^{-1/2}\log n)^d)$, since the density of Q on a small compact neighborhood of $C(\mu)$ is bounded. □

Remark 3.6. As mentioned at the beginning of this section, F is twice continuously differentiable if Q has a twice continuously differentiable density. We expect that the proof can be extended to the case where Q has a smooth density only in a neighborhood of $C(\mu)$. In the case of S^1 this is known under the assumption of just continuity of the density at μ (see [25] or the proof in [10] or [12]). It is for this reason we have not assumed in Theorem 3.3 and Corollary 3.5 that Q has a smooth density, although the Fréchet function is assumed to be twice continuously differentiable in a neighborhood $C(\mu)$.

Remark 3.7. Although it is curious that the proof of Theorem 3.3 does not hold for $d = 2$, the authors expect that a proof of Corollary 3.5 for the case $d = 2$ may be given using polar coordinates. For the moment the CLT for S^2 is derived only under the support restriction of Corollary 2.3.

Remark 3.8. Suppose $\mathcal{G}$ is a Lie group of isometries on S^d, $d > 2$, acting *freely* on S^d Then the projection $\pi : S^d \to S^d/\mathcal{G}$ is a *Riemannian submersion* on S^d onto its quotient space $M = S^d/\mathcal{G}$ ([23], pp. 63-65, 97-99). Let Q be a probability measure on S^d with a twice continuously differentiable density and a Karcher or intrinsic mean μ. Let $\tilde{\mu}$ be the projection of μ. Then, in local coordinates, the differential of the Fréchet function on M vanishes at $\tilde{\mu}$, because π is smooth and the differential of the Fréchet function on S^d vanishes at μ. If $\tilde{\mu}$ is a Karcher or intrinsic mean of $\tilde{Q}$, then the delta method provides a CLT for the corresponding sample Fréchet mean $\tilde{\mu}_n$ in local coordinates. If $\tilde{\mu}$ is just a local minimum, one can still use the CLT for two sample problems (see [9, 10]). One may also explore the opposite route for a probability $\tilde{Q}$ on M with a density and a unique intrinsic/Karcher mean $\tilde{\mu}$ and a probability Q, among a family of distributions with smooth densities on S^d whose projection on M is $\tilde{Q}$, such that Q satisfies the hypothesis of Corollary 3.5 with $\pi(\mu) = \tilde{\mu}$. One may then apply the CLT on S^d to derive one on $S^d/\mathcal{G}$. As an example consider the antipodal map $g(p) = -p$, and $\mathcal{G} = \{g, \text{identity}\}$. Let $\tilde{Q}$ be a probability on $M = S^d/\mathcal{G} = \mathbb{R}P^d$ (the real projective space) thought of as a probability on the upper hemisphere vanishing smoothly at the boundary and with a unique intrinsic mean $\tilde{\mu} = \{\mu, -\mu\}$, where μ is the Karcher mean of Q (restricted to the hemisphere). This opens a way for CLTs on Kendall's shape spaces as well.

Remark 3.9. Instead of defining the Fréchet mean restricted to the squared distance ρ^2, one may define it with respect to ρ^α, $\alpha \geq 1$, in (1.1), and derive Theorems 2.2, 3.3 if the assumptions hold with respect to ρ^α in place of ρ^2. Note that Proposition 3.1 extends easily to this case.

Remark 3.10. As indicated in Remark 3.8, one of the significances of a CLT on S^d is that it may provide a route to intrinsic CLTs on $S^d/\mathcal{G}$, the space of orbits under a *Lie group* $\mathcal{G}$ *of isometries* of S^d. Such spaces include the so-called *axial spaces* (or *real projective spaces* $\mathbb{R}P^d$) and Kendall type shape spaces (after omitting a singular set) which are important in *shape-based image analysis*. For the latter spaces S^d is the so-called *preshape sphere* (see, e.g., [10], p. 82). Observe that the hypothesis (i) of Theorem 3.3 may not hold in all such quotient spaces. For example, on $\mathbb{R}P^d$ one only has the order $O(\epsilon)$ in hypothesis (i) in Theorem 3.3, since the cut locus of a point in $\mathbb{R}P^d$ is isomorphic to $\mathbb{R}P^{d-1}$. For Kendall's planar shape space, identified as the *complex projective space* $\mathbb{C}P^{k-2}$, of dimension $d = 2k - 4$, the volume measure of $C\,(B(\mu;\epsilon))$ is $O(\epsilon^2)$, since the cut locus of a point of $\mathbb{C}P^{k-2}$ is isomorphic to $\mathbb{C}P^{k-3}$. For these facts refer to [23], Section 2.114, pp. 102, 103.

