A Nonparametric Theory of Statistics on Manifolds

Rabi Bhattacharya

Dedicated to Friedrich Götze on the Occasion of his Sixtieth Birthday

Abstract An expository account of the recent theory of nonparametric inference on manifolds is presented here, with outlines of proofs and examples. Much of the theory centers around Fréchet means; but functional estimation and classification methods using nonparametric Bayes theory are also indicated. Applications in paleomagnetism, morphometrics and medical diagnostics illustrate the theory.

Keywords Fréchet mean • intrinsic inference • extrinsic inference • shape spaces

2010 *Mathematics Subject Classification.* Primary 62G20; Secondary 62G05, 62G10, 62H35, 62P10

1 Introduction

Statistical inference on manifolds such as circles and spheres has a long history, dating back at least to early twentieth century. But a great deal of activity was inspired by the seminal 1953 paper of R.A. Fisher on the shifts of the earth's magnetic poles over geological time scales. Statistical inference on landmarks based shape manifolds, which are of special interest in this article, came later and owes

R. Bhattacharya (✉)
Department of Mathematics, The University of Arizona, Tucson, AZ 85721, USA
e-mail: rabi@math.arizona.edu

P. Eichelsbacher et al. (eds.), *Limit Theorems in Probability, Statistics and Number Theory*, Springer Proceedings in Mathematics & Statistics 42,
DOI 10.1007/978-3-642-36068-8_9,

much of its development to the pioneering work of Kendall [32–34], providing the appropriate geometric foundation for these spaces, and to Bookstein [13–15], who in a somewhat different vein created methodologies for applications of statistics of shapes to biology and medical imaging. We must also mention the work of Karcher [31] on the uniqueness of Fréchet means of probability measures on Riemannian manifolds, and the work of Ziezold [46] on the almost sure convergence properties of Fréchet mean sets on metric spaces. Parametric inference for these spaces grew quite rapidly during the past two decades or so. In addition to the work already mentioned, important contributions were made by many authors, such as Kent [36, 37], Goodall [26], Dryden and Mardia [18], Prentice and Mardia [43], and others. A comprehensive account of this theory with extensive references to original work until 1998 may be found in the book by Dryden and Mardia [19].

In the present article we provide an expository account of the recent nonparametric theory on general manifolds, with special emphasis on shape manifolds. This theory is largely based on the notion of the *Fréchet mean* of a probability measure Q, namely, the minimizer, if unique, of the expected squared distance from a point on the manifold. If the distance on the manifold M is the geodesic distance with respect to a Riemannian metric, the Fréchet mean is said to be *intrinsic*. If the distance is the Euclidean distance inherited from an embedding of M in a Euclidean space, then the Fréchet mean is called *extrinsic.* Hendriks and Landsman [27, 28], provided asymptotics of the extrinsic mean on regular submanifolds of Euclidean spaces, with the embedding given by the inclusion map. Independently of this, a theory of extrinsic inference for Fréchet means on general manifolds originated in the 1998 dissertation of Patrangenaru, and further developed in [10, 11]. The latter articles also provided a general theory of intrinsic inference. While the emphasis in applications in the latter articles are to the sphere S^d and Kendall's planar shape spaces, embeddings of projective shape spaces and $3D$ shape spaces and inference for Fréchet means on them are developed in [2,3,8,9,39]. Further progress in both intrinsic and extrinsic inference may be found in [4, 5] and in the monograph Bhattacharya and Bhattacharya [6]. Our goal here is to present the core of this emerging field in a reasonably accessible manner.

Because references to Bhattacharya and Patrangenaru and Bhattacharya and Bhattacharya occur frequently, we would henceforth refer to them as BP and BB, respectively.

Here is an outline of the contents of the paper. In Sect. 2, basic properties of Fréchet means on metric spaces are established, including consistency (Theorem 2.1), and a general result on the asymptotic distribution of sample Fréchet means on manifolds (Theorem 2.5). The latter turns out to be crucial for intrinsic inference developed in later sections. Consistency and asymptotic distribution of extrinsic sample means are established in Sect. 3 (Theorems 3.1, 3.3), while Sect. 4 provides the corresponding results for intrinsic sample means (Theorem 4.1). The groundwork for statistical inference on general manifolds is laid in Sect. 5, including the construction of confidence regions and two-sample and match pair tests, based on the asymptotic Normal and chisquare distributions derived in earlier sections. Section 6 describes the geometries of landmarks based shape spaces. Here an observation, called a k-ad, consists of k landmarks, chosen with expert help, on an object of interest such as a brain scan, or some other digital image. The goal may be medical diagnosis, classifying biological species and subspecies, or computer vision/robotics. Because of differences in equipments used and/or their positioning relative to the object while recording images, etc., one considers for analysis the k-ad modulo an appropriate Lie group of transformations. In particular, the *similarity shape* of a k-ad is its orbit (or maximal invariant) under Euclidean rigid motions of translation and rotation, as well as scaling. The space of such shapes of k-ads in $\mathbb{R}^m$ is *Kendall's shape space* $\Sigma_m^k (k > m)$. For $m = 2$, it is more convenient for analytical and computational purposes to represent the k points of the k-ad in $\mathbb{R}^2$ as points in the complex plane. The planar shape space Σ_2^k can then be identified with the *complex projective space* $\mathbb{C}P^{k-2}$, which is a manifold of considerable interest in differential geometry. Its natural Riemannian structure is described in Sect. 6.1.1. Sect. 6.1.2 considers the intrinsic geometry of Σ_m^k in dimensions $m > 2$. Unfortunately, here the Lie group action is not free, resulting in orbits of different dimensions in different regions of Σ_m^k. If one removes the regions of singularity, the manifold is no longer complete in the Riemannian metric, and its curvature grows unboundedly as one approaches the singular sets, making inference difficult. For some recent progress in overcoming this in extending principal components analysis to Riemannian manifolds, see [29]. Sub-section 6.1.2 is devoted to the extrinsic geometry of Σ_2^k under the so-called *Veronese-Whitney embedding*, which is equivariant under the unitary group $SU(k-1)$.

As a matter of *notation*, a k-ad x in $\mathbb{R}^m$ is represented as an $m \times k$ *matrix*, with the k points appearing as k column vectors in $\mathbb{R}^m$. The *transpose* of a matrix A is expressed as A^t.

Section 6.2 defines the so-called *reflection similarity shape* $r\sigma(x)$ of a k-ad x in $\mathbb{R}^m$, identified with the orbit of the centered and scaled k-ad z under the group $O(m)$ of all orthogonal transformations. When restricted to the non-singular part of Σ_m^k using only k-ads each of which is of full rank m, the *reflection-similarity shape* space $R\Sigma_m^k$ is a manifold, although not complete. But its extrinsic analysis is facilitated by the embedding $r\sigma(x) \to z^t z$ into the space $S(k, \mathbb{R})$ of all symmetric $k \times k$ matrices, or into $S(k-1, \mathbb{R})$ if one reduces the k-ad to a $(k-1)$-ad by Helmertization to remove translation. Here z is the *preshape* of x obtained by scaling (to norm 1) the translated, or Helmertized k-ad. This new shape space and its embedding were originally introduced by Bandulasiri and Patrangenaru [8], and also arrived at independently by Dryden et al. [20]. The geometry and extrinsic inference for it was further developed in [2, 3, 6, 9]. This is a significant step in the analysis of 3D shapes. In the remaining two Sects. 6.3 and 6.4 we introduce affine and projective shapes. These are of much importance in problems of scene recognition and machine vision.

A proper extrinsic analysis requires a good *equivariant embedding*, whereby a reasonably large isometric group action on a Riemannian manifold M is replicated on its image (in an N-dimensional Euclidean space E^N), by the action of a subgroup of the general linear group $GL(N, \mathbb{R})$, via a group homomorphism. Often this latter is also a group of isometries on the image of M under the embedding when endowed with the metric tensor induced from E^N. This helps preserve much of the geometry of M. In view of this, in most examples of data analysis the results of extrinsic and intrinsic inference turn out to be nearly identical although they are based on different methodologies. The embeddings of the shape spaces considered in this article are equivariant under appropriately large group actions.

To illustrate the general theory, in Sect. 7 we develop in some detail intrinsic and extrinsic inference procedures on two specific manifolds—the sphere S^d and the planar shape space Σ_2^k. Section 8 provides a brief introduction to density estimation and classification using the nonparametric Bayes theory. Finally, Sect. 9 provides three examples of data analysis using the nonparametric theory presented in this article, and contrasts these, where possible, with results of parametric inference carried out in the literature. As is well recognized, nonparametric methods provide inference whose validity is model independent, while parametric models may be miss-specified and lead to conclusions not quite right. However, this advantage is often accompanied by larger confidence regions and smaller powers of tests. It, therefore, comes as a pleasant surprise that in most examples where data are available and for which parametric inference has been carried out, the model-independent procedures for shape spaces described in this article yield sharper inference-narrower confidence regions and much smaller p-values -than their parametric counterparts.

Finally, mention should also be made of the work of Ellingson et al. [21] for the estimation of the extrinsic mean of distributions of planar contours representing continuous planar shapes, via an infinite dimensional version of the Veronese-Whitney embedding of Σ_2^k.

We conclude this section with a sketch of the estimation of the extrinsic mean on $M = S^d$. Here the embedding J is the inclusion map of S^d into $\mathbb{R}^{d+1}$. The extrinsic mean μ_E of Q on S^d is given by $\mu_E = \mu^J/|\mu^J|$, where μ^J is the mean of Q viewed as a measure on $\mathbb{R}^{d+1}$. We assume $\mu^J \neq 0$, which is the necessary and sufficient condition for the uniqueness of the extrinsic mean on S^d. The extrinsic sample mean of i.i.d. observations $X_1, \cdots, X_n$ is, similarly, $\hat{\mu}_E = \bar{X}/|\bar{X}|$, where $\bar{X} = (X_1 + \cdots + X_n)/n$. It is easy to check that when $\hat{\mu}_E$ and μ_E are viewed as vectors in $\mathbb{R}^{d+1}$, $\hat{\mu}_E$ is asymptotically Normal $N(\mu_E, \Sigma/n)$, where the $(d+1) \times (d+1)$ matrix Σ is singular, since $\hat{\mu}_E$ lies nearly on $T_{\mu_E}(S^d)$-the tangent space of S^d at μ_E. The tangential component of $\hat{\mu}_E$, expressed in d coordinates with respect to a chosen orthonormal basis of $T_{\mu_E}(S^d)$, has the asymptotic distribution $N(0, \Sigma_1/n)$. Here Σ_1 is a $d \times d$ matrix which is nonsingular if the covariance matrix of Q (on R^{d+1}) is nonsingular. One may use this result for estimation and testing on S^d. For details of this and for intrinsic inference on S^d see Example 7.1.

2 Asymptotic Distribution Theory for Fréchet Means

Let (M, ρ) be a metric space and Q a probability measure on the Borel sigma-field of M. Consider a *Fréchet function* of Q defined by

$$F(x) = \int \rho^{\alpha}(x, y) Q(dy), \quad x \in M, \tag{1}$$

for some $\alpha \geq 1$. We will be mostly concerned with the case $\alpha = 2$. Assume that F is finite at least for one x. A minimizer of F, if unique, serves as a measure of location of Q. In general, the set C_Q of minimizers of F is called the *Fréchet mean set* of Q. In the case the minimizer is unique, one says that the *Fréchet mean exists* and refers to it as the *Fréchet mean* of Q. If $X_1, \cdots, X_n$ are i.i.d observations with common distribution Q, the Fréchet mean set and the Fréchet mean of the empirical $Q_n = 1/n \sum_{1 \leq j \leq n} \delta_{X_j}$ are named the *sample Fréchet mean set* and the *sample Fréchet mean*, respectively. For a reason which will be clear from the result below, in the case the Fréchet mean of Q exists, a (every) measurable selection from C_{Q_n} is taken to be a sample Fréchet mean.

The following is a general result on Fréchet mean sets C_Q and C_{Q_n} of Q and Q_n and *consistency* of the sample Fréchet mean.

Theorem 2.1 (Ziezold [46], BP [10], BB [6]). *Let M be a metric space such that every closed and bounded subset of M is compact. Suppose $\alpha \geq 1$ in (1) and $F(x)$ is finite for some x. Then (a) the Fréchet mean set C_Q is nonempty and compact, and (b) given any $\epsilon > 0$, there exists a positive integer valued random variable $N = N(\omega, \epsilon)$ and a P-null set $\Omega(\epsilon)$ such that*

$$C_{Q_n} \subseteq C_Q^{\epsilon} = \{x \in M : \rho(x, C_Q) < \epsilon\} \; \forall n \geq N, \forall \omega \in (\Omega(\epsilon))^c. \tag{2}$$

(c) *In particular, if the Fréchet mean of Q exists then the sample Fréchet mean, taken as a measurable selection from C_{Q_n}, converges almost surely to it.*

Remark 2.2. Unfortunately, it does not seem possible in general to estimate the Fréchet mean set C_Q consistently by C_{Q_n}, that is, the Hausdorff distance between the two does not necessarily go to zero with probability one, as n goes to infinity. Consider, for example, the simple case of $M = S^d$, with ρ as the chord distance and $\alpha = 2$. Take an absolutely continuous Q for which C_Q is not a singleton, as would be the case for the uniform distribution in particular. It is easy to see that the sample Fréchet mean set C_{Q_n} is, with probability one, a singleton.

Unless stated otherwise, we will assume in this article that the manifold M is *connected and satisfies the property that its closed bounded subsets are compact*. Obviously this is true if M is compact. The assumption also holds for all Riemannian manifolds which are complete under the geodesic distance, by the Hopf-Rinow theorem (see [17], pp. 146–147).

Remark 2.3. It has been shown by Karcher [31] for the case $\alpha = 2$ in (1) that, if the Fréchet function of Q is finite, then on a Riemannian manifold M with non-positive sectional curvature the Fréchet mean always exists as a unique minimizer.

We give a proof of Theorem 2.1 for a compact metric M, which is the case in many of the applications of interest here. Part (a) is then trivially true. For part (b), for each $\epsilon > 0$, write

$$\eta = \inf\{F(x) : x \in M\} \equiv F(q)\ \forall\ q \in C_Q,$$
$$\eta + \delta(\epsilon) = \inf\{F(x) : x \in M \backslash C_Q^\epsilon\}. \tag{3}$$

If $C_Q^\epsilon = M$, then (2) trivially holds. Consider the case $C_Q^\epsilon \neq M$, so that $\delta(\epsilon) > 0$.

