

LARGE SAMPLE THEORY OF INTRINSIC AND EXTRINSIC SAMPLE MEANS ON MANIFOLDS. I

BY RABI BHATTACHARYA¹ AND VIC PATRANGENARU

Indiana University and Georgia State University

Sufficient conditions are given for the uniqueness of intrinsic and extrinsic means as measures of location of probability measures Q on Riemannian manifolds. It is shown that, when uniquely defined, these are estimated consistently by the corresponding indices of the empirical \hat{Q}_n . Asymptotic distributions of extrinsic sample means are derived. Explicit computations of these indices of \hat{Q}_n and their asymptotic dispersions are carried out for distributions on the sphere S^d (directional spaces), real projective space $\mathbb{R}P^{N-1}$ (axial spaces) and $\mathbb{C}P^{k-2}$ (planar shape spaces).

1. Introduction. The aim of this article is to develop nonparametric statistical inference procedures for measures of location of distributions on general manifolds, which are complete as metric spaces. Although the main applications are to distributions on (i) spheres S^d (spaces of directions), (ii) real projective spaces $\mathbb{R}P^{N-1}$ (axial spaces) and (iii) complex projective spaces $\mathbb{C}P^{k-2}$ (planar shape spaces), a general theory for both compact and noncompact manifolds is sought. In this introduction a summary of the main results is presented, along with a brief review of the literature on the subject.

A natural index of location for a probability measure Q on a metric space M with the distance ρ is the so-called *Fréchet mean* which minimizes $F(p) = \int \rho^2(p, x)Q(dx)$, if there is a unique minimizer. In general, the set of all minimizers is called the *Fréchet mean set*. In the case M is a d -dimensional connected C^∞ Riemannian manifold with a metric tensor g and geodesic distance d_g , we will assume that (M, d_g) is complete and we will refer to the Fréchet mean (set) as the *intrinsic mean (set)*. We say that the *intrinsic mean exists* if there is a unique minimizer, and denote it by $\mu_1(Q)$. It is shown in Theorem 2.1 that (i) the intrinsic mean set is compact, (ii) for each point m in the intrinsic mean set, the Euclidean mean of the distribution on the tangent space at m of the inverse of the exponential map is zero and (iii) in the case of simply connected M of nonpositive curvature, the intrinsic mean exists if F is finite; a particular case of this result, when M is a Bookstein's shape space of labeled triangles, with a Riemannian metric of constant negative curvature is due to Le and Kume (2000). From a result of Karcher (1977) it follows that if the distribution is

Received March 1999; revised October 2001.

¹Supported by NSF Grant DMS-00-73865 and a Guggenheim Fellowship.

AMS 2000 subject classifications. Primary 62H11; secondary 62H10.

Key words and phrases. Fréchet mean, intrinsic mean, extrinsic mean, consistency, equivariant embedding, mean planar shape.

sufficiently concentrated then the intrinsic mean exists. For planar shape spaces $\mathbb{C}P^{k-2}$, a useful necessary and sufficient condition for the existence of an intrinsic mean is proved by Le (1998) for distributions Q which are absolutely continuous (w.r.t. the volume measure) with a density that is a function only of the distance from a given point.

An important question on the estimation of location is that of *consistency*. Theorem 2.3 says that if M is a metric space such that all closed bounded subsets of M are compact then, with probability 1, given any $\varepsilon > 0$, the Fréchet sample mean set based on a random sample from Q will be within a distance less than ε from the Fréchet mean set of Q for all sufficiently large sample sizes. Thus if M is a complete Riemannian manifold and if the intrinsic mean exists then, almost surely, all measurable choices from the intrinsic sample mean set converge uniformly to the intrinsic mean of Q . In particular, this generalizes and strengthens the strong consistency result for compact M that follows from an earlier result of Ziezold (1977). [Also see Kent and Mardia (1997) and Le (1998).]

Much of the literature in the field deals with special cases of what we call the *extrinsic mean*, perhaps because of the technical difficulties involved in proving the existence of an intrinsic mean and in computing the intrinsic sample mean, even when it exists. To define an extrinsic mean of Q with respect to an embedding j of M in a Euclidean space (\mathbb{R}^k, d_0) , consider first the set of all points p of \mathbb{R}^k such that there is a *unique* point x in $j(M)$ having the smallest distance from p , that is, satisfying $d_0(p, j(M)) = d_0(p, x)$. Such points p are called *nonfocal*, and points which are not nonfocal are called *focal*. For example, the only focal point of S^d in \mathbb{R}^{d+1} is the origin. For an embedding j of M in \mathbb{R}^k a probability measure Q on M is said to be *nonfocal* if, when viewed as a measure on \mathbb{R}^k via j , its mean μ is a nonfocal point. The *extrinsic mean* $\mu_E(Q)$ of a nonfocal Q is the j -preimage of the projection $P_{j(M)}(\mu)$ of μ on $j(M)$, that is, $j(\mu_E(Q))$ is the point of $j(M)$ closest to μ . One may show that the set of all focal points is closed and has Lebesgue measure zero in \mathbb{R}^k (Theorem 3.2). Being thus guaranteed that most probability measures on M are nonfocal, one proceeds to show that the *extrinsic sample mean* $\bar{X}_E = \mu_E(\hat{Q}_n)$ based on a random sample is a *strongly consistent* estimate of the (population) extrinsic mean $\mu_E(Q)$ of a nonfocal Q (Theorem 3.4). Here \hat{Q}_n is the empirical distribution of the random sample. As far as the estimation of the intrinsic mean $\mu_I(Q)$ is concerned, Theorem 3.3(b) in the present article proves that under an equivariant embedding one has $\mu_I(Q) = \mu_E(Q)$ provided M is a compact two point homogeneous space other than a round sphere, and Q is nonfocal and invariant under the subgroup of isometries leaving a given point fixed. In particular, under the assumed symmetries, the extrinsic sample mean is a strongly consistent estimator of the intrinsic mean $\mu_I(Q)$ if the latter exists.

As indicated above, for an embedding of M in an Euclidean space \mathbb{R}^k , the extrinsic mean $\mu_E(Q)$ exists under broad verifiable conditions. The next important task, beyond consistency, is to derive the asymptotic distribution of the

extrinsic sample mean and use this to construct confidence regions for $\mu_E(Q)$ and, therefore, of $\mu_I(Q)$ when the intrinsic and extrinsic means coincide. A general method is presented for this. Let \bar{X} be the sample mean, when the observations X_i are viewed as points in the ambient space \mathbb{R}^k . In Theorem 3.6, the projection $H(\bar{X})$ of \bar{X} on the tangent space to M at $\mu_E(Q)$ is shown to be asymptotically normal centered at $\mu_E(Q)$, and a computation of the asymptotic dispersion is given. One derives bootstrap-based confidence regions for $\mu_E(Q)$ (Corollary 3.7) by Efron's percentile method [Efron (1982)] with a coverage error $O_p(n^{-1/2})$ for general Q , which is particularly useful in those cases where the asymptotic dispersion matrix is difficult to compute. Note once again that, under the hypothesis of symmetry in Theorem 3.3(b), if the intrinsic mean exists then the above confidence regions apply to it.

Finally, Section 4 applies the preceding theory to (i) real projective spaces $\mathbb{R}P^{N-1}$, or the *axial spaces*, and to (ii) complex projective spaces $\mathbb{C}P^{k-2}$, or the *planar shape spaces*. Under the so-called *Veronese–Whitney embedding*, the explicit formulas for the extrinsic mean of a nonfocal distribution on an axial space are given in Theorem 4.2. For planar shape spaces the corresponding results are presented in Theorem 4.4. It is also pointed out in Example 4.3 that inconsistent Procrustes estimators in some parametric models arise when Q is focal, thus clarifying an issue raised in Dryden and Mardia (1998), page 280. As an application both intrinsic the extrinsic (*Procrustean*) sample means are computed using some data from Bookstein (1991) on children with the so-called Apert syndrome. This is presented in graphical form. Also, the extrinsic sample mean of 13 complete observations of Apert data is used to estimate a missing landmark in one incomplete observation. The data here are quite concentrated, which makes the extrinsic sample mean almost indistinguishable from the intrinsic sample mean.

We now briefly mention some of the earlier literature on statistical inference on Riemannian manifolds. In parametric statistical inference, the *information matrix* has been used as a Riemannian metric on the parameter space ever since Rao (1945). For more recent treatments and advances in this direction, we refer to Amari (1985), Barndorff-Nielsen and Cox (1994), Burbea and Rao (1982), Efron (1975) and Oller and Corcuera (1995). Pioneering work on directional analysis was carried out by G. S. Watson beginning in the 1950s [see Watson (1983) and Mardia and Jupp (1999) and the references in both]. Some classes of semiparametric models were analyzed by Beran (1979), Watson (1983) and others. Statistical analysis for axial and shape spaces similar in spirit to the inference for extrinsic means presented here may be found in Kent (1992), Kent and Mardia (1997), Le (1998) and Prentice and Mardia (1995). Nonparametric bootstrap methods for inference on extrinsic means of axes have been employed in Beran and Fisher (1998) and in Fisher, Hall, Jing and Wood (1996). The recent books by Dryden and Mardia (1998) and Kendall, Barden, Carne and Le (1999) are good sources for readable accounts of various methodologies in the field, emphasizing their applications.