4. Real data examples

4.1. Kendall's planar shape space (Corpus callosum shapes of normal and ADHD children). We consider a planar shape data set, which involves measurements of a group typically developing children and a group of children suffering the ADHD (attention deficit hyperactivity disorder). ADHD is one of the most common psychiatric disorders for children that can continue through adolescence and adulthood. Symptoms include difficulty staying focused and paying attention, difficulty controlling behavior, and hyperactivity (over-activity). ADHD in general has three subtypes: (1) ADHD-hyperactive-impulsive, (2) ADHD-inattentive, (3) combined hyperactive-impulsive and inattentive (ADHD-combined) [39]. ADHD-200 Dataset (`http://fcon_1000.projects.nitrc.org/indi/adhd200/`) is a data set that records both anatomical and resting-state functional MRI data of 776 labeled subjects across eight independent imaging sites, 491 of which were obtained from typically developing individuals and 285 in children and adolescents with ADHD (ages: 7-21 years old). The corpus callosum shape data are extracted using the CCSeg package, which contains 50 landmarks, with 50 landmarks on the contour of the corpus callosum of each subject (see [27]). After quality control, 647 CC shape data out of 776 subjects were obtained, which included 404 (n_1) typically developing children, 150 (n_2) diagnosed with ADHD-combined, 8 (n_3) diagnosed with ADHD-hyperactive-impulsive, and 85 (n_4) diagnosed with ADHD-inattentive. Therefore, the data lie in the space Σ_2^{50}, which has a high dimension of $2 \times 50 - 4 = 96$. To provide a better picture of the data, we give displays of the landmark data by making the scatter plots of the landmarks selected from the contours of the CC midsections for the 243 young individuals diagnosed with ADHD. See Figure 1.

We carry out *extrinsic two-sample tests* based on Corollary 2.5 between the group of typically developing children and the group of children diagnosed with ADHD-combined, and also between the group of typically developing children and ADHD-inattentive children. We construct test statistics based on the asymptotic distribution of the extrinsic mean for the planar shapes.

The p-value for the two-sample test between the group of typically developing children and the group of children diagnosed with ADHD-combined is 5.1988×10^{-11}, which is based on the asymptotic chi-squared distribution given in Corollary 2.5. The p-value for the test between the group of typically developing children and the group ADHD-Inattentive children is smaller than 10^{-50}. It has been suggested the small p-values may result from the high dimension of the data. An alternative approach may perhaps be based on neighborhood testing in the context of Hilbert manifolds in which the shape contour is treated as an infinite-dimensional object [20, 37, 38].

The planar shape data and the codes used for computing the p-values can be found in `http://www.stat.duke.edu/~ll162/research/planar.zip`.

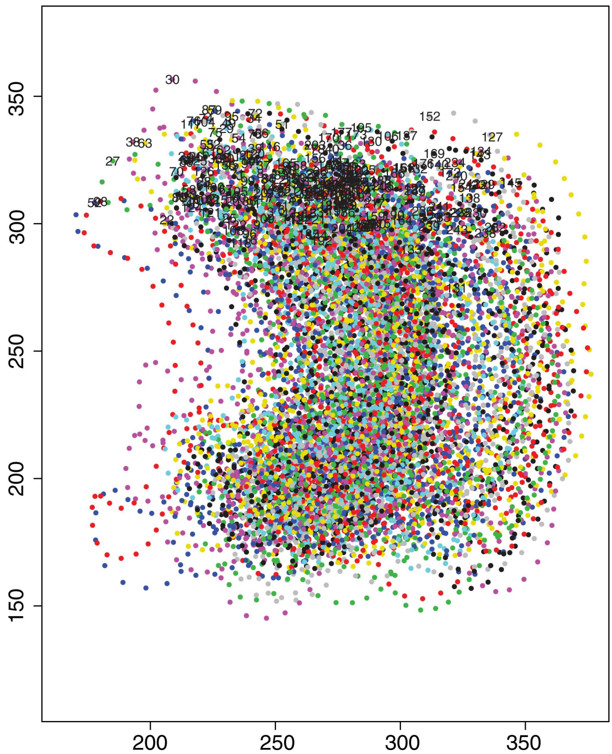

FIGURE 1. Raw landmarks from the contour of the Corpus Callosum for 243 ADHD children.