Let $F_n(x)$ be the Fréchet function of Q_n, namely,

$$F_n(x) = \frac{1}{n} \sum_{1 \le j \le n} \rho^\alpha(x, X_j).$$

Now use the elementary inequality,

$$|\rho^\alpha(x, y) - \rho^\alpha(x', y)| \le \alpha\rho(x, x')[\rho^{\alpha-1}(x, y) + \rho^{\alpha-1}(x', y)] \le c\alpha\rho(x, x'),$$

with $c = 2\max\{\rho^{\alpha-1}(x, y), x, y \in M\}$, to obtain

$$|F(x) - F(x')| \le c\alpha\rho(x, x'),\ |F_n(x) - F_n(x')| \le c\alpha\rho(x, x'),\ \forall x, x'. \tag{4}$$

For each $x \in M\backslash C_Q^\epsilon$ find $r = r(x, \epsilon) > 0$ such that $c\alpha\rho(x, x') < \delta(\epsilon)/4\ \forall x'$ within a distance r from x. Let $m = m(\epsilon)$ of these balls with centers $x_1, \cdots, x_m$ (in $M\backslash C_Q^\epsilon$) cover $M\backslash C_Q^\epsilon$. By the SLLN, there exist integers $N_i = N_i(\omega)$ such that, outside a P-null set $\Omega_i(\epsilon)$, $|F_n(x_i) - F(x_i)| < \delta(\epsilon)/4\ \forall\ n \ge N_i (i = 1, \cdots, m)$. Let $N' = \max\{N_i : i = 1, \cdots, m\}$. If $n > N'$, then for every i and all x in the ball with center x_i and radius $r(x_i, \epsilon)$,

$$F_n(x) > F_n(x_i) - \delta(\epsilon)/4 > F(x_i) - \delta(\epsilon)/4 - \delta(\epsilon)/4$$
$$\ge \eta + \delta(\epsilon) - \delta(\epsilon)/2 = \eta + \delta(\epsilon)/2.$$

Next choose a point $q \in C_Q$ and find $N'' = N''(\omega)$, again by the SLLN, such that, if $n \ge N''$ then $|F_n(q) - F(q)| < \delta(\epsilon)/4$ and, consequently, $F_n(q) < \eta + \delta(\epsilon)/4$, outside of a P-Null set $\Omega''(\epsilon)$. Hence (2) follows with $N = \max\{N', N''\}$ and $\Omega(\epsilon) = \{\cup\Omega_i(\epsilon) : i = 1, \cdots, m\} \cup \Omega''(\epsilon)$. Part (c) is an immediate consequence of part (b).

Remark 2.4. For a compact metric space M, the conclusions of Theorem 2.1 hold for a *generalized Fréchet function* F by letting the integrand in (1) be an arbitrary continuous function $f(x, y)$ on $M \times M$ instead of $\rho^\alpha(x, y)$. Only a slight modification of the above proof is required for this.

For noncompact M, the proof of Theorem 2.1 is a little more elaborate and may be found in [6] or, for the case $\alpha = 2$, in [10].

We now proceed to derive the asymptotic distribution of sample Fréchet means on a d-dimensional differentiable manifold M. Let Q be a probability measure on M such that $Q(U) = 1$ for some open subset U of M which is C^2 diffeomorphic to an open set V of $\mathbb{R}^d$.

Consider a generalized Fréchet function F on U:

$$F(p) = \int_U f(p, p')Q(dp'), \; p \in U, \tag{5}$$

where $f : U \times U \to \mathbb{R}$, and the integral is finite for all p in U. Assume that F is twice differentiable in a neighborhood of the minimizer μ of F, assumed unique, and let μ_n be a consistent Fréchet sample mean. Let $\phi : U \to V$ be a C^2 diffeomorphism. Write $h(x, y) = f(\phi^{-1}x, \phi^{-1}y)$ for $x, y \in V$. Then $\nu = \phi(\mu)$ and $\nu_n = \phi(\mu_n)$ are the Fréchet minimizers of $Q \circ \phi^{-1}$ and $Q_n \circ \phi^{-1}$, respectively, of the Fréchet functions

$$H(x) = \int_V h(x, y)Q \circ \phi^{-1}(dy), \tag{6}$$

$$H_n(x) = \int_V h(x, y)Q_n \circ \phi^{-1}(dy) = \frac{1}{n}\sum_{j=1}^{n} h(x, Y_j), \; x \in V,$$

where $Y_j = \phi(X_j)$. Write $\psi^r(x, y) = D_r h(x, y) = (\partial/\partial x_r)h(x, y)$ ($r = 1, \cdots, d$) and and let D stand for the *gradient*. For example, $D\psi^r(x, y)$ is the vector $(D_1\psi^r(x, y), \cdots, D_d\psi^r(x, y))$. By assumption, H is twice differentiable in a neighborhood of $\phi(\mu)$ and a Taylor expansion yields

$$\begin{aligned} 0 &= \frac{1}{\sqrt{n}} \sum_{1 \le j \le n} \psi^r(\nu_n, Y_j) \\ &= \frac{1}{\sqrt{n}} \sum_{1 \le j \le n} \psi^r(\nu, Y_j) + \left[\frac{1}{n} \sum_{1 \le j \le n} D\psi^r(\nu, Y_j) + \epsilon_{n,r} \right] \cdot \sqrt{n}(\nu_n - \nu), \end{aligned} \tag{7}$$

where $\cdot$ denotes inner product in $\mathbb{R}^d$ and, for some $\theta_{n,r}$ lying on the line segment joining ν_n and ν,

$$\epsilon_{n,r} = \frac{1}{n} \sum_{1 \le j \le n} D\psi^r(\theta_{n,r}, Y_j) - \frac{1}{n} \sum_{1 \le j \le n} D\psi^r(\nu, Y_j).$$

The following result, which is a slight extension of Theorem 2.1 in [11], now follows from (7).

Theorem 2.5. *Let Q be a probability measure on a d-dimensional manifold M. Assume that*

(i) there exists an open subset U of M such that $Q(U) = 1$,
(ii) for a given function f on $U \times U$, the generalized Fréchet function F of Q in (5) is finite and has a unique minimizer μ in U, and there is a neighborhood of μ on which $p \to f(p, p')$ is twice continuously differentiable for every p',
(iii) there exists a C^2-diffeomorphism $\phi : U \to V$ where V is an open subset of $\mathbb{R}^d$ such that for the function $\psi^r(x, y) = D_r h(x, y) = (\partial/\partial x_r) f(\phi^{-1}x, \phi^{-1}y)$ on $V \times V$ one has $E(\psi^r(\nu, Y_1))^2 < \infty \; \forall \; r = 1, \cdots, d$, with $\nu = \phi(\mu)$ and Y_1 having the distribution $Q \circ \phi^{-1}$,
(iv) $\sup\{E|D\psi^r(\nu, Y_1) - D\psi^r(y, Y_1)| : |y - \nu| \le \epsilon\} \to 0$, as $\epsilon \downarrow 0$, and, finally,
(v) the $d \times d$ matrix $\wedge = ((ED_s\psi^r(\nu, Y_j))) \equiv ((ED_s D_r h(\nu, Y_j)))$ is nonsingular.

Then $\nu_n = \phi(\mu_n)$ has the asymptotic distribution given by

$$\sqrt{n}(\nu_n - \nu) \to N(0, \wedge^{-1}\Sigma\wedge) \text{ in distribution as } n \to \infty, \tag{8}$$

where Σ is the covariance matrix of $(\psi^r(\nu, Y_j), r = 1, \cdots, d)$.

Remark 2.6. Suppose M is a Riemannian manifold and Q a probability on M. If $q \in M$ and $C(q)$ is the cut locus of q (see Sect. 4 for definition), and if $Q(M \backslash C(q)) = 1$, then one may take U in Theorem 2.5 to be $M \backslash C(q)$. The inverse exponential map on $M \backslash C(q)$ may be taken to be the required diffeomorphism ϕ on $U = M \backslash C(q)$ onto its image V in the tangent space $T_q M$.

Note that $Q(M \backslash C(q)) = 1$ if Q is absolutely continuous with respect to a volume measure on M (see [24], p. 141).

Remark 2.7. On a Riemannian manifold M the Fréchet mean of Q for the case $f(p, p') = \rho^2(p, p')$ with geodesic distance ρ is called the *intrinsic mean* of Q. For manifolds M of nonnegative curvature, a recent criterion due to Afsari [1] under which Q is known to have an intrinsic mean is that the support of Q lie in a geodesic ball of radius $r^*/2$ where $r^* = \min\{inj(M), \pi/\sqrt{\bar{C}}\}$, $inj(M)$ being the injectivity radius of M (see Sect. 4), and $\bar{C}$ the least upper bound of sectional curvatures of M (see [31, 38]). Hence one may take U in Theorem 2.5 to be this geodesic ball in this case. For manifolds of non-positive curvature, the intrinsic mean always exists provided the Fréchet function is finite [31].

Remark 2.8. On a general differentiable manifold M, it is often useful and convenient to consider the *extrinsic mean* of Q which is the minimizer, if unique, with respect to the Euclidean distance ρ induced by an appropriate equivariant embedding of M in a Euclidean space E^N. For the case $\alpha = 2$ in (1), a broad verifiable necessary and sufficient condition for the existence of a unique minimizer is often available (see the next section). If the assumptions of Theorem 2.5 hold then one may still apply it to the extrinsic sample mean, as would be the case, e.g., of the sphere $S^d = \{x \in \mathbb{R}^{d+1} : |x|^2 = 1\}$ with the embedding given by the inclusion map in $\mathbb{R}^{d+1}$, if one takes ϕ to be the inverse exponential map on $S^d \backslash \{-p_0\}$ for a suitable point p_0. But a more broadly applicable CLT for the sample Fréchet mean is provided in the next section (see Theorem 3.3).

3 Asymptotic Distribution of the Extrinsic Sample Mean on a Manifold

Let M be a d-dimensional differentiable manifold and Q a probability measure on it. Consider an embedding $J : M \to E^N$, where E^N is an N-dimensional real vector space, which we may identify with $\mathbb{R}^N$. The extrinsic distance ρ_E on M with respect to the embedding is given by the induced Euclidean distance on $J(M) : \rho_E(p, q) = |J(p) - J(q)|$, where $|.|$ denotes the norm on E^N and $<, >$ denotes the inner product. Letting $Q^J = Q \circ J^{-1}$ denote the induced distribution on E^N, and μ^J its mean, the Fréchet function on the image $J(M)$ of M is given by

$$\begin{aligned} F^J(x) &= \int |x-y|^2 Q^J(dy) = \int |x - \mu^J - (y-\mu^J)|^2 Q^J(dy) \qquad (9) \\ &= |x-\mu^J|^2 + \int |y-\mu^J|^2 Q^J(dy) + 2 < x-\mu^J, \int (y-\mu^J) Q^J(dy) > \\ &= |x-\mu^J|^2 + \int |y-\mu^J|^2 Q^J(dy), \qquad (x \in J(M)) \end{aligned}$$

the integration being over E^N. The last sum is minimized (over $J(M)$) by taking x as the *orthogonal projection* $P(\mu^J)$ of μ^J on $J(M)$, i.e., the point in $J(M)$, if unique, which is at the minimum Euclidean distance from μ^J. Hence we have the following useful result.

Theorem 3.1 (Patrangenaru [40], Hendriks and Landsman [28], BP [10]). *Assume that the projection $P(\mu^J)$ is unique. Then the extrinsic mean of Q is $\mu_E = J^{-1}P(\mu^J)$.*

Remark 3.2. It is known that the set of points x of non-uniqueness of the projection $x \to P(x)$ on E^N (onto $J(M)$) has Lebesgue measure zero [10]. As an example, consider the case $M = S^d$, and the embedding in $\mathbb{R}^{d+1}$ given by the inclusion map. Then the only point of non-uniqueness of the projection map P is the origin 0 in $\mathbb{R}^{d+1}$, in which case the extrinsic mean set is all of S^d. The projection P in this case is defined by $P(x) = x/|x|$ for $x \neq 0$. Thus the extrinsic mean of Q on the sphere is $\mu_E = \mu^J/|\mu^J|$, which exists if and only if the Euclidean mean μ^J of the induced distribution Q^J on E^N is nonzero.

We now derive the asymptotic distribution of the extrinsic sample mean $\hat{\mu}_E$. Let $Y_i = J(X_i)$, where X_i $(i = 1, \cdots, n)$ are i.i.d. observations from the distribution Q on M. The mean of the probability $Q_n^J = \frac{1}{n}\sum_{i=1}^n \delta_{Y_i}$ induced on E^N by the empirical $Q_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ on M is $\overline{Y} = \frac{1}{n}(Y_1 + \cdots + Y_n) = \int y Q_n^J(dy)$. Then $J(\widehat{\mu}_E) = P(\overline{Y})$. By calculus, and the CLT,

$$n^{1/2}\left[P(\bar{Y}) - P(\mu^J)\right] = n^{1/2}[(\text{Jacob } P)_{\mu^J}(\overline{Y} - \mu^J)] + o_p(1) \to N(0, C), \quad (10)$$

in distribution as $n \to \infty$. Here $(\text{Jacob } P)_x$ is the $N \times N$ Jacobian matrix of P (at x) considered as a transformation on $\mathbb{R}^N \approx E^N$ into $\mathbb{R}^N$, and $C = (\text{Jacob } P)_{\mu^J}\Sigma(\text{Jacob} P)^t_{\mu^J}$, Σ being the $N \times N$ covariance matrix of Y_1. Since P maps a neighborhood V of μ^J into the image manifold $J(M)$ of dimension d (smaller than N), the rank of $(\text{Jacob } P)_{\mu^J}$ is d, and the asymptotic distribution in (10) is singular. For purposes of inference it is therefore important to consider the differential $d_{\mu^J}P$ of P at μ^J as a map on the N-dimensional tangent space $T_{\mu^J}(\mathbb{R}^N) \approx \mathbb{R}^N$ into the d-dimensional tangent space $T_{P(\mu^J)}(J(M))$ of the manifold $J(M)$ at $P(\mu^J)$, rather than as a map on $T_{\mu^J}(\mathbb{R}^N)$ into $T_{P(\mu^J)}(\mathbb{R}^N)$ as considered in (10). Consider a standard basis, or frame, $\{e_i : i = 1, \cdots, N\}$ of $E^N \approx \mathbb{R}^N$ (in which Y_j's are expressed) and an orthonormal basis (frame) $\{F_1(y), \cdots, F_d(y)\}$ of the tangent space $T_y(J(M))$ for y in a neighborhood of $P(\mu^J)$ in $J(M)$.

