Abstract

This article provides an exposition of some recent developments in nonparametric inference on manifolds, along with a brief account of an emerging theory on data analysis on stratified spaces. Much of the theory is developed around the notion of Fréchet means with applications, for the most part, to landmark based shape spaces. A number of applications are illustrated with real data in such areas as paleomagnetism, morphometrics and medical diagnostics. Connections to scene recognition and machine vision are also explored.
1 Introduction

Manifolds are metric spaces that look locally, but not globally, like a numerical space $\mathbb{R}^p$. Statistics on simple manifolds such as the unit spheres $S^d$ in $\mathbb{R}^{d+1}$ has a long history which, according to Watson (1983), goes all the way back to a paper by D. Bernoulli (1734), where a test for uniformity of distributions of planetary orbits in the Solar System was first considered. A test by von Mises (1918) for the uniformity of a distribution on the unit circle using a parametric model led to the rejection of the hypothesis that the "fractional parts" $x - [x]$ of atomic weights $x$ of atoms of elements are uniformly distributed. In a seminal paper, R.A. Fisher (1953) used what is now called the von Mises-Fisher distribution on $S^2$ to convincingly demonstrate for the first time that the Earth’s magnetic poles had shifted over geological times. This work inspired a great deal of activity on what is called directional statistics on spheres and axial spaces. The book by Watson (1983) contains a wealth of results on the asymptotics of the center of mass and other statistics based on samples from a class of distributions on the sphere including the von Mises-Fisher distribution, or the Langevin model as Watson calls it. N. I. Fisher et al. (1987) and Mardia and Jupp (2000) may be consulted for a comprehensive treatment of the theory on directional statistics and diverse applications, as well as extensive references to earlier work.

The present article develops nonparametric inference on arbitrary manifolds, with main applications to shape spaces. The systematic study of the geometry of direct similarity shape spaces was initiated by D.G. Kendall (1977), (1984), (1989), which we will review in Section 3. Around the same time, in a different vein, Bookstein (1978),(1986), (1991) developed inference procedures for distinguishing shape distributions with applications in morphometrics and medical imaging. Kendall’s notions of direct similarity shapes were used for much of the development of inference on these spaces, especially for planar images, by many researchers including J.T. Kent, K.V. Mardia, I.L. Dryden, C. R. Goodall, M.J. Prentice and others. A comprehensive account of this largely parametric theory based around the notion of certain types of mean similarity shapes may be found in the monograph by Dryden and Mardia (1998). Also see Kent(1992),(1994). As manifolds, Kendall’s direct similarity planar shape spaces, that are at the center of this monograph, are complex projective spaces $\mathbb{C}P^{k-2}$, where $k$ is the number
of points of the 2D configurations whose shapes are considered. These are quotient spaces of the action of the Lie group of direct similarities on $\mathbb{C}^k \setminus \Delta$, where $\Delta$ is the diagonal of $\mathbb{C}^k$.

Turning to nonparametric inference on more general manifolds, we need some basic definitions borrowed from differential geometry. A Lie group is a differentiable manifold $K$ which has in addition an algebraic structure of a group, for which the group multiplication $\circ : K \times K \to K$, and the operation of taking the inverse $k \to k^{-1}$ are differentiable functions between manifolds. Assume $K$ is a Lie group with unit $1_K$, and $M$ is a manifold. Consider a differentiable function $\alpha : K \times M \to M$, and for each element $k \in K$, define the function $\alpha_k : M \to M$, given by

$$\alpha_k(p) = k \cdot p = \alpha((k, p)).$$

The function $\alpha$ is a left action on $M$ if

$$k \cdot (h \cdot p) = (k \circ h) \cdot p, \forall k, h \in K, \forall p \in M, \quad 1_K \cdot p = p, \forall p \in M.$$  

A homogeneous space is a manifold $M$ that admits a left action $\alpha$ by a Lie group $K$ that is transitive on $M$, that is, for any pair $p_1, p_2$ of points in $M$, there is a group element $k \in K$, such that $\alpha_k(p_1) = p_2$. A simple example of a homogeneous space is the sphere $S^2$, with the transitive action given by the group of rotations $SO(3)$. Beran (1968), (1969) developed nonparametric tests for uniformity of a distribution on a large class of homogeneous spaces, and applied it to the circle. Giné (1975) derived tests for uniformity on general compact Riemannian manifolds. Hendriks (1990) considered the orthonormal basis of eigenfunctions of the Laplace-Beltrami operator for an expansion of the density of a distribution on an arbitrary compact or homogeneous Riemannian manifold, and applied this to estimate the density using i.i.d. observations from the distribution. Healy et al. (1998) developed a method of density estimation on the sphere $S^2$, by considering observations from a distribution on $S^2$ as those obtained by perturbations of the true distribution with i.i.d. rotations, and using deconvolution by Fourier analysis on $S^2$ and $SO(3)$ to estimate the true distribution/density. Although this is an interesting and significant method, and has the potential for extension to more general contexts, it may be noted that not all i.i.d. observations
from a density on $S^2$ may be viewed in this way. In Sections 3 and 6 we will mention some more recent work on the topic of density estimation on manifolds. References to other earlier work will appear later as we develop the main themes of the present article.

In the rest of the introduction we provide an expository outline of this paper. In a lighter vein our presentation may be labeled “In praise of the mean”. For the purpose of discrimination among distributions $Q$ on a metric space $(S, \rho)$ - a problem which arises in many contexts - one may compare the means $\mu = \mu(Q)$ of these distributions defined as the minimizers of the Fréchet functions $F(x)$, which are expected squared distances from a variable point on $S$ for these distributions:

$$\mu \in \text{argmin}_{x \in S} F(x), \text{ with } F(x) = \int_S \rho^2(x, y) Q(dy).$$  \hspace{1cm} (1.3) \hfill \\

While this notion coincides with the usual definition of the mean $\mu = \int yQ(dy)$ of $Q$ on a vector space $S$, its extension to general non-linear spaces has far reaching consequences. The first formal definition of such a mean on a metric space is due to Fréchet (1948). As usual, the probability measures $Q$ considered in the article are defined on the Borel sigma-field of $S$. We will often refer to $\mu$ in (1.3) as the Fréchet mean of $Q$, if the minimizer is unique. In general, the set of minimizers will be called the Fréchet mean set of $Q$. When applied to the empirical distribution $Q_n = (1/n) \sum_{i=1}^n \delta_{X_i}$ of $n$ independent observations $X_i$ each with distribution $Q$, defined on a probability space $(\Omega, \mathcal{F}, P)$, the corresponding quantities are referred to as the Fréchet sample mean $\mu_n$ and the Fréchet sample mean set. Much of this article is devoted to the use of the Fréchet sample mean $\mu_n$ as an estimator of the Fréchet mean $\mu$ of $Q$, and for the two-sample problem of discrimination between two distributions. These seemingly simple tasks, however, pose a number of challenges, varying with the structure of the topological space $S$.

First, what distance $\rho$ to choose on $S$? Second, given an appropriate distance $\rho$, what conditions on $Q$ guarantee a unique minimizer? Without uniqueness the inference based on the Fréchet sample mean can hardly proceed; under reasonable conditions the uniqueness of the minimizer implies consistency of $\mu_n$. Next comes the problem of deriving the asymptotic distribution of $\mu_n$ which lies at the heart of statistical inference for $\mu$. For the most part, in this article our emphasis is on a $d$-dimensional differentiable manifold $S = M$, or
more generally, on a *stratified space*. In particular, we consider in some de-
tail the direct similarity shape spaces introduced by Kendall (1977), (1984),
a more suitable analog of these for analyzing 3D images, and affine and pro-
jective shape spaces appropriate for scene recognition and machine vision.
These spaces and their geometries are described in detail in Sections 3 - 4.
Given a Riemannian structure $g$ on $M$ the distance $\rho$ may be taken to be
the geodesic distance $\rho_g$ derived from this metric tensor. The correspond-
ing Fréchet mean, when it exists, is called the *intrinsic mean* ( under $g$),
and is labeled $\mu_I$. There is no known broad criterion for the uniqueness of
the minimizer in (1.3) under such a distance that would satisfy the needs
of nonparametric statistical inference. Basically, except for a very special
result of Le (see Kendall et. al. (1999), p. 211 ) on the complex projective
space with the Fubini-Study metric requiring that the density with respect
to the volume measure depends on the geodesic distance from some point
and decrease monotonically with this distance, uniqueness has been derived
only for distributions with support contained in a rather small geodesic ball.
The first significant result under such a support condition is due to Karcher
(1977), which was followed by extensions due to Le (2001), Kendall (1990)
, Groisser (2004) and Afsari (2011)). Although one expects that uniqueness
holds for “most” distributions $Q$ in some sense, the theoretical impediment
in the use of the intrinsic sample mean for inference cannot be ignored. Even
assuming uniqueness, the task of computing the intrinsic sample mean and
one- and two-sample statistics for intrinsic analysis is rather time con-
suming. To have some idea of the difficulties in finding readily verifiable conditions for
uniqueness of the intrinsic mean, one may look at the simplest case, namely
the circle $S^1$, for which a necessary and sufficient condition for uniqueness of
the intrinsic mean of an absolutely continuous $Q$ is presented in Bhattacharya
and Bhattacharya (2012), Chapter 5. Also see Huckemann and Hotz (2012))
for some further results for $S^1$. As an alternative to a geodesic distance one
may consider a distance $\rho = \rho_J$ induced by an embedding $J$ of $M$ in a Eu-
clidean space $E^N$. We will refer to this distance as the *extrinsic distance*
(under $J$). Consider now the metric space $J(M)$ with the Euclidean distance
and the image probability measure $QoJ^{-1}$ induced by $J$ on it. The Fréchet
mean of this probability on $J(M)$ is easily shown to be the *projection $P(\mu^J)$
on it, if unique, of the mean of $QoJ^{-1}$ in $E^N$, where $QoJ^{-1}$ is now viewed as
a probability on $E^N$. Then the Fréchet mean on $M$ is $\mu_J = J^{-1}(P(\mu^J))$. We
will refer to this mean as the *extrinsic mean* of \( Q \) under the embedding \( J \). When the embedding \( J \) is clearly specified by the context, the extrinsic mean will be simply labeled as \( \mu_\mathcal{E} \), without specifying \( J \). The Fréchet minimizer (1.3) under the extrinsic distance is unique if and only if the projection \( P(\mu^J) \) is unique. This provides an explicit and precise criterion for the uniqueness of the Fréchet minimizer and also a method for computing it. As an example, if \( M = S^d = \{ x \in \mathbb{R}^{d+1} : \|x\| = 1 \} \), and \( J \) is the inclusion map, namely, \( J(x) = x \), then \( P(x) = \frac{x}{\|x\|} \) is unique for all \( x \neq 0 \). The only point in \( \mathbb{R}^{d+1} \) whose projection is not unique is \( x = 0 \), and the set of minimizers in (1.3) in the case when \( \mu^J = 0 \) is the entire sphere \( S^d \). This is in contrast with the situation involving the intrinsic mean. Even on \( S^d \) with the Riemannian structure on the tangent space of \( S^d \) induced by the Euclidean scalar product there is no known broad criterion for uniqueness of the Fréchet minimizer under the geodesic distance, aside from the very restrictive support assumptions mentioned above. The general notion of the extrinsic mean on an arbitrary differentiable manifold \( M \) and the generally verifiable criterion for uniqueness mentioned was given in Patrangenaru (1998), independently of some results of Hendriks and Landsman (1996), (1998), who considered a similar notion for submanifolds \( M \) of a Euclidean space \( E^N \). The embedding chosen by the last named authors was the inclusion map. The general theory of extrinsic and intrinsic means on manifolds for estimation and two-sample tests is further developed in Bhattacharya and Patrangenaru (2003), (2005), Crane and Patrangenaru (2011) and Bhattacharya and Bhattacharya (2012).