2. Intrinsic means and moments of a probability measure on a Riemannian manifold. Let (M, g) be a d -dimensional connected and complete Riemannian manifold, that is, M is a d -dimensional C^∞ connected manifold with a complete Riemannian metric g . Denote by d_g the (geodesic) distance under g . We consider M -valued random variables X , that is, measurable maps on a probability space (Ω, \mathcal{A}, P) into (M, \mathcal{B}) , where \mathcal{B} denotes the Borel sigma-algebra of M . All probability measures on M below are defined on \mathcal{B} . Note that every closed bounded subset of M is compact [do Carmo (1992), pages 146–149].

For the following definition we consider, more generally, a metric space (M, d) and a probability measure Q on the Borel sigma-algebra \mathcal{B} of M .

DEFINITION 2.1. Let Q be a probability measure on the metric space (M, d) . The *Fréchet mean set* of Q is the set of all minimizers of the map F on M defined by

$$(2.1) \quad F(p) = \int d^2(p, x)Q(dx), \quad p \in M.$$

If there is a unique minimizer, this is called the *Fréchet mean* of Q . If M is a Riemannian manifold, the Fréchet mean (set) w.r.t. the geodesic distance $d = d_g$ is defined to be the *intrinsic mean (set)* of Q ; if the minimizer is unique, the intrinsic mean will be labeled $\mu_1(Q)$. If X is an M -valued random variable having distribution Q , then the above are also referred to as the *Fréchet*, or *intrinsic mean (set)* of X , as the case may be.

Riemannian manifolds are “curved,” so that geodesics starting at a point p may meet for a second time in the cut locus of p . Technical details on cut locus and normal coordinates are as follows. If the manifold is complete, the *exponential map* at q is defined on the tangent space $T_q M$ by $\exp_q v = \gamma(1)$, where $\gamma : t \rightarrow \gamma(t)$ is the geodesic with $\gamma(0) = q$, $\dot{\gamma}(0) = v$. An open set $U \subset M$ is said to be a *normal neighborhood* of q ($\in U$), if \exp_q is a diffeomorphism on a neighborhood V of the origin of $T_q M$ onto U , with V such that $tv \in V$ for $0 \leq t \leq 1$, if $v \in V$. Suppose $U = \exp_q V$ is a normal neighborhood of q . Then (x^1, x^2, \dots, x^d) are said to be the *normal coordinates* of a point $p \in U$ w.r.t. a fixed orthobasis (v_1, v_2, \dots, v_d) of $T_q M$ if $p = \exp_q (x^1 v_1 + x^2 v_2 + \dots + x^d v_d)$.

Let $v \in T_q M$ be such that $g(v, v) = 1$. The set of numbers $s > 0$, such that the geodesic segment $\{\exp_q tv : 0 \leq t \leq s\}$ is minimizing is either $(0, \infty)$ or $(0, r(v)]$, where $r(v) > 0$. We will write $r(v) = \infty$ in the former case. If $r(v)$ is finite, then $\exp_q r(v)v$ is the *cut point of q in the direction v* [Kobayashi and Nomizu (1996), page 98]. Let $S_q M = \{v \in T_q M : g(v, v) = 1\}$; then the largest open subset of M in which a normal coordinate system around q is defined is $\exp_q(V(q))$, where $V(q) = \{tv : 0 \leq t < r(v), v \in S_q M\}$. The *cut locus* of q is $C(q) = \exp_q\{r(v)v : v \in S_q M, r(v) \text{ finite}\}$ [Kobayashi and Nomizu (1996), page 100]. Note that $C(q)$ has volume measure 0, and M is the disjoint union

of $\exp_q(V(q))$ and $C(q)$. The *injectivity radius at the point q* is $r_q = \inf\{r(v) : v \in \mathcal{S}_q M\}$.

EXAMPLE 2.1. For the d -dimensional unit sphere, $M = \mathbb{S}^d = \{p \in \mathbb{R}^{d+1} : \|p\| = 1\}$, with the Riemannian metric induced by the Euclidean metric on \mathbb{R}^{d+1} , the exponential map is given by

$$(2.2) \quad \exp_p(v) = \cos(\|v\|)p + \sin(\|v\|)\|v\|^{-1}v, \quad v \in T_p \mathbb{S}^d, v \neq 0.$$

Also, $V(p) = \{v \in T_p \mathbb{S}^d : \|v\| < \pi\}$ and $C(p) = -p$. We may now determine the exponential map when M is a real (complex) projective space $\mathbb{R}P^d$ ($\mathbb{C}P^{d/2}$ for d even) of constant (constant holomorphic) curvature. In this case M is a quotient of a round sphere S , and the projection map $\pi : S \rightarrow M$ is a Riemannian submersion. If we denote by \exp the exponential map for both the sphere and projective space, we have $\exp_{\pi(p)} d\pi(v) = \pi(\exp_p(v))$. If $p \in \mathbb{S}^d$, then since $\mathbb{R}P^d$ is homogeneous, for $[p] \in \mathbb{R}P^d$, we may assume without loss of generality that $p = (1, 0, \dots, 0)$. Then $C([p]) = \{[q] : q = (0, q^1, \dots, q^d) \in \mathbb{S}^d\} = \mathbb{R}P^{d-1}$ is the *projective hyperplane from infinity* of the point $[p]$. Similarly, we may assume that the point $[p] \in \mathbb{C}P^{d/2}$ is represented by $p = (1, 0, \dots, 0)$ and in this case $C([p])$ is $\mathbb{C}P^{d/2-1}$, the *complex projective hyperplane at infinity* of the point $[p]$.

If $Q(C(q)) = 0$, we will denote by $\lambda_Q = \lambda_{Q,q}$ the *image measure of Q under \exp_q^{-1}* on $M \setminus C(q)$. We will suppress q in $\lambda_{Q,q}$.

THEOREM 2.1. Assume (M, g) is a complete connected Riemannian manifold. Let $I(Q)$ be the intrinsic mean set of Q and set $C(Q) = \bigcup_{q \in I(Q)} C(q)$. (a) If there is a point p on M such that $F(p)$ is finite, then the intrinsic mean set is a nonempty compact set. (b) If $q \in I(Q)$ and $Q(C(Q)) = 0$, then

$$(2.3) \quad \int_{V(q)} v \lambda_Q(dv) = 0.$$

(c) Suppose (M, g) has nonpositive curvature, and M is simply connected. Then every probability measure Q on M has an intrinsic mean, provided $F(p)$ is finite for some p .

PROOF. (a) It follows from the triangle inequality (for d_g) that if $F(p)$ is finite for some p , then F is finite and continuous on M . To show that a minimizer exists, let l denote the infimum of F and let $p_n \in M$ be such that $F(p_n) \rightarrow l$ as $n \rightarrow \infty$. By the triangle and the Schwarz inequalities, and by integration w.r.t. Q , one has

$$(2.4) \quad \begin{aligned} d_g^2(p_n, p_1) &\leq 2d_g^2(p_n, x) + 2d_g^2(x, p_1) \quad \forall x \in M, \\ d_g^2(p_n, p_1) &\leq 2(F(p_n) + F(p_1)). \end{aligned}$$

Hence since $F(p_n)$ ($n \geq 1$) is a bounded sequence, so is p_n ($n \geq 1$). By completeness of M , p_n has a subsequence converging to some point p^* . Then

$F(p^*) = l$, so that p^* is a minimizer. Also the inequalities (2.4) applied this time to p^* and an arbitrary minimizer m show that $d_g^2(m, p^*) \leq 4l$. In other words, the set of minimizers is bounded. It is also a closed set, since its complement is clearly open, proving compactness of the intrinsic mean set. To prove (b), note that $\exp_q(V(q))$ has Q -probability 1. Consider an arbitrary point x in $\exp_q(V(q))$; then with probability 1 there is a unique geodesic, say $\gamma_{x,\mu}$ joining x and μ with $\gamma_{x,\mu}(0) = x$, $\gamma_{x,\mu}(1) = \mu$. Also let $\mu_v(t)$ be the geodesic starting at μ [$\mu_v(0) = \mu$] with tangent vector v [$(d\mu_v(t)/dt)(0) = v$]. Let $\alpha_{v,x}$ be the angle made by the vectors tangent to these geodesics at μ . Then [see Helgason (1978), page 77, and Oller and Corcuera (1995), Proposition 2.10]

$$(2.5) \quad d_\mu F(v) = 2 \int d_g(x, \mu) \|v\| \cos(\alpha_{v,x}) Q(dx).$$

Select a point $q \in I(Q)$ and write the integral in (2.5) in normal coordinates on $T_q M$. If $\mu \in I(Q)$, then μ is a critical point of F . Then we select $\mu = q$, and evaluate the right-hand side of (2.5) at $v = v_i = \frac{\partial}{\partial x^i}$. Note that given that \exp_q is a radial isometry, the right-hand side of (2.5) in this case is $2 \int x^i \lambda_Q(dx)$, where x^i are the normal coordinates of an arbitrary point of $\exp_q(V(q))$. Then in such coordinates, (2.5) becomes (2.3).