Theorem 3.3 (BP [11], BB [6]). *Assume that the extrinsic mean is unique and the projection operator P is continuously differentiable in a neighborhood of μ^J. Then one has*

$$n^{1/2}(d_{\mu^J}P)(\overline{Y} - \mu^J) \to N(0, \Gamma) \text{ in distribution,} \quad (11)$$

with $\Gamma = B\Sigma B^t$, and

$$n^{1/2}(d_{\bar{Y}}P)(\overline{Y} - \mu^J) \to N(0, \Gamma) \text{ in distribution.} \quad (12)$$

Here $B = B(\mu^J) = ((b_{ij}(\mu^J)))$ is the $d \times N$ matrix of $d_{\mu^J}P$ with respect to an orthonormal basis $\{e_i : i = 1, \cdots, N\}$ of $T_{\mu^J}E^N \approx \mathbb{R}^N$ and a smooth orthonormal basis $\{F_1(P(\mu^J)), \cdots, F_d(P(\mu^J))\}$ of $T_{P(\mu^J)}(J(M))$, i.e., for y in a neighborhood of $P(\mu^J)$ in $J(M)$, $(d_xP)e_i = \sum_j b_{ji}(P(x))F_j(P(x))$.

Note that (12) follows from (11) using a Slutsky type argument.

Remark 3.4. If, for $z \in J(M)$, one views $T_z(J(M))$ as a subspace of $T_z(E^N)$ spanned by $\{e_i : i = 1, \cdots, N\}$, then $B(P(y)) = F(P(y))(\text{Jacob } P)_y$, where the $d \times N$ matrix $F(P(y))$ has row vectors $F_1(P(y)), \cdots, F_d(P(y))$ which form an orthonormal basis of $T_{P(y)}(J(M))$.

Remark 3.5. The matrix Γ in (11) is nonsingular if the support of the distribution of $(d_{\mu^J}P)(Y_i - \mu^J)$ does not lie in a subspace of $T_{P(\mu^J)}(J(M))$ (of dimension smaller than d). In particular, this is the case if Q has an absolutely continuous component with respect to the volume measure on M.

Example 3.1. Consider the sphere $S^d = \{x \in \mathbb{R}^{d+1} : |x|^2 = 1\}$ and the inclusion map as the embedding J. Then $P(x) = x/|x|$ $(x \neq 0)$. It is not difficult to check that the Jacobian matrix of the projection, considered as a map on $\mathbb{R}^{d+1}$ into $\mathbb{R}^{d+1}$, is given by

$$(\text{Jacob } P)_x = |x|^{-1}[I_{d+1} - |x|^{-2}(xx^t)], \ (x \neq 0). \tag{13}$$

Let $A(x)$ be a $d \times (d+1)$ matrix whose rows form an orthonormal basis of $T_x(S^d) = \{v \in \mathbb{R}^{d+1} : x^t v = 0\}$. Then the differential of $P(x)$, as a map on $\mathbb{R}^{d+1}$ into S^d, is expressed in coordinates of this basis as $(d_x P)u = A(x)(\text{Jacob } P)_x u (u \in \mathbb{R}^{d+1})$. The left sides of (11) and (12) are then obtained by letting $x = \mu^J$ and $\overline{X}$, respectively, and $u = \overline{Y} - \mu^J$. For $d = 2$, and $x = (x^1, x^2, x^3)^t \neq (0, 0, \pm 1)^t$, and $x^3 \neq 0$, one may choose the two rows of $A(x)$ as $(-x^2, x^1, 0)/\sqrt{(x^2)^2 + (x^1)^2}$ and $((x^1, x^2, -((x^2)^2 + (x^1)^2)/x^3)/c$, where c normalizes the second vector to unity. For $x = (0, 0, \pm 1)$, one may simply take the basis vectors of $T_x(S^2)$ as (1,0,0), and (0,1,0). If $x^3 = 0$ and $x^1 \neq 0$, $x^2 \neq 0$, then the second vector in the basis may be taken as (0,0,1). Permuting the indices, all cases are now covered.

4 Asymptotic Distribution of the Intrinsic Sample Mean and the Role of Curvature

In this section we apply Theorem 2.5 to the *intrinsic mean* μ_I on a Riemannian manifold M with metric tensor g. That is, μ_I is the Fréchet mean with respect to the geodesic distance $\rho = \rho_g$ (with $\alpha = 2$ in (1)).

On the tangent space $T_p(M)$ at p of a complete Riemannian manifold M, one defines the *exponential map* $Exp_p : T_p(M) \to M$, by letting $Exp_p(v)$ be the point $q = \gamma(|v|)$ reached at time $t = |v|$ by the unit speed geodesic $\gamma(t) = \gamma(t; p, v)$ with $\gamma(0) = p$ and initial speed $\dot{\gamma}(0) = v/|v|$ if $v \neq 0$, and $Exp_p(0) = p$. For each unit vector v in $T_p(M)$, let $t_0 = t_0(p, v)$ be the supremum of all t such that the unit speed geodesic $\gamma(.; p, v)$ is length minimizing on $[0, t]$. Then $\gamma(t_0; p, v)$ is called a *cut point* of p and the set of all cut points of p (as v varies over all unit vectors in $T_p(M)$) is called the *cut locus* of p and denoted by $C(p)$. For $q \in M \backslash C(p)$, the inverse $Exp_p^{-1}(q)$ of the exponential map is defined as $v = v(q) \in T_p(M)$ such that $Exp_p(v) = q$. It is known that Exp_p^{-1} is a diffeomorphism on $M \backslash C(p)$ onto its image in $T_p(M)$, which is homeomorphic to an open ball in $T_p(M)$ with center 0 (p. 271) [17]. The quantity $inj(M) = \inf\{\rho_g(p, C(p)); p \in M\}$ is called the *injectivity radius* of M. The inverse exponential map $Exp_p^{-1}(q)$ gives rise to the so-called *normal coordinates* of q (with pole p), $q \in M \backslash C(p)$, when expressed in terms of an orthonormal basis of $T_p(M)$.

Let Q be a probability with support contained in a geodesic ball $B_r(p)$ of radius r centered at p. If a unique minimizer of the Fréchet function $F(q) = \int \rho_g^2(q, q')Q(dq')$, $q \in B_r(p)$, exists (in $B_r(p)$), it is called a *local intrinsic mean* of Q in $B_r(p)$. We will denote by $\bar{C}$ the least upper bound of sectional curvatures of M, if this l.u.b. is positive, and zero if the l.u.b is negative or zero. Part (a) of the following theorem, which is an extension of Theorem 2.2 in [11], and Theorem 4.2 in [4], now follows from Theorem 2.5. Part (b) is derived in [5]. For its notation we use the order $A \geq B$ for symmetric $d \times d$ matrices A,B to mean that $A - B$ is nononegative definite. The function f appearing in (16) is defined as

$$f(x) = \begin{cases} 1 & \text{if } \bar{C} = 0 \\ \sqrt{\bar{C}}x \cos(\sqrt{\bar{C}}x)/\sin(\sqrt{\bar{C}}x) & \text{if } \bar{C} > 0 \\ \sqrt{\bar{C}}x \cosh(\sqrt{\bar{C}}x)/\sinh(\sqrt{\bar{C}}x) & \text{if } \bar{C} < 0 \end{cases} \tag{14}$$

with $\bar{C}$, as defined earlier, the l.u.b. of the sectional curvatures of M if positive, or zero otherwise. Theorem 2.5 is a CLT for the local intrinsic sample mean μ_n around the local intrinsic mean μ_I of a probability Q, based on i.i.d. observations $X_1, \cdots, X_n$ with common distribution Q. Actually we look at vector valued $Y_i = \phi(X_i)$, where ϕ is the inverse exponential map Exp_p^{-1} on an appropriate open subset of $T_p(M)$, and derive a CLT for $\nu_n = \phi(\mu_n)$ around $\nu = \phi(\mu_I)$. Estimation of μ_I is then achieved via ϕ^{-1}.

Theorem 4.1. *Let Q have support in a geodesic ball $B_r(p)$ with $\overline{B_r(p)} \subset M \backslash C(p)$.*

Assume the following conditions (A1)–(A5)*:*

(A1) The local intrinsic mean μ_I exists in $B_r(p)$.
(A2) Let ϕ denote the inverse exponential map Exp_p^{-1}, $h(z, y) = \rho_g^2(\phi^{-1}z, \phi^{-1}y)$, with $z, y \in V = Exp_p^{-1}(B_r(p))$ expressed in normal coordinates with respect to an orthonormal basis of $T_p(M)$; then $z \to h(z, y)$ is twice continuously differentiable for all y.
(A3) With $\psi^{(r)}(z, y) \equiv D^r h(z, y) = (\partial/\partial z^r) d_g^2(\phi^{-1}z, \phi^{-1}y)$, one has $E(\psi^{(r)}(\nu, Y_1))^2 < \infty \; \forall r = 1, \cdots, d$, where $\nu = Exp_p^{-1}(\mu_I)$ and Y_1 has the distribution $Q \circ \phi^{-1}$.
(A4) One has $\sup\{E|D(\psi^{(r)}(y, Y_1) - D(\psi^{(r)}(\nu, Y_1)| : |y - \nu| \leq \epsilon\} \to 0$ as $\epsilon \downarrow 0$.
(A5) $\Lambda = ((ED_s\psi^{(r)}(\nu, Y_1))) \equiv ((\{ED_sD_rh(z, Y_1)\}_{z=\nu}))$ is nonsingular.

Then,

(a) Denoting by μ_n the local intrinsic sample mean, $\phi(\mu_n)$ has the asymptotic distribution given by

$$\sqrt{n}\,[\phi(\mu_n) - \phi(\mu_I)] \to N\left(0, \Lambda^{-1}\tilde{\Sigma}\Lambda^{-1}\right) \tag{15}$$

in distribution as $n \to \infty$, where $\tilde{\Sigma} = Cov(\{\psi^{(r)}(\nu, Y_1) : r = 1, \cdots, d\})$.
(b) If one takes $p = \mu_I$, then $\nu = 0$, and
(i) $\psi^{(r)}(0, y) = -2y^r$
(ii) $E(Y_1) = \int yQ \circ \phi^{-1}(dy) = 0$,

(iii) $\tilde{\Sigma} = 4Cov(Y_1) = 4E(Y_1 Y_1^t)$,
(iv) The matrix $\Lambda = ((\Lambda_{rs}))_{1 \le r,s \le d}$ *satisfies the order relation*

$$\Lambda \ge ((2E(([1 - f(|Y_1|)]/|Y_1|^2)Y_1^r Y_1^s + f(|Y_1|)\delta_{rs})))_{1 \le r,s \le d}, \quad (16)$$

with equality in (16) *in the case of constant sectional curvature.*

Remark 4.2. If Q has a density component with respect to the volume measure on M, then $\tilde{\Sigma}$ is nonsingular.

Remark 4.3. It has been proved by W.S. Kendall [35] that if the support of Q is contained in $B_{r^*/2}(p)$ where $r^* = \min\{inj(M), \pi/\sqrt{C}\}$, then a local Fréchet mean μ_I of Q exists in $B_{r^*/2}(p)$. The result of Afsari [1] shows that this μ_I is the global minimizer on M. If, in addition, the support of Q is contained in $B_{r^*/2}(\mu_I)$, then all the assumptions of Theorem 4.1 are satisfied [5, 6]. For manifolds with nonpositive curvature, the central limit theorem (15) holds for all Q, provided the Fréchet function of Q is finite and $E|Y_1|^2 < \infty$ (see [11], Remark 2.2)

Example 4.1. The sphere $S^d = \{x \in \mathbb{R}^{d+1} : |x|^2 = 1\}$ is a compact Riemannian manifold under the metric induced by the inclusion map. Its geodesics are the big circles, the geodesic starting at p with an initial velocity v being given by $\gamma(t; p, v) = (\cos t|v|)p + (\sin t|v|)v/|v|$, with $v \in T_p(S^d) = \{v \in \mathbb{R}^{d+1} : p^t v = 0\}$. The cut locus of p is $C(p) = \{-p\}$. The exponential map and its inverse are given by

$$Exp_p(0) = p, Exp_p(v) = \cos(|v|)p + \sin(|v|)v/|v|, \ v \ne 0, \ (v \in T_p(S^d)); \quad (17)$$
$$Exp_p^{-1}(p) = 0, Exp_p^{-1}(q) = \arccos(p^t q)/(1 - (p^t q)^2)^{1/2}[q - (p^t q)p], \ (q \ne p, -p).$$

The *geodesic distance* between p and q is $\rho_g(p, q) = \arccos(p^t q) \in [0, \pi]$, so that the injectivity radius is $inj(S^d) = \pi$. Because of isotropy, the sectional curvature is the same for every section of $T_p(S^d)$, for all p, and the unit sphere has therefore the constant curvature 1. Thus the quantity r^* appearing in Remark 4.3 has the value π, so that the conclusions of Theorem 4.1 hold if the support of Q is contained in $B_{\pi/2}(p)$ as well as in $B_{\pi/2}(\mu_I)$. Here the function f in (14) is $f(u) = u(\cos u)/(\sin u)$. The normal coordinates $y^1, \cdots, y^d$ at μ_I of $x = Exp_{\mu_I}(y)$, where y is expressed as $y = y^1 v_1 + \cdots + y^d v_d$ with respect to an orthonormal basis $\{v_r : r = 1, \cdots, d\}$ of $T_{\mu_I}(S^d)$, are now given by (see (17)):

$$y_r = \arccos(\mu_I^t x)/(1 - (\mu_I^t x)^2)^{1/2} x^t v_r, \ (r = 1, \cdots, d), x \in S^d. \quad (18)$$

Now $\Lambda_{r,s}$ is computed from its definition in (A5), with Y_{1r} given by the right hand side of (18) obtained by substituting X_1 (with distribution Q) for x, and, similarly, Y_{1s} is obtained by changing r to s.

5 Nonparametric Inference on General Manifolds

Theorems 2.5 and 3.3 allow us to construct nonparametric confidence regions for intrinsic and extrinsic means of probability measures Q on a manifold M, and to carry out nonparametric two-sample tests for the equality of such means of two distributions Q_1 and Q_2 on M. The latter tests are really meant to distinguish Q_1 from Q_2. On high dimensional spaces, such as the shape spaces of main interest here, the means are generally good indices for this purpose, as the data examples in Sect. 9 show.