Continuing on the subject of the proper choice of a distance on a differentiable manifold \( M \), one may note that there are in general many possible Riemannian structures, as well as many possible embeddings. It is not clear which one is best for statistical inference based on the intrinsic or the extrinsic sample mean. For intrinsic means, one criterion consists in selecting a Riemannian structure with the largest group of isometries of the manifold. For extrinsic means, one good criterion is based on the notion of an equivariant embedding \( J \) under the action \( \alpha \) of a Lie group \( \mathcal{K} \) on \( M \). The embedding \( J \) is *equivariant* if there exists a Lie group action \( A : \mathcal{K} \times E^N \), such that \( J(\alpha(k,x)) = A(k,J(x)), \forall k \in \mathcal{K}, \forall x \in M \). Larger the group \( \mathcal{K} \), greater is this preservation of geometry. In the special case when \( M \) has a Riemannian metric, the group \( \mathcal{K} \) is often a group of isometries. One may then seek \( A \) such
that \( \{A_k, k \in \mathcal{K}\} \) is also a group of isometries, when \( J(M) \) has the induced Riemannian structure from \( E^N \). This turns out to be the case with a number of applications of importance in this article. A simple example is provided by the inclusion map \( J \) on \( S^d \) as mentioned above. The group \( \mathcal{K} \) here is the special orthogonal group \( SO(d+1) \), and \( A \) is the restriction of the action of \( GL(d+1, \mathbb{R}) \) on \( \mathbb{R}^{d+1} \) to \( SO(d+1) \times \mathbb{R}^{d+1} \). It may be noted, however, that the extrinsic distance is not the same as the (intrinsic) distance under the embedding. For \( S^d \), the extrinsic distance is the chord distance, whereas the geodesic distance under the induced Riemannian structure from \( E^{d+1} \) is the so-called arc distance, namely the shortest arc length between two points on \( S^d \) measured along a great circle passing through them.

We have indicated some analytical as well as computational advantages of the extrinsic mean over the intrinsic mean. It will turn out, in several data examples that we consider in this article, that the intrinsic sample mean and the extrinsic sample mean are very close to each other and the nonparametric tests based on them yield nearly identical values. However, as we shall see in the next section devoted to the asymptotic distribution theory of Intrinsic and extrinsic sample means, the methodologies for the intrinsic and extrinsic inferences are quite different. The preceding paragraph points to an explanation of this welcome convergence of intrinsic and extrinsic analysis.

Here is an outline of the contents of the rest of this paper. In Section 2, we give some general asymptotic results for Fréchet means on manifolds, along with an application to paleomagnetism. Sections 3 and 4 are dedicated to the application of the general results in section 2 to planar similarity shapes of \( k \)-ads, and 3D reflection similarity shapes. Section 5 includes applications of the general results in Section 2 to projective shape spaces and to the analysis of 3D scenes from digital camera images. The last section focuses on a related new direction of research on data analysis on sample spaces with a manifold stratification.

In the rest of the article we will abbreviate Bhattacharya and Patrangenaru as \( BP \) and Bhattacharya and Bhattacharya as \( BB \).
2 Large sample properties of Frechet Means

The first significant results on the almost sure convergence properties of Frechet mean sets on general metric spaces appear to have been derived by Ziezold (1977). In BP (2003) these results were strengthened under the following topological assumption on the metric space $(S, \rho)$.

Closed bounded subsets of $S$ are compact. \hspace{1cm} (2.1)

It may be noted that all compact metric spaces satisfy (2.1). In addition, according to the Hopf-Rinow Theorem (Do Carmo (1992, pp. 146,147)) all Riemannian manifolds which are complete in the geodesic distance satisfy (2.1). We will denote by $C_Q$ and $C_{Q_n}$ the Frechet mean sets of $Q$ and $Q_n$, respectively. For $C \subset S$ and $\varepsilon > 0$, define $C^{\varepsilon} = \{ x \in S : \rho(x, C) < \varepsilon \}$.

**Theorem 2.1.** (Ziezold (1977), BP (2003)). Assume (2.1) and finiteness of the Frechet function $F$ in (1.3). Then (a) $C_Q$ is nonempty and compact, and (b) for each $\varepsilon > 0$ there exists an integer $N_\varepsilon = N_\varepsilon(\omega)$ and a $P$-null set $F_\varepsilon$ such that $C_{Q_n} \subset (C_Q)^{\varepsilon}$, $\forall n \geq N_\varepsilon$ and $\forall \omega \in \Omega \setminus F_\varepsilon$.

In view of (b), in the case $C_Q$ is a singleton, we will take the Frechet sample mean to be any measurable selection from $C_{Q_n}$, and denote the Frechet means of $Q$ and $Q_n$ as $\mu$ and $\mu_n$, respectively. We will from now on say that a Frechet mean $\mu$ of $Q$ exists if $C_Q = \{ \mu \}$ is a singleton. One then has the following immediate consequence of Theorem 2.1.

**Corollary 2.2.** (Consistency). Under the assumptions of Theorem 2.1, if the Frechet mean $\mu$ of $Q$ exists then the Frechet sample mean $\mu_n$ converges almost surely to the Frechet mean $\mu$, as $n \to \infty$.

**Remark 2.3.** If the Frechet minimizer in (1.3) is not unique, i.e. $C_Q$ is not a singleton, it is generally not possible to estimate $C_Q$ by $C_{Q_n}$ in any reasonable sense. For, in general, one does not have a reverse set relation interchanging $Q$ and $Q_n$ in part (b) of Theorem 2.1. One may find easy examples where $C_{Q_n}$ is a singleton (almost surely), but $C_Q$ is not. See Remark 2.6 in BP(2003) for an example.

We now turn to the asymptotic distribution theory of the Frechet sample mean. The following result is a slight, but useful, extension of Theorem 2.1.
in BP(2005). Also see BB(2012), pp. 28-30. Denote by $X_i, i = 1, \ldots, n$, i.i.d observations with common distribution $Q$ on $(S, \rho)$. For a differentiable function $f$ on an open subset of $\mathbb{R}^d$ denote $D_r f(y)$ as the partial derivative of $f$ at $y$ with respect to the $r$-th coordinate.

**Theorem 2.4.** (BP (2005)) Let $(S, \rho)$ be a metric space, and $Q$ a probability measure on $S$ such that there exists an open subset $U$ of $S$ with the following properties: (1) $Q(U) = 1$; (2) the Fréchet mean $\mu$ of $Q$ exists and belongs to $U$; (3) there is a homeomorphism $\varphi : U \to V$ for some open subset $V$ of $\mathbb{R}^d$, with the property that $h$ defined on $V \times V$ by $h(x, y) = \rho^2(\varphi^{-1}(x), \varphi^{-1}(y))$ is such that for every $x$ in $U$, the function $y \to h(x, y)$ is twice continuously differentiable in a neighborhood of $\nu = \varphi(\mu)$; (4) $E[D_r h(\varphi(X_1), \nu)^2] < \infty$; (5) $E[\sup_{\{\|y - \nu\| < \varepsilon\}} D_r D_s h(\varphi(X_1), \nu) - D_r D_s h(\varphi(X_1), y)] \to 0$ as $\varepsilon \to 0, \forall r, s$; and (6) the $d \times d$ matrix defined by $\Lambda = E((D_r D_s h(\varphi(X_1); \nu))_{r,s=1,\ldots,d}$ is nonsingular. Then

$$\sqrt{n}(\varphi(\mu_n) - \varphi(\mu)) \xrightarrow{\mathcal{L}} N_d(0, \Lambda^{-1}\Sigma(\Lambda^t)^{-1}),$$

where $\Sigma$ is the covariance matrix of $D_r h(\varphi(X_1), \nu)_{r=1,\ldots,d}$.

The proof of Theorem 2.4 is patterned after that of the CLT for M-estimators (See BP(2005) and BB (2012), pp. 28-30). By replacing $\Lambda$ and $\Sigma$ by their sample counterparts $\hat{\Lambda}$ and $\hat{\Sigma}$, respectively, and assuming that $\Sigma$ is nonsingular, one obtains the following:

$$n[\varphi(\mu_n) - \varphi(\mu)]^T(\hat{\Sigma}^{-1}\hat{\Lambda}^t)[\varphi(\mu_n) - \varphi(\mu)] \xrightarrow{\mathcal{L}} \chi^2_d,$$

where $\chi^2_d$ is the chi-square distribution with $d$ degrees of freedom. From (2.3) one obtains a confidence region for $\nu$ (and of $\mu = \varphi^{-1}(\nu)$).

**Remark 2.5.** Although Theorem 2.4 is applicable to both extrinsic and intrinsic sample means, the CLT for the extrinsic sample mean given below holds under broader conditions. Thus the main use of this result is for the intrinsic sample mean, in which case $(S, \rho) = (M, \rho_g)$ is a Riemannian manifold with the geodesic distance induced by a Riemannian structure $g$ on $M$(See BP (2005), Corollary 2.1 and Theorems 2.2, 2.3). Also, one may take $\mu$ to be a local minimizer restricted to the metric space $U$ or its closure, rather than the global minimizer on $S$ which may not exist (See BB (2008) and BB( 2012), Section 8.4).
Specializing Theorem 2.4 to the case of a \(d\)-dimensional Riemannian manifold \((S, \rho) = (M, \rho_g)\), we now state an important corollary.

**Corollary 2.6.** Let \((S, \rho) = (M, \rho_g)\). Assume also the following: (1) The intrinsic mean \(\mu_I\) of \(Q\) exists and belongs to \(U = M \setminus C(\mu_I), C(\mu_I)\) being the cut locus of \(\mu_I\). Let \(\varphi = \text{Exp}_{\mu_I}^{-1}\) be the inverse of the exponential map at \(\mu_I\) defined on \(U\), and let \(V\) be its image in the tangent space \(T_{\mu_I}(M)\). The conditions (4)-(6) of Theorem 2.4 hold. Then (2.2) and (2.3) hold.

**Remark 2.7.** In the context of Corollary 2.6, the curvature of a Riemannian manifold has an important bearing on the asymptotic dispersion of the intrinsic sample mean given in (2.3) in normal coordinates (See BB(2008)).

We next turn to a broadly applicable CLT for extrinsic analysis. We begin with a general result concerning the computation of the extrinsic mean of a probability measure on a metric space \((S, \rho)\). Consider a one-to-one measurable map \(J\) on \(S\) into a real vector space \(H\) (a finite or infinite dimensional Hilbert space for the case when \((S, \rho)\) is a Hilbert manifold, see (Ellingson et. al.(2012)) ). Let \(P(x)\) denote the projections of \(x \in H\) onto \(J(S) \subset H\). In particular, if the set \(P(x)\) is a singleton, we will denote the distance minimizing point in \(J(S)\) also as \(P(x)\). Let \(\|x\|\) denote the norm of \(x \in H\). For a given probability measure \(Q\) on \(S\), we will assume that the mean \(\mu_J = \int_H x(QoJ^{-1})(dx)\) of the induced probability \(QoJ^{-1}\) on \(H\) is finite.

**Theorem 2.8.** (Patrangenaru (1998), BP(2003)) Let \(Q\) be a probability measure on a metric space \((S, \rho)\), and \(J\) a one-to-one measurable map on \(S\) into a real Hilbert space \(H\). The extrinsic mean of \(Q\) on \(S\) with respect to the induced distance \(\rho_J(p, q) = \|J(p) - J(q)\|\) \((p, q \in S)\) exists iff the mean \(\mu_J\) of the induced probability \(QoJ^{-1}\) on \(H\) has a unique projection onto \(J(S) \subset H\), and then the extrinsic mean of \(Q\) is given by \(J^{-1}P(\mu_J)\).

**Notation.** In order to simplify notation, one may sometimes express the extrinsic mean as \(\mu_E = P(\mu_J)\), instead of \(\mu_E = J^{-1}(P(\mu_J))\).

**Remark 2.9.** Let \(S = M\) be a \(d\)-dimensional manifold and \(J\) an embedding of \(M\) into an Euclidean space \(E^N\). A point \(x \in E^N\) is said to be non-focal if the projection \(P(x)\) is unique, i.e., a singleton. Otherwise it is said to be focal. The set of focal points \(x\) of \(E^N\) has Lebesgue measure zero (BP(2003)).
To derive the asymptotic distribution of the extrinsic sample mean $\mu_{n,E}$, consider a random sample $(X_1, ..., X_n)$ from $Q$, and write $Y_i = J(X_i)$. Then the mean of $\hat{Q}_n^j = \frac{1}{n} \sum_{i=1}^n \delta_{Y_i}$ is $\bar{Y}$. The projection $P$ is defined uniquely on a neighborhood $V$ of $\mu^j$ in $E^N$ into $J(M) \subset E^N$. Considering it as a differentiable map on $V$ into $E^N$, one may use the so-called delta method to show that

$$\sqrt{n}[P(\bar{Y}) - P(\mu^j)] = \sqrt{n}[(GradP)_{\mu^j}(\bar{Y} - \mu^j)] + o_P(1) \xrightarrow{c} N_N(0,C_{\text{asn}} \to \infty),$$

(2.4)

where $C = (GradP)_{\mu^j} \Sigma(GradP)_{\mu^j}^T$. Here $\Sigma$ is the $N \times N$ covariance matrix of $Y_1$ and $(GradP)_{\mu}$ is the $N \times N$ Jacobian matrix of $P$ at $y$ with respect to the standard basis (frame) of the tangent space $T_y(E^N) \equiv \mathbb{R}^N$. Since the image of $P$ restricted to a neighborhood $V$ of $\mu^j$ lies in the $d$-dimensional manifold $J(M)$, the $N \times N$ dispersion matrix $C$ in (2.4) is singular, if $d < N$. Although the $N$-dimensional random vector $(GradP)_{\mu^j}$ lies in $T_{P(\mu^j)}(J(M))$ considered as a subspace of $T_{P(\mu^j)}(E^N) \equiv \mathbb{R}^N$, for statistical inference it is more convenient to express it in terms of a basis of $T_{P(\mu^j)}(J(M))$. Therefore, we choose a convenient moving orthonormal basis (frame) $y \to (F_1(y), F_2(y), \ldots, F_d(y))$ of the subspace $T_y(J(M)) \subset T_y(E^N) \equiv \mathbb{R}^N$ (for $y$ in a neighborhood of $P(\mu^j)$ in $J(M)$). The following proposition then follows from (2.4).