For part (c) of the theorem, we adapt the proof of Kobayashi and Nomizu (1996), Theorem 9.1, to our situation as follows. By part (a) there is a point q in the intrinsic mean set. By a classical result due to J. Hadamard [see Helgason (1978), page 74], since M is simply connected and complete, $C(q) = \emptyset$, and we define a map G on M by

$$(2.6) \quad G(p) = \int_M \|\exp_q^{-1}(p) - v\|^2 \lambda_Q(dv).$$

Since on a simply connected manifold of nonpositive curvature \exp_q is expanding, we have $G(p) \leq F(p)$. On the other hand by part (b), $G(p) = G(q) + \|\exp_q^{-1}(p)\|^2$ and, since \exp_q is a radial isometry, $F(q) = G(q)$. Therefore, q is in fact the unique minimizer of F . \square

REMARK 2.1. If M has nonpositive curvature and is not simply connected, the intrinsic mean does not exist in general. If M is flat a sufficient condition for the existence of the intrinsic mean is that the support of Q is contained in a geodesically convex open normal neighborhood of M and $F(p)$ is finite for some p . In general, if the infimum of the injectivity radii is a positive number $r(M)$ and the *scalar curvature* of (M, g) is bounded from above by $(\pi/r(M))^2$ and if the support of Q is contained in a closed geodesic ball \overline{B}_ρ of radius $\rho = r(M)/4$, then the intrinsic mean exists. To see this note that, when restricted to the closed geodesic ball $\overline{B}_{2\rho}$, F has a unique minimum at some point in \overline{B}_ρ [see Karcher (1977), Theorem 1.2]. Clearly, this minimum value is no more than ρ^2 . On the other hand, if $p \in (\overline{B}_{2\rho})^c$, then $F(p) \geq d_g^2(p, \overline{B}_\rho) > \rho^2$. This proves the

uniqueness of the minimum of F in M , when the support of Q is contained in \underline{B}_ρ . Necessary and sufficient conditions for the existence of the intrinsic mean of absolutely continuous radially distributed probability measures on $\mathbb{C}P^{d/2}$ are given in Le (1998) and Kendall, Barden, Carne and Le (1999).

REMARK 2.2. Mean values of a random variable on a manifold were defined in Oller and Corcuera (1995) and previously in Emery and Mokobodzki (1991), as exponential barycenters. The mean values in the sense of Oller and Corcuera (1995) turn out to be critical points of F , while the intrinsic means defined here are minimizers of F . This explains, for example, why in Oller and Corcuera (1995) the von Mises distribution on S^d is found to have two mean values, while in fact there is only one intrinsic mean. Note that the density at $x \in S^d$ of the *von Mises distribution* w.r.t. the volume form is a constant multiple of $\exp(mx)$.

In the case of a Riemannian manifold, the points in the intrinsic mean set are points of local minima of F and are therefore *Karcher means* [Kendall (1990) and Le (1998)].

REMARK 2.3. If $C(q)$ has Q -measure zero, for some $q \in M$, an *intrinsic moment w.r.t. a given set of normal coordinates* of an arbitrary order $s = (s^1, \dots, s^d) \in \mathbb{Z}_+^d$ can be defined by $\int x^s \lambda_Q(dx)$ where $x^s = (x^1)^{s^1} \dots (x^d)^{s^d}$, if the latter is finite.

REMARK 2.4. As the proof shows, part (a) of Theorem 2.1 holds for the Fréchet mean set of a probability measure Q on any metric space M with the property that all closed bounded subsets of M are compact.

For the structure of probability measures which are invariant under a group of isometries one has the following simple result.

PROPOSITION 2.2. *Suppose K is a group of isometries of (M, g) which leaves the measure Q invariant. Then the intrinsic mean set is left invariant by K . In this case Q induces a quotient measure on the space of orbits M/K and the mean set of Q is a union of orbits.*

PROOF. An isometry τ of (M, g) is a diffeomorphism of M such that $g(d\tau(v), d\tau(v)) = g(v, v)$ for all $v \in T_q M$. Since $d_g(p, q) = d_g(\tau(p), \tau(q))$ for all $p, q \in M$, if Q is invariant under τ then one has $F(\tau(p)) = F(p)$ [see (2.1)]. In particular, this is true when p is a minimizer of F and $\tau \in K$. The claim follows from these observations. \square

EXAMPLE 2.2. If Q is rotationally symmetric on S^d (such as the von Mises measure), then the intrinsic mean set of Q is a union of parallel $(d - 1)$ -dimensional spheres or poles of the axis of rotation, since the space of orbits is one-dimensional. Let $SO(d)$ be the *special orthogonal group* (or group of rotations).

The $SO(d)$ invariant measures on S^d depend on one function of one real variable, as shown in Watson [(1983), Section 4.2]. The uniform distribution on a compact Riemannian manifold whose density is $1/\text{vol}(M)$ w.r.t. the volume measure is an example of an invariant distribution. Recall that the volume measure of a Riemannian manifold in a local chart is given by

$$(2.7) \quad \text{vol}(A) = \int_A \det \left(g_x \left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j} \right) \right)^{1/2} \lambda(dx),$$

where λ is the Lebesgue measure. The intrinsic mean set of the uniform distribution is M .

DEFINITION 2.2. Let X_1, \dots, X_n be independent random variables with a common distribution Q on a metric space (M, d) , and consider their *empirical distribution* $\hat{Q}_n = \frac{1}{n} \sum_{k=1}^n \delta_{X_k}$. The *Fréchet sample mean (set)* is the Fréchet mean (set) of \hat{Q}_n , that is, the (set of) minimizer(s) m of $p \rightarrow \frac{1}{n} \sum_{j=1}^n d^2(X_j, p)$. If M is a Riemannian manifold, then the Fréchet sample mean (set) of \hat{Q}_n for the distance $d = d_g$ is called the *intrinsic sample mean (set)*.

The following result establishes the strong consistency of the Fréchet sample mean as an estimator of the Fréchet mean of the underlying distribution.

THEOREM 2.3. Let Q be a probability measure on a metric space (M, d) such that every closed bounded subset of M is compact. Assume F is finite on M . (a) Then, given any $\varepsilon > 0$, there exist a P -null set N and $n(\omega) < \infty \forall \omega \in N^c$, such that the Fréchet (sample) mean set of $\hat{Q}_n = \hat{Q}_{n,\omega}$ is contained in the ε -neighborhood of the Fréchet mean set of Q for all $n \geq n(\omega)$. (b) If the Fréchet mean of Q exists then every measurable choice from the Fréchet (sample) mean set of \hat{Q}_n is a strongly consistent estimator of the Fréchet mean of Q .

PROOF. (a) We will first prove that for every compact subset K of M one has

$$(2.8) \quad \sup_{p \in K} |F_{n,\omega}(p) - F(p)| \rightarrow 0 \quad \text{a.s. as } n \rightarrow \infty,$$

$$F_{n,\omega}(p) := \int d^2(x, p) \hat{Q}_{n,\omega}(dx) \equiv \frac{1}{n} \sum_{j=1}^n d^2(X_j, p).$$

To prove (2.8) first observe that for a given $p_0 \in K$ one has, in view of the strong law of large numbers (SLLN) applied to $\frac{1}{n} \sum_{j=1}^n d(X_j, p_0)$,

$$(2.9) \quad \sup_{p \in K} \frac{1}{n} \sum_{j=1}^n d(X_j, p) \leq \frac{1}{n} \sum_{j=1}^n d(X_j, p_0) + \sup_{p \in K} d(p, p_0)$$

$$\leq \int d(x, p_0) Q(dx) + 1 + \text{diam}(K) = A, \text{ say,}$$

which holds for all $n \geq n_1(\omega)$, where $n_1(\omega) < \infty$ outside a P -null set N_1 . Fix $\varepsilon' > 0$. From (2.9) one obtains, using the inequality $|d^2(X_j, p) - d^2(X_j, p')| \leq \{d(X_j, p) + d(X_j, p')\}d(p, p')$, the bound

$$(2.10) \quad \sup_{\{p, p' \in K : d(p, p') < \delta_1\}} |F_{n, \omega}(p) - F_{n, \omega}(p')| \leq 2A\delta_1 = \varepsilon'/3$$

$$\forall n \geq n_1(\omega) \ (\omega \notin N_1),$$

where $\delta_1 := A/6\varepsilon'$. For the next step in the proof of (2.8), let $\delta_2 > 0$ be such that $|F(p) - F(p')| < \varepsilon'/3$ if $p, p' \in K$ and $d(p, p') < \delta_2$. Let $\delta = \min\{\delta_1, \delta_2\}$, and $\{q_1, q_2, \dots, q_r\}$ be a δ -net in K , that is, $\forall p \in K$, there exists $q(p) \in \{q_1, \dots, q_r\}$ such that $d(p, q(p)) < \delta$. By the SLLN, there exist a P -null set N_2 and $n_2(\omega) < \infty$ $\forall \omega \notin N_2$ such that

$$(2.11) \quad \max_{i=1,2,\dots,r} |F_{n, \omega}(q_i) - F(q_i)| < \varepsilon'/3 \quad \forall n \geq n_2(\omega) \ (\omega \notin N_2).$$

Note that by (2.10), (2.11) and the fact that $|F(q(p)) - F(p)| < \varepsilon'/3 \ \forall p \in K$, one has

$$\begin{aligned} & \sup_{p \in K} |F_{n, \omega}(p) - F(p)| \\ & \leq \sup_{p \in K} |F_{n, \omega}(p) - F_{n, \omega}(q(p))| + \sup_{p \in K} |F_{n, \omega}(q(p)) - F(q(p))| \\ & \quad + \sup_{p \in K} |F(q(p)) - F(p)| \\ & < \varepsilon'/3 + \varepsilon'/3 + \varepsilon'/3 = \varepsilon' \quad \forall n \geq n(\omega) := \max\{n_1(\omega), n_2(\omega)\}, \end{aligned}$$

outside the P -null set $N_3 = N_1 \cup N_2$. This proves (2.8).