For the construction of an extrinsic confidence region for the extrinsic mean μ_E of Q one may use the corresponding region for μ^J using (11) or (12) and then transform by J^{-1}. In general (12) is simpler to use. The following asymptotic chisquare distribution is an easy consequence:

$$n\left[(d_{\bar{Y}}P)(\bar{Y}-\mu^J)\right]^t(\hat{B}\hat{\Sigma}\hat{B}^t)^{-1}\left[(d_{\bar{Y}}P)(\bar{Y}-\mu^J)\right]\to\chi^2_d \text{ in distribution,} \quad (19)$$

where χ^2_d is the chisquare distribution with d degrees of freedom. Here $\hat{B}=B(\bar{Y})$ estimates $B=B(\mu^J)$, and $\hat{\Sigma}$ is the sample covariance matrix of $Y_1,\cdots,Y_n$. The statistic does not depend on the choice of the orthonormal basis of $T_{\bar{Y}}(J(M))$ for computing $\hat{B}$. The relation (19) may be used to construct a confidence region for the extrinsic mean μ_E. Bootstrapping, which leads to a smaller order of coverage error in the case of an absolutely continuous Q, may not always be feasible if N is large and the sample size n is not sufficiently large to ensure that, with high probability, the bootstrap sample is not degenerate.

Turning to the (local) intrinsic mean μ_I of Q, (15) leads to the asymptotic chisquare distribution

$$n[\phi(\mu_n)-\phi(\mu_I)]^t\hat{\Lambda}\hat{\tilde{\Sigma}}^{-1}\hat{\Lambda}[\phi(\mu_n)-\phi(\mu_I)]\to\chi^2(d) \quad (20)$$

in distribution as $n\to\infty$, where $\hat{}$ denotes an estimate with Q replaced by the empirical Q_n; that is, the distribution $Q\circ\phi^{-1}$ of Y_1 is replaced by $Q_n\circ\phi^{-1}=n^{-1}\sum_{1\le i\le n}\delta_{Y_i}$. This leads to a confidence region for μ_I. One arbitrariness here is the choice of the point p in computing ϕ. It seems reasonable to take p close to μ_n. Another idea is to use $p=\mu_I$, in which case $\phi(\mu_I)=0$. To use (20) in this case to obtain a confidence region would be computationally more intensive. It would involve finding those values p such that, with $\phi=Exp_p^{-1}$, the left side in (20) (with $\phi(\mu_I)=0$) is smaller than $\chi^2_d(1-\alpha)$, the $(1-\alpha)$-th quantile of χ^2_d. This requires computing the quantities in (20), including $\phi(\mu_n)$, for each p. But, unlike the case of the extrinsic mean where the ambient vector space E^N has generally a large dimension and the bootstrap estimate of the covariance matrix Σ tends to be singular, $\tilde{\Sigma}$ in (20) is a $d\times d$ matrix. If Q is absolutely continuous, which is a reasonable assumption in most shape data, the bootstrap construction of the confidence region will tend to have a smaller coverage error than the one using χ^2_d.

We next consider the the two-sample problem of distinguishing two distributions Q_1 and Q_2 on M, based on two independent samples of sizes n_1 and n_2, respectively: $\{Y_{j1} = J(X_{j1}) : j = 1, \cdots, n_1\}, \{Y_{j2} = J(X_{j2}) : j = 1, \cdots, n_2\}$. Hence the proper null hypothesis is $H_0 : Q_1 = Q_2$. For high dimensional M it is often sufficient to test if the two Fréchet means are equal. For the extrinsic procedure, again consider an embedding J into E^N. Write μ_i for μ_i^J for the population means and $\bar{Y}_i$ for the corresponding sample means on E^N $(i = 1, 2)$. Let $n = n_1 + n_2$, and assume $n_1/n \to p_1, n_2/n \to p_2 = 1 - p_1, 0 < p_i < 1 (i = 1, 2)$, as $n \to \infty$. If $\mu_1 \neq \mu_2$ then $Q_1 \neq Q_2$. One may then test $H_0 : \mu_1 = \mu_2 (= \mu, \text{say})$. Since N is generally quite large compared to d, the direct test for $H_0 : \mu_1 = \mu_2$ based on $\bar{Y}_1 - \bar{Y}_2$ is generally not a good test. Instead, we compare the two extrinsic means μ_{E1} and μ_{E2} of Q_1 and Q_2 and test for their equality. This is equivalent to testing if $P(\mu_1) = P(\mu_2)$. Then, by (12), assuming H_0,

$$n^{1/2} d_{\bar{Y}} P(\bar{Y}_1 - \bar{Y}_2) \to N(0, B(p_1\Sigma_1 + p_2\Sigma_2)B^t) \tag{21}$$

in distribution, as $n \to \infty$.

Here $\bar{Y} = p_1\bar{Y}_1 + p_2\bar{Y}_2$ is the *pooled estimate* of the common mean $\mu_1 = \mu_2 = \mu$, say, $B = B(\mu)$ (see (11)), and Σ_1, Σ_2 are the covariance matrices of Y_{j1} and Y_{j2}. This leads to the asymptotic chisquare statistic below:

$$n[d_{\bar{Y}} P(\bar{Y}_1 - \bar{Y}_2)]^t [\hat{B}(p_1\hat{\Sigma}_1 + p_2\hat{\Sigma}_2)\hat{B}^t]^{-1} [d_{\bar{Y}} P(\bar{Y}_1 - \bar{Y}_2)] \to \chi_d^2 \tag{22}$$

in distribution, as $n \to \infty$.

Here $\hat{B} = B(\bar{Y})$, $\hat{\Sigma}_i$ is the sample covariance matrix of Y_{ji}. One rejects the null hypothesis H_0 at a level of significance $1 - \alpha$ if and only if the observed value of the left side of (22) exceeds $\chi_d^2(1 - \alpha)$.

For the two-sample intrinsic test, let μ_{I1}, μ_{I2} denote the intrinsic means of Q_1 and Q_2 and consider $H_0 : \mu_{I1} = \mu_{I2}$. Denoting by μ_{n1}, μ_{n2} the intrinsic sample means, (15) implies that, under H_0,

$$n^{1/2}[\phi(\mu_{n1}) - \phi(\mu_{n2})] \to N(0, p_1\Lambda_1^{-1}\tilde{\Sigma}_1\Lambda_1^{-1} + p_2\Lambda_2^{-1}\tilde{\Sigma}_2\Lambda_2^{-1}) \tag{23}$$

in distribution,

where $\phi = Exp_p^{-1}$ for some convenient p in M, and Λ_i, $\tilde{\Sigma}_i$ are as in Theorem 4.1 with the empirical Q_{ni} in place of $Q_i (i = 1, 2)$. For p choose μ_n on the geodesic from μ_{n1} to μ_{n2} with $P_g(\mu_n, \mu_{n1}) = p_2 P_g(\mu_{n1}, \mu_{n2})$, and with this choice we write $\hat{\phi}$ for ϕ. The test then rejects $H_0 : Q_1 = Q_2$, if

$$n[\hat{\phi}(\mu_{n1}) - \hat{\phi}(\mu_{n2})]^t [p_1\hat{\Lambda}_1^{-1}\hat{\tilde{\Sigma}}_1\hat{\Lambda}_1^{-1} + p_2\hat{\Lambda}_2^{-1}\hat{\tilde{\Sigma}}_2\hat{\Lambda}_2^{-1}]^{-1} [\hat{\phi}(\mu_{n1}) - \hat{\phi}(\mu_{n2})] > \chi_d^2(1 - \alpha). \tag{24}$$

Finally, consider a *match pair problem* with i.i.d. observations (X_{j1}, X_{j2}) having the distribution Q on the product manifold $M \times M$. If J is an embedding of M into E^N, then $\tilde{J}(x, y) = (J(x), J(y))$ is an embedding of $M \times M$ into $E^N \times E^N$. Let μ_{E1}, μ_{E2} be the extrinsic means of the (marginal) distributions Q_1 and Q_2 of X_{j1} and X_{j2}, respectively. Once again, we are interested in testing $H_0 : Q_1 = Q_2$ by checking if $\mu_{E1} = \mu_{E2}$. Note that the extrinsic mean of Q is $\tilde{\mu}_E = (\mu_{E1}, \mu_{E2})$. If $\bar{Y}_1$, $\bar{Y}_2$ are the sample means of $Y_{j1} = J(X_{j1})$, $Y_{j2} = J(X_{j2})$, $j = 1, \cdots, n$, on E^N with $E(Y_{j1}) = \mu_1$ and $E(Y_{j2}) = \mu_2$, and $\bar{\tilde{Y}} = (\bar{Y}_1, \bar{Y}_2)$, then the extrinsic sample mean in the image space $\tilde{J}(M \times M)$ is $(P(\bar{Y}_1), P(\bar{Y}_2))$. Also, write $\bar{Y} = (\bar{Y}_1 + \bar{Y}_2)/2$. Under H_0, $\mu_1 = \mu_2 = \mu$, say, and one has

$$n^{1/2} d_{\bar{Y}} (P(\bar{Y}_1) - P(\bar{Y}_2)) \to N(0, \Sigma_{11} + \Sigma_{22} - \Sigma_{12} - \Sigma_{21}). \tag{25}$$

On the right, Σ_{11} and Σ_{22} are the $d \times d$ covariance matrices of $(d_\mu P)(Y_{j1} - \mu_1)$ and $(d_\mu P)(Y_{j2} - \mu_2)$, while Σ_{12} is the $d \times d$ cross covariance matrix of $(d_\mu P)(Y_{j1} - \mu_1)$ and $(d_\mu P)(Y_{j2} - \mu_2)$, and $\Sigma_{21} = \Sigma_{12}^t$. As above, one derives a chisquare test for H_0, using (25) and sample estimates of the covariance matrices.

6 Geometry of Shape Spaces and Equivariant Embeddings

The manifolds of main interest to us are shape spaces of landmarks based k-ads. A k-ad is a set of k labeled landmarks, $k > m$, not all the same, measured on an object or scene of interest. In general, the k-ad $(x_1, \cdots, x_k)$ is a k-tuple of points in $\mathbb{R}^m$, represented as an $m \times k$ matrix, although only $m = 2$ and 3 are of practical interest for the most part. The shape of a k-ad is the k-ad modulo a Lie group of transformations or, equivalently, it is the maximal invariant, or orbit, of the k-ad under this group. The appropriate Lie group depends on the particular statistical goal and the way the measurement of a k-ad may vary, for example, because of differences in equipment, the position and angle from which the observations are taken or recorded, etc.

6.1 Kendall's Similarity Shape Space Σ_m^k

The similarity shape of a k-ad $x = (x_1, \cdots, x_k)$ in $\mathbb{R}^m$, not all points the same, is its orbit under the group generated by translations, scaling and rotations. Writing $\bar{x} = (x_1 + \cdots + x_k)/k$, $<\bar{x}> = (\bar{x}, \cdots, \bar{x})$, the effect of translation is removed by looking at $(x_1 - \bar{x}, \cdots, x_k - \bar{x}) = x - <\bar{x}>$, which lies in the $mk - m$ dimensional hyperplane L of $\mathbb{R}^{mk}$ made up of $m \times k$ matrices with the m row sums all equal to zero. To get rid of scale, one looks at $u = (x - <\bar{x}>)/|x - <\bar{x}>|$, where $|.|$ is the usual norm in $\mathbb{R}^{mk}$. This translated and scaled k-ad is called the *preshape* of the k-ad. It lies on the unit sphere in L, and is isomorphic to $S^{m(k-1)-1}$. An alternative representation of the preshape is obtained as $p = xH/|xH|$, where H is the $k \times (k-1)$ Helmert matrix comprising $k-1$ column vectors forming an orthonormal basis of $1^{\perp}$, namely, the subspace of $\mathbb{R}^k$ orthogonal to $(1, \cdots, 1)^t$. A standard H has the j-th column given by $(a(j), \cdots, a(j), -ja(j), 0, \cdots, 0)^t$, where $a(j) = [j(j+1)]^{-1/2}$ $(j = 1, \cdots, k-1)$. Then p is an $m \times (k-1)$ matrix of norm one. The shape $\sigma(x) = \sigma(p)$ of x is then identified with the orbit of p under all rotations:

$$\sigma(x) = \sigma(p) = \{Ap : A \in SO(m)\}, \tag{26}$$
$$[SO(m) = \{\mathrm{A} : AA^t = I_m, det(A) = 1\}].$$

$SO(m)$ is called the *special orthogonal group* acting on $\mathbb{R}^m$. The set of all shapes $\sigma(x)$ is Kendall's *similarity shape* space Σ_m^k, $R > m$.

6.1.1 Intrinsic Geometry of Σ_2^k

For the case $m = 2$, it is convenient to regard a k-ad $x = ((x_1, y_1), \cdots, (x_k, y_k))$ as a k-tuple $z = (z_1, \cdots, z_k)$ of numbers $z_1 = x_1 + iy_1, \cdots, z_k = x_k + iy_k$ in the complex plane $\mathbb{C}$, and let $p = (z - <\bar{z}>)/|z - <\bar{z}>|$. Then the shape of x, or z, is identified with the orbit O_p,

$$\sigma(z) = \sigma(p) = \{e^{i\theta} p : \theta \in (-\pi, \pi]\} = O_p. \tag{27}$$

One may equivalently, consider the shape as the orbit $\{\lambda((z - <\bar{z}>) : \lambda \in \mathbb{C}\}$. That is, the shape of x, or z, is identified with a complex line passing through the origin in the subspace of $\mathbb{C}^k$ of complex dimension $k-1$ defined by $\tilde{L} = \{q = (q_1, \cdots, q_k) \in \mathbb{C}^k \backslash \{0\} : q_1 + \cdots + q_k = 0\} \approx \mathbb{C}^{k-1} \backslash \{0\}$. The shape space is then identified with the *complex projective space* $\mathbb{C}P^{k-2}$, of (real) dimension $2k-4$.