**Proposition 2.10.** (BP(2005), BB(2012)). Assume that the extrinsic mean $\mu_J = P(\mu^j)$ exists. Also assume that $E\|Y_1\|^2 < \infty$. Let $B$ denote the $d \times N$ matrix whose rows are $F_i^T(P(\mu^j)), i = 1, \ldots, d$, form an orthonormal basis of $T_{P(\mu^j)}(J(M))$ regarded as a subspace of $T_{P(\mu^j)}(E^N) \equiv \mathbb{R}^N$. Then

$$\sqrt{n}[B(GradP)_{\mu^j}(\bar{Y} - \mu^j)] \xrightarrow{c} N_d(0,BCB^T) as n \to \infty.$$  

(2.5)

Also, writing $B_n$ for the $d \times N$ matrix whose rows are $F_i^T(P(\mu^j)), i = 1, \ldots, d$, one has

$$\sqrt{n}[B_n(GradP)_{\bar{Y}}(\bar{Y} - P(\mu^j))] \xrightarrow{c} N_d(0,BCB^T) as n \to \infty.$$  

(2.6)

**Proof.** The relation (2.5) follows from (2.2), while (2.6) follows from (2.5) and a Slutsky type argument.

**Remark 2.11.** The extrinsic covariance matrix of $Q$ w.r.t. the orthogonal basis $F_1(\mu^j), \ldots, F_d(\mu^j)$ is $\Sigma_J = BCB^T$. Note that (2.5) expresses the
asymptotic multivariate normal distribution \( N_d(\gamma, n^{-1}\Sigma_J) \) of the coordinates \((Z_1, \ldots, Z_d)_T\) of the random vector \((\text{GradP})_{\mu_j}\) with respect to the basis \(F_i(P(\mu_j)), i = 1, \ldots, d\), with \(\gamma = B(\text{GradP})_{\mu_j}(\mu^j)\). Under broad conditions, the extrinsic covariance matrix \(\Sigma_J\) is a nonsingular \(d \times d\) matrix. For this it is sufficient to require that the distribution of \((\text{GradP})_{\mu_j}(Y_i)\) does not have support on a \((d - 1)\)-dimensional subspace of \(T_{P(\mu_j)}(J(M))\).

The following important consequence of Proposition 2.10 is now simple to derive if one recalls the relation (2.4).

**Theorem 2.12.** (BP(2005), BB(2012)). Let the hypothesis of Theorem 2.5 hold, and assume \(\Sigma_J = BCB^T\) is nonsingular. Then (a) one has the following:

\[
\sqrt{n}[B_n(P(\bar{Y}) - P(\mu^j))] \xrightarrow{L} N_d(0, \hat{\Sigma}_J) \quad \text{as} \quad n \to \infty, \quad \text{and}
\]

\[
n[B_n(P(\bar{Y}) - P(\mu^j))]^T(\hat{\Sigma}_J)^{-1}[B_n(P(\bar{Y}) - P(\mu^j))] \xrightarrow{L} N_d(0, \Sigma_J), \quad \text{as} \quad n \to \infty. \quad (2.7)
\]

Here \(\hat{\Sigma}_J\) is the extrinsic sample covariance matrix \(\Sigma_J\) given by

\[
\hat{\Sigma}_J = B_n(\text{GradP})_Y \hat{\Sigma}(\text{GradP})_Y^TB_n^T,
\]

\(\hat{\Sigma}\) being the sample covariance matrix of \(Y_i, i = 1, \ldots, n\). Also, (b) a confidence region for \(P(\mu^j)\) of asymptotic level \(1 - \alpha\) is given by

\[
C_{1-\alpha} = \left\{ \nu \in J(M), n[B_n(P(\bar{Y}) - \nu)]^T(\hat{\Sigma}_J)^{-1}[B_n(P(\bar{Y}) - \nu)] \leq \chi^2_d(1 - \alpha) \right\}.
\]

**Example 2.13.** Let \(M = S^d = \{x \in \mathbb{R}^{d+1} : \|x\| = 1\}\) be the \(d\)-dimensional unit sphere centered at the origin of the orthogonal coordinate system. The natural embedding of \(S^d\) into \(\mathbb{R}^{d+1}\) is the inclusion map \(J(x) = x\). The projection is uniquely defined as \(P(x) = \frac{x}{\|x\|}, \forall x \neq 0\), and the origin 0 is the only focal point. The \((d + 1) \times (d + 1)\) Jacobian matrix of \(P\) (considered as a map on \(\mathbb{R}^{d+1}\) into \(\mathbb{R}^{d+1}\)) is given by

\[
(\text{GradP})_x = \|x\|^{-1}[I_{d+1} - |x|^{-2}(xx^T)](x \neq 0), \quad (2.10)
\]
where \( I_{d+1} \) is the \((d + 1) \times (d + 1)\) identity matrix. The tangent space
\( T_x(S^d) = \{ y \in \mathbb{R}^{d+1} : x^T y = 0 \} \). One may choose an orthonormal basis
\((F_1(x), \ldots, F_d(x))\) of this tangent space, and let \( B(y) \) be the \( d \times (d + 1)\) matrix
with rows \( F_i(y), 1 \leq i \leq d \). If a probability \( Q \) on \( S^d \) is given such that,
considered as a probability on \( \mathbb{R}^{d+1} \), its mean is \( \mu \neq 0 \), then the extrinsic mean
of \( Q \) is \( \mu_E = P(\mu) = \mu/\|\mu\| \). One takes \( B = B(\mu) \) in (2.7). If \( X_1, \ldots, X_n \)
are i.i.d. with distribution \( Q \), then regarding these as elements of \( \mathbb{R}^{d+1} \), one
takes \( Y_i = X_i, \forall i = 1, \ldots, n \) and \( \bar{Y} = \bar{X} \). Assuming \( \bar{X} \neq 0 \), \( P(\bar{X}) = \bar{X}/\|\bar{X}\| \)
is the extrinsic sample mean \( \mu_E \), say. One may now apply (2.9) to obtain a
confidence region for \( \mu_E \).

**Application 2.14.** (Paleomagnetism). We consider an application of the
results described above in this section to the estimation of the mean
direction of the earth’s magnetic pole. In a seminal paper, R.A. Fisher (1953)
used a parametric model, now known as the von Mises-Fisher distribution,
observations on the magnetic pole from fossilized rock samples. Its density
with respect to the uniform distribution on the sphere \( S^2 \) is given by
\[
   f(x; \mu, \tau) = C(\tau \exp(\tau < x, \mu >), x \in S^2, (\mu \in S^2, \tau \geq 0). \tag{2.11}
\]
Fisher thought of \( \mu \) as the true mean direction. The parameter \( \tau \) measures
concentration of the distribution around \( \mu \); the larger \( \tau \), the larger the concentra-
tion. One can check using the rotational invariance of this density around
\( \mu \) that, in our terminology, \( \mu \) is both the intrinsic mean with respect to the
godesic distance as well as the extrinsic mean with respect to the Euclidean
distance inherited from \( \mathbb{R}^3 \) via the embedding by the inclusion map (Also see
Theorem 3.3 in BP(2003)). The maximum likelihood estimate of \( \mu \), based on i.i.d.
observations \( X_1, \ldots, X_n \) turns out to be \( \bar{X}/\|\bar{X}\| \), which is the same as the extrinsic sample mean \( \mu_E \).
Using \( n = 9 \) observations from the Icelandic lava flow of 1947 - 1948 See
Table 2.1 in BB(2012)), the MLE, or the sample extrinsic mean was found to
be \( \hat{\mu}_E = (.1346, .2948, .9449)^T \). Here \((0, 0, 1)\) denotes the earth’s geographic
North. From a second set of 45 observations from the Quaternary period, the
MLE turned out to be \( \hat{\mu}_E = (.0172, -.2978, -.9545)^T \), almost a complete
reversal of magnetic polarity. This was a study of immense importance in
the field of paleomagnetism, providing compelling evidence for the first time
of the shifting of the earth’s magnetic polarity over geological time scales.
It also provided important clues on the tectonic movements of the Earth’s crust over such periods (See Irving (1964)). The confidence regions on $S^2$, of asymptotic level .95, based on the asymptotic distribution of the MLE in the parametric model

$$C(0.95) = \{ p \in S^2, \rho_g(\hat{\mu}_E, p) \leq 0.1536 \}, \quad (2.12)$$

and the nonparametric confidence region are shown on the left hand side in Figure 1.

It is seen that Fisher’s confidence region for $\mu_{E_1}$ (a geodesic ball of radius .1536 around $\hat{\mu}_E$) is about 10% larger in area (on $S^2$) than the nonparametric region based on (2.9)(See BB(2012), pp. 8-12, for further details). Since we are unable to get hold of the second data set provided by J. Hospers to Fisher, to make a corresponding comparison of the confidence regions by the parametric method using Fisher’s model and our nonparametric confidence region. However, from another set of 31 observations from the Jurassic period found in Irving (1963), the sample extrinsic mean is obtained as $\hat{\mu}_E = (,1346, 2984, 9449)^T$. Once again the confidence region based on the MLE using the model (2.11) (namely, a geodesic ball with center $\hat{\mu}_E$ and radius .1475) is seen to be about 10% larger than that given by (2.9) (See Figure 1). We finally note that the results of nonparametric intrinsic and ex-
tristic inferences in this example are very close to each other. For example, for the first data set analyzed by Fisher, the geodesic distance between the intrinsic and extrinsic sample means is .0007 (See BB(2012), pp.11, 69-71).

We next turn to two-sample tests to discriminate between two distributions on a $d$ dimensional manifold $M$. Let $X_{a,j} : j = 1, \ldots, n_a, a = 1, 2$ be two independent random samples drawn from distributions $Q_a, a = 1, 2$ on $M$, and let $J$ be an embedding of $M$ into $E^N$. Denote by $\mu_a$ the mean of the induced probability $Q_a \circ J^{-1}$ and $\Sigma_a$ its covariance matrix ($a = 1, 2$). Then the extrinsic mean of $Q_a$ is $\mu_{E,a} = J^{-1}(P(\mu_a))$, assuming $Q_a$ is nonfocal. Write $Y_{a,j} = J(X_{a,j}), j = 1, \ldots, n_a, a = 1, 2$ and let $Y_a, a = 1, 2$ be the corresponding sample means. Assuming finite second moments of $Y_a$, $a = 1, 2$, which is automatic if $M$ is compact, one has, by Theorem 2.12

$$\sqrt{n_a}B_a(P(\bar{Y}_a) - P((\mu_a)]) \xrightarrow{c} \mathcal{N}_d(0, \Sigma_{J,a}), a = 1, 2,$$

(2.13)

where $\Sigma_{J,a}$ is the extrinsic covariance matrix of $Q_a$, and $B_a$ are same as in Proposition 2.10 and Theorem 2.12, but with $Q$ replaced by $Q_a$ ($a=1,2$).