To complete the proof of (a), fix $\varepsilon > 0$. Let C be the (compact) Fréchet mean set of Q , $\ell := \min\{F(p) : p \in C\}$. Write $C^\varepsilon := \{p : d(p, C) < \varepsilon\}$. It is enough to show that there exist $\theta(\varepsilon) > 0$ and $n(\omega) < \infty \ \forall \omega$, outside a P -null set N such that

$$(2.12) \quad \begin{aligned} F_{n, \omega}(p) & \leq \ell + \theta(\varepsilon)/2 & \forall p \in C, \\ F_{n, \omega}(p) & \geq \ell + \theta(\varepsilon) & \forall p \in M \setminus C^\varepsilon, \ \forall n \geq n(\omega) \ (\omega \notin N). \end{aligned}$$

For (2.12) implies that $\min\{F_{n, \omega}(p) : p \in M\}$ is not attained in $M \setminus C^\varepsilon$ and, therefore, the Fréchet mean set of $\hat{Q}_{n, \omega}$ is contained in C^ε , provided $n \geq n(\omega)$ ($\omega \notin N$). To prove (2.12) we will first show that *there exist a compact set D containing C and $n_3(\omega) < \infty$ outside a P -null set N_3 such that both $F(p)$ and $F_{n, \omega}(p)$ are greater than $\ell + 1 \ \forall p \in M \setminus D$, for all $n \geq n_3(\omega)$ ($\omega \notin N_3$)*. If M is compact then this is trivially true, by taking $M = D$. So assume M is noncompact. Fix $p_0 \in C$ and use the inequality $d(x, q) \geq |d(q, p_0) - d(x, p_0)|$ to get

$$\int d^2(x, q) Q(dx) \geq \int \{d^2(q, p_0) + d^2(x, p_0) - 2d(q, p_0)d(x, p_0)\} Q(dx)$$

or

$$(2.13) \quad F(q) \geq d^2(q, p_0) + F(p_0) - 2d(q, p_0)F^{1/2}(p_0).$$

Similarly, using $\hat{Q}_{n,\omega}$ in place of Q ,

$$(2.14) \quad F_{n,\omega}(q) \geq d^2(q, p_0) + F_{n,\omega}(p_0) - 2d(q, p_0)F_{n,\omega}^{1/2}(p_0).$$

Since M is unbounded, one may take q at a sufficiently large distance Δ from C such that, by (2.13), $F(q) > \ell + 1$ on $M \setminus D$, where $D := \{q : d(q, C) \leq \Delta\}$. Since $F_{n,\omega}(p_0) \rightarrow F(p_0)$ a.s., by (2.14) one may find a P -null set N_3 and $n_3(\omega) < \infty$ such that $F_{n,\omega}(q) > \ell + 1$ on $M \setminus D \quad \forall n \geq n_3(\omega) \quad (\omega \notin N_3)$. This proves the italicized statement above.

Finally, let $D_\varepsilon := \{p \in D : d(p, C) \geq \varepsilon\}$. Then D_ε is compact and $\ell_\varepsilon := \min\{F(p) : p \in D_\varepsilon\} > \ell$, so that there exists $\theta = \theta(\varepsilon)$, $0 < \theta(\varepsilon) < 1$, such that $\ell_\varepsilon > \ell + 2\theta$. Now apply (2.8) with $K = D$ to find $n_4(\omega) < \infty$ outside a P -null set N_4 such that $\forall n \geq n_4(\omega)$, one has (i) $F_{n,\omega}(p) \leq \ell + \theta/2 \quad \forall p \in C$ and (ii) $F_{n,\omega}(p) > \ell + \theta \quad \forall p \in D_\varepsilon$. Since $F_{n,\omega}(p) > \ell + 1$ on $M \setminus D \quad \forall n \geq n_3(\omega)$ ($\forall \omega \notin N_3$), one has $F_{n,\omega}(p) > \ell + \theta \quad \forall p \in D_\varepsilon \cup (M \setminus D) = M \setminus C^\varepsilon$ if $n \geq n(\omega) := \max\{n_3(\omega), n_4(\omega)\}$ for $\omega \notin N$, where $N = N_3 \cup N_4$. This proves (2.12), and the proof of part (a) is complete.

Part (b) is an immediate consequence of part (a). \square

REMARK 2.5. A theorem of Ziezold (1977) for general separable (pseudo) metric spaces implies the conclusion of part (b) of Theorem 2.3 for *compact metric spaces* M , but not for noncompact M . In metric spaces such that all closed bounded subsets are compact, the present theorem provides (i) strong consistency for Fréchet sample means and (ii) *uniform convergence* to the Fréchet mean of Q of arbitrary measurable selections from the sample mean set. This applies to both intrinsic and extrinsic means of Q and \hat{Q}_n on manifolds.

REMARK 2.6. Under the hypothesis of Theorem 2.3(a), the Hausdorff distance between the intrinsic sample mean set and the intrinsic mean set does not in general go to 0, as the following example shows. Consider n independent random variables X_1, \dots, X_n with the same distribution on the unit circle, that is, absolutely continuous w.r.t. the uniform distribution. Then with probability 1, we may assume that for $i \neq j$, $X_i \neq X_j$. Assume $X_j = e^{i\theta_j}$ and let $X_j^* = e^{i\theta_j^*} = -X_j$, where the arguments θ_j^* are in the increasing order of their indices. $F(e^{i\theta})$ is periodic with period 2π and is a piecewise quadratic function; on each interval $[\theta_j^*, \theta_{j+1}^*]$, $F(e^{i\theta}) = \sum_{k=1}^n (2\pi \varepsilon_{k,j} + (-1)^{\varepsilon_{k,j}} (\theta - \theta_k))^2$ where $\varepsilon_{k,j} \in \{0, 1\}$. Therefore, the points of local minima have the form $\frac{1}{n} \sum_{k=1}^n (\theta_k + 2\pi \varepsilon_{j,k} (-1)^{\varepsilon_{j,k}})$ and each local minimum value $m_j = m_j(\theta_1, \theta_2, \dots, \theta_n)$ is a quadratic form in $\theta_1, \dots, \theta_n$. Since $\varepsilon_{k,j} \in \{0, 1\}$, there are at most 2^n such possible distinct quadratic polynomials. Given that each of the variables θ_j^* is continuous, the

probability that there is a fixed pair of indices $i \neq j$, such that $m_i(\theta_1, \dots, \theta_n) = m_j(\theta_1, \dots, \theta_n)$ is 0. This shows that, with probability 1, all the local minima are distinct and the intrinsic sample mean exists. On the other hand, the intrinsic mean set of the uniform measure on the circle is the whole circle, proving that in this case the Hausdorff distance between the intrinsic sample mean set and the intrinsic mean set is π with probability 1.

REMARK 2.7. The computation of the intrinsic mean set of a probability measure on a nonflat manifold M often involves nonstandard numerical algorithms, even if M has a Riemannian metric of maximum degree of mobility. For this reason, in the next section we will focus on a different approach to indices of location of probability measures on manifolds.

3. Extrinsic means of distributions on submanifolds. Since most of the literature on directional and shape analysis is concerned with parametric inference (e.g., MLEs, likelihood ratios, etc.), there has not been much emphasis on intrinsic analysis. For purposes of nonparametric or semiparametric inference, however, statistical analysis of intrinsic indices such as the intrinsic mean is very important. But it is generally not easy to prove the existence (i.e., uniqueness) of the intrinsic mean. Also intrinsic means, when they exist, are often very difficult to compute. On the other hand, a manifold can be also looked at as a submanifold of some Euclidean space, and a probability measure on it can be regarded as a probability measure in that ambient linear space. Such an approach has been employed in directional analysis in Mardia and Jupp (1999), Watson (1983) and Fisher, Hall, Jing and Wood (1996), and in shape analysis in Kent (1992), Dryden and Mardia (1993) and Le (1998).

In this section we give a general treatment of the notion of an extrinsic mean, and of statistical inference for it. We will also show that under special structures of invariance and symmetry the intrinsic and extrinsic means coincide, and therefore the extrinsic sample mean, which is easier to compute, can be used as a consistent estimator of the intrinsic sample mean.

Assume M is a closed submanifold of the Euclidean space $\mathbb{E}^k = (\mathbb{R}^k, d_0)$ where d_0 denotes the Euclidean distance, $d_0(x, y) = \|y - x\|$. Let Q be a probability measure on M . Let G^c be the set of nonfocal points of M in E^k . The *projection map* $P_M : G^c \rightarrow M$ is defined as $P_M(p) = x$ if $d_0(p, M) = d_0(p, x)$.

In this case, the Fréchet function is defined on M by

$$(3.1) \quad F_0(p) = \int_M \|p - x\|^2 Q(dx).$$

DEFINITION 3.1. The *extrinsic mean set* of Q is the set of all minimizers of F_0 on M . If there is a unique minimizer, this is called the *extrinsic mean* of Q and will be labeled $\mu_E(Q)$.