Note that $\{e^{i\theta} : \theta \in (-\pi, \pi]\}$ is a 1-dimensional compact group $G(\approx S^1)$ of isometries of the preshape sphere $\mathbb{C}S^{k-1} = \{q = (q_1, \cdots, q_k) : |q| = 1, q_1 + \cdots + q_k = 0\}$. By Helmertization, we will use the *representation* of $\mathbb{C}S^{k-1}$ as $\{p = (p_1, \cdots, p_{k-1}) \in \mathbb{C}^{k-1} : |p| = 1\}$, which is isomorphic to S^{2k-3}, and $\Sigma_2^k = \mathbb{C}S^{k-1}/G$. Recall that the metric tensor on $S^{2k-3} \approx \mathbb{C}S^{k-1}$ is that inherited from the inclusion map into $\mathbb{R}^{2(k-1)} = \{(x_1, y_1, x_2, y_2, \cdots, x_{k-1}, y_{k-1}) : (x_j, y_j) \in \mathbb{R}^2 \ \forall j\} \approx \mathbb{C}^{k-1} = \{(z_1, z_2, \cdots, z_{k-1}) : z_j = x_j + iy_j \in \mathbb{C} \ \forall j\}$, namely, $\langle v, w \rangle = Re(vw^*)$, when v,w are expressed as complex $1 \times (k-1)$ matrices (row vectors) in $\mathbb{C}S^{k-1}$. The projection map is then $\pi : p \to \sigma(p)$. The vertical subspace V_p is obtained by differentiating the curve $\theta \to e^{i\theta} p$, say at $\theta = 0$, yielding ip. That is, $V_p = \{cip : c \in \mathbb{R}\}$. Thus the horizontal subspace is $H_p = \{\tilde{v} : Re(p\tilde{v}^*) = 0, Re((ip)\tilde{v}^*) = 0\} = \{\tilde{v} : p\tilde{v}^* = 0\}$. The geodesics $\gamma(t; \sigma(p), v)$ for $v = (d_p\pi)\tilde{v}$ (for $\tilde{v}$ in H_p), and the exponential map $Exp_{\sigma(p)}$ on Σ_2^k are specified by this isometry between $T_{\sigma(p)}(\Sigma_2^k)$ and H_p for all shapes $\sigma(p)$

(see Example 4.1). Thus, identifying vectors v in H_p with vectors v in $T_{\sigma(p)}(\Sigma_2^k)$, one obtains

$$T_{\sigma(p)}(\Sigma_2^k) = \{v = (d_p\pi)\tilde{v} : \forall v \text{ such that } p\tilde{v}^* = 0\} \tag{28}$$
$$Exp_{\sigma(p)}0 = \sigma(p),\ Exp_{\sigma(p)}v = \sigma(\cos(|\tilde{v}|)p + \sin(|\tilde{v}|)\tilde{v}/|\tilde{v}|)\ (v \neq 0,\ p\tilde{v}^* = 0);$$
$$\gamma(t;\sigma(p),v) = \sigma((\cos t)p + (\sin t)\tilde{v}/|\tilde{v}|),\ (t \in \mathbb{R}, p\tilde{v}^* = 0),\ v \neq 0.$$

Denoting by ρ_{gs} and ρ_g the geodesic distances on $\mathbb{C}S^{k-1}$ and Σ_2^k, respectively, and recalling that (see Example 4.1) $\rho_{gs}(p,q) = \arccos(Repq^*)$, one has

$$\begin{aligned}\rho_g(\sigma(p),\sigma(q)) &= \inf\{\rho_{gs}(p',q') : p' \in O_p, q' \in O_q\} \\ &= \inf\{\arccos(Ree^{i\theta}pq^*) : \theta \in [0,2\pi)\} \\ &= \arccos(|pq^*|) \in [0,\pi/2].\end{aligned} \tag{29}$$

It follows that the geodesics are periodic with period π, and the cut locus of $\sigma(p)$ is $\{\sigma(q) : \text{all q such that} \arccos(|pq^*|) = \pi/2\}$, and that the injectivity radius of Σ_2^k is $\pi/2$. The inverse exponential map is given by $Exp^{-1}_{\sigma(p)}(\sigma(q)) = v$, where $v = (d_p\pi)\tilde{v}$ ($\tilde{v} \in H_p$), and $\tilde{v}$ satisfies (Use (17) with the representation of S^{2k-3} as $\mathbb{C}S^{k-1}$)

$$\begin{aligned}\tilde{v} &= Exp_p^{-1}(qe^{i\theta}) \\ &= [\arccos(Re(pq^*e^{-i\theta})](1 - [Re(pq^*e^{-i\theta})]^2)^{-1/2}qe^{i\theta} - (pq^*e^{-i\theta})p,\end{aligned} \tag{30}$$

where θ is so chosen as to minimize $\rho_{gs}(p,qe^{i\theta}) = \arccos(Re(pq^*e^{-i\theta}))$. That is, $(pq^*e^{-i\theta}) = |pq^*|$, or $e^{i\theta} = pq^*/|pq^*|$ (for $pq^* \neq 0$, i.e., for $\sigma(q)$ not in $C(\sigma(p))$.

Hence, writing $\rho = (\arccos)\rho_g(\sigma(p),\sigma(q))$, $\rho \neq 0$, one has

$$\begin{aligned}\tilde{v} &= [\arccos(|pq^*|)](1 - |pq^*|^2)^{-1/2}\{(pq^*/|pq^*|)q - |pq^*|p\} \\ &= [\rho/\sin\rho]\{qe^{i\theta} - (\cos\rho)p\}\ (e^{i\theta} = pq^*/\cos\rho).\end{aligned} \tag{31}$$

This horizontal vector $\tilde{v}(\in H_p)$ represents $Exp^{-1}_{\sigma(p)}(\sigma(q)) = v$.

The sectional curvature of Σ_2^k at a section generated by two orthonormal vector fields $\tilde{W}_1$ and $\tilde{W}_2$ is $1 + 3\cos^2\phi$ where $\cos\phi = \langle U_1, iU_2\rangle$, U_1 and U_2 being the horizontal lifts of $\tilde{W}_1$ and $\tilde{W}_2$ (see [17]).

6.1.2 Extrinsic Geometry of Σ_2^k Induced by an Equivariant Embedding

One problem with carrying out an intrinsic analysis of the Fréchet mean is that no broad sufficient condition is known for its existence (i.e., of the uniqueness of the minimizer of the corresponding Fréchet function). Also, often such an analysis, assuming uniqueness, is computationally much more intensive than an extrinsic analysis. However, for an extrinsic analysis to be very effective one should choose a good embedding which retains as many geometrical features of the shape manifold as possible without making it cumbersome. Let Γ be a Lie group acting on a differentiable manifold M, and denote by $GL(N, \mathbb{R})$ the linear group of nonsingular transformations on a Euclidean space E^N of dimension N onto itself. An embedding J on M into E^N is said to be Γ-*equivariant* if there exists a group homomorphism $\Phi : \gamma \to \phi_\gamma$ of Γ into $GL(N, \mathbb{R})$ such that $J(\gamma p)) = \phi_\gamma(Jp)$ $\forall p \in M, \gamma \in \Gamma$. Often, when there is a natural Riemannian structure on M, Γ is a group of isometries of M. Consider the so-called *Veronese-Whitney embedding J* of Σ_2^k into the (real) vector space $S(k-1, \mathbb{C})$ of all $(k-1) \times (k-1)$ Hermitian matrices $B = B^*$, defined by

$$J\sigma(p) = p^* p \;\; [\sigma(p) = \{e^{i\theta} p, \theta \in [0, 2\pi), p \in \mathbb{C}S^{k-1}]. \qquad (32)$$

The Euclidean inner product on $S(k-1, \mathbb{C})$, considered as a real vector space, is given by $\langle B, C \rangle = Re(\text{Trace}(BC^*))$. Let $SU(k-1)$ denote the special unitary group of all $(k-1) \times (k-1)$ unitary matrices A (i.e., $A^*A = I, det(A) = 1$) acting on $S(k-1, \mathbb{C})$ by $B \to A^*BA$. Then the embedding (32) is Γ-equivariant, with $\Gamma = \{\gamma_A : A \in SU(k-1)\}$ and the group action on Σ_2^k given by: $\gamma_A \sigma(p) = \sigma(pA)$. For $J\sigma(pA) = A^* p^* pA = \phi(\gamma_A)(J\sigma(p))$, say, where the group homomorphism on Γ onto $SU(k-1)$ is given by $\gamma_A \to \phi(\gamma_A) : \phi(\gamma_A)B = A^*BA$. Note that $SU(k-1)$ is a group of isometries of $S(k-1, \mathbb{C})$. If Σ_2^k is given the metric tensor inherited from $S(k-1, \mathbb{C})$ by the embedding (32), then the embedding is isometric as well as equivariant.

A *size-and-shape similarity shape* $s\sigma(z)$ is defined for Helmertized k-ads $z = (z_1, \cdots, z_{k-1})$ as its orbit under $SO(m)$. An equivariant embedding for it is $s\sigma(z) \to z^*z/|z|$, on the *size-and-shape-similarity shape* space $S\Sigma_2^k$ into $S(k-1, \mathbb{C})$.

6.2 *Reflection Similarity Shape Space $R\Sigma_m^k$, $m > 2$*

For $m > 2$, let $\tilde{N}S^{m(k-1)-1}$ be the subset of the centered preshape sphere $S^{m(k-1)-1}$ whose points p span $\mathbb{R}^m$, i.e., which, as $m \times (k-1)$ matrices, are of full rank. We define the *reflection similarity shape* of the k-ad as

$$r\sigma(p) = \{Ap : A \in O(m)\} \; (p \in \tilde{N}S^{m(k-1)-1}), \qquad (33)$$

where $O(m)$ is the set of all $m \times m$ orthogonal matrices $A : AA^t = I_m$, $det(A) = \pm 1$. The set $\{r\sigma(p) : p \in \tilde{N}S^{m(k-1)-1}\}$ is the *reflection similarity shape space* $R\Sigma_m^k = \tilde{N}S^{m(k-1)-1}/O(m)$. Since $\tilde{N}S^{m(k-1)-1}$ is an open subset of the sphere $S^{m(k-1)-1}$, it is a Riemannian manifold. Also $O(m)$ is a compact Lie group acting on it. Hence there is a unique Riemannian structure on $R\Sigma_m^k$ such that the projection map $p \to O(p)$ is a Riemannian submersion.

We next consider a useful embedding of $R\Sigma_m^k$ into the vector space $S(k-1, \mathbb{R})$ of all $(k-1) \times (k-1)$ real symmetric matrices (see [3, 8, 9, 20]). Define

$$J(r\sigma(p)) = p^t p \ (p \in \tilde{N}S^{m(k-1)-1}), \tag{34}$$

with p an $m \times (k-1)$ matrix with norm one. Note that the right side is a function of $r\sigma(p)$. Here the elements p of the preshape sphere are Helmertized. To see that this is an embedding, we first show that J is one-to-one on $R\Sigma_m^k$ into $S(k-1, \mathbb{R})$. For this note that if $J(r\sigma(p))$ and $J(r\sigma(q))$ are the same, then the Euclidean distance matrices $((|p_i - p_j|))_{1 \le i \le j \le k-1}$ and $((|q_i - q_j|))_{1 \le i \le j \le k-1}$ are equal. Since p and q are centered, by geometry this implies that $q_i = Ap_i (i = 1, \cdots, k-1)$ for some $A \in O(m)$, i.e., $r\sigma(p) = r\sigma(q)$. We omit the proof that the differential dJ is also one-to-one. It follows that the embedding is equivariant with respect to a group action isomorphic to $O(k-1)$.

For $m > 2$, a *size-and-reflection shape* $sr\sigma(z)$ of a Helmertized k-ad z in $\mathbb{R}^m$ of full rank m is given by its orbit under the group $O(m)$. The space of all such shapes is the size-and-reflection shape space $SR\Sigma_m^k$. An $O(k-1)$-equivariant embedding of $SR\Sigma_m^k$ into $S(k-1, \mathbb{R})$ is : $J(sr\sigma(z)) = z^t z/|z|$.

6.3 Affine Shape Space $A\Sigma_m^k$

Let $k > m + 1$. Consider the set of all k-ads in $\mathbb{R}^m$, with full rank m as $m \times k$ matrices. The affine shape of a k-ad x may be identified with its orbit under all affine transformations:

$$\sigma(x) = \{Ax + c : A \in GL(m, \mathbb{R}), c \in \mathbb{R}^m\}. \tag{35}$$

If the k-ad is centered as $u = x - < \bar{x} >$, then the affine shape of x, or of u, is given by

$$\sigma(x) = \sigma(u) = \{Au : A \in GL(m, \mathbb{R})\}, \ (u \text{ centered } k - ad \ \text{ of rank } m). \tag{36}$$

The space of all such affine shapes is the *affine shape space* $A\Sigma_m^k$. Note that two Helmertized k-ads u and v (as $m \times (k-1)$ matrices of full rank) have the same shape if and only if the rows of u and v span the same m-dimensional subspace of $\mathbb{R}^{k-1}$. Hence we can identify $A\Sigma_m^k$ with the Grasmannian $G_m(k-1)$, namely, the set of all m-dimensional subspaces of $\mathbb{R}^{k-1}$ (Sparr [45]. For the Grassmann manifold, refer to

Boothby [16], pp. 63, 168, 362, 363). For extrinsic analysis on $A\Sigma_m^k \approx G_m(k-1)$, consider the embedding of $A\Sigma_m^k$ into $S(k-1, \mathbb{R})$ given by

$$J(\sigma(u)) = FF^t, \tag{37}$$

where $F = (f_1 \cdots f_m)$ is a $(k-1) \times m$ matrix and $\{f_1, \cdots, f_m\}$ is an orthonormal basis of the m-dimensional subspace L, say, of $\mathbb{R}^{k-1}$ spanned by the rows of u. Note that the $(k-1) \times (k-1)$matrix FF^t is idempotent and is the matrix of orthogonal projection of $\mathbb{R}^{k-1}$ onto L. It is independent of the orthonormal basis chosen. The embedding is $O(k-1)$-equivariant under the group action $\sigma(u) \to \sigma(uO)$ ($O \in O(k-1)$) on $A\Sigma_m^k$, with $O(k-1)$ acting on $S(k-1, \mathbb{R})$ by $A \to OAO^t$.

6.4 Projective Shape Space $P\Sigma_m^k$

First, recall that the real projective space $\mathbb{R}P^m$ is the space of all lines through the origin in $\mathbb{R}^{m+1}$. Its elements are $[p] = \{\lambda p : \lambda \in \mathbb{R}\backslash\{0\}\}$ for all $p \in \mathbb{R}^{m+1}\backslash\{o\}$. It is also conveniently represented as the quotient S^m/G where G is the two-point group $\{e, -e\}$, e being the identity map and $-ep = -p$ ($p \in S^m$). That is, a line through p is identified with $\{p/|p|, -p/|p|\}$ ($p \in \mathbb{R}^{m+1}\backslash\{o\}$). As a consequence, there is a unique Riemannian metric tensor on $\mathbb{R}P^m = S^m/G$ such that $p \to \{p, -p\}$ is a Riemannian submersion, with $\langle u, v\rangle_{\mathbb{R}P^m} = u^t v$ for all vectors u, v in $T_{[p]}\mathbb{R}P^m$. The geodesic distance is given by $\rho_g([p],[q]) = \arccos(|p^t q|) \in [0, \pi/2]$, and the cut locus of $[p]$ is $C([p]) = \{[q] : \cos(|p^t q|) = \pi/2\}$, so that the injectivity radius of $\mathbb{R}P^m$ is $\pi/2$. Its sectional curvature is constant $+1$ (as it is of S^m). The exponential map of $T_{[p]}\mathbb{R}P^m$ (and its inverse on $\mathbb{R}P^m\backslash(C([p]))$ can be easily expressed in terms of those for the sphere S^m. We will use [ ] for both representations.