That is, $B_a(y)$ is the $d \times N$ matrix of an orthonormal basis (frame) of $T_y(J(M)) \subset T_y(E^N) \approx \mathbb{R}^N$ for $y$ in a neighborhood of $P(\mu_a)$, and $B_a = B_a(P(\mu_a))$. Similarly, $C_a = (GradP)_{\mu_a}\Sigma_a(GradP)_{\mu_a}^T (a = 1, 2)$. The null hypothesis $H_0 : \mu_{E,1} = \mu_{E,2}$, say, is equivalent to $H_0 : P(\mu_1) = P(\mu_2) = \pi$, say. Then, under the null hypothesis, letting $B = B(\pi)$, one has $B_1 = B_2 = B$, and

$$[B(\frac{1}{n_1}C_1 + \frac{1}{n_2}C_2)B^T]^{-1/2}B[P(\bar{Y}_1) - P(\bar{Y}_2)] \xrightarrow{c} \mathcal{N}_d(0, I_d),$$

as $n_1 \to \infty, n_2 \to \infty$. (2.14)

For statistical inference one estimates $C_a$ by $\hat{C}_a = (GradP)_{\hat{Y}_a}\hat{\Sigma}_a(GradP)_{\hat{Y}_a}^T$ where $\hat{\Sigma}_a$ is the sample covariance matrix of sample $a (a = 1, 2)$. Also $B$ is replaced by $\hat{B} = B(\hat{\pi})$ where $\hat{\pi}$ is a sample estimate of $\pi$. Under $H_0$, both $P(\hat{Y}_1)$ and $P(\hat{Y}_2)$ are consistent estimates of $\pi$, but we take a “pooled estimate”

$$\hat{\pi} = P(\frac{1}{n_1 + n_2}(n_1P(\bar{Y}_1) + n_2P(\bar{Y}_2))).$$

(2.15)

We, therefore, have the following result.
Theorem 2.15. (BB(2012)). Assume the extrinsic sample covariance matrix $\hat{\Sigma}_{E,a}$ is nonsingular for $a = 1, 2$. Then, under $H_0 : \mu_{E,1} = \mu_{E,2}$, one has:

$$\lim_{n_1 \to \infty, n_2 \to \infty} \frac{L}{\chi^2_d} = \chi^2_d(1 - \alpha),$$

as $n_1 \to \infty, n_2 \to \infty$. \hfill (2.16)

For the two-sample intrinsic test, let $\mu_{I1}, \mu_{I2}$ denote the intrinsic means of $Q_1$ and $Q_2$, assumed to exist, and consider $H_0 : \mu_{I1} = \mu_{I2}$. Denoting by $\hat{\mu}_{I1}, \hat{\mu}_{I2}$ the sample intrinsic means, (2.2) implies that, under $H_0$,

$$\left(\frac{1}{n_1} \Lambda_1^{-1} + \frac{1}{n_2} \Lambda_2^{-1}\right)^{-1/2}[\varphi_p(\hat{\mu}_{I1}) - \varphi_p(\hat{\mu}_{I2})] \xrightarrow{\text{L}} N_d(0, I_d),$$

as $n_1 \to \infty, n_2 \to \infty$. \hfill (2.17)

where $\varphi_p = \exp_p^{-1}$ for some convenient $p$ in $M$, and $\Lambda_a, \Sigma_a$ are as in Theorem 2.4 with $Q_a$ in place of $Q(a = 1, 2)$. One simple choice for $p$ is the pooled estimate lying on the distance minimizing geodesic connecting $\mu_{I1}, \mu_{I2}$ at a distance $\frac{\rho_g(\mu_{I1}, \mu_{I2})}{n}$ from $\mu_{I1}$. With this choice we write $\hat{\varphi}$ for $\varphi_p$. Let $\hat{\Sigma}_a$ and $\hat{\Lambda}_a$ be the same as $\hat{\Sigma}, \hat{\Lambda}$ in Theorem 2.4 and Corollary 2.6, but using the $a$-th sample and with $\hat{\varphi}$ for $\varphi(a = 1, 2)$. We thus arrive at the following.

Theorem 2.16. The test to reject $H_0 : \mu_{I1} = \mu_{I2}$ iff

$$\left[\hat{\varphi}(\hat{\mu}_{I1}) - \hat{\varphi}(\hat{\mu}_{I2})\right]^T\left[\frac{1}{n_1} \hat{\Lambda}_1^{-1} \hat{\Sigma}_1 \Lambda_1^{-1} + \frac{1}{n_2} \hat{\Lambda}_2^{-1} \hat{\Sigma}_2 \Lambda_2^{-1}\right]^{-1}\left[\hat{\varphi}(\hat{\mu}_{I1}) - \hat{\varphi}(\hat{\mu}_{I2})\right] > \chi^2_d(1 - \alpha),$$

has asymptotic level of significance $\alpha$. \hfill (2.18)

Next, consider a match pair problem with i.i.d. observations $(X_{ij}, X_{j2}), j = 1, \ldots, n$, 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 $\mu_{E,a}$ be the extrinsic means of the (marginal) distributions $Q_a, (a = 1, 2)$. Once again, we are interested
in testing $H_0 : \mu_{E,1} = \mu_{E,2} = \mu_{E}$, say. Note that the extrinsic mean of $Q$ is $\bar{\mu}_E = (\mu_{E,1}, \mu_{E,2})$. If for $a = 1, 2$, $\bar{Y}_a$ is the sample mean of $Y_{aj} = J(X_{aj})$, $j = 1, \ldots, n$ on $E^N$, with $E(Y_a) = \mu_a$, then the extrinsic sample mean is $\bar{\mu}_E = (J^{-1}(P(\mu_1)), J^{-1}(P(\mu_2)))$. Let $B$ denote the $d \times N$ matrix of $d$ orthonormal basis vectors of $T_J(\mu_E)(J(M)) \subset T_J(\mu_E)(E^N)$. If $P$ is the projection operator on the set of all nonfocal points in $E^N$ into $J(M)$, then the projection on $\bar{J}(M)$ is $\bar{P} = P \times P : E^N \times E^N \rightarrow J(M) \times J(M)$. Under $H_0$,

$$n^{1/2}B(\bar{P}(\bar{Y}_1) - \bar{P}(\bar{Y}_2)) \overset{L}{\rightarrow} \mathcal{N}_d(0, \Sigma_{J,1} + \Sigma_{J,2} - \Sigma_{J,12} - \Sigma_{J,21}), \quad \text{as } n \rightarrow \infty. \quad (2.19)$$

On the right, $\Sigma_{J,a}$ is the extrinsic covariance matrix of $X_{a1}$, while $\Sigma_{J,ab} = B(\text{Grad}P)_{\mu_a} \text{Cov}(Y_{a1}, Y_{b1})(\text{Grad}P)_{\mu_b} B^T, \ a \neq b$. From (2.19) we derive the following result by the usual Slutsky type argument.

**Theorem 2.17.** The test which rejects $H_0$ iff

$$n^{1/2}B(\bar{P}(\bar{Y}_1) - \bar{P}(\bar{Y}_2))^T[\hat{\Sigma}_{J,1} + \hat{\Sigma}_{J,2} - \hat{\Sigma}_{J,12} - \hat{\Sigma}_{J,21}]\hat{B}(\bar{P}(\bar{Y}_1) - \bar{P}(\bar{Y}_2)) \geq \chi_d^2(1 - \alpha) \quad (2.20)$$

has the asymptotic confidence coefficient $1 - \alpha$. Here $\hat{B}$ is the $d \times N$ matrix whose rows are orthonormal basis vectors of $T_J(\bar{\mu}_E)(J(M))$, with $J(\bar{\mu}_E) = P(\frac{1}{2} [P(\bar{Y}_1) + P(\bar{Y}_2)])$. Also, $\hat{C}_a, \hat{\Sigma}_a$ are as in (2.18) \(\square\)

Our final topic in this section concerns the use of Efron’s bootstrap (Efron (1979)) for the approximation of true coverage probabilities for confidence regions for the Fréchet means and for p-values of two-sample tests for equality of Fréchet means. That the bootstrap outperforms the traditional approximation of the distribution of the standardized sample mean by the CLT when the underlying distribution is absolutely continuous was first proved by Singh (1981). Extensions to studentized, or pivoted, (multivariate) sample means were obtained by Babu and Singh (1984), who derived asymptotic expansions of the distribution. Indeed, such an expansion holds for more general statistics which are smooth functionals of sample means or which admit a stochastic expansion by a polynomial of sample means (Bhattacharya (1987)). This is derived from refinements of the multivariate CLT by asymptotic expansions (Bhattacharya (1977), Bhattacharya and Ghosh (1978), Chandra and Ghosh (1979)). As a consequence, the chi-square approximation of the distributions
of statistics appearing in (2.3) and (2.9), for example, has an error of the order \( O(n^{-1}) \), whereas the bootstrap approximation of these distributions has an error \( O(n^{-2}) \) (See, e.g., Beran (1987), Hall (1987), (1992), Bhattacharya and Qumsiyeh (1989)). Therefore, bootstrapping the statistics such as appearing in (2.3) and (2.9) would lead to smaller coverage errors than the classical chisquare approximations. In the two-sample case with sample sizes \( m \) and \( n \), the error of approximation of the distributions of asymptotically chisquare statistics is \( O(N^{-2}) \) where \( N = \min\{m, n\} \). In the present context of generally high-dimensional manifolds, often the bootstrap approximation of the covariance is singular, which sometimes makes the bootstrap approximation of the distributions either not feasible, or subject to further errors in case of rather arbitrary augmentations. We briefly describe now bootstrapping procedures for two-sample, or match pair, testing problems on manifolds. Consider, for example, the test \( H_0 : P(\mu_1) = P(\mu_2) \) for the equality of extrinsic means of two distributions on \( M \). Let \( \theta = P(\mu_1) - P(\mu_2) \), and write \( T(X|\theta) \) for the statistic on the left of (2.16), but with \( P(\bar{Y}_1) - P(\bar{Y}_2) \) replaced by \( P(\bar{Y}_1) - P(\bar{Y}_2) - \theta \). Whatever be the true \( \theta \), the asymptotic distribution of \( T(X|\theta) \) is chisquare with \( d \) degrees of freedom. Hence a bootstrap-based test of asymptotic level \( \alpha \) is to reject \( H_0 \) iff \( T(X|0) > c_{1-\alpha}^* \), where \( c_{1-\alpha}^* \), is the \((1-\alpha)\)th quantile of the bootstrapped values \( T(X|\bar{Y}_1) - P(\bar{Y}_2) \). It follows that the p-value of the test is given by: p-value = proportion of bootstrapped values \( T(X^*|P(\bar{Y}_1) - P(\bar{Y}_2)) \) which exceed the observed value of \( T(X|0) \). Similar bootstrapping procedures apply to testing the equality of sample intrinsic means, and to match pair problems.

As pointed out above, for a manifold of a high dimension \( d \), the bootstrap version of the covariance matrix, as appears, for example, within square brackets \([\ldots] \) is often singular if the sample size is not very large. Suppose the classical chisquare approximation is not considered reliable enough, especially if the p-value that it provides is only marginally small. Then, instead, one may consider the very conservative Bonferroni procedure of looking individually at each of the \( d \) contrasts provided, for example, by a principal component analysis (PCA) of the sample covariance matrix, and compute \( d \) p-values for the \( d \) tests for the \( d \) population contrasts, using the bootstrap approximations of the \( d \) distributions of sample contrasts. The estimated p-value is then set at \( d \)-times the minimum of these p-values. It should be pointed out that one should not just pick only the principal components which account
for the largest variability. If the variability due to a particular contrast is small, while its true population mean value is nonzero, then the corresponding test will bring out the difference with high probability, i.e., with a small p-value (See BB(2012), p. 16).

3 Kendall’s Planar Shape Spaces $\Sigma^k_2$

An important development in statistics on manifolds was due to David G. Kendall (1984) who introduced the spaces $\Sigma^k_m$ of direct similarity shapes of $k$-ads (ordered sets of $k$ points) in $\mathbb{R}^m$ as and orbifold. To be specific, one considers only $k$-ads in which the $k$ points are not all equal, and removes translation by centering the $k$-ad $x = (x^1, \ldots, x^k)$ to

$$\xi = (x^1 - \bar{x}, \ldots, x^k - \bar{x}).$$

(3.1)

Note that the set of all translated $k$-ads lie in a vector subspace $L^m_k$ in $(\mathbb{R}^m)^k$ of dimension $pk - p$, $L^m_k = \{\xi = (\xi^1, \ldots, \xi^k) \in (\mathbb{R}^m)^k : \xi^1 + \cdots + \xi^k = 0\}$. The size-and-shape $s_\sigma(x)$ is the orbit of $\xi = (\xi^1, \ldots, \xi^k)$ under the action of $SO(m)$. If the size is not relevant, its effect is removed by scaling $\xi$ to unit size as $u = \frac{\xi}{||\xi||}$. The transformed quantity $u$ is called a preshape, and the set $S(L^m_k)$ of all preshapes comprises a manifold of dimension $mk - m - 1$, namely, the unit sphere in $L^m_k$ called the preshape sphere. Note that $S(L^m_k) \sim S^{mk-m-1}$, the unit sphere centered at the origin in $\mathbb{R}^{mk-m}$. Finally, the direct similarity shape $\sigma(x)$ of a $k$-ad is the orbit, of $u = (u^1, \ldots, u^k)$ under all rotations in $\mathbb{R}^m$. That is,

$$s_\sigma(x) = \{A\xi = (A\xi^1, \ldots, A\xi^k) : A \in SO(m)\},$$

$$\sigma(x) = \{Au = (Au^1, \ldots, Au^k) : A \in SO(m)\},$$

(3.2)

Thus the $m$-dimensional size-and-direct-similarity-shape-space is $S\Sigma^k_m = L^m_k / SO(m) \sim (\mathbb{R}^{mk-m}\setminus\{0\})/SO(p)$ and the Kendall shape space or direct similarity shape space is the compact quotient space $\Sigma^k_m = S(L^m_k)/SO(m) \sim S^{mk-m-1} / SO(m)$.