PROPOSITION 3.1. *Assume μ is the mean of Q as a probability measure on \mathbb{R}^k . Then (a) the extrinsic mean set is the set of all points $m \in M$, with $d_0(\mu, m) = d_0(\mu, M)$, and (b) if $\mu_E(Q)$ exists then μ exists and is nonfocal and $\mu_E(Q) = P_M(\mu)$.*

PROOF. (a) If $p, x \in M$, then $\|p - x\|^2 = \|p - \mu\|^2 + 2\langle p - \mu, \mu - x \rangle + \|\mu - x\|^2$ and if we integrate this identity over M w.r.t. Q , given that $\int_M x Q(dx) = \int_{\mathbb{R}^k} x Q(dx) = \mu$, we get

$$(3.2) \quad F_0(p) = \|p - \mu\|^2 + \int_M \|\mu - x\|^2 Q(dx).$$

In particular, for any points $p, m \in M$, $F_0(p) - F_0(m) = d_0^2(\mu, p) - d_0^2(\mu, m)$ and (a) follows by selecting m to be a minimizer of F_0 . (b) If $\mu_E(Q)$ exists then μ exists and from part (a) it follows that the distance from an arbitrary point on M to μ has the unique minimizer $\mu_E(Q)$, that is, μ is nonfocal and since $d_0(\mu_E(Q), \mu) = d_0(\mu, M)$, $\mu_E(Q) = P_M(\mu)$. \square

THEOREM 3.2. *The set of focal points of a submanifold M of E^k is a closed subset of \mathbb{E}^k of Lebesgue measure 0.*

PROOF. A point p is nonfocal, with $d_0(p, M) = r$, if and only if the (hyper)sphere $S(p, r)$ of radius r centered at p has a unique point x in common with M . In this case the interior of the ball $B(p, r)$ is included in $\mathbb{E}^k \setminus M$ and $T_x M \subseteq T_x S(p, r)$; x is the point of absolute minimum of the function L_p defined on M by $L_p(y) = d_0^2(p, y)$. Let $u = (u^1, \dots, u^d)$ be coordinates of points $y = y(u)$ on M , with $y(0) = x$. In Milnor [(1963), page 36] it is shown that x is a degenerate critical point of L_p if and only if p is a focus. Moreover, from the computations in Milnor [(1963), page 35] it follows that if K_1, K_2, \dots, K_s are the nonzero principal curvatures of M at the point x and $|t| < \min\{|K_1|^{-1}, |K_2|^{-1}, \dots, |K_s|^{-1}\}$ for any unit vector v in $v_x M$, the normal space at the point x , the matrix

$$((\partial y / \partial u^i(0))(\partial y / \partial u^j(0)) - t v(\partial^2 y / \partial u^i \partial u^j(0)))$$

is positive definite. In particular, since $r < \min\{|K_1|^{-1}, |K_2|^{-1}, \dots, |K_s|^{-1}\}$, the matrix

$$((\partial y / \partial u^i(0))(\partial y / \partial u^j(0)) - (p - y(0))(\partial^2 y / \partial u^i \partial u^j(0)))$$

is positive definite. There are a neighborhood N of p and an open neighborhood U of 0 such that, for any $u \in U$ and $q \in N$, the matrix of the second partial derivatives of $L_q(y(u))$, namely,

$$((\partial y / \partial u^i(u))(\partial y / \partial u^j(u)) - (q - y(u))(\partial^2 y / \partial u^i \partial u^j(u)))$$

is positive definite. Since the manifold topology of M coincides with the induced topology, one may assume that there is a ball $B(x, \varepsilon)$, such that $y(U) = M \cap B(x, \varepsilon)$. Let ε be as small as necessary. Since x is the only common point of M and $S(p, r)$, and the set $M \setminus \text{Int } B(x, \varepsilon)$ is closed, there is a number δ , $r > \delta > 0$, such that $d_0(p, M \setminus \text{Int } B(x, \varepsilon)) = r + \delta$. Let $q \in \text{Int } B(p, \delta/2)$ and $z \in M \setminus \text{Int } B(x, \varepsilon)$. Then $d_0(q, z) > |d_0(q, p) - d_0(p, z)| > r + \delta - \delta/2 > d_0(q, x)$. It follows that $d_0(q, M) = d_0(q, M \cap \text{Int } B(x, \varepsilon))$. If $y \in \text{Int } B(x, \varepsilon) \setminus \{x\}$ is such that $d_0^2(q, y) = d_0^2(q, M)$, it follows by the positive definiteness of the displayed matrix above that y is an isolated point of minimum of L_q , proving that the set of nonfocal points is open.

Let $G(\infty)$ be the set of foci of M , and let G be the set of focal points. It is known [Milnor (1963), page 33] that $G(\infty)$ has Lebesgue measure zero. If x is a point on M , we define $G(x)$ to be the set of all points f in E^k such that there is at least another point $x' (\neq x)$ on M with $d_0(x, f) = d_0(x', f) = d_0(f, M)$. Another description of $G(x)$ is as the set of all centers f of spheres of E^k that are tangent to M at least at two points, one of which is x , and whose interiors are disjoint from M . The tangent space $T_x M$ is included in the tangent space at x to such a sphere. Therefore the normal line at x to such a sphere is included in the normal space $\nu_x M$, which means that a point in $G(x)$ is in $\nu_x M$. We show that on each ray starting at x in $\nu_x M$ (x is the zero element, if $\nu_x M$ is regarded as a vector space) there is at most one point in $G(x)$. Indeed if f_1, f_2 are two distinct points on such a ray starting at x , assume f_1 is closer to x than f_2 . Let x', x'' be such that $d_0(x', f_1) = d_0(x, f_1) = d_0(f_1, M)$, $d_0(x'', f_2) = d_0(x, f_2) = d_0(f_2, M)$. Then x' is a point of M in the interior of $S(f_2, d_0(f_2, M))$, a contradiction. Given that $G(x)$ intersects the radii coming out of x in $\nu_x M$ at most at one point, the Lebesgue measure of $G(x)$ in $\nu_x M$ is zero.

Let NM be the disjoint union of $\nu_x M$, $x \in M$. NM is the normal bundle of M and it is a manifold of dimension k . We define the map $N : NM \rightarrow \mathbb{R}^k$ by $N(x, \nu_x) = x + \nu_x$. One may show [Milnor (1963)] that the critical values of N are the foci of M . Therefore if $f = N(x, \nu_x)$ is a focal point that is not a focus, f is a regular value of N . Thus, if λ represents the Lebesgue volume form in \mathbb{R}^k , then $N^* \lambda$ is a volume form on $N^{-1}(\mathbb{R}^k \setminus G(\infty))$, and the Lebesgue measure of $G \setminus G(\infty)$ is

$$\lambda(G \setminus G(\infty)) \leq \int_{NM \setminus G(\infty)} N^* \lambda.$$

If we apply Fubini's theorem integrating over the base M the integral in each fiber (normal space $\nu_x M$), we see that the integrand in $\nu_x M$ is a volume form that is a multiple $C(x)$ of the Lebesgue measure in $\nu_x M$. Therefore

$$\int_{NM \setminus G(\infty)} N^* \lambda = \int_M C(x) \left(\int_{G(x)} \lambda_x(dv) \right) \text{vol}_M(dx),$$

which is zero since

$$\int_{G(x)} \lambda_x(dv) = \lambda_x(G(x)) = 0. \quad \square$$

Now we extend the notion of extrinsic means to embeddings of manifolds.

DEFINITION 3.2. Assume Q is a probability measure on M and $j: M \rightarrow \mathbb{R}^k$ is an embedding, such that $j(M)$ is closed. We say that Q is *nonfocal w.r.t. j* if Q regarded as a measure $j(Q)$ on \mathbb{R}^k has a mean $\mu(j(Q)) = \int_{\mathbb{R}^k} x(j(Q))(dx)$ which is a nonfocal point of $j(M)$. The extrinsic mean of a probability measure Q which is nonfocal w.r.t. j is $\mu_E^j(Q) := j^{-1}(P_{j(M)}(\mu(j(Q))))$.

Since the embedding j is assumed to be given, we will normally drop the superscript j and write $\mu_E(Q)$ for $\mu_E^j(Q)$.

DEFINITION 3.3. A *Riemannian embedding* is an embedding $j: M \rightarrow \mathbb{R}^k$ which pulls back the induced Riemannian structure on $j(M)$ to the Riemannian structure of M . A Riemannian embedding is said to be *equivariant* at a point p of M , if every isometry of $j(M)$ that keeps $j(p)$ fixed is the restriction of an Euclidean isometry. A *two point homogeneous space* is a Riemannian manifold such that for each two pairs of its points $(p, q), (p', q')$ with $d_g(p, q) = d_g(p', q')$, there is an isometry τ with $\tau(p) = p'$ and $\tau(q) = q'$.

To simplify notation, we will often write p for $j(p)$ and M for $j(M)$, in case of a Riemannian embedding j . It is known that M is a two-point homogeneous space if and only if, for each $p \in M$, the *isotropy group* H_p of all isometries of M which keep p fixed is *transitive* on every *geodesic sphere* $S(p, r) := \{x \in M : d_g(x, p) = r\}$, $r > 0$ [Chavel (1993), page 147]. That is, given $q, q' \in S(p, r)$, there exists $h \in H_p$ such that $h(q) = h(q')$.

The following theorem links the intrinsic mean of Q on a Riemannian manifold with its extrinsic mean under an embedding which is equivariant at a point p . Note that for every $h \in H_p$ the *differential* dh maps $T_p M$ into itself.

THEOREM 3.3. *Let $j: M \rightarrow \mathbb{R}^k$ be a Riemannian embedding which is equivariant at p . Assume that $0 \equiv (p, 0)$ is the only fixed point of $T_p M$ under the family of maps $\{dh : h \in H_p\}$. Assume also that Q is a probability measure on M which is invariant under H_p , and $\mu(j(Q))$ is finite and nonfocal. (a) Then either $\mu_E(Q) = p$ or $\mu_E(Q) \in C(p)$, the cut locus of p . The same holds for the intrinsic mean $\mu_I(Q)$ if it exists. (b) If, in addition to the hypothesis above, M is a compact two point homogeneous space other than the round sphere and $\mu_I(Q)$ exists, then $\mu_I(Q) = \mu_E(Q) = p$.*

PROOF. (a) The mean $\mu(j(Q))$ of $j(Q)$, regarded as a measure on the ambient Euclidean space, is invariant under each Euclidean isometry \hat{h} , say, which extends $h \in H_p$. For Q , as a measure on the Euclidean space, is invariant under $\hat{h} \forall h \in H_p$, due to the equivariant embedding at p and the invariance of Q on M under H_p . It now follows that $\mu_E(Q)$ is invariant under H_p . Suppose now that $\mu_E(Q) \neq p$. We will show that in that case $\mu_E(Q) \in C(p)$. If this is not so then $\mu_E(Q) \in \exp_p(V(p))$. Then there exists a unique minimizing geodesic joining p and $\mu_E(Q)$. Because of uniqueness this geodesic, say, γ , is left invariant by the isometries $h \in H_p$. Then $\dot{\gamma}(0)$ is invariant under $dh \forall h \in H_p$, contradicting the hypothesis that 0 is the only invariant vector in T_pM under $\{dh : h \in H_p\}$.