4.2. **Positive definite matrices with application to diffusion tensor imaging.** We consider $\mathrm{Sym}^+(p)$, the space of $p \times p$ positive definite matrices. Let $A \in \mathrm{Sym}^+(p)$ which follows a distribution Q. The Euclidean metric of A is given by $\|A\|^2 = \mathrm{Trace}(A)^2$. Since $\mathrm{Sym}^+(p)$ is an open convex subset of $\mathrm{Sym}(p)$, the space of all $p \times p$ symmetric matrices, the mean of Q with respect to the Euclidean distance is given by the Euclidean mean

$$\mu_E = \int AQ(dA). \tag{4.1}$$

Another metric for $\mathrm{Sym}^+(p)$ is the log-*Euclidean metric* [2]. Let $J \equiv \log : \mathrm{Sym}^+(p) \to \mathrm{Sym}(p)$ be the inverse of the *exponential map* $B \to e^B$, $\mathrm{Sym}(p) \to \mathrm{Sym}^+(p)$, which is the matrix exponential of B. J is a diffeomorphism. The log Euclidean distance is given by

$$\rho_{LE}(A_1, A_2) = \|\log(A_1) - \log(A_2)\|. \tag{4.2}$$

Note that J is an embedding on $\mathrm{Sym}^+(p)$ onto $\mathrm{Sym}(p)$ and, in fact, it is an equivariant embedding under the group action of $\mathrm{GL}(p, \mathbb{R})$, the general linear group of $p \times p$ nonsingular matrices. The extrinsic mean of Q under J is given by

$$\mu_{E,J} = \exp(\int (\log(A))Q(dA)). \tag{4.3}$$

Also, this is the *intrinsic mean* of Q under the bi-invariant metric of $\mathrm{Sym}^+(p)$ as a Lie group under multiplication: $A_1 \circ A_2 = \exp(\log(A_1) + \log(A_2))$. Since it is also the metric inherited from the vector space $\mathrm{Sym}(p)$, $\mathrm{Sym}^+(p)$ has zero sectional curvature. Another commonly used metric tensor on $\mathrm{Sym}^+(p)$ is the *affine metric*: $\langle\langle Y, Z\rangle\rangle_A = \mathrm{Trace}(A^{-1}YA^{-1}Z)\ \forall Y, Z \in \mathrm{Sym}(p)$. It is known that, with this metric, $\mathrm{Sym}^+(p)$ has nonpositive curvature [33]. We do not use this in our DTI data example, because it is computation intensive and yields results that are often indistinguishable from those using the log-Euclidean metric [40].

Theorem 2.2 applies to sample Fréchet means under both the Euclidean and log-Euclidean distances. Let $X_1, \ldots, X_{n_1}$ be an i.i.d. sample from Q_1 on $\mathrm{Sym}^+(p)$ and $Y_1, \ldots, Y_{n_2}$ be an i.i.d. sample from distribution Q_2 on $\mathrm{Sym}^+(p)$, with $\bar{X}$ and $\bar{Y}$ their corresponding sample means. Consider the case $p = 3$: $\bar{X}$ and $\bar{Y}$ are the sample mean vectors of dimension 6 for the 6 distinct values of the vectorized data. Let Σ_X and Σ_Y be the sample covariance matrices. For testing the two-sample hypothesis $H_0 : Q_1 = Q_2$, use the test statistic $(\bar{X} - \bar{Y})\Sigma^{-1}(\bar{X} - \bar{Y})^T$ with $\Sigma = (1/n_1\Sigma_X + 1/n_2\Sigma_Y)$, which has the asymptotic chi-square distribution $\chi^2(6)$. A similar test statistic is used for the log-Euclidean distance after taking matrix-log of the data.

$\mathrm{Sym}^+(3)$, the space of 3×3 positive definite matrices, has important applications in diffusion tensor imaging (DTI). Diffusion tensor imaging provides measurements of 3×3 diffusion matrices of molecules of water in tiny voxels in the white matter of the brain. When there are no barriers, the diffusion matrix is isotropic. When a trauma occurs, due to an injury or a disease, this highly organized structure, due to axon (nerve fiber) bundles and their myelin sheaths (electrically insulating layers), is disrupted and anisotropy decreases. Statistical analysis of DTI data using two- and multiple-sample tests is important in investigating brain diseases such as autism, schizophrenia, Parkinson's disease and Alzheimer's disease. There has been a growing body of work on DTI data analysis [18, 29, 40].