The so-called *Veronese-Whitney embedding* of $\mathbb{R}P^m$ into $S(m+1, \mathbb{R})$ is given by

$$J([p]) = pp^t, \; (p = (p_1, \cdots, p_{m+1})^t \in S^m). \tag{38}$$

It is clearly $O(m+1)$-equivariant, with the group action on $\mathbb{R}P^m$ as : $A[p] = [Ap]$ ($A \in O(m+1)$).

Turning to landmarks based projective shapes, assume $k > m+2$. *A frame of* $\mathbb{R}P^m$ is a set of $m+2$ ordered points $([p_1], \cdots, [p_{m+2}])$ such that every subset of $m+1$ of these points spans $\mathbb{R}P^m$, i.e., every subset of $m+1$ points of $\{p_1, \cdots, p_{m+2}\}$ spans $\mathbb{R}^{m+1}$. The *standard frame* of $\mathbb{R}P^m$ is $([e_1], [e_2], \cdots, [e_{m+1}], [e_1 + e_2 + \cdots + e_{m+1}])$, where e_i ($\in \mathbb{R}^{m+1}$) has 1 in the ith position and zeros elsewhere. A k-ad $y = (y_1, \cdots, y_k) = ([p_1], \cdots, [p_k]) \in (\mathbb{R}P^m)^k$ is in *general position* if there exist $i_1 < i_2 < \cdots < i_{m+2}$ such that $(y_{i_1}, \cdots, y_{i_{m+2}})$ is a frame of $\mathbb{R}P^m$. A *projective transformation* α on $\mathbb{R}P^m$ is defined by

$$\alpha[p] = [Ap], \; (p \in \mathbb{R}^{m+1}\backslash\{0\}) \tag{39}$$

where $A \in GL(m+1, \mathbb{R})$. The usual operation of matrix multiplication on $GL(m+1, \mathbb{R})$ then leads to a corresponding group of projective transformations on $\mathbb{R}P^m$. This is the *projective group* $PGL(m)$. Note that, for a given A in $GL(m+1, \mathbb{R})$, cA determines the same element of $PGL(m)$ for all $c \neq 0$. The *projective shape* of a k-ad $y = (y_1, \cdots, y_k) = ([p_1], \cdots, [p_k]) \in (\mathbb{R}P^m)^k$ in general position is its orbit under $PGL(m)$:

$$\sigma(y) = \{\alpha y \equiv (\alpha[p_1], \cdots, \alpha[p_k]) : \alpha \in PGL(m)\}, \qquad (40)$$
$$(y = ([p_1], \cdots, [p_k] \text{ in general position}).$$

The *projective shape space* $PG\Sigma_m^k$ is the set of all projective shapes of k-ads in general position. Following Mardia and Patrangenaru [39] and Patrangenaru et al. [41], we will consider a particular dense open subset of $PG\Sigma_m^k$. Fix a set of $m+2$ indices $I = \{i_j : j = 1, \cdots, m+2\}$, $1 \leq i_1 < i_2 < \cdots < i_{m+2} \leq k$. Define $PG_I\Sigma_m^k$ as the set of shapes $\sigma(y)$ in $PG\Sigma_m^k$, $y = (y_1, \cdots, y_k) = ([p_1], \cdots, [p_k])$, such that every subset of $m+1$ points of $\{[p_{i_j}], j = 1, \cdots, m+2\}$ spans $\mathbb{R}P^m$.

The shape space $PG_I\Sigma_m^k$ (with $I = \{1, 2, \cdots, m+2\}$) may be identified with $(\mathbb{R}P^m)^{k-m-2}$ (see [39]). It has been shown in [12] that the full projective shape space $PG\Sigma_m^k$ in a differentiable manifold.

7 Inference on Shape Spaces

In this section we indicate how Theorems 2.5, 3.3, 4.1 and the inference procedures for general manifolds described in Sect. 5 may be applied to shape spaces, using the sphere S^d and the planar shape space Σ_2^k as illustrations.

For intrinsic analysis, consider the function $h(z, y) = \rho_g^2(Exp_p z, Exp_p y)$ for z, y in T_pM, with an appropriate choice of p. One first needs to express explicitly the quantities $D_r h(z, y)$, $D_r D_s h(z, y)$ in normal coordinates at p, i.e., at $z = 0 \equiv Exp_p^{-1} p$. (See Theorem 4.1.) For this let $\gamma(s)$ be a geodesic starting at p, and $m \in M$. Define the *parametric surface* $c(s, t) = Exp_m(tExp_m^{-1}\gamma(s))$, $s \in [0, \epsilon)$, $\epsilon > 0$ small. Note that $c(s, 0) = m$ for all s, $c(s, 1) = \gamma(s)$, and that, for all fixed $s \in [0, \epsilon)$, $t \to c(s, t)$ is a geodesic starting at m and reaching $\gamma(s)$ at $t = 1$. Writing $T(s, t) = (\partial/\partial t)c(s, t)$, $S(s, t) = (\partial/\partial s)c(s, t)$, one then has $S(s, 0) = 0$, $S(s, 1) = \dot{\gamma}(s)$. Also, $\langle T(s, t), T(s, t)\rangle$ does not depend on t and, therefore,

$$\rho_g^2(\gamma(s), m) = \int_0^1 \langle T(s, t), T(s, t)\rangle dt. \qquad (41)$$

Differentiating this respect to s and recalling the symmetry $(D/\partial s)T(s, t) = (D/\partial t)S(s, t)$ on a parametric surface (see [17, p. 68, Lemma 3.4]), and $(D/\partial t)T(s, t) = 0$, one has

$$\begin{aligned}(d/ds)\rho_g^2(\gamma(s), m) &= 2\int_0^1 \langle (D/\partial s)T(s,t), T(s,t)\rangle dt \\ &= 2\int_0^1 \langle (D/\partial t)S(s,t), T(s,t)\rangle dt = 2\int_0^1 (d/dt)\langle S(s,t), T(s,t)\rangle dt \\ &= 2\langle S(s,1), T(s,1)\rangle = -2\langle \dot{\gamma}(s), Exp_{\gamma(s)}^{-1} m\rangle.\end{aligned} \tag{42}$$

Setting $s = 0$ in (42) and letting $\dot{\gamma}(0) = v_r$, with $\{v_r : r = 1, \cdots, d\}$ an orthonormal basis of T_pM, one shows that the normal coordinates y_r of m (i.e., the coordinates of $y = Exp_p^{-1}m$ with respect to $\{v_r : r = 1, \cdots, d\}$) satisfy

$$-2y^r \equiv -2\langle Exp_p^{-1}m, v_r\rangle = [(d/ds)\rho_g^2(\gamma(s), m)]_{s=0}. \tag{43}$$

From this one gets

$$D_r h(0, y) = -2y^r (r = 1, \cdots, d). \tag{44}$$

If $Q(C(p)) = 0$, then writing $\tilde{Q}$ for the distribution induced from Q by the map Exp_p^{-1} on T_pM, the Fréchet function may be expressed as

$$F(q) = \int \rho_g^2(q, m)Q(dm) = \int h(z, y)\tilde{Q}(dy) = \tilde{F}(z), \ (z = Exp_p^{-1}q). \tag{45}$$

Since a (local) minimum of this is attained at $q = \mu_I$, $\tilde{F}$ must satisfy a first order condition $D_r\tilde{F}(z) = 0$ at $z = \nu$. In particular, letting $p = \mu_I$ and, consequently, $\nu = 0$, one has $\int D_r h(0, y)\tilde{Q}(dy) = 0$, so that (44) yields

$$\int y^r \tilde{Q}(dy) = 0 \ (r = 1, \cdots, d), \ (\tilde{Q} = Q \circ \phi^{-1}, \phi = Exp_{\mu_I}^{-1}). \tag{46}$$

Note that (44) and (46) are the relations stated in Theorem 4.1(b)(i),(ii).

By Theorem 4.1, the asymptotic distribution of the sample intrinsic mean μ_n is that of $\phi^{-1}(\nu_n)$, where $\phi = Exp_p^{-1}$, and (see (7))

$$\sqrt{n}(\nu_n - \nu) \simeq \Lambda^{-1}[(1/\sqrt{n}) \sum_{1\leq j\leq n} Dh(\nu, Y_j)], \ (\Lambda_{rs} = ED_rD_sh(\nu, Y_1), 1 \leq r, s \leq d), \tag{47}$$

with $Y_j = \phi(X_j)$, where X_j are i.i.d. with distribution Q. By (44), the right side of (47) simplifies to $\Lambda^{-1}[-2(1/\sqrt{n})\sum_{1\leq j\leq n} Y_j]$, if $p = \mu_I$ (and $\nu = 0$).

Example 7.1 (Confidence region for the intrinsic/extrinsic mean of Q on the sphere S^d). Let μ_I be the intrinsic mean of Q on S^d. Given n i.i.d. observations $X_1, \cdots, X_n$ on S^d with common distribution Q, let μ_n be the intrinsic sample mean. Write $\phi = Exp_{\mu_I}^{-1}$, and $\phi_p = Exp_p^{-1}$, so that $\phi_{\mu_I} = \phi$. By Theorem 4.1,

$$\sqrt{n}[\phi(\mu_n) - \phi(\mu_I)) = \sqrt{n}\phi(\mu_n) \to N(0, \Lambda^{-1}\tilde{\Sigma}\Lambda^{-1}) \text{ in distribution as } n \to \infty, \tag{48}$$

where the $d \times d$ matrices Λ and $\tilde{\Sigma}$ are given by

$$\tilde{\Sigma} = 4Cov(\phi(X_1)), \tag{49}$$

$$\Lambda_{rs} = 2E[(1-(X_1^t\mu_I)^2)^{-1}\{1-(1-(X_1^t\mu_I)^2)^{-1/2}\cdot(X_1^t\mu_I)\arccos(X_1^t\mu_I)\}(X_1^t\nu_r)(X_1^t\nu_s)$$
$$+ (1-(X_1^t\mu_I)^2)^{-1/2}\cdot(X_1^t\mu_I)(\arccos(X_1^t\mu_I)))\delta_{rs}], 1 \le r, s \le d.$$

Here $\{\nu_r : 1 \le r \le d\}$ is an orthonormal basis of $T_{\mu_I}S^d$.

A confidence region for μ_I, of asymptotic level $1-\alpha$, is then given by

$$\{p \in S^d : n\phi_p(\mu_n)^t\hat{\Lambda}_p\hat{\tilde{\Sigma}}_p^{-1}\hat{\Lambda}_p\phi_p(\mu_n) \le \chi_d^2(1-\alpha)\}, \tag{50}$$

where Λ_p, $\tilde{\Sigma}_p$ are obtained by replacing μ_I by p in the expressions for Λ and $\tilde{\Sigma}$ in (49). The 'hat' ($\hat{}$) indicates that the expectations are computed under the empirical Q_n, rather than Q. As mentioned in Sect. 5, it would be computationally simpler to choose a particular $p = p_0$, say, and let $\phi = Exp_{p_0}^{-1}$. Then (20) yields a simpler confidence region:

$$\{p \in S^d : n[\phi(\mu_n) - \phi(p)]^t\hat{\Lambda}_{p_0}\tilde{\Sigma}_{p_0}^{-1}\hat{\Lambda}_{p_0}[\phi(\mu_n)) - \phi(\mu_p)] \le \chi_d^2(1-\alpha)\}. \tag{51}$$

We now turn to the distribution of the extrinsic mean $\bar{X}/|\bar{X}|$. The $(d+1)\times(d+1)$ Jacobian matrix $(Jacob)_xP$ of the projection map $P : x \to x/|x|$, viewed as a map on $\mathbb{R}^{d+1}\backslash\{0\}$ into $\mathbb{R}^{d+1}$, is given by (13). Let $B(x)$ be the $d \times (d+1)$ matrix of the differential d_xP (on $T_x\mathbb{R}^{d+1}$ into $T_{P(x)}S^d = \{u \in \mathbb{R}^{d+1} : P(x)^tu = 0\}$) whose d rows form an orthonormal basis of $T_{P(x)}S^d$. Then the differential of the projection map is

$$(d_xP)u = [B(x)(Jacob)_xP]u. \tag{52}$$

If $\mu = EX_1 \neq 0$, then, by (19), a confidence region for the extrinsic mean $\mu/|\mu|$ is given by

$$\{x/|x| \in S^d : n[(d_{\bar{X}}P)(\bar{X}-x)]^t(\hat{B}\hat{\Sigma}\hat{B}^t)^{-1}[(d_{\bar{X}}P)(\bar{X}-x)] \le \chi_d^2(1-\alpha)\}. \tag{53}$$

Here $\hat{B} = B(\bar{X})$, $\Sigma = Cov(X_1)$, and $\hat{\Sigma}$ is obtained by replacing Q by Q_n in computing expectations.