In the case $m = 2$, the Kendall shape space $\Sigma^k_2$ has a particularly nice manifold structure. Indeed if we regarding each $k$-ad in the plane as an
ordered set of \( k \) complex numbers \( z = (z^1, \ldots, z^k) \), the centered \( k \)-ad
\[
ζ = (ζ^1, \ldots, ζ^k)
\]
\[
ζ^j = z^j - z, \forall j = 1, \ldots, k,
\]
has the same direct similarity shape as \( z = (z^1, \ldots, z^k) \), and lies in the complex hyperplane
\[
L_k = \{ ζ ∈ C^k : ζ^1 + \cdots + ζ^k = 0 \} \sim C^{k-1}.
\]
It we consider a second \( k \)-ad \( z' = (z'^1, \ldots, z'^k) \), having the same Kendall shape as the \( k \)-ad \( z = (z^1, \ldots, z^k) \), they differ by a translation, followed by a rotation and a scaling. That is
\[
z'^j = ρe^{iθ} z^j + b, ∀ j = 1, \ldots, k, ρ > 0, b ∈ C, θ.
\]
From (3.5), the centered \( k \)-ad \( ζ' = (ζ'^1, \ldots, ζ'^k) \), \( ζ'^j = z'^j - z', ∀ j = 1, \ldots, k \), satisfies to the following:
\[
ζ'^j = ρe^{iθ} ζ^j, ∀ j = 1, \ldots, k, ρ > 0.
\]
If in (3.6) we set \( λ = ρe^{iθ} \), then \( λ ∈ C^* = C\setminus\{0\} \), and from this equation, we see that the two \( k \) ads have the same Kendall shape, if and only if
\[
ζ' = λζ.
\]
Note that \( ζ', ζ \) are nonzero complex vectors in \( L_k \), and (3.7), shows that they differ by a nonzero scalar multiple \( λ \). We obtained the following

**Theorem 3.1.** The \( k \)-ads \( z, z' \) have the same Kendall shape if and only if the centered \( k \)-ads \([ζ'] = [ζ] ∈ P(L_k)\), therefore the Kendall planar shape space \( Σ^k_2 \) can be identified with the complex projective space \( P(L_k) \). Moreover, since \( L_k \) has complex dimension \( k - 1 \), \( CP(L_k) \simeq P(C^{k-1}) = CP^{k-2} \), therefore Kendall planar shape analysis is data analysis on \( CP^{k-2} \).

In its representation as \( CP^{k-2} \), the so-called Veronese-Whitney embedding of Kendall’s planar shape space into the real vector space \( S(k - 1, C) \) of \( (k - 1) \times (k - 1) \) Hermitian matrices is given by
\[
J(σ(u)) = u^*u.
\]
The Euclidean inner product on the real vector space $S(k-1, \mathbb{C})$ is given by $<B, C> = \text{Re}(\text{Trace}(BC^*))$. Let $SU(k-1, \mathbb{C})$ 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 $\alpha(A, B) = A^*BA$. Then the embedding (3.8) is $SU(k-1, \mathbb{C})$-equivariant, with the group action $\gamma$ of $SU(k-1, \mathbb{C})$ on $\Sigma_2^k, 2k$ given by: $\gamma(A, \sigma(u)) = \sigma(Au)$. Note that $SU(k-1, \mathbb{C})$ is a group of isometries of $S(k-1, \mathbb{C})$ If $\Sigma_2^k$ is given the metric tensor inherited from $S(k-1, \mathbb{C})$ by the embedding (3.8), then the embedding is isometric as well as $SU(k-1, \mathbb{C})$-equivariant. The extrinsic distance induced by the embedding $J$ in (3.8) is given by

$$\rho_J(\sigma(p), \sigma(q)) = (2(1 - |pq^*|^2)^{1/2}. \quad (3.9)$$

This is referred to as the Procrustean distance $\rho$ in Kendall et al. (1999), and it is $\sqrt{2}$-times the distance referred to as the full Procrustes distance $d_F$ in Dryden and Mardia (1998), p. 65. We next turn to extrinsic analysis on $\Sigma_2^k$, using the embedding (3.8).

**Proposition 3.2.** (BP(2003)). The extrinsic mean $\mu_E$ of $Q$ on $\Sigma_2^k$ under the Veronese-Whitney embedding (3.8) exists, as the unique minimizer of the Frechet function under the induced Euclidean distance, if and only if the largest eigenvalue $\lambda_{k-1}$ of the mean $\mu_J$ of $Q \circ J^{-1}$ on $S(k-1, \mathbb{C})$ is simple. In this case, if $u$ is a unit eigenvector in the one-dimensional eigenspace of $\lambda_{k-1}$, the image $J(\mu_E)$ of the extrinsic mean is given by $u^*u$.

Assuming that the largest eigenvalue of $\mu_J$ is simple, one may now obtain the asymptotic distribution of the sample extrinsic mean $\mu_{n,E}$, namely, that of $J(\mu_{n,E}) = v_nv_n$, where $v_n$ is a unit eigenvector of $\frac{1}{n} \sum_{i=1}^n J(X_i)$ corresponding to its largest eigenvalue. Here $X_1, \ldots, X_n$ are i.i.d. observations on $\Sigma_2^k$. For details, as well as for how to apply Theorem 2.12 for constructing a confidence region for the extrinsic mean see BP(2005).

For intrinsic analysis on $\Sigma_2^k \sim \mathbb{C}P^{k-2}$, one may first use the metric tensor on the preshape sphere in $L_2^kS^{2k-3}$ inherited from the inclusion map in $\mathbb{R}^{2k-2}$. The metric tensor on $\mathbb{C}P^{k-2} = S^{2k-3}/SO(2)$ is a Riemannian submersion (See, e.g., Gallot et al. (1990), p.63). The tangent space at a point $\sigma(p)$ of $\Sigma_2^k, p \in S^{2k-3}$, may be identified with the so-called horizontal subspace of $T_pS^{2k-3}$ orthogonal to the direction of the orbit of $p$ under the action of $SO(2)$. In particular the geodesic distance between the shapes $\sigma(p), \sigma(q)$ of
two points $p, q \in S^{2k-3}$, is given by
\[
\rho_g(\sigma(p), \sigma(q)) = \arccos|pq^*| \in [0, \pi/2].
\] (3.10)

This is the same as the so-called *procrustes distance* in Dryden and Mardia (1998), p.68, and the geodesic distance in Kendall et. al. (1999), relation (9.2), p. 205. One may now proceed to apply corollary 2.6. Two sample set intrinsic test provided by Theorem 2.16 may now be applied.

**Application 3.3.** (Brain scan shapes of schizophrenic and normal children). $k = 13$ landmarks were recorded on the midsagittal slice of the brain scan of each of $n_1 = 14$ schizophrenic children and $n_2 = 14$ normal children (Bookstein (1991). The objective is to detect schizophrenia via possible changes in shape. The shape space is $\Sigma_{k}^{13}$. Intrinsic and extrinsic sample means were found to be within a geodesic distance $O(10^{-5})$ from each other. The chisquare tests for equality of mean shapes provided by Theorems 2.15, 2.16 yield nearly identical $p$-values $O(10^{-11})$ (see BB(2012), pp 15, 16, 106-109). These are smaller by many orders of magnitude compared to those of tests based on parametric model assumptions in Dryden and Mardia (1998), Example 7.4. These $p$-values are 0.01 (Goodall’s F test), 0.04 (Bookstein’s Monte Carlo test), and 0.66 (Hotelling’s $T^2$ test).

**Application 3.4.** (Shapes of male and female gorilla skulls). Here $k = 8$ landmarks are chosen on gorilla skulls: $n_1 = 29$ male skulls, $n_2 = 30$ female.
Mean shapes for two group of children, along with pooled sample mean

Figure 3: The sample extrinsic means for the 2 groups along with the pooled sample mean.

Figure 4: 8 landmarks from skulls of 30 females (red) and 29 male gorillas skulls (see Dryden and Mardia (1998)). Data are displayed in Figure 4. The shape space is $\Sigma_8$. Once again, the intrinsic and extrinsic tests for the null hypothesis of equal mean shapes are very close, each yielding a $p$-value less than $10^{-16}$. A parametric test (Hotelling’s $T^2$ gives a $p$-value 0.0001 (see Dryden and Mardia (1998), Example 7.2). Further, a Bayesian nonparametric method on manifolds with Dirichlet priors was developed and applied to these data by Bhattacharya and Dunson (2010) and BB(2012). Here 25 observations of each gender were randomly chosen to form training samples. The classification rule used Bayes factors to compare the posterior probability of each of the remaining 9 observations to be that of a male or a female. One female was on the borderline, and one male was classified as
Figure 5: Estimated shape densities of gorillas: female(solid), male(dot). Estimate(red), 95% C.R.(blue,green). Densities evaluated at a dense grid of points drawn from the unit speed geodesic starting at female extrinsic mean in direction of male extrinsic.

female (see Figure 5).

4 Kendall’s Shape Spaces $\Sigma^k_m$ and Reflection Similarity Shape Spaces $R\Sigma^k_m$

We have defined Kendall’s shape space $\Sigma^k_m$ of $k$-ads in $m$-dimension in the preceding section. For $m > 2$, the group of isometries $SO(m)$ on the preshape sphere $S^{m(k-1)-1}$ does not act freely on $\Sigma^k_m$, and hence the orbits under the group have different dimensions over different regions of $S^{m(k-1)}$. For example, in the case $m = 3$, if the points in a $k$-ad are collinear, then the points in its preshape lie on a line in $\mathbb{R}^3$ passing through the origin and the 1-dimensional subgroup in $SO(3)$ of rotations around this line keep the $k$-ad fixed. The orbit of this preshape under $SO(3)$ has dimension smaller than that of $SO(3)$, where as for preshapes with non-collinear points the orbits are of the same dimension as that of $SO(3)$. In general, $k$-ads lying on hyperplanes of dimension $m - 2$ or less have orbits of dimensions smaller than that of $SO(m)$. Hence $\Sigma^k_m$ is not a manifold.

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 $m \times k$ matrices, are of full
rank. We define the reflection similarity shape of the $k$-ad $p$ as

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

(4.1)

where $O(m)$ is the set of all $m \times m$ orthogonal matrices $A$. The set $\{r\sigma(p) : p \in \tilde{N}S^{m(k-1)-1}\}$ is the reflection similarity shape space $R\Sigma^k_m = \tilde{N}S^{m(k-1)-1}/O(m)$. Since $\tilde{N}S^{m(k-1)-1}$ may be identified with 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 $R\Sigma^k_m$ such that the projection map $p \rightarrow r\sigma(p)$ is a Riemannian submersion. However, the difficulties in carrying out intrinsic inference on this space are similar to those for $N\Sigma^k_m$. We next consider a useful embedding of $R\Sigma^k_m$ into the vector space $S(k, \mathbb{R})$ of all $k \times k$ real symmetric matrices first obtained by Bandulasiri and Patrangenaru (2005), and later independently found by Dryden et al. (2008). It was further developed in Bandulasiri et al. (2009) and A. Bhattacharya (2008a, 2008b). Define

$$J(r\sigma(p)) = p^tp,$$

(4.2)

with $p$ an $m \times k$ 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 show that $J$ is one-to-one on $R\Sigma^k_m$ into $S(k, \mathbb{R})$, note that if $J(r\sigma(p))$ and $J(r\sigma(q))$ are the same, then the Euclidean distance matrices $((|p_i - p_j|))_{1 \leq i \leq j \leq k}$ and $((|q_i - q_j|))_{1 \leq i \leq j \leq k}$ are equal. Since $p$ and $q$ are centered, by geometry this implies that $q_i = Ap_i (i = 1, \ldots, k)$ for some $A \in O(m)$, i.e., $r\sigma(p) = r\sigma(q)$. It is not difficult to see that the embedding is equivariant with respect to a group action isomorphic to $O(k - 1)$.

Extrinsic inference on $R\Sigma^k_m$ proceeds from the following result on the extrinsic mean.

**Theorem 4.1.** (A. Bhattacharya (2008b)). Let the eigenvalues of the mean $\mu^J$ of $Q \circ J^{-1}$ under the embedding (4.2) be $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_k$, with corresponding eigenvectors $v_1, \ldots, v_k$ with squared norms $\|v_j\|^2 = \frac{\lambda_j}{\lambda_1 + \lambda_2 + \ldots + \lambda_k}$, $j = 1, \ldots, k$. Then the extrinsic mean $\mu_E$ exists if and only if $\lambda_m > \lambda_{m+1}$, and then $J(\mu_E) = P(\mu^J) = (v_1 \ldots v_m)(v_1 \ldots v_m)^t$.