We now consider a diffusion tensor imaging (DTI) data set consisting of 46 subjects with 28 HIV+ subjects and 18 healthy controls. Diffusion tensors were extracted along the fiber tract of the splenium of the corpus callosum. The DTI data for all the subjects are registered in the same *atlas space* based on arc lengths, with 75 features obtained along the fiber tract of each subject. This data set has been studied in a regression setting in [43]. Our results are new and do not follow from [43]. We carry out two-sample tests between the control group and the HIV+ group for each of the 75 sample points along the fiber tract. Therefore, 75 tests are performed in total. Two types of tests are carried out based on the Euclidean distance and the log-Euclidean distance.

The simple Bonferroni procedure for testing H_0 yields a p-value equal to 75 times the smallest p-value which is of order 10^{-7}. To identify sites with significant differences, the 75 p-values are ordered from the smallest to the largest with a *false discovery rate* of $\alpha = 0.05$; 58 sites are found to yield significant differences using the Euclidean distance and 47 using the log-Euclidean distance (see [7]).

Remark 4.1. Extremely small p-values such as of the order $O(10^{-5})$ or smaller, computed using the chi-square approximation, are subject to coverage errors. They simply indicate that the p-value is extremely small. With such large observed values of the statistic, von Bahr's inequality [42], showing the tail probability under H_0 to be smaller than $o(n^{-r})$ for every $r > 0$, may perhaps be used as a justification.

Acknowledgements

The authors are grateful to the two referees for their reviews. Their constructive suggestions and criticism helped improve the paper. The authors are indebted to Professor Susan Holmes for a helpful discussion. They thank Professor Hongtu Zhu for kindly providing the data sets used in Section 4. This work was partially supported by the NSF grants DMS 1406872 and IIS 1546331.

References

[1] Bijan Afsari, *Riemannian L^p center of mass: existence, uniqueness, and convexity*, Proc. Amer. Math. Soc. **139** (2011), no. 2, 655–673, DOI 10.1090/S0002-9939-2010-10541-5. MR2736346

[2] A. Arsigny, P. Fillard, X. Pennec, and N. Ayache, *Log-Euclidean metrics for fast and simple calculus on diffusion tensors*, Magn. Reson. Med. **56** (2006), no. 2, 411–421.

[3] Ananda Bandulasiri, Rabi N. Bhattacharya, and Vic Patrangenaru, *Nonparametric inference for extrinsic means on size-and-(reflection)-shape manifolds with applications in medical imaging*, J. Multivariate Anal. **100** (2009), no. 9, 1867–1882, DOI 10.1016/j.jmva.2009.03.007. MR2543073

[4] A. Bandulasiri and V. Patrangenaru, *Algorithms for nonparametric inference on shape manifolds*, Proc. of JSM 2005, MN (2005), 1617–1622.

[5] Dennis Barden, Huiling Le, and Megan Owen, *Central limit theorems for Fréchet means in the space of phylogenetic trees*, Electron. J. Probab. **18** (2013), no. 25, 25, DOI 10.1214/EJP.v18-2201. MR3035753

[6] Bojan Basrak, *Limit theorems for the inductive mean on metric trees*, J. Appl. Probab. **47** (2010), no. 4, 1136–1149. MR2752884

[7] Yoav Benjamini and Yosef Hochberg, *Controlling the false discovery rate: a practical and powerful approach to multiple testing*, J. Roy. Statist. Soc. Ser. B **57** (1995), no. 1, 289–300. MR1325392

[8] Abhishek Bhattacharya, *Statistical analysis on manifolds: a nonparametric approach for inference on shape spaces*, Sankhyā **70** (2008), no. 2, Ser. A, 223–266. MR2551814

[9] A. Bhattacharya and R. N. Bhattacharya, *Statistics on Riemannian manifolds: asymptotic distribution and curvature*, Proc. Amer. Math. Soc. **136** (2008), 2957–2967.

[10] Abhishek Bhattacharya and Rabi Bhattacharya, *Nonparametric inference on manifolds*, With applications to shape spaces, Institute of Mathematical Statistics (IMS) Monographs, vol. 2, Cambridge University Press, Cambridge, 2012. MR2934285

[11] R. Bhattacharya and L. Lin, *A central limit theorem for Fréchet means*, arXiv1306.5806 (2013).

[12] R. N. Bhattacharya, *Smoothness and convexity of the Fréchet function on a Riemannian manifold, uniqueness of the intrinsic mean, and nonsingularity of the asymptotic dispersion of the sample Fréchet mean*, unpublished manuscipt (2007).