Example 7.2 (Inference on the planar shape space Σ_2^k). To apply Theorem 4.1, we use (47) where $\phi = Exp_{\sigma(p)}^{-1}$ and p is a suitable point in $\mathbb{C}S^{k-1}$. To derive a computable expression for Λ, write the geodesic γ in the parametric surface $c(s,t)$ as $\gamma = \pi \circ \tilde{\gamma}$, where $\tilde{\gamma}$ is a geodesic in $\mathbb{C}S^{k-1}$ starting at $\tilde{\mu} \in \pi^{-1}\{\mu_I\}$. Then, with $\tilde{T}(s,1) = (d_{\gamma(s)}\pi^{-1})T(s,1)$,

$$(d/ds)\rho_g^2(\gamma(s),m) = 2 < T(s,1), \dot{\gamma}(s) >= 2 < \tilde{T}(s,1), \dot{\tilde{\gamma}}(s) >, \tag{54}$$
$$(d^2/ds^2)\rho_g^2(\gamma(s),m) = 2 < D_s\tilde{T}(s,1), \dot{\tilde{\gamma}}(s) > .$$

The final inner products are in $T\mathbb{C}S^{k-1}$, namely, $\langle\tilde{v},\tilde{w}\rangle = Re(\tilde{v}\tilde{w}^*)$. Note that $\tilde{T}(s,1) = -Exp^{-1}_{\tilde{\gamma}(s)}q,\ q \in \pi^{-1}m$, may be expressed by (30) and (31) as

$$\tilde{T}(s,1) = -(\rho(s)/\sin\rho(s))[e^{i\theta(s)}q - (\cos\rho(s))\tilde{\gamma}(s)], \tag{55}$$

where $\rho(s) = \rho_g(\gamma(s),m)$ and $e^{i\theta(s)} = (1/\cos\rho(s))\tilde{\gamma}(s)q^*$. The covariant derivative $D_s\tilde{T}(s,1)$ is the projection of $(d/ds)\tilde{T}(s,1)$ onto $H_{\tilde{\gamma}(s)}$. Since $\langle\tilde{\mu},\dot{\tilde{\gamma}}(0)\rangle = 0$, (54) then yields

$$[(d^2/ds^2)\rho_g^2(\gamma(s),m)]_{s=0} = 2\langle[(d/ds)\tilde{T}(s,1)]_{s=0}, \dot{\tilde{\gamma}}(0)\rangle. \tag{56}$$

Differentiating (55) one obtains

$$\begin{aligned}[(d/ds)\tilde{T}(s,1)]_{s=0} &= [(d/ds)(\rho(s)\cos\rho(s))/\sin\rho(s))]_{s=0}\tilde{\mu} \\ &+ [(\rho(s)cos\rho(s))/sin\rho(s))]_{s=0}\dot{\tilde{\gamma}}(0) - [(d/ds)(\rho(s)/(\cos\rho(s))(\sin\rho(s))]_{s=0}(\tilde{\mu}q^*)q \\ &- [\rho(s)/(\cos\rho(s))(\sin\rho(s))]_{s=0}(\dot{\tilde{\gamma}}(0)q*)q.\end{aligned} \tag{57}$$

From (54), $2\rho(s)\dot{\rho}(s) = 2\langle\tilde{T}(s,1),\tilde{\gamma}'(s)\rangle$, which along with (55) leads to

$$[(d/ds)\rho(s)]_{s=0} = -(1/\sin r)\langle(\tilde{\mu}q^*/\cos r)q, \dot{\tilde{\gamma}}(0)\rangle,\ (r = \rho_g(m,\mu_I)). \tag{58}$$

One then gets (see BB [5,6])

$$\begin{aligned}&\langle[(d/ds)\tilde{T}(s,1)]_{s=0}, \dot{\tilde{\gamma}}(0)\rangle = \{(r\cos r)/(\sin r)\}|\dot{\tilde{\gamma}}(0)|^2 \\ &- \{(1/\sin^2 r) - (r\cos r)/\sin^3 r\}(Re(x))^2 + r/((\sin r)(\cos r))(Im(x))^2, \\ &(x = e^{i\theta}q\dot{\tilde{\gamma}}(0)^*, e^{i\theta} = \tilde{\mu}q^*/\cos r).\end{aligned} \tag{59}$$

One can check that the right side of (59) depends only on $\pi(\tilde{\mu})$ and not any particular choice of $\tilde{\mu}$ in $\pi^{-1}\{\mu_I\}$.

Now let $\{\nu_1,\cdots,\nu_{k-2}, i\nu_1,\cdots,i\nu_{k-2}\}$ be an orthonormal basis of $T_{\sigma(p)}\Sigma_2^k$ where we identify Σ_2^k with $\mathbb{C}P^{k-2}$, and choose the unit vectors $\nu_r = (\nu_r^1,\cdots,\nu_r^{k-1})$, $r = 1,\cdots,k-2$, to have zero imaginary parts and satisfy the conditions $p^*\nu_r = 0$, $\nu_r^t\nu_s = 0$ for $r \neq s$.

Suppose now that $\sigma(p) = \mu_I$, i.e., $\gamma(0) = \mu_I$. If $\dot{\gamma}(0) = v$, then $\gamma(s) = Exp_{\mu_I}(sv)$, so that $\rho_g^2(\gamma(s),m) = h(sv,y)$ with $y = Exp^{-1}_{\mu_I}m$. Then, expressing v in terms of the orthonormal basis,

$$[(d^2/ds^2)\rho_g^2(\gamma(s),m)]_{s=0} = [(d^2/ds^2)h(sv,y)]_{s=0} = \Sigma v_i v_j D_i D_j h(0,y). \tag{60}$$

Integrating with respect to Q now yields

$$\sum v_i v_j \Lambda_{ij} = E[(d^2/ds^2)\rho_g^2(\gamma(s), X)]_{s=0}, \quad (X \text{ with distribution } Q). \tag{61}$$

This identifies the matrix Λ from the calculations (56) and (59). To be specific, consider independent observations $X_1, \cdots, X_n$ from Q, and let $Y_j = Exp_{\mu_I}^{-1} X_j (j = 1, \cdots, n)$. In normal coordinates with respect to the above basis of $T_{\mu_I}\Sigma_2^k$, one has the following coordinates of Y_j:

$$(Re(Y_j^1), \cdots, Re(Y_j^{k-2}), Im(Y_j^1), \cdots, Im(Y_j^{k-2})) \in \mathbb{R}^{2k-4}. \tag{62}$$

Writing

$$\Lambda = \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{pmatrix}$$

in blocks of $(k-2) \times (k-2)$ matrices, one arrives at the following expressions of the elements of these matrices, using (59)–(62). Denote $\rho_g^2(\mu_I, X_1) = h(0, Y_1)$ by ρ. Then

$$\begin{aligned}(\Lambda_{11})_{rs} &= 2E[\rho(\cot\rho)\delta_{rs} - (1/\rho^2)(1-\rho\cot\rho)(ReY_1^r)(ReY_1^s) \\ &\quad + \rho^{-1}(\tan\rho)(ImY_1^r)(ImY_1^s)]; \\ (\Lambda_{22})_{rs} &= 2E[\rho(\cot\rho)\delta_{rs} - (1/\rho^2)(1-\rho\cot\rho)(ImY_1^r)(ImY_1^s) \\ &\quad + \rho^{-1}(\tan\rho)(ReY_1^r)(ReY_1^s)]; \\ (\Lambda_{12})_{rs} &= 2E[\rho(\cot\rho)\delta_{rs} - (1/\rho^2)(1-\rho\cot\rho)(ReY_1^r)(ImY_1^s) \\ &\quad + \rho^{-1}(\tan\rho)(ImY_1^r)(ReY_1^s)]; \\ (\Lambda_{21})_{rs} &= (\Lambda_{12})_{sr}.(r, s = 1, \cdots, k-2).\end{aligned} \tag{63}$$

One now arrives at the CLT for the intrinsic sample mean μ_n by Theorem 4.1, or the relation (20). A two-sample test for $H_0 : Q_1 = Q_2$, is then provided by (30).

We next turn to extrinsic analysis on Σ_2^k, using the embedding (34). Let μ^J be the mean of $Q \circ J^{-1}$ on $S(k-1, \mathbb{C})$. To compute the projection $P(\mu^J)$, let T be a unitary matrix, $T \in SU(k-1)$ such that $T\mu^J T^* = D = diag(\lambda_1, \cdots, \lambda_{k-1})$, $\lambda_1 \leq \cdots \leq \lambda_{k-2} \leq \lambda_{k-1}$. For $u \in \mathbb{C}S^{k-1}$, $u^*u \in J(\Sigma_2^k)$, write $v = Tu^*$. Then $Tu^*uT^* = vv^*$, and

$$\begin{aligned}||u^*u - \mu_J||^2 = ||vv^* - D||^2 &= \sum_{i,j} |v_i v_j - \lambda_j \delta_{ij}|^2 \\ &= \sum_j (|v_j|^2 + \lambda_j^2 - 2\lambda_j |v_j|^2) \\ &= \sum_j \lambda_j^2 + 1 - 2\sum_j \lambda_j |v_j|^2,\end{aligned} \tag{64}$$

which is minimized on $J(\Sigma_2^k)$ by $v = (v^1, \cdots, v^{k-1})$ for which $v^j = 0$ for $j = 1, \cdots, k-2$, and $|v^{k-1}| = 1$. That is, the minimizing u^* in (64) is a unit eigenvector of μ^J with the largest eigenvalue λ_{k-1}, and $P(\mu^J) = u^*u$. This projection is unique if and only if the largest eigenvalue of μ^J is simple, i.e., $\lambda_{k-2} < \lambda_{k-1}$.

Assuming that the largest eigenvalue of μ^J is simple, one may now obtain the asymptotic distribution of the sample extrinsic mean $\mu_{n,E}$, namely, that of $J(\mu_{n,E}) = v_n^* v_n$, where v_n is a unit eigenvector of $\bar{\tilde{X}} = \sum \tilde{X}_j/n$ corresponding to its largest eigenvalue. Here $\tilde{X}_j = J(X_j)$, for i.i.d observations $X_1, \cdots, X_n$ on Σ_2^k. For this purpose, a convenient orthonormal basis (frame) of $T_p S(k-1, \mathbb{C}) \approx S(k-1, \mathbb{C})$ is the following:

$$v_{a,b} = 2^{-1/2}(e_a e_b^t + e_b e_a^t) \;\text{ for } a < b, v_{a,a} = e_a e_a^t; \tag{65}$$

$$w_{a,b} = i2^{-1/2}(e_a e_b^t - e_b e_a^t) \;\text{ for } b < a \; (a, b = 1, \cdots, k-1),$$

where e_a is the column vector with all entries zero other than the a-th, and the a-th entry is 1. Let $U_1, \cdots, U_{k-1}$ be orthonormal unit eigenvectors corresponding to the eigenvalues $\lambda_1 \leq \cdots \leq \lambda_{k-2} < \lambda_{k-1}$. Then choosing $T = (U_1, \cdots, U_{k-1}) \in SU(k-1)$ $T\mu^J T^* = D = diag(\lambda_1, \cdots, \lambda_{k-1})$, such that the columns of $Tv_{a,b}T^*$ and $Tw_{a,b}T^*$ together constitute an orthonormal basis of $S(k-1, \mathbb{C})$. It is not difficult to check that the differential of the projection operator P satisfies

$$(d_{\mu^J} P)Tv_{a,b}T^* = \begin{cases} 0 & \text{if } 1 \leq a \leq b < k-1, \;\text{ or } a = b = k-1, \\ (\lambda_{k-1} - \lambda_a)^{-1} T v_{a,k-1} T^* & \text{if } 1 \leq a < k-1, b = k-1; \end{cases} \tag{66}$$

$$(d_{\mu^J} P)Tw_{a,b}T^* = \begin{cases} 0 & \text{if } 1 \leq a \leq b < k-1, \\ (\lambda_{k-1} - \lambda_a)^{-1} T w_{a,k-1} T^* & \text{if } 1 \leq a < k-1. \end{cases}$$

To check these, take the projection of a linear curve $c(s)$ in $S(k-1, \mathbb{C})$ such that $\dot{c}(0)$ is one of the basis elements $v_{a,b}$, or $w_{a,b}$, and differentiate the projected curve with respect to s. It follows that $\{Tv_{a,k-1}T^*, Tw_{a,k-1}T^* : a = 1, \cdots, k-2\}$ form an orthonotmal basis of $T_{P(\mu^J)} J(\Sigma_2^k)$. Expressing $\tilde{X}_j - \mu^J$ in the orthonormal basis of $S(k-1, \mathbb{C})$, and $d_{\mu^J} P(\tilde{X}_j - \mu^J)$ with respect to the above basis of $T_{P(\mu^J)} J(\Sigma_2^k)$, one may now apply Theorem 3.3.

For a two-sample test for $H_0 : Q_1 = Q_2$, one may use (22), as explained in Sect. 5.

8 Nonparameric Bayes for Density Estimation and Classification on a Manifold

8.1 Density Estimation

Consider the problem of estimating the density q of a distribution Q on a Riemannian manifold (M, g) with respect to the volume measure λ on M. According to Ferguson [22], given a finite non-zero base measure α on a measurable space $(\mathcal{X}, \Sigma)$, a random probability P on the class $\mathcal{P}$ of all probability measures on $\mathcal{X}$ has the Dirichlet distribution D_α if for every measurable partition $\{B_1, \ldots, B_k\}$ of $\mathcal{X}$, the D_α - distribution of $(P(B_1), \ldots, P(B_k)) = (\theta_1, \ldots, \theta_k)$, say, is Dirichlet with parameters $(\alpha(B_1), \ldots, \alpha(B_k))$. Sethuraman [44] gave a very convenient "stick breaking" representation of the random P. To define it, let $u_j(j = 1, \ldots)$ be an i.i.d. sequence of $beta(1, \alpha(\mathcal{X}))$ random variables, independent of a sequence $Y_j(j = 1, \ldots)$ having the distribution $G = \frac{\alpha}{\alpha(\mathcal{X})}$ on $\mathcal{X}$. Sethuraman's representation of the random probability with the Dirichlet prior distribution D_α is

$$P \equiv \sum w_j \delta_{Y_j}, \tag{67}$$

where $w_1 = u_1, w_j = u_j(1 - u_1)\ldots(1 - u_{j-1})(j = 2, \ldots)$, and δ_{Y_j} denotes the Dirac measure at Y_j. As this construction shows, the Dirichlet distribution assigns probability one to the set of all discrete distributions on $\mathcal{X}$, and one cannot retrieve a density estimate from it directly. The Dirichlet priors constitute a conjugate family, i.e., the posterior distribution of a random P with distribution D_α, given observations $X_1, \ldots, X_n$ from P is $D_{\alpha+\sum_{1\le i\le n}\delta_{X_i}}$. A general method for Bayesian density estimation on a manifold (M, g) may be outlined as follows. Suppose that q is continuous and positive on M. First find a parametric family of densities $m \to K(m; \mu, \tau)$ on M where $\mu \in M$ and $\tau > 0$ are "location" and "scale" parameters, such that K is continuous in its arguments, $K(\cdot; \mu, \tau)d\lambda(\cdot)$ converges to δ_μ as $\tau \downarrow 0$, and the set of all "mixtures" of $K(\cdot; \mu, \tau)$ by distributions on $M \times (0, \infty)$ is dense in the set $C_\lambda(M)$ of all continuous densities on M in the supremum distance, or in $L^1(d\lambda)$. The density q may then be estimated by a suitable mixture. To estimate the mixture, use a prior D_β with full support on the set of all probabilities on the space $M \times (0, \infty)$ of "parameters" (μ, τ). A draw from the prior may be expressed in the form (67), where u_j are i.i.d. $beta(1, b)$ with $b = \beta(M \times (0, \infty))$, independent of $Y_j = (m_j, t_j)$, say, which are i.i.d. $\frac{\beta}{b}$ on $M \times (0, \infty)$. The corresponding random density is then obtained by integrating the kernel K with respect to this random mixture distribution,

$$\sum w_j K(m; m_j, t_j). \tag{68}$$

Given M-valued (Q-distributed) observations $X_1, \ldots X_n$, the posterior distribution of the mixture measure is Dirichlet D_{β_X}, where $\beta_X = \beta + \sum_{1 \leq i \leq n} \delta_{Z_i}$, with $Z_i = (X_i, 0)$. A draw from the posterior distribution leads to the random density in the form (68), where u_j are i.i.d. $beta(1, b+n)$, independent of (m_j, t_j) which are i.i.d. $\frac{\beta_X}{(b+n)}$. One may also consider using a somewhat different type of priors such as $D_\alpha \times \pi$ where D_α is a Dirichlet prior on M, and π is a prior on $(0, \infty)$, e.g., gamma or Weibull distribution.