Theorem 4.1 was independently obtained by Bandulasiri et. al. (2009a). The general results of section 2 (See Theorems 2.12, 2.15, 2.17) may now be used to obtain confidence regions and two-sample and match-pair tests.
**Application 4.2.** (Glaucoma detection). To detect any shape change of the inner eye due to glaucoma, 3D images of the optic nerve head (ONH) of both eyes of 12 mature rhesus monkeys were recorded. One of the eyes was subjected to increased intraocular pressure (IOP). \(k = 5\) landmarks of the inner eye were measured on each eye. For this match pair experiment, the manifold is \(\mathbb{R} \Sigma_k^5 \times \mathbb{R} \Sigma_k^5\). The null hypothesis is that the (extrinsic) mean lies on the diagonal of this product manifold (BP(2005), BB(2009)). The \(p\)-value of the nonparametric chi-square test is \(1.55 \times 10^{-5}\) (BB(2012)).

For \(m > 2\), a size-and-reflection shape \(s \sigma(z)\) 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 \(S \Sigma_m^k\). An \(O(k - 1)\)-equivariant embedding of \(SR \Sigma_m^k\) into \(S(k - 1, \mathbb{R})\) is: \(J(s \sigma(z)) = z^t z\). For extrinsic inference in this case refer to Bandulasiri et al. (2009).

**Application 4.3.** (Glaucomatous shape-and-reflection-shape change) For the 3D images of the optic nerve head (ONH) data in application 4.2, Bandulasiri et al. (2009)) gave a 95% bootstrap c. r. in the mean difference between the size-and-reflection-shape coordinates, showing that the extrinsic means are significantly different in the treated vs glaucomatous eye.

## 5 Projective shape spaces \(P \Sigma_m^k\) and analysis of 3D scenes from digital camera images

The huge amounts of digital images available today, pose challenging questions on the type of data that can be retrieved from such images to provide an accurate information about the scene pictured, and on the statistical methodology that needs to be developed to analyze such data. It is impossible to extract Kendall shape data from digital images, in the absence of knowledge on the internal camera parameters. With similarity shape analysis out of discussion, researchers in computer vision, geometry and statistics had to come with better way of understanding the scene from its camera images. Note that even for an almost flat scene, like a face of a building seen from partial closeups in the Figure 6, the similarity shape of the scene has little to do with the similarity of the scene that was imaged, since parallelism is not preserved in digital camera pictures. Nevertheless a straight line on a
Figure 6: Partial views of a side of a building in Atlanta, GA.

A planar scene looks like a straight line in its picture as well, as one may notice in any of the images in Figure 6, and the cross ratio of four points collinear points is invariant under ideal digital camera image acquisition process from the 3D world to the 2D camera film, which is based on a central projection principle. An image point captured on the camera film can be identified with a line going through the pinhole that includes the corresponding real point on the scene, leading to the notion of real projective plane \( \mathbb{R}P^2 \), as set of all lines going through the origin of \( \mathbb{R}^3 \). Formally the \( m \)-dimensional real projective space \( \mathbb{R}P^m \), is the set of orbits of the action of the multiplicative group \( (\mathbb{R}, \cdot) \) on \( \mathbb{R}^{m+1} \setminus \{0\} \), given by the scalar multiplication, and the orbit of \( (x_1, \ldots, x_m, x_{m+1}) \in \) is labeled \( [x] = [x^1 : \cdots : x^{m+1}] \). The numerical space \( \mathbb{R}^m \) can be embedded in \( \mathbb{R}P^m \), preserving collinearity. An example of such an affine embedding is

\[
  h((u^1, \ldots, u^m)) = [u^1 : \ldots : u^m : 1] = [\tilde{u}],
\]

(5.1)

where \( \tilde{u} = (u^1, \ldots, u^m, 1)^T \).

Images of the same flat scene differ by a projective transformation, which depends on the camera location and on the internal camera parameters. Two configurations of points are said to have the same projective shape if they differ by a projective transformation. Recall that a projective transformation...
α : ℝP^m → ℝP^m in m dimensions is associated with a nonsingular matrix
A ∈ GL(m + 1, ℝ), and is given by:

\[ \alpha([x^1 : \cdots : x^{m+1}]) = [A(x^1 \cdots x^{m+1})^T]. \]  (5.2)

In affine coordinates (inverse of the affine embedding (5.1)), the projective
transformation (5.2) is given by \( v = f(u) \), with

\[ v^j = \frac{a_{m+1}^j + \sum_{i=1}^{m} a_i^j u^i}{a_{m+1}^m + \sum_{i=1}^{m} a_i^m u^i}, \forall j = 1, \ldots, m \]  (5.3)

where \( \det((a_i^j)_{i,j=1,\ldots,m+1}) \neq 0 \). These basic elements of projective geometry
motivated the focus of the first papers on data analysis from digital camera
images to be on estimating invariants from a 2D flat scene. These include the
early computer vision work (Mundy and Zisserman (1992), Mundy et. al.
(1993), Maybank and Beardsley (1994), Lentz and Meer (1994)) and, later,
in statistics, work by Mardia et al. (1996) or Goodall and Mardia (1999).

Patrangenaru (1999) proposed a new approach to projective shape analysis,
based on the idea of registration of a generic configuration, with respect to a
projective frame selected from the points of this configuration. A projective
frame (projective basis (Hartley, 1993)) in \( \mathbb{R}P^m \) is an ordered set of \( m + 2 \)
projective points in general position. An example of projective frame in
\( \mathbb{R}P^m \) is the standard projective frame \( ([e_1], \ldots, [e_{m+1}], [e_1 + \ldots + e_{m+1}]) \). In
projective shape analysis it is preferable to employ coordinates invariant with
respect to the group \( \text{PGL}(m) \) of projective transformations. A projective
transformation takes a projective frame to a projective frame, and its action
on \( \mathbb{R}P^m \) is uniquely determined by its action on a projective frame, therefore
if we define the projective coordinate(s) of a point \( p \in \mathbb{R}P^m \) w.r.t. a projective
frame \( \pi = (p_1, \ldots, p_{m+2}) \) as being given by

\[ p^\pi = \beta^{-1}(p), \]  (5.4)

where \( \beta \in \text{PGL}(m) \) is a projective transformation taking the standard pro-
jective frame to \( \pi \). These coordinates have automatically the invariance prop-
erty, therefore given a \( k \)-ad in \( \mathbb{R}P^m \), that includes a projective frame \( \pi \), which
for convenience is selected to be given by the first \( m + 2 \) landmarks of this
\( k \)-ad \( (p_1, \ldots, p_k) \), the resulting projective shape of this \( k \)-ads is in a one to
one correspondence with the projective coordinates \( (p_{m+3}^\pi, \ldots, p_k^\pi) \). Thus
Proposition 5.1. (Patrangenaru (2001)) The projective shape space $P_I \Sigma^k_m$ of $k$-ads in $\mathbb{RP}^m$, containing a projective frame for a given $(m + 2)$-tuple of landmark indices $I = (i_1, \ldots, i_{m+2})$ can be identified with $(\mathbb{RP}^m)^{k-m-2}$.

A useful illustration of Proposition 5.1, due to Buibas et al. (2012) is given in Figure 7, showing that the projective frame methodology, makes possible a statistical analysis of large 2D scenes. The essence of Proposition 5.1 is that projective shape data analysis on $P_I \Sigma^k_m$ is multivariate axial data analysis, allowing solutions to various testing problems, by using non-parametric methods from directional data analysis as developed by Watson (1983), Fisher and Hall (1992), Fisher et. al. (1996). This nonparametric methodology was used in Patrangenaru (2001) who addressed the issue of detecting a spatial scene from its 2D images, using a test for total variance of a projective shape. The general nonparametric theory for statistics on manifolds (see BP (2002, 2003, 2005)) led to a projective frame based statistical analysis of projective shapes of configurations in arbitrary dimensions, without any distributional assumptions. The projective shape space $P\Sigma^k_m = P_{(1, \ldots, m+2)} \Sigma^k_m$ is identified with $(\mathbb{RP}^m)^q$, where $q = m - k - 2$. Mardia and Patrangenaru (2005) considered the Veronese-Whitney equivariant embedding $j_k : P\Sigma^k_m = (\mathbb{RP}^m)^q \rightarrow (S(m+1))^q$ defined by

$$j_k([x_1], \ldots, [x_q]) = (x_1x_1^T, \ldots, x_qx_q^T), \quad (5.5)$$

Figure 7: Digital images of a planar scene. Top: prior to projective frame registration. Bottom: after projective frame registration (note the alignment after registration)
where \( x_s \in \mathbb{R}^{m+1}, x_s^T x_s = 1, \forall s = 1, \ldots, q. \)

**Remark 5.2.** The embedding \( j_k \) in (5.5) so far yields the fastest known computational algorithms in projective shape analysis. Basic axial statistics related to Watson’s method of moments such as the sample mean axis (Watson(1983)) and extrinsic sample covariance matrix (Prentice(1984)) can be expressed in terms of \( j_{m+3} = j \).

A random projective shape \( Y \) of a \( k \)-ad in \( \mathbb{R}P^m \) is given in axial representation by the multivariate random axes

\[
(Y^1, \ldots, Y^q), \quad Y^s = [X^s], \quad (X^s)^T X^s = 1, \quad \forall s = 1, \ldots, q. \tag{5.6}
\]

Patrangenaru et al. (2010) showed that the extrinsic mean \( \mu_E \) of the random projective shape \( (Y^1, \ldots, Y^q) \) exists if \( \forall s = 1, \ldots, q \), the largest eigenvalue of \( E(X^s(X^s)^T) \) is simple, and

\[
\mu_E = ([\gamma_1(m+1)], \ldots, [\gamma_q(m+1)]), \tag{5.7}
\]

where \( \lambda_s(a) \) and \( \gamma_s(a) \), \( a = 1, \ldots, m+1 \) are the eigenvalues in increasing order, and the corresponding unit eigenvector of \( E(X^s(X^s)^T) \).

If \( Y_r, r = 1, \ldots, n \) are i.i.d.r.o.’s (independent identically distributed random objects) from a population of projective shapes (in its multi-axial representation), for which the mean shape \( \mu_E \) exists, from a general consistency theorem for extrinsic means on manifolds in BP(2003) it follows that the extrinsic sample mean \( [\Sigma]_{jk,n} \) is a strongly consistent estimator of \( \mu_E \). In the multivariate axial representation

\[
Y_r = ([X^s_1], \ldots, [X^s_q]), \quad (X^s_r)^T X^s_r = 1; \quad s = 1, \ldots, q. \tag{5.8}
\]

Let \( J_s \) be the random symmetric matrix given by

\[
J_s = n^{-1}\Sigma_{r=1}^n X^s_r (X^s_r)^T, \quad s = 1, \ldots, q, \tag{5.9}
\]

and let \( d_s(a) \) and \( g_s(a) \) be the eigenvalues in increasing order and the corresponding unit eigenvector of \( J_s \), \( a = 1, \ldots, m+1 \). Then the sample mean projective shape in its multi-axial representation is given by

\[
\Sigma_{jk,n} = ([g_1(m+1)], \ldots, [g_q(m+1)]). \tag{5.10}
\]

30
Assume $\Sigma$ be the covariance matrix of $j_k(Y^1, \ldots, Y^q)$ regarded as a random vector in $(S(m+1))^q$, and let $P = P_{j_k} : (S(m+1))^q \to j_k((\mathbb{R}P^m)^q)$ be the projection on $j_k((\mathbb{R}P^m)^q)$. From (??) it follows that the extrinsic covariance matrix of $(Y^1, \ldots, Y^q)$ is given by

$$
\Sigma_E = \left[ e(s,a)(P(\mu)) \cdot D_\mu P(r e^b_a) \right]_{(s=1,\ldots,q),(a=1,\ldots,m)} \cdot \Sigma
$$

$$
\cdot \left[ e(s,a)(P(\mu)) \cdot D_\mu P(r e^b_a) \right]^T_{(s=1,\ldots,q),(a=1,\ldots,m)}.
$$

(5.11)

Assume $Y_1, \ldots, Y_n$ are i.i.d.r.o.’s from a $j_k$-nonfocal probability measure on $(\mathbb{R}P^m)^q$ and $\mu_E$ in (5.7) is the extrinsic mean of $Y_1$. We arrange the pairs of indices $(s,a), s = 1, \ldots, q; a = 1, \ldots, m$, in their lexicographic order. Then, from equation (2.8), the extrinsic sample covariance matrix estimator of $S_{E,n}$, has the entries

$$
S_{E,n,(s,a),(t,b)} = n^{-1}(d_s(m+1) - d_s(a))^{-1}(d_t(m+1) - d_t(b))^{-1} \cdot \sum_{r=1}^n (g_s(a)^T X^*_r)(g_t(b)^T X^*_r)(g_s(m+1)^T X^*_r)(g_t(m+1)^T X^*_r).
$$

(5.12)

$S_{E,n}$ is a strongly consistent estimator of the population extrinsic covariance matrix in (5.11). In preparation for an asymptotic distribution of $\overline{Y}_{j_k,n}$ we set

$$
D_s = (g_s(1) \ldots g_s(m)) \in \mathcal{M}(m+1,m;\mathbb{R}), s = 1, \ldots, q.
$$

(5.13)

If $\mu = ([\gamma_1], \ldots, [\gamma_q])$, where $\gamma_s \in \mathbb{R}^{m+1}$, then, for $s = 1, \ldots, q$, we define a Hotelling’s $T^2$ type-statistic

$$
T(\overline{Y}_{j_k,n};\mu) = n(\gamma_1^T D_1, \ldots, \gamma_q^T D_q) S_{E,n}^{-1}(\gamma_1^T D_1, \ldots, \gamma_q^T D_q)^T.
$$

(5.14)

**Theorem 5.3.** Assume $(Y_r)_{r=1,\ldots,n}$ are i.i.d.r.o.’s on $(\mathbb{R}P^m)^q$, and $Y_1$ is $j_k$-nonfocal, with $\Sigma_E > 0$. Let $\lambda_s(a)$ and $\gamma_s(a)$ be the eigenvalues in increasing order and corresponding unit eigenvectors of $E[X^a_1(X^a_1)^T]$. If $\lambda_s(1) > 0$, for $s = 1, \ldots, q$, then $T(\overline{Y}_{E,n};\mu_E)$ converges weakly to $\chi^2_{mq}$.