[13] Rabi N. Bhattacharya, Marius Buibaş, Ian L. Dryden, Leif A. Ellingson, David Groisser, Harrie Hendriks, Stephan Huckemann, Huiling Le, Xiuwen Liu, James S. Marron, Daniel E. Osborne, Vic Pătrângenaru, Armin Schwartzman, Hilary W. Thompson, and Andrew T. A. Wood, *Extrinsic data analysis on sample spaces with a manifold stratification*, Advances in mathematics, Ed. Acad. Române, Bucharest, 2013, pp. 241–251. MR3203429

[14] Rabi Bhattacharya and Vic Patrangenaru, *Large sample theory of intrinsic and extrinsic sample means on manifolds. I*, Ann. Statist. **31** (2003), no. 1, 1–29, DOI 10.1214/aos/1046294456. MR1962498

[15] Rabi Bhattacharya and Vic Patrangenaru, *Large sample theory of intrinsic and extrinsic sample means on manifolds. II*, Ann. Statist. **33** (2005), no. 3, 1225–1259, DOI 10.1214/009053605000000093. MR2195634

[16] Louis J. Billera, Susan P. Holmes, and Karen Vogtmann, *Geometry of the space of phylogenetic trees*, Adv. in Appl. Math. **27** (2001), no. 4, 733–767, DOI 10.1006/aama.2001.0759. MR1867931

[17] Manfredo Perdigão do Carmo, *Riemannian geometry*, translated from the second Portuguese edition by Francis Flaherty, Mathematics: Theory & Applications, Birkhäuser Boston, Inc., Boston, MA, 1992. MR1138207

[18] Ian L. Dryden, Alexey Koloydenko, and Diwei Zhou, *Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging*, Ann. Appl. Stat. **3** (2009), no. 3, 1102–1123, DOI 10.1214/09-AOAS249. MR2750388

[19] Ian L. Dryden, Alfred Kume, Huiling Le, and Andrew T. A. Wood, *A multi-dimensional scaling approach to shape analysis*, Biometrika **95** (2008), no. 4, 779–798, DOI 10.1093/biomet/asn050. MR2461211

[20] Leif Ellingson, Vic Patrangenaru, and Frits Ruymgaart, *Nonparametric estimation of means on Hilbert manifolds and extrinsic analysis of mean shapes of contours*, J. Multivariate Anal. **122** (2013), 317–333, DOI 10.1016/j.jmva.2013.08.010. MR3189326

[21] J. Felsenstein, *Evolutionary trees from dna sequences: A maximum likelihood approach*, Journal of Molecular Evolution **17**, no. 6, 368–376.

[22] Maurice Fréchet, *Les éléments aléatoires de nature quelconque dans un espace distancié* (French), Ann. Inst. H. Poincaré **10** (1948), 215–310. MR0027464

[23] Sylvestre Gallot, Dominique Hulin, and Jacques Lafontaine, *Riemannian geometry*, 2nd ed., Universitext, Springer-Verlag, Berlin, 1990. MR1083149

[24] S. Holmes, *Statistical approach to tests involving phylogenetics*, in Mathematics of Evolution and Phylogeny (Gascuel, O., editor), OUP, Oxford, 2005.

[25] T. Hotz and S. Huckemann, *Intrinsic means on the circle: uniqueness, locus and asymptotics*, Ann. Inst. Statist. Math. **67** (2015), no. 1, 177–193, DOI 10.1007/s10463-013-0444-7. MR3297863

[26] Thomas Hotz, Stephan Huckemann, Huiling Le, J. S. Marron, Jonathan C. Mattingly, Ezra Miller, James Nolen, Megan Owen, Vic Patrangenaru, and Sean Skwerer, *Sticky central limit theorems on open books*, Ann. Appl. Probab. **23** (2013), no. 6, 2238–2258, DOI 10.1214/12-AAP899. MR3127934

[27] C. Huang, M. Styner, and H. T. Zhu, *Penalized mixtures of offset-normal shape factor analyzers with application in clustering high-dimensional shape data*, J. Amer. Statist. Assoc., to appear (2015).