Consistency of the posterior is generally established by checking full Kullback-Liebler support of the prior D_β (see [25], pp. 137–139). Strong consistency has been established for the planar shape spaces using the complex Watson family of densities (with respect to the volume measure or the uniform distribution on Σ_2^k) of the form $K([z]; \mu, \tau) = c(\tau) exp \frac{|z * \mu|^2}{\tau}$ in [6, 7], where it has been shown, by simulation from known distributions, that, based on a prior $D_\beta \times \pi$ chosen so as to produce clusters close to the support of the observations, the Bayes estimates of quantiles and other indices far outperform the kernel density estimates of Pelletier [42], and also require much less computational time than the latter. In moderate sample sizes, the nonparametric Bayes estimates perform much better than even the MLE (computed under the true model specification)!

8.2 Classification

Classification of a random observation to one of several groups is one of the most important problems in statistics. This is the objective in medical diagnostics, classification of subspecies and, more generally, this is the target of most image analysis. Suppose there are r groups or populations with a priori given relative sizes or proportions $\pi_i (i = 1, \ldots, r), \sum \pi_i = 1$, and densities $q_i(x)$ (with respect to some sigma-finite measure). Under $0 - 1$ loss function, the average risk of misclassification (i.e., the Bayes risk) is minimized by the rule: Given a random observation X, classify it to belong to group j if

$\pi_j q_j(X) = max\{\pi_i q_i(X) : i = 1, \ldots, r\}$. Generally, one uses sample estimates of π_i-s and q_i-s, based on random samples from the r groups (training data). Nonparametric Bayes estimates of q_i-s on shapes spaces perform very well in classification of shapes, and occasionally identify outliers and misclassified observations (see, [6, 7]).

9 Examples

In this section we apply the theory to a number of data sets available in the literature.

Example 9.1 (Paleomagnetism). The first statistical confirmation of the shifting of the earth's magnetic poles over geological times, theorized by paleontologists based on observed fossilised magnetic rock samples, came in a seminal paper by R.A. Fisher [23]. Fisher analyzed two sets of data—one recent (1947–1948) and another old (Quaternary period), using the so-called *von Mises-Fisher model*

$$f(x;\mu,\tau) = c(\tau)\exp\{\tau x^t\mu\}(x \in S^2), \tag{69}$$

Here $\mu(\in S^2)$, is the *mean direction*, extrinsic as well as intrinsic ($\mu = \mu_I = \mu_E$), and $\tau > 0$ is the concentration parameter. The maximum likelihood estimate of μ is $\hat{\mu} = \bar{X}/|\bar{X}|$, which is also our sample extrinsic mean. The value of the MLE for the first data set of $n = 9$ observations turned out to be $\hat{\mu} = \hat{\mu}_E = (.2984, .1346, .9449)$, where (0,0,1) is the geographic north pole. Fisher's 95% confidence region for μ is $\{\mu \in S^2 : \rho_g(\hat{\mu},\mu) \leq 0.1536)\}$. The sample intrinsic mean is $\hat{\mu}_I = (.2990, .1349, .9447)$, which is very close to $\hat{\mu}_E$.The nonparametric confidence region based on $\hat{\mu}_I$, as given by (50), and that based on the extrinsic procedure (53), are nearly the same, and both are about 10% smaller in area than Fisher's region. (See [6], Chap. 2.)

The second data set based on $n = 29$ observations from the Quaternary period that Fisher analyzed, using the same parametric model as above, had the MLE $\hat{\mu} = \bar{X}/|\bar{X}| = (.0172, -.2978, -.9545)$, almost antipodal of that for the first data set, and with a confidence region of geodesic radius .1475 around the MLE. Note that the two confidence regions are not only disjoint, they also lie far away from each other. This provided the first statistical confirmation of the hypothesis of shifts in the earth's magnetic poles, a result hailed by paleontologists (see [30]). Because of difficulty in accessing the second data set, the nonparametric procedures could not be applied to it. But the analysis of another data set dating from the Jurassic period, with $n = 33$, once again yielded nonparametric intrinsic and extrinsic confidence regions very close to each other, and each about 10 % smaller than the region obtained by Fisher's parametric method (see [6], Chap. 5, for details).

Example 9.2 (Brain scan of schizophrenic and normal patients). We consider an example from Bookstein [15] in which 13 landmarks were recorded on a midsagittal two-dimensional slice from magnetic brain scans of each of 14 schizophrenic patients and 14 normal patients. The object is to detect the deformation, if any, in the shape of the k-ad due to the disease, and to use it for diagnostic purposes. The shape space is Σ_2^{13}. The intrinsic two-sample test (22) has an observed value 95.4587 of the asymptotic chisquare statistic with 22 degrees of freedom, and a p-value 3.97×10^{-11}. The extrinsic test based on (24) has an observed value 95.5476 of the chisquare statistic and a p-value 3.8×10^{-11}. The calculations made use of the analytical computations carried out in Example 7.2. It is remarkable, and reassuring, that completely different methodologies of intrinsic and extrinsic inference essentially led to the same values of the corresponding asymptotic chisquare statistics (a phenomenon observed in other examples as well). For details of these calculations and others we refer to [6]. This may also be contrasted with the results of parametric inference in the literature for the same data, as may be found in [19], pp. 146, 162–165. Using a isotropic Normal model for the original landmarks data, and after removal of "nuisance" parameters for translation, size and rotation, an F-test known as Goodall's F-test (see [26]) gives a p-value .01. A Monte Carlo test based permutation test obtained by 999 random assignments of the data into two groups and computing Goodall's F-statistic, gave a p-value .04. A Hotelling's T^2 test in the tangent space of the pooled sample mean had a p-value .834. A likelihood ratio test based on the isotropic offset Normal distribution on the shape space has the value 43.124 of the chisquare statistic with 22 degrees of freedom, and a p-value .005.

Example 9.3 (Glaucoma detection- a match pair problem in 3D). Our final example is on the $3D$ reflection similarity shape space $R\Sigma_3^k$. To detect shape changes due to glaucoma, data were collected on twelve mature rhesus monkeys.

One of the eyes of each monkey was treated with a chemical agent to temporarily increase the intraocular pressure (IOP). The increase in IOP is known to be a cause of glaucoma. The other eye was left untreated. Measurements were made of five landmarks in each eye, suggested by medical professionals. The data may be found in [11]. The match pair test based on (25) yielded an observed value 36.29 of the asymptotic chisquare statistic with degrees of freedom 8. The corresponding p-value is 1.55×10^{-5} (see [6], Chap. 9). This provides a strong justification for using shape change of the inner eye as a diagnostic tool to detect the onset of glaucoma. An earlier computation using a different nonparametric procedure in [11] provided a p-value .058. Also see [9] where a 95 % confidence region is obtained for the difference between the extrinsic size-and-shape reelection shapes between the treated and untreated eyes.

Acknowledgements The author wishes to thank the referee for helpful suggestions. This research is supported by NSF grant DMS 1107053.

References

1. B. Afsari, Riemannian L^p center of mass: Existence, uniqueness, and convexity. Proc. Am. Math. Soc. **139**, 655–673 (2011)
2. A. Bhattacharya, Nonparametric statistics on manifolds with applications to shape spaces. Ph.D. Thesis, University of Arizona (2008)
3. A. Bhattacharya, Statistical analysis on manifolds: a nonparametric approach for inference on shape spaces. Sankhya, **70**, 43 (2008)
4. A. Bhattacharya, R.N. Bhattacharya, Nonparametric statistics on manifolds with application to shape spaces. Pushing the Limits of Contemporary Statistics: Contributions in honor of J.K. Ghosh. IMS Collections **3**, 282–301 (2008)
5. A. Bhattacharya, R.N. Bhattacharya, Statistics on Riemannian manifolds: asymptotic distribution and curvature. Proc. Am. Math. Soc. **136**, 2957–2967 (2008)
6. A. Bhattacharya, R.N. Bhattacharya, *Nonparametric Inference on Manifolds with Applications to Shape Spaces.* IMS Monograph, No. 2 (Cambridge University Press, Cambridge, 2012)
7. A. Bhattacharya, D. Dunson, Nonparametric Bayesian density estimation on manifolds with applications to planar shapes. Biometrika **97**, 851–865 (2010)

8. A. Bandulasiri, V. Patrangenaru, Algorithms for nonparametric inference on shape manifolds, in *Proceedings of JSM 2005, MN* (2005), pp. 1617–1622
9. A. Bandulasiri, R.N. Bhattacharya, V. Patrangenaru, Nonparametric inference on shape manifolds with applications in medical imaging. J. Multivariate Anal. **100**, 1867–1882 (2009)
10. R.N. Bhattacharya, V. Patrangenaru, Large sample theory of intrinsic and extrinsic sample means on manifolds. Ann. Stat. **31**, 1–29 (2003)
11. R.N. Bhattacharya, V. Patrangenaru, Large sample theory of intrinsic and extrinsic sample means on manifolds-II. Ann. Stat. **33**, 1225–1259 (2005)
12. R.N. Bhattacharya, J. Ramirez, B. Polletta, Manifold structure of the projective shape space. (Unpublished, 2009)
13. F. Bookstein, *The Measurement of Biological Shape and Shape Change*. Lecture Notes in Biomathematics (Springer, Berlin, 1978)
14. F.L. Bookstein, Size and shape spaces of landmark data (with discussion). Stat. Sci. **1**, 181–242 (1986)
15. F.L. Bookstein, *Morphometric Tools for Landmark Data: Geometry and Biology* (Cambridge University Press, Cambridge, 1991)
16. W.M. Boothby, *An Introduction to Differentiable Manifolds and Riemannian Geometry*, 2nd edn. (Academic, New York, 1986)
17. M. Do Carmo, *Riemannian Geometry* (Birkhäuser, Boston, 1992)
18. I.L. Dryden, K.V. Mardia, Size and shape analysis of landmark data. Biometrica **79**, 57–68 (1992)
19. I.L. Dryden, K.V. Mardia, *Statistical Shape Analysis* (Wiley, New York, 1998)
20. I.L. Dryden, A. Kume, H. Le, A.T.A. Wood, The MDS model for shape: an alternative approach. *Biometrika* **95**(4), 779–798 (2008)
21. L. Ellingson, F.H. Ruymgaart, V. Patrangenaru, Nonparametric estimation for extrinsic mean shapes of planar contours. (to appear, 2013)
22. T. Ferguson, A Bayesian analysis of some nonparametric problems. Ann. Stat. **1**, 209–230 (1973)
23. R.A. Fisher, Dispersion on a sphere. Proc. R. Soc. Lond. Ser. A **217**, 295–305 (1953)
24. S. Gallot, D. Hulin, J. Lafontaine, *Riemannian Geometry*. Universitext (Springer, Berlin, 1990)
25. J.K. Ghosh, R.V. Ramamoorthi, *Bayesian Nonparametrics* (Springer, New York, 2003)
26. C.R. Goodall, Procrustes methods in the statistical analysis of shape (with discussion). J. R. Stat. Soc. Ser B. **53**, 285–339 (1991)
27. H. Hendriks, Z. Landsman, Asymptotic tests for mean location on manifolds. C.R. Acad. Sci. Paris Sr. I Math. **322**, 773–778 (1996)
28. H. Hendriks, Z. Landsman, Mean location and sample mean location on manifolds: asymptotics, tests, confidence regions. J. Multivariate Anal. **67**, 227–243 (1998)
29. S. Huckemann, T. Hotz, A. Munk, Intrinsic shape analysis: geodesic PCA for Riemannian manifolds modulo isometric Lie group actions (with discussions). Stat. Sinica **20**, 1–100 (2010)
30. E. Irving, *Paleomagnetism and Its Application to Geological and Geographical Problems* (Wiley, New York, 1964)
31. H. Karcher, Riemannian center of mass and mollifier smoothing. Comm. Pure Appl. Math. **30**, 509–554 (1977)
32. D.G. Kendall, The diffusion of shape. Adv. Appl. Probab. **9**, 428–430 (1977)
33. D.G. Kendall, Shape manifolds, procrustean metrics, and complex projective spaces. Bull. Lond. Math. Soc. **16**, 81–121 (1984)
34. D.G. Kendall, A survey of the statistical theory of shape. Stat. Sci. **4**, 87–120 (1989)
35. W.S. Kendall, Probability, convexity, and harmonic maps with small image I: uniqueness and fine existence. Proc. Lond. Math. Soc **61**, 371–406 (1990)
36. J.T. Kent, New directions in shape analysis. in *The Art of Statistical Science*, ed. by K.V. Mardia. pp. 115–127 (Wiley, New York, 1992)
37. J.T. Kent, The complex Bingham distribution and shape analysis. J. Roy. Stat. Soc. Ser. B **56**, 285–299 (1994)

38. H. Le, Locating Fréchet means with application to shape spaces. Adv. Appl. Prob. **33**, 324–338 (2001)
39. K.V. Mardia, V. Patrangenaru, Directions and projective shapes. Ann. Stat. **33**, 1666–1699 (2005)
40. V. Patrangenaru, Asymptotic statistics on manifolds and their applications. Ph.D. Thesis, Indiana University, Bloomington (1998)
41. V. Patrangenaru, X. Liu, S. Sugathadasa, Nonparametric 3D projective shape estimation from pairs of 2D images-I, in memory of W.P.Dayawansa. J. Multivariate Anal. **101**, 11–31 (2010)
42. B. Pelletier, Kernel density estimation on Riemannian manifolds. Stat. Probab. Lett. **73**(3), 297–304 (2005)
43. M.J. Prentice, K.V. Mardia, Shape changes in the plane for landmark data. Ann. Stat. **23**, 1960–1974 (1995)
44. J. Sethuraman, A constructive definition of Dirichlet priors. Stat. Sinica **4**, 639–650 (1994)
45. G. Sparr, Depth computations from polyhedral images, in *Proceedings of 2nd European Conference on Computer Vision*, ed. by G. Sandini (Springer, New York, 1994), pp. 378–386
46. H. Ziezold, On expected figures and a strong law of large numbers for random elements in quasi-metric spaces, in *Transactions of the Seventh Prague Conference on Information Theory, Statistical Functions, Random Processes and of the Eighth European Meeting of Statisticians*, vol. A. (Tech. Univ. Prague, Prague (1974), pp. 591–602, Reidel, Dordrecht, 1977