Suppose next that $\mu_I(Q)$ exists. Then $\mu_I(Q)$ is invariant under H_p , since $F(y) = F(hy)$, $\forall h \in H_p$, due to the invariance of Q under H_p . The same argument as above now shows that either $\mu_I(Q) = p$ or $\mu_I(Q) \in C(p)$.

(b) It follows from a classification theorem due to Wang (1952) that besides the round spheres, there are only four types of two-point homogeneous spaces, namely, the real projective spaces, complex projective spaces, quaternionic projective spaces and the Cayley projective planes [see also Helgason (1978), page 535]. It is known from Warner (1965) that for any point $p \in M$, $C(p)$ is a strong deformation retract of $M \setminus \{p\}$, and, in particular, $C(p)$ has the homotopy type of $M \setminus \{p\}$. On the other hand, if M is one of the two-point homogeneous spaces other than a sphere given by Wang's classification, then the cohomology of $M \setminus \{p\}$ is not trivial. This shows that in this case $M \setminus \{p\}$ is not homotopically trivial and therefore $C(p)$ is also not homotopically trivial. This implies that if M is not a sphere, $C(p)$ has at least two points q, q' . Moreover, since the isotropy group H_p is transitive on the geodesic sphere $S(p, d_g(p, q))$, we may assume that $d_g(p, q) = d_g(p, q') = r$. Hence if $\mu_E(Q) \in C(p)$ there exists $q' \in C(p) \setminus \{\mu_E(Q)\}$ such that $d_g(p, \mu_E(Q)) = d_g(p, q')$. By the transitivity of H_p on $S(p, r)$, there exists $h \in H_p$ such that $h(\mu_E(Q)) = q'$, contradicting the invariance of $\mu_E(Q)$. By (a), $\mu_E(Q) = p$.

The same argument applies to $\mu_I(Q)$ if it exists. \square

EXAMPLE 3.1. Let Q be a probability measure on a sphere, with $\mu(j(Q)) \neq 0$, such that the group leaving Q invariant is the stabilizer of a given point p . Then $\mu_E(Q)$ is either p or the antipodal point of p on the sphere. The same is true of $\mu_I(Q)$ if it exists. Such examples of probability distributions are given in Watson [(1983), page 136] and Fisher (1993), including the von Mises distributions. Another example of an invariant distribution is given by the *Dimroth–Watson distribution* on the real projective plane $\mathbb{R}P^2$, whose Radon–Nykodim derivative at the point $[x]$ w.r.t. the volume measure of a constant curvature Riemannian structure on $\mathbb{R}P^2$ is proportional to $\exp[k(p \cdot x)^2]$, $x \in S^2$, and is $O(2)$ invariant. A general $O(2)$ invariant measure with a density on $\mathbb{R}P^2$ has the Radon–Nykodim derivative w.r.t. the volume form at the point $[x]$ proportional to $f((p \cdot x)^2)$, $x \in S^2$,

where f is a density of an absolutely continuous positive measure on a finite interval. An example of equivariant embedding of $\mathbb{R}P^2$ furnished with a Riemannian structure with constant curvature into the space of symmetric matrices $S(3, \mathbb{R})$ is provided by the Veronese-like map $j[u] = uu^t$. The Euclidean distance d_0 on $S(3, \mathbb{R})$ is given by $d_0^2(A, B) := \text{Tr}((A - B)(A - B))$. As such if u, v are in \mathbb{S}^2 , $d_0^2(j[u], j[v]) = \text{Tr}(uu^t - vv^t)(uu^t - vv^t) = \text{Tr}(uu^t uu^t - 2uu^t vv^t + vv^t vv^t) = 2(1 - (u \cdot v)^2)$. The fact that the embedding is equivariant follows from the action of isometries of $O(3)$ on $S(3, \mathbb{R})$, by simultaneous left and right multiplication. By Theorem 3.3(b), if the intrinsic mean direction of a rotationally invariant measure on $\mathbb{R}P^2$ (the space of directions in the three-dimensional Euclidean space) exists, it is the same as the extrinsic mean. Note that proportional distances yield the same Fréchet mean set, and therefore the intrinsic mean sets of a probability measure on $\mathbb{R}P^2$ obtained after scaling \mathbb{S}^2 to different radii are all the same. Finally, note that Kobayashi (1968) gave a general construction of an isometric embedding of a compact symmetric space, which can be used to provide an equivariant embedding of any two-point homogeneous space (including the Cayley plane) into a Euclidean space.

DEFINITION 3.4. Assume $X = (X_1, \dots, X_n)$ are i.i.d. M -valued random variables whose common distribution is a nonfocal measure Q on (M, j) and the function $p \rightarrow \sum_{r=1}^n \|j(p) - j(X_r)\|^2$ has a unique minimizer on M ; this minimizer is the *extrinsic least squares sample mean*. If the mean $\overline{j(X)}$ of the sample $j(X) = (j(X_1), \dots, j(X_n))$ is a nonfocal point, the *extrinsic sample mean* is

$$(3.3) \quad \overline{X}_E := j^{-1}(P_{j(M)}(\overline{j(X)})) \equiv \mu_E^j(\hat{Q}_n),$$

where $\hat{Q}_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$ is the empirical distribution.

From now on, we will occasionally omit the embedding, that is we assume M is a submanifold of the Euclidean space and j is the inclusion map. To ease notational complexity in this case, we will often write X_i for $j(X_i)$ and \overline{X} for $\overline{j(X)} = \frac{1}{n} \sum_{i=1}^n j(X_i)$.

THEOREM 3.4. Assume Q is a nonfocal probability measure on the manifold M and $X = \{X_1, \dots, X_n\}$ is a random sample from Q . (a) If the sample mean \overline{X} is a nonfocal point then the least squares sample mean equals the extrinsic sample mean \overline{X}_E . (b) \overline{X}_E is a strongly consistent estimator of $\mu_E(Q)$.

PROOF. (a) If \overline{X} is a nonfocal point then by Proposition 3.1, applied to the empirical \hat{Q}_n , the extrinsic least squares sample mean is the extrinsic sample mean. (b) By the SLLN, \overline{X} converges to $\mu[j(Q)]$ almost surely. Since F^c is open, by Proposition 3.2, and the projection P_M from F^c to M is continuous, $j^{-1}(P_M(\overline{X}))$ converges to $\mu_E(Q)$ almost surely. \square

In particular, from Theorem 3.4 we get the following consequence.

REMARK 3.1. If Q is focal, the extrinsic mean set has at least two points. Therefore by Theorem 2.3(a) the extrinsic sample mean set may have more than one point, and a selection from the extrinsic sample mean set sequence may not have a limit.

COROLLARY 3.5. Assume M , Q and the equivariant embedding j are as in Theorem 3.3(b). Then the extrinsic least squares sample mean is a strongly consistent estimator of the intrinsic mean of Q .

We now consider a method for constructing confidence regions for extrinsic means $\mu_E(Q)$ on regular submanifolds. For the formulas below we omit summation symbols for any repeated index in the same monomial.

Let H be the projection on the affine subspace $\mu_E(Q) + T_{\mu_E}M$. We would like to determine the asymptotic distribution of $H(\bar{X})$. While $H(\bar{X})$ is not the same as $P_M(\bar{X})$, its asymptotic distribution is easier to compute. For large samples the extrinsic sample mean is close to the extrinsic mean and, therefore, $H(\bar{X})$ and $P_M(\mu)$ will be close to each other. When M is a linear variety, the two maps coincide. Thus for concentrated data the delta method for H gives a good estimate of the distribution of the extrinsic sample mean. Assume that around $P_M(\mu)$ the implicit equations of M are $F^1(x) = \dots = F^c(x) = 0$, where F^1, \dots, F^c are functionally independent. Then $\bar{X} - H(\bar{X})$ is in $\nu_{P_M(\mu)}M$, the orthocomplement of $T_{P_M(\mu)}M$; thus it is a linear combination of the gradients $\text{grad}_{P_M(\mu)} F^1, \dots, \text{grad}_{P_M(\mu)} F^c$. We need to evaluate the differential of the map H at μ , in terms of F^1, \dots, F^c . Set $\nu_\alpha = \|\text{grad}_{P_M(\mu)} F^\alpha\|^{-1} \text{grad}_{P_M(\mu)} F^\alpha$ ($\alpha = 1, \dots, c$) and

$$h_{\alpha\beta}(\mu) = \nu_\alpha \nu_\beta,$$

$$(h^{\alpha\beta}(\mu))_{\alpha,\beta=1,\dots,c} = ((h_{\alpha\beta}(\mu))_{\alpha,\beta=1,\dots,c})^{-1}.$$

Then $x - H(x) = t^\beta(x, \mu) \nu_\beta$ where $t^\beta(x, \mu) = h^{\alpha\beta}(\mu)(x - P_M(\mu)) \nu_\alpha$. Therefore, $H(x) = x + h^{\alpha\beta}(\mu)((P_M(\mu) - x) \nu_\alpha) \nu_\beta$, $d_\mu H(v) = v - h^{\alpha\beta}(\mu)(v \nu_\alpha) \nu_\beta$, that is,

$$(3.4) \quad G_i^j = \frac{\partial H^j}{\partial x^i}(\mu) = \delta_{ij} - h^{\alpha\beta}(\mu) \nu_\alpha^i \nu_\beta^j,$$

where $\delta_{ij} = 1$ or 0 according as $i = j$ or $i \neq j$. By the delta method we arrive at the following theorem.