[28] Stephan Huckemann, Jonathan C. Mattingly, Ezra Miller, and James Nolen, *Sticky central limit theorems at isolated hyperbolic planar singularities*, Electron. J. Probab. **20** (2015), no. 78, 34. MR3371437

[29] Sungkyu Jung, Armin Schwartzman, and David Groisser, *Scaling-rotation distance and interpolation of symmetric positive-definite matrices*, SIAM J. Matrix Anal. Appl. **36** (2015), no. 3, 1180–1201, DOI 10.1137/140967040. MR3379023

[30] H. Karcher, *Riemannian center of mass and mollifier smoothing*, Comm. Pure Appl. Math. **30** (1977), no. 5, 509–541. MR0442975

[31] Wilfrid S. Kendall, *Probability, convexity, and harmonic maps with small image. I. Uniqueness and fine existence*, Proc. London Math. Soc. (3) **61** (1990), no. 2, 371–406, DOI 10.1112/plms/s3-61.2.371. MR1063050

[32] Wilfrid S. Kendall and Huiling Le, *Limit theorems for empirical Fréchet means of independent and non-identically distributed manifold-valued random variables*, Braz. J. Probab. Stat. **25** (2011), no. 3, 323–352, DOI 10.1214/11-BJPS141. MR2832889

[33] Christophe Lenglet, Mikaël Rousson, Rachid Deriche, and Olivier Faugeras, *Statistics on the manifold of multivariate normal distributions: theory and application to diffusion tensor MRI processing*, J. Math. Imaging Vision **25** (2006), no. 3, 423–444, DOI 10.1007/s10851-006-6897-z. MR2283616

[34] Robby G. McKilliam, Barry G. Quinn, and I. Vaughan L. Clarkson, *Direction estimation by minimum squared arc length*, IEEE Trans. Signal Process. **60** (2012), no. 5, 2115–2124, DOI 10.1109/TSP.2012.2186444. MR2954196

[35] John W. Milnor, *Topology from the differentiable viewpoint*, based on notes by David W. Weaver. Princeton Landmarks in Mathematics, Princeton University Press, Princeton, NJ, 1997. Revised reprint of the 1965 original. MR1487640

[36] Daniel Osborne, Vic Patrangenaru, Leif Ellingson, David Groisser, and Armin Schwartzman, *Nonparametric two-sample tests on homogeneous Riemannian manifolds, Cholesky decompositions and diffusion tensor image analysis*, J. Multivariate Anal. **119** (2013), 163–175, DOI 10.1016/j.jmva.2013.04.006. MR3061421

[37] D. Osborne, V. Patrangenaru, M. Qiu, and H. W. Thompson, *Nonparametric data analysis methods in medical imaging*, John Wiley & Sons, Ltd, 2015, pp. 182–205.

[38] V. Patrangenaru and L. Ellingson, *Nonparametric statistics on manifolds and their applications*, Texts in Statistical Science, Chapman & Hall/CRC, 2015.

[39] J. R. Ramsay, *Current status of cognitive-behavioral therapy as a psychosocial treatment for adult attention-deficit/hyperactivity disorder*, Curr Psychiatry Rep. **9(5)** (2007), 427–433.

[40] A. Schwartzman, *Lognormal distributions and geometric averages of positive definite matrices*, arXiv e-prints (2014).

[41] Karl-Theodor Sturm, *Probability measures on metric spaces of nonpositive curvature*, Heat kernels and analysis on manifolds, graphs, and metric spaces (Paris, 2002), Contemp. Math., vol. 338, Amer. Math. Soc., Providence, RI, 2003, pp. 357–390, DOI 10.1090/conm/338/06080. MR2039961

[42] Bengt von Bahr, *On the central limit theorem in R_k*, Ark. Mat. **7** (1967), 61–69 (1967). MR0224128

[43] Ying Yuan, Hongtu Zhu, Weili Lin, and J. S. Marron, *Local polynomial regression for symmetric positive definite matrices*, J. R. Stat. Soc. Ser. B. Stat. Methodol. **74** (2012), no. 4, 697–719, DOI 10.1111/j.1467-9868.2011.01022.x. MR2965956

[44] Herbert Ziezold, *On expected figures and a strong law of large numbers for random elements in quasi-metric spaces*, Transactions of the Seventh Prague Conference on Information Theory, Statistical Decision Functions, Random Processes and of the Eighth European Meeting of Statisticians (Tech. Univ. Prague, Prague, 1974), Reidel, Dordrecht, 1977, pp. 591–602. MR0501230

Department of Mathematics, The University of Arizona, Tucson, Arizona 85721
E-mail address: `rabi@math.arizona.edu`

Department of Statistics and Data Sciences, The University of Texas at Austin, Austin, Texas 78712
E-mail address: `lizhen.lin@austin.utexas.edu`