If $Y_1$ is a $j_k$-nonfocal population on $(\mathbb{R}P^m)^q$, since $(\mathbb{R}P^m)^q$ is compact, it follows that $j_k(Y_1)$ has finite moments of any order. According to Bhattacharya and Ghosh (1978), this along with an assumption of a nonzero absolutely continuous component, suffices to ensure a coverage error of order $O(n^{-1})$ of the pivotal statistic $T(\overline{Y}_{E,n};\mu_E)$. If the sample size is large enough.
relative to the dimension of the manifold, in order to have a valid bootstrap version of \( S_{E,n}^{-1} \), the theory provides a coverage error of order \( O_P(n^{-2}) \) for the bootstrap estimate of the distribution of \( T(\overline{Y}_{E,n}; \mu_E) \) (also see Chandra and Ghosh(1979), Bhattacharya and Qumsieh(1989), Hall(1992)).

**Corollary 5.4.** Let \( Y_r = ([X_1^r], \ldots, [X_q^r]), X_{st}^T X_{st} = 1, s = 1, \ldots, q, r = 1, \ldots, n \), be i.i.d.r.o.'s from a \( j_k \)-nonfocal distribution on \( (\mathbb{R}P^m)^q \) which has a nonzero absolutely continuous component, and with \( \Sigma_E > 0 \). For a random resample with repetition \( (Y_1^*, \ldots, Y_n^*) \) from \( (Y_1, \ldots, Y_n) \), consider the eigenvalues of \( \frac{1}{n} \sum_{r=1}^n X_r^* X_r^{*T} \) in increasing order and corresponding unit eigenvectors \( d_1^*(a) \) and \( g_1^*(a), a = 1, \ldots, m + 1 \). Let \( S_{E,n}^* \) be the matrix obtained from \( S_{E,n} \), by substituting all the entries with \( * \)-entries. Then the bootstrap distribution function of the statistic

\[
T(\overline{Y}_E^*; \overline{Y}_E) = n(g_1(m + 1)^T D_1^*, \ldots, g_q(m + 1)^T D_q^*) S_{E,n}^{*-1}
\]

\[
(g_1(m + 1)^T D_1^*, \ldots, g_q(m + 1)^T D_q^*)^T
\]

approximates the true distribution of \( T(\overline{Y}_E; \mu_E) \) given by (5.14), with an error of order \( O_P(n^{-2}) \).

Theorem 5.3 and Corollary 5.4 are useful in estimation and testing for mean projective shapes. From Theorem (5.3) we derive a large sample confidence region for \( \mu_E \).

**Corollary 5.5.** Assume \( (Y_r)_{r=1,\ldots,n} \) are i.i.d.r.o.'s from a \( j_k \)-nonfocal probability distribution on \( (\mathbb{R}P^m)^q \), and \( \Sigma_E > 0 \). An asymptotic \( (1 - \alpha) \)-confidence region for \( \mu_E = [\nu] \) is given by \( R_\alpha(Y) = \{[\nu] : T(\overline{Y}_{j_k,n}; [\nu]) \leq \chi^2_{mq}(1 - \alpha)\} \), where \( T(\overline{Y}_E, [\nu]) \) is given in (5.14). If the probability measure of \( Y_1 \) has a nonzero-absolutely continuous component w.r.t. the volume measure on \( (\mathbb{R}P^m)^q \), then the coverage error of \( R_\alpha(Y) \) is of order \( O(n^{-1}) \).

For small samples the coverage error could be quite large, and the bootstrap analogue in Corollary 5.4 is preferable, provided the sample size is large enough to have a large percentage of resamples with repetition, for which \( S_{E,n}^{*-1} \) in (5.15) makes sense.

**Corollary 5.6.** Under the hypotheses of Corollary 5.4, The corresponding \( 100(1 - \alpha) \)% confidence region for \( \mu_E \) is

\[
C_{n,\alpha}^* := j_k^{-1}(U_{n,\alpha}^*)
\]

(5.16)
with \( U_{n,\alpha}^* \) given by

\[
U_{n,\alpha}^* = \{ \mu \in j_k((\mathbb{R}P^m)^q) : T(\mathcal{F}_{j_k,n}; \mu) \leq c^*_{1-\alpha} \},
\]  

(5.17)

where \( c^*_{1-\alpha} \) is the upper \( 100(1-\alpha)\% \) point of the values of \( T(\mathcal{F}_{j_k,n}; \mathcal{F}_E) \) given by (??). The region given by (5.16)-(5.17) has coverage error \( O_P(n^{-2}) \).

Remark 5.7. If \( \Sigma_E \) is singular one may also use a method for constructing nonpivotal bootstrap confidence regions for \( \mu_E \) using Corollary 5.1 of BP(2003).

The projective frame approach was used in a more general context of an analysis of projective shapes of curves in Paige et. al. (2005) and in Munk et. al.(2008). It was also helpful in projective shape density estimation (Lee et. al. (2004)) in combination with a kernel density estimation on Riemannian manifolds as developed by Pelletier (2005). For practical purposes, using extrinsic sample means on projective shape spaces, Patrangenaru and Patrangenaru (2004) reconstructed 2D scenes from their partial views (see the almost planar scene in Figure 8 reconstructed from the partial views in Figure 6). Until 2010, projective shape analysis from digital camera images was limited to 2D data analysis, while the need was for an understanding of 3D scenes. Note that Longuet-Higgins (1981) gave an algorithm for the 3D reconstruction of a point configuration from two ideal pinhole cameras of known internal camera parameters. A camera is noncalibrated if the internal camera parameters are unknown. The following is a key result.

Theorem 5.8. (Faugeras(1992), Hartley, Gupta and Chang (1992)) The 3D reconstruction problem from two noncalibrated camera images has a solution
in terms of the realization of the fundamental matrix $F = T \times R$. Any two solutions can be obtained from each other by a 3D projective transformation.

**Remark 5.9.** Initially, computer vision experts regarded the projective appearance of the reconstruction of a 3D scene from two camera outputs as a complication that can not be circumvented if the cameras are noncalibrated, and digital imaging data widely available on the internet were ignored in 3D scene analysis. Theorem 5.8 was restated in Sughatadasa (2006) from the shape analysis perspective as follows:

**Theorem 5.10.** (Patrangenaru et. al. (2010)). A spatial $\mathcal{R}$ reconstruction of a 3D configuration $\mathcal{C}$ can be obtained in absence of occlusions from two of its ideal camera views. Any such 3D reconstruction $\mathcal{R}$ of $\mathcal{C}$, has the same 3D projective shape as $\mathcal{C}$.

**Application 5.11.** (Face identification). A landmark based example of face recognition based on Theorem 5.10 using a data set extracted from the BBC program “Tomorrow’s World” that was considered in Mardia and Patrangenaru (2005) may be found in Patrangenaru et. al. (2010). The eight facial landmarks considered there shown to be noncoplanar by Balan et al. (2009), thus requiring a 3D projective shape analysis, unlike the analysis, based on fewer (six) landmarks in Mardia and Patrangenaru (2005).

A two sample test for equality of mean 3D projective shape is based on a test for `mean projective shape change. Unlike other shape spaces, 3D projective shape spaces have a *Lie group structure*, borrowed for the projective unit quaternions (Crane and Patrangenaru (2011)). Indeed if a real number $x$ is identified with $(0, 0, 0, x) \in \mathbb{R}^4$, and if we label the quadruples $(1, 0, 0, 0), (0, 1, 0, 0), (0, 0, 1, 0)$ by $\rightarrow i, \rightarrow j, \rightarrow k$, respectively, then the multiplication table given by

<table>
<thead>
<tr>
<th></th>
<th>$\rightarrow i$</th>
<th>$\rightarrow j$</th>
<th>$\rightarrow k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rightarrow i$</td>
<td>-1</td>
<td>$\rightarrow k$</td>
<td>-$\rightarrow j$</td>
</tr>
<tr>
<td>$\rightarrow j$</td>
<td>-$\rightarrow k$</td>
<td>-1</td>
<td>$\rightarrow i$</td>
</tr>
<tr>
<td>$\rightarrow k$</td>
<td>$\rightarrow j$</td>
<td>-$\rightarrow i$</td>
<td>-1</td>
</tr>
</tbody>
</table>

can be extended by linearity to the *quaternion multiplication* $\cdot$ and $(\mathbb{R}^4, +, \cdot)$ has a structure of noncommutative field, often labeled by $\mathbb{H}$. Note that if
\( h, h' \in \mathbb{H} \), then \( \| h \cdot h' \| = \| h \| \| h' \| \), therefore the three dimensional sphere inherits a group structure, the \textit{group of quaternions of norm one}.

Moreover since \( \mathbb{R}P^3 \) is the quotient \( \mathbb{R}^4 \setminus \{0\} / \sim \) where \( \forall x \in \mathbb{R}^4 \setminus \{0\}, x \sim \lambda x, \lambda \in \mathbb{R}^* \), and since \( \lambda(x \cdot y) = (\lambda x) \cdot y = x \cdot (\lambda y), \forall \lambda \in \mathbb{R} \) it follows that the multiplication

\[
[x] \cdot [y] =: [x \cdot y], \quad (5.18)
\]

is a well defined \textit{Lie group} operator on \( \mathbb{R}P^3 \). Note that if \( h = t + x \overrightarrow{i} + y \overrightarrow{j} + z \overrightarrow{k} \), its \textit{conjugate} is \( \overline{h} = t - x \overrightarrow{i} - y \overrightarrow{j} - z \overrightarrow{k} \), and it turns out that the inverse of \( h \in \mathbb{H} \) is given by

\[
h^{-1} = \|h\|^{-2} h, \quad (5.19)
\]

On the other hand, the projective shape space \( P \Sigma_3^k \) is homeomorphic to \( M = (\mathbb{R}P^3)^q \), where \( q = k - 5 \). Therefore with this identification, the projective shape space \( P \Sigma_3^k \sim (\mathbb{R}P^3)^q \) \textit{inherits a Lie group structure} from the group structure of \( \mathbb{R}P^3 \). The multiplication in \( (\mathbb{R}P^3)^q \) is given by

\[
([h_1], \ldots, [h_q]) \cdot ([h'_1], \ldots, [h'_q]) := ([h_1] \cdot [h'_1], \ldots, [h_q] \cdot [h'_q]) = ([h_1 \cdot h'_1], \ldots, [h_q \cdot h'_q]).
\]

If we use the spherical representation in Patrangenaru et. al. (2010), the identity element is given by \( 1_{(\mathbb{R}P^3)^q} = ([0 : 0 : 0 : 1], \ldots, [0 : 0 : 0 : 1]) \), and given a point \( h = ([h_1], \ldots, [h_q]) \in (\mathbb{R}P^3)^q \), from (5.19), its inverse is \( h^{-1} = \overline{h} = ([\overline{h}_1], \ldots, [\overline{h}_q]) \).