THEOREM 3.6. Let $\{X_k\}_{k=1,\dots,n}$ be a random sample from a nonfocal distribution Q on the submanifold M , given in a neighborhood of $\mu_E(Q)$ by the equations $F^1(x) = \dots = F^c(x) = 0$. Assume Q has mean μ and covariance matrix Σ as a distribution in the ambient numerical space. If G is the matrix given by (3.4), then $n^{1/2}(H(\bar{X}) - P_M(\mu))$ converges weakly to $N(0, G\Sigma G^t)$ in the tangent space of M at the extrinsic mean $\mu_E(Q) = P_M(\mu)$ of Q .

Sometimes the matrix $\Gamma = G\Sigma G^t$ may be difficult to compute and one may use nonpivotal bootstrap, that is, Efron's percentile bootstrap to obtain a confidence region for $\mu_E(Q)$. We state this as follows [see Efron (1982)]:

COROLLARY 3.7. *Under the hypothesis of Theorem 3.6, one may construct an asymptotic $(1 - \alpha)$ -confidence region for $\mu_E(Q) = P_M(\mu)$, using the bootstrapped statistic $n^{1/2}(\overline{H}(\overline{X}^*) - \overline{H}(\overline{X}))$. Here \overline{H} is the projection on the affine subspace $\overline{X}_E + T_{\overline{X}_E}M$ and \overline{X}^* is the mean of a random sample with repetition of size n from the empirical \hat{Q}_n considered as a probability measure on the Euclidean space in which M is embedded.*

REMARK 3.2. Suppose F is finite and Q is nonfocal. By Theorem 3.2, there exists $\delta > 0$ such that \overline{X} is nonfocal if $\|\overline{X} - \mu\| < \delta$. Since $P(\|\overline{X} - \mu\| \geq \delta) = O(n^{-1})$, one may define \overline{X}_E to be any measurable selection from the sample extrinsic mean set if \overline{X} is focal. Theorem 3.6 and a corresponding version of Corollary 3.7 hold for this \overline{X}_E .

EXAMPLE 3.2. Let $M = S^d$, j the usual embedding (inclusion) in \mathbb{R}^{d+1} and Q a nonfocal probability measure on S^d , that is, $\mu = \int_{\mathbb{R}^{d+1}} x Q(dx) \neq 0$. Let $m = P_M\mu$. Then $H(\overline{X}) - m = \overline{X} - m - \{(\overline{X} - m) \cdot m\}m = \overline{X} - (\overline{X} \cdot m)m = \overline{X} - \mu - \{(\overline{X} - \mu) \cdot m\}m$. Hence $\sqrt{n}(H(\overline{X}) - m)$ converges in distribution to a d -dimensional normal distribution supported by the tangent space $T_m S^d$ identified with $\{x \in \mathbb{R}^{d+1} : xm = 0\}$. As a measure on \mathbb{R}^{d+1} this normal distribution has mean 0 and covariance matrix $\Gamma := \Sigma + (m^t \Sigma m)mm^t - 2\Sigma(mm^t)$, where Σ is the covariance matrix of Q viewed as a measure on \mathbb{R}^{d+1} . An asymptotic $(1 - \alpha)$ -confidence region for m may now be constructed using the estimate of Γ obtained by replacing in its expression (i) Σ by the sample covariance matrix S and (ii) m by $\overline{X}/|\overline{X}|$. Alternatively, one may use the bootstrap procedure of Corollary 3.7.

4. Means of distributions on axial spaces, planar shape spaces and their Veronese–Whitney embeddings. The space of all directions in \mathbb{R}^N , or axial space, is an $(N - 1)$ -dimensional real projective space. It is the space $\mathbb{R}P^{N-1}$ of equivalence classes on a round sphere in \mathbb{R}^N with antipodal points identified. As such this space carries a Riemannian structure of constant positive curvature, since the antipodal map is an isometry of the round sphere. This is the space of elliptic geometry, and the total group of isometries $SO(N)$ has maximum mobility. Given that this elliptic space is locally isometric with a round sphere, if the support of a distribution w.r.t. the stabilizer of a point [which is a subgroup of $SO(N)$ isomorphic to $SO(N - 1)$] has small diameter, then the intrinsic mean exists (see Remark 2.1).

In this section we would like to consider the general situation when the distribution is not concentrated. Note that for N odd, $\mathbb{R}P^{N-1}$ cannot be embedded in \mathbb{R}^N .

Usually, in directional statistics, one regards an axial distribution as one corresponding to an \mathbb{S}^{N-1} -valued random variable X such that X and $-X$ have the same distribution [Watson (1983), Chapter 5; Fisher, Hall, Jing and Wood (1996); and Beran and Fisher (1998)]. One may show that the Veronese–Whitney map defined in Section 3 for $N = 3$, and given for arbitrary N by the same formula $j([u]) = u u^t$ ($\|u\| = 1$), is an equivariant embedding of $\mathbb{R}P^{N-1}$ into a $\frac{1}{2}N(N+1)$ -dimensional Euclidean space. To see this, let $S(N, \mathbb{R})$ denote the set of all real $N \times N$ symmetric matrices. Since the Euclidean distance d_0 between two symmetric matrices is

$$(4.1) \quad d_0(A, B) = \text{Tr}((A - B)(A - B)^t) = \text{Tr}((A - B)^2),$$

the group $O(N)$ acts as a group of isometries of $(S(N, \mathbb{R}), d_0)$ via

$$(4.2) \quad T(A) = T A T^t$$

and leaves $M = j(\mathbb{R}P^{N-1})$ invariant. It is known that $O(N)$ acts transitively on \mathbb{S}^{N-1} , that is, if $u, v \in \mathbb{R}^N$, $\|u\| = \|v\| = 1$, there is a $T \in O(N)$ such that $Tu = v$. Then $T(j[u]) = T u u^t T^t = v v^t = j([v])$, showing that $O(N)$ acts transitively on M . The stabilizer of this action is $O(N-1)$. Therefore $\mathbb{R}P^{N-1}$ with the Riemannian metric induced by j is a homogeneous space, with a group of isometry of largest dimension. From Theorem 3.1 in Kobayashi (1972), it turns out that with this metric $\mathbb{R}P^{N-1}$ has constant positive curvature, and j is an equivariant embedding of $\mathbb{R}P^{N-1}$ into $S(N, \mathbb{R})$. Also $M = j(\mathbb{R}P^{N-1})$ is included in the set $S_+(N, \mathbb{R})$ of symmetric nonnegative definite N by N real matrices. $S_+(N, \mathbb{R})$ is convex, so the mean under Q of matrices in $S_+(N, \mathbb{R})$ is a matrix in $S_+(N, \mathbb{R})$. Therefore, we are interested in determining only the projection of a semipositive matrix on M . If A is in $S(N, \mathbb{R})$ and T is an orthogonal matrix, then $d_0(A, M) = d_0(T(A), M)$. Given A in $S_+(N, \mathbb{R})$, there is T in $O(N)$ such that $T(A) = \text{diag}(\eta_a)_{a=1, \dots, N} = D$, and the entries of D are all nonnegative, in increasing order. Let $x = (x^a)$ be a unit vector in \mathbb{R}^N . After elementary computations we get

$$(4.3) \quad d_0^2(D, j([x])) = 1 + \sum \eta_a^2 - 2 \sum \eta_a (x^a)^2 \geq d_0^2(D, j([e_N])),$$

where e_N is the eigenvector of D of unit length corresponding to the highest eigenvalue. Note that if η_N has multiplicity two or more, then for any $t \in [0, 2\pi]$ and for any unit vector $x = (x^a) \in \mathbb{R}^N$, we have

$$d_0^2(D, j[x]) \geq d_0^2(D, j[\cos t e_{N-1} + \sin t e_N]) = d_0^2(D, j[e_N])$$

and D is focal. If η_N is simple, that is, has multiplicity one, then $d_0^2(D, j[x]) \geq d_0^2(D, j[e_N])$ and the equality holds only if $[x] = [e_N]$. In this last case, D is a nonfocal and $P_M(D) = j([e_N])$. We will call such an eigenvector of length 1 a *highest eigenvector* of D . One then obtains the following:

PROPOSITION 4.1. *The set F of the focal points of $M = j(\mathbb{R}P^{N-1})$ in $S_+(N, \mathbb{R})$ is contained in the set of matrices in $S_+(N, \mathbb{R})$ whose largest eigenvalues are of multiplicity at least 2. The projection $P_M: S_+(N, \mathbb{R}) \setminus F \rightarrow M$ associates to each nonnegative definite symmetric matrix A with a highest eigenvalue of multiplicity one, the matrix $j([m])$ where m is a highest unit eigenvector of A .*

If Q is a probability measure on $\mathbb{R}P^{N-1}$, assume $[X], \|X\| = 1$ is a $\mathbb{R}P^{N-1}$ -valued random variable with distribution Q . As a consequence of Corollary 3.5 and Proposition 4.1 we obtain the theorem:

THEOREM 4.2. *Assume $[X_r], \|X_r\| = 1, r = 1, \dots, n$, is a random sample from a probability measure Q on $\mathbb{R}P^{N-1}$. Then (a) Q is nonfocal if the highest eigenvalue of $E[X_1 X_1^t]$ is simple and in this case $\mu_E(Q) = [m]$, where m is a unit eigenvector of $E[X_1 X_1^t]$ corresponding to this eigenvalue. (b) Under the hypothesis of (a) the extrinsic sample mean $[\bar{X}]_E$ is a strongly consistent estimator of $\mu_E(Q)$.*

Note that when it exists, $[\bar{X}]_E$ is given by $[\bar{X}]_E = [m]$, where m is a unit eigenvector of $S_n = \frac{1}{n} \sum_{r=1}^n X_r X_r^t$. It may be noted that in this case $[\bar{X}]_E$ is also the MLE for the mean of a Bingham distribution [Prentice (1984) and Kent (1992)] and for the mean of the Dimroth–Watson distribution, whose density function at $[x]$ is proportional to $\exp(k(\mu \cdot x)^2)$, where k is a constant. For these or more general parametric families, MLE asymptotics or bootstrap methods [Fisher and Hall (1992)] are commonly used. Nonparametric techniques of estimation of extrinsic means will be presented in a forthcoming second part of this article.