**Remark 5.12.** The Lie group structure on \( (\mathbb{R}P^3)^q \), allows a one sample test for means means approach to tests for means in matched pairs, which is an alternative to Theorem 2.15. If \( X \) and \( Y \) are paired random objects on a Lie group \( (G, \circ) \). The \textit{change from X to Y} is the random object \( C =: X^{-1} \circ Y \). We say that \textit{there is no mean change} from \( X \) to \( Y \) if the mean change from \( X \) to \( Y \) is the identity of the group \( G \), that is the null hypothesis is

\[
H_0 : \mu_C = 1_G. \quad (5.21)
\]

One may now apply the Corollaries 5.5, 5.6 to the 3D case. The unit element in \( (\mathbb{R}P^3)^q \) will be labeled \( 1_q \). Given two paired random objects, \( H_1, H_2 \) in their spherical representation on \( (\mathbb{R}P^3)^q \), we set \( Y = \overline{H}_1 H_2 \), and let \( \mu_E \) be the extrinsic mean of \( Y \). Then testing the existence of mean 3D
projective shape change from $H_1$ to $H_2$ amounts to the hypothesis testing problem

$$H_0 : \mu_E = 1_q \text{ vs. } H_1 : \mu_E \neq 1_q.$$  (5.22)

Assume $(H_{1,r}, H_{2,r})_{r=1,\ldots,n}$ are identically distributed random objects from paired distributions on $(\mathbb{R}P^3)^q$, such that $Y_1 = H_{1,1}H_{2,1}$ has a $j_k$-nonfocal probability distribution on $(\mathbb{R}P^3)^q$. Testing the hypothesis (5.22) in the case $m = 3$, at level $\alpha$, amounts to finding a $1 - \alpha$ confidence region for $\mu_E$ given by Corollary 5.5. If the sample is small and the extrinsic sample covariance matrix is degenerate, checking if $1_q$ is in a $1 - \alpha$ confidence region, amounts to finding the upper $\frac{2}{q}$ cutoffs for the bootstrap distributions of the test statistics $T^*_s$, $s = 1, \ldots, k - 5$, and checking if the values of $T_s$, for $\mu_E = 1_q$ are all in the corresponding confidence intervals (see Corollary 5.6).

**Application 5.13.** (glaucomatous projective shape change) Crane and Patrangenaru (2011) carried out an analysis for the matched pairs problem in Applications 4.2, 4.3, using projective shape change. In this application they extracted the matched configurations of 9 landmarks from stereo images of the eye fundus (see Figure 9), to allow for 3D projective shape reconstructions. Three cutoffs $T^*_s$, $s = 1, 2, 4$ for the bootstrap percentiles were found to

![Figure 9: Nine anatomical landmarks of the Optic Nerve Head region](image)
be much smaller than the corresponding statistics \( T_s \), showing a significant mean projective shape change.

6 Some Future Directions

Analysis of Fréchet means on manifolds is today an established area. However many of the sample spaces in modern data analysis, including Kendall shape spaces in dimension 3 or higher are not just manifolds, they do have a manifold stratification (are stratified spaces) though.

Definition 6.1. A filtration by closed subsets \( F_i, i = 0,1, \ldots \) of a metric space, such that the difference between successive members \( F_i \) and \( F_{i-1} \) of the filtration is either empty or a smooth submanifold of dimension \( i \), is called a stratification. The connected components of the difference \( F_i \setminus F_{i-1} \) are the strata of dimension \( i \).

Data analysis on stratified spaces is still in its very early stage. Key examples of stratified sample spaces that are not manifolds, considered so far include

- shape spaces, representing orbits of point configurations under actions of groups of transformations like groups of direct similarities, affine groups, or projective general linear groups (for example, see Kendall et. al. (1999), Patrangenaru and Mardia (2003), Patrangenaru et. al. (2010) for direct similarities, affine transformations, and projective transformations, respectively);

- spaces of positive semidefinite matrices, arising as data points in diffusion tensor imaging (seeArsigny et al (2006), Basser and Pierpaoli (2006), Schwartzman et. al. (2008), for example); and

- tree spaces, representing metric phylogenetic trees on fixed sets of taxa (see Billera et. al. (2001), Wang and Marron(2007) and Owen and Provan (2011), for example).

Huckemann et.al. (2010) were first to analyze data on a stratified space (3D Kendall shape data of stems of fir trees, in the neighborhood of a singular point on \( \Sigma^3 \)). Groisser and Tagare (2009) gave a stratification of
affine shape spaces of $k$-ads in a finite dimensional Euclidean spaces. Fréchet means on sample spaces with negative curvature exhibit a new phenomenon; simulations of Fréchet means on such spaces, such as the tree space $T_4$ in Patrangenaru and Ellingson (2011), or on higher dimensional tree spaces in Skwerer (2012), show that Fréchet sample means often stick to singularities. Recall that a tree with $p$ leaves ($p$-tree) is a connected, simply connected graph, with a distinguished vertex, labeled $o$, called the root, and $p$ vertices of degree 1, called leaves, that are labeled from 1 to $p$. The other (internal) vertices have degree 3 or higher. In addition, we assume that with all interior edges have positive lengths. (An edge of a $p$-tree is called interior if it is not connected to a leaf). Now consider a tree $T$, with interior edges $e_1, \ldots, e_r$ of lengths $l_1, \ldots, l_r$ respectively. If $T$ is binary, then $r = p - 2$, otherwise $r < n-2$. The vector $(l_1, \ldots, l_r)^T$ specifies a point in the positive open orthant $(0, \infty)^r$. A $p$-tree has the maximal possible number of interior edges (namely $p-2$) and thus determines the largest possible dimensional orthant, when it is a binary tree; in this case the orthant is $p - 2$-dimensional. The orthant corresponding to each tree which is not binary appears as a boundary face of the orthants corresponding to at least three binary trees; in particular the origin of each orthant corresponds to the (unique) tree with no interior edges. We construct the space $T_p$ by taking one $p-2$-dimensional orthant for each of the $(2p-3)!!$ possible binary trees, and gluing them together along their common faces. Tree spaces $T_p$ with $p$ small dimension are displayed in Figure 10 (see Billera et al. (2001)).

Figure 10: Tree spaces $T_3, T_4, T_5$.

Phylogenetic trees with $p$ leaves are points on a metric space $T_p$ that has
p − 2 dimensional stratification. In particular the space $T_3$ of trees with 3 leaves is 3-spider, union of three line segments with a common end (see Figure 11, left). Assume $X_i, i = 1, \ldots, n$ are i.i.d.r.o.'s on a spider $S_p$, of legs $L_a, a = 1, \ldots, p$, and center $C$. Assume the intrinsic mean $\mu_{X_i,t}$ exists and the Fréchet variance is finite. Any probability measure $Q$ on $S_p$ decomposes uniquely as a weighted sum of probability measures $Q_k$ on the legs $L_k$ and an atom $Q_0$ at $C$. More precisely, there are nonnegative real numbers \{w_k\}_{k=0}^p summing to 1 such that, for any Borel set $A \subseteq S_p$, the measure $Q$ takes the value
\[
Q(A) = w_0Q_0(A \cap C) + \sum_{k=1}^p w_kQ_k(A \cap L_k).
\] (6.1)
We will consider the nontrivial case when the moments $\mu_a = E(Q_a), a = 1, \ldots, p$ are all positive.

**Theorem 6.2.** (Hotz et. al. (2010)). (i). If $\exists a \in \overline{1,p}$ such that $\mu_a > \sum_{b \neq a} \mu_b$ then $\mu_{X_i,t} \in L_a$ and $\tilde{X}_{n,F} \in L_a$, and $\sqrt{n}(\tilde{X}_{n,F} - \mu_{X_i,t})$ has asymptotically a normal distribution with finite mean. (ii) If $\exists a \in \overline{1,p}$ such that $\mu_a = \sum_{b \neq a} \mu_b$, then asymptotically, after folding the legs $L_b, b \neq a$ into one half line opposite to $L_a$, $\sqrt{n}(\tilde{X}_{n,t})$ has asymptotically the distribution of the absolute value of a normal distribution with finite mean. (iii) If $\forall a \in \overline{1,p}, \mu_a < \sum_{b \neq a} \mu_b, \mu_{X_i,t} = C$ and there is $n_0$ s.t. $\forall n \geq n_0, \tilde{X}_{n,t} = 0$ a.s..

**Remark 6.3.** Under the assumptions of Theorem 6.2 (iii), we say that the sample mean is sticky. Theorem 6.2 was recently extended to C. L. T. on open books (Hotz et. al (2013)).

The stickiness of the Fréchet sample mean at the star tree on $T_3$ was one of the first reasons for studying the asymptotics of Fréchet sample means of distributions on tree spaces (see Figure 11, center and right ). The explosion of genetic data available through molecular biology has made tree-building even more popular. The data that biologists use, usually comes from one homogenous sequence, which in the biologist language means problems between gene trees and genes that are made from one tree. The gene sequence might be about 200 base pairs long. One of the problems that has occurred in the last 40 years is that biologists believe that the way evolution works is that there would only be one species tree. Different genes have different histories,
so you get different gene trees. Putting them together is a statistical problem that helps study the evolutionary process. The process of phylogenetic tree building is based on changes in the DNA sequence describing branching of species, as shown pictorially in Figures 12 and 13. For more details on tree building see Billera et al. (2001). The biological interpretation of the intrinsic mean of a population of phylogenetic trees with $p$ leaves being at the star tree is that the phylogenies in the family are too diverse to offer a plausible evolutionary scenario. For computational algorithms on $(T_p, \rho_0)$, see Owen and Provan (2011). Unlike the intrinsic mean tree, which is unique due to the hyperbolicity of $(T_p, \rho_0)$, if we canonically embedded $T_p$ in $\mathbb{R}^p$ as a $p - 2$ dimensional stratified space, the extrinsic mean set reflects all average evolutionary trees from a given family of phylogenetic trees, setting more mean evolution scenarios. Extrinsic data analysis should be further pursued on stratified spaces, for two additional reasons. One is computational: it is well documented that computations of extrinsic sample means on manifolds are faster than their intrinsic counterparts (Bhattacharya et. al. (2012)). Secondly the asymptotic behavior of extrinsic sample means on stratified spaces is much easier to understand than that of intrinsic sample means. Both intrinsic and extrinsic sample means can be sticky; if a probability measure $Q$ on stratified space $M$ of dimension $m$ has its Fréchet mean $\mu$ on the singular part $M_{m-1}$ of $M$, the sample mean of the independent identically distributed random objects $X_1, X_2, \ldots, X_1 \sim Q$ on $M$ is sticky if for $n$ large
enough, the Fréchet sample mean lies in the singular part as well, that is $\mu_n \in M_{m-1}$. Stickiness was first observed to occur in distributions on trees (Basrak(2010)), and later on, in open books. The proof of the sticky extrinsic CLT is fairly elementary (see Bhattacharya et al.(2012)); in case of Fréchet means based on the chord distance of an embedding of the open book, a heuristic explanation can be given as follows: think of a soft paperback book $M$, where the infinitely thin leaves are hanging down, so the mean vector of the distribution on $M$ sits somewhere underneath the spine, but not too far from it. The Fréchet mean, projection of the open book of the mean
Figure 13: Length of Edges on a Phylogenetic Tree (from S. Holmes presentation posted in the Trees Working Group-SAMSI-Fall 2010)

vector will be on the spine, which is closer to the mean vector than the leaves, and, given a large sample from this distribution, by the by LLN, the sample mean vector will be also closer to the spine than to the leaves, therefore the extrinsic sample mean, will be on the spine as well, and by the classical CLT and the delta method, the extrinsic sample mean will have asymptotically a Gaussian distribution on the spine. The proof of stickiness of intrinsic means on open books is more involved (Hotz et. al.(2012)), but the result can be similarly interpreted: think of a soft hardback book $M$, where the infinitely thin hard leaves, so the intrinsic mean of the distribution on $M$ sits somewhere on the spine. Given a large sample from this distribution, by the by sticky LLN, the intrinsic sample mean, will be on also the spine, and will be asymptotically a Gaussian distribution on the spine. For details see Hotz et. al. (2012).

**Remark 6.4.** The immediate objective on Fréchet mean analysis is proving stickiness CLT’s on stratified spaces in general, developing inference for Fréchet means on stratified spaces, and applying these results in various contexts of object data analysis.

**Acknowledgements**

Many thanks go to our colleagues of the SAMSI Working Group (2010-2011) on sampling from stratified spaces, especially to Steve Marron and the group
co-leaders Ezra Miller and Stephan Huckemann, and to Xiuwen Liu and Leif Ellingson for related joint work. We are also grateful to Susan Holmes for giving us access to her talks on phylogenetic trees in the Working Group of Trees at SAMSI in Fall 2010. The authors gratefully acknowledge support from NSF grants DMS-DMS-0806011, 1107053 and DMS-0805977, 1107053. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References


50