We turn now to planar shape spaces [see Kendall (1984)].

DEFINITION 4.1. *A planar k -ad is an ordered set (z_1, z_2, \dots, z_k) of k points in the Euclidean plane at least two of which are distinct. Two k -ads (z_1, z_2, \dots, z_k) and $(z'_1, z'_2, \dots, z'_k)$ are said to have the *same shape* if there is a direct similarity T in the plane, that is, a composition of a rotation, a translation and a homothety such that $T(z_j) = z'_j$ for $j = 1, \dots, k$. Having the same shape is an equivalence relationship in the space of planar k -ads, and the set of all equivalence classes of k -ads is called the *planar shape space of k -ads*, or the space \sum_2^k . Without loss of generality one may assume that two k -ads that have the same shape also have the same center of mass, that is, $\sum z_j = \sum z'_j = 0$, and they have the same shape if there is a composition of a transformation T which keeps the origin fixed, and is a rotation followed by a homothety such that $T(z_j) = z'_j$ for $j = 1, \dots, k-1$. Such a transformation T is determined by a nonzero complex number, that is to say, the two k -ads with center of mass 0 have the same shape if there is a $z \in \mathbb{C} \setminus \{0\}$ such that $zz_j = z'_j$ for $j = 1, \dots, k-1$. Thus the shape equivalence class of a planar*

k -ad is uniquely determined by a point in $\mathbb{C}P^{k-2}$, that is to say, Σ_2^k is identified with $\mathbb{C}P^{k-2}$. Kendall (1984) pointed out that there is no unique way to identify Σ_2^k with $\mathbb{C}P^{k-2}$ and indeed our method of identification differs from Kendall's method.

Tests appropriate for mismatch of shapes of k -ads were introduced in Sibson (1978) based on the so-called *Procrustes statistic*. The *Procrustean distance*, in our terminology, is the distance induced by the Euclidean distance on $\mathbb{C}P^{k-2}$ via a quadratic Veronese–Whitney embedding into a unit sphere of the linear space $S(k-1, \mathbb{C})$ of *selfadjoint complex matrices of order $k-1$* . In order to define $j: \mathbb{C}P^{k-2} \rightarrow S(k-1, \mathbb{C})$ it is useful to note that $\mathbb{C}P^{k-2} = \mathbb{S}^{2k-3}/\mathbb{S}^1$, where \mathbb{S}^{2k-3} is the space of complex vectors \mathbb{C}^{k-1} of norm 1, and the equivalence relation on \mathbb{S}^{2k-3} is by multiplication with scalars in \mathbb{S}^1 (complex numbers of modulus 1). If $z = (z^1, z^2, \dots, z^{k-1})$ is in \mathbb{S}^{2k-3} , we will denote by $[z]$ the equivalence class of z in $\mathbb{C}P^{k-2}$. The *Veronese–Whitney* (or simply *Veronese*) map is in this case $j([z]) = z z^*$ where, if z is considered as a column vector, z^* is the adjoint of z , that is, the conjugate of the transpose of z . The Euclidean distance in the space of Hermitian matrices $S(k-1, \mathbb{C})$ is $d_0^2(A, B) = \text{Tr}((A - B) \times (A - B)^*) = \text{Tr}((A - B)^2)$.

Kendall (1984) (see his Theorem 1) has shown that the Euclidean distance in $S(k-1, \mathbb{C})$ induces via j a Riemannian structure on $\mathbb{C}P^{k-2}$, which is known in literature as the *Fubini–Study metric* and has a highest dimensional group of isometries on $\mathbb{C}P^{k-2}$ among all the Riemannian metrics on this manifold. The isometry group is the special unitary group $SU(k-1)$ of all $(k-1) \times (k-1)$ complex matrices A , with $A^*A = I$, $\det(A) = 1$. By analogy with the action of the orthogonal group in the real projective space, one may show that the group $SU(k-1)$ acts transitively as a group of isometries and up to a scaling factor, j is an equivariant embedding of $\mathbb{C}P^{k-2}$ into the space of self adjoint matrices $S(k-1, \mathbb{C})$.

Since $M = j(\mathbb{C}P^{k-2})$ is $SU(k-1)$ invariant, the techniques used for $j(\mathbb{R}P^{N-1})$ can be adapted to determine the focal points of M in $S_+(k-1, \mathbb{C})$, the space of nonnegative definite self-adjoint $(k-1) \times (k-1)$ complex matrices. We are then led to the following:

PROPOSITION 4.3. *The focal points of $M = j(\mathbb{C}P^{N-1})$ in $S_+(k-1, \mathbb{C})$ are the nonnegative definite symmetric matrices with the highest eigenvalue of multiplicity at least 2. The projection $P_M: S_+(k-1, \mathbb{C}) \setminus F \rightarrow M$ associates to each matrix $A \in S_+(k-1, \mathbb{C})$ with a highest eigenvalue of multiplicity 1, the matrix $j([m])$, where m is a highest unit eigenvector of A .*

The following result, which follows from Theorem 3.4 and Proposition 4.3, addresses the question of consistency of Procrustes estimators [see Dryden and Mardia (1998), page 280].

THEOREM 4.4. *Let Q be a probability distribution on $\mathbb{C}P^{k-2}$ and let $\{\{Z_r\}, \|Z_r\| = 1\}_{r=1, \dots, n}$ be a random sample from Q . (a) Q is nonfocal iff λ , the largest eigenvalue of $E[Z_1 Z_1^*]$, is simple and in this case $\mu_E(Q) = [m]$, where m is an eigenvector of $E[Z_1 Z_1^*]$ corresponding to λ , with $\|m\| = 1$. (b) The extrinsic sample mean $\overline{[Z]}_E$ is a consistent estimator of $\mu_E(Q)$ iff λ is simple.*

EXAMPLE 4.1. The Dryden–Mardia distribution on $\mathbb{C}P^{k-2}$ is induced by a \mathbb{C}^{k-1} -valued random variable Z which has a multivariate normal distribution with mean ν and covariance matrix $\sigma^2 I_{2k-2}$. The variable X on $\mathbb{C}P^{k-2}$ corresponding to Z is $X = [Z] = \{\lambda Z, \lambda \in \mathbb{C}^*\}$. Kent and Mardia (1997) showed that $E(j(X)) = \alpha I_{k-1} + \beta \nu \nu^*$, where $\alpha > 0$, $\alpha + \beta > 0$. If we write this quadratic form w.r.t. orthogonal coordinates with the first axis along ν , we notice that as a matrix, $E(j(X))$ is conjugate with a diagonal matrix whose diagonal entries are all α except for the entry $\alpha + \beta$ in the upper left corner, showing that $E(j(X))$ is nonfocal for j . By Theorem 4.4 the extrinsic mean of the Dryden–Mardia distribution exists and the extrinsic sample mean is a consistent estimator of the extrinsic mean.

EXAMPLE 4.2. The complex unit sphere is $\mathbb{C}S^{k-2} = \{z \in \mathbb{C}^{k-1} \mid \|z\| = 1\}$. Kent (1994) defines on $\mathbb{C}S^{k-2}$ the complex Bingham distribution associated with a Hermitian matrix as a parameter by the probability density function

$$(4.4) \quad f_A(z) = C(A)^{-1} \exp(z^* A z), \quad z \in \mathbb{C}S^{k-2}.$$

This density is constant along the orbit of z via the action of $\mathbb{C}S^0$ given by $(e^{i\theta}, z) \mapsto e^{i\theta} z$. The space of orbits is $\mathbb{C}P^{k-2}$ and the image of the volume measure of $\mathbb{C}S^{k-2}$ on $\mathbb{C}P^{k-2}$ in this projection is the volume measure associated with the Fubini–Study metric. Therefore $f_A(z)$ induces a probability density function on $\mathbb{C}P^{k-2}$, which we call the density of the *complex Bingham distribution for planar shapes*, given by

$$(4.5) \quad g_A([z]) = f_A(z), \quad [z] \in \mathbb{C}S^{k-2}.$$

Assume λ_A is the largest eigenvalue of A and let V_A be the eigenspace corresponding to λ_A . Then the extrinsic mean set of the complex Bingham distribution for planar shapes is the set $\{[\mu] \mid \mu \in V_A \setminus \{0\}\}$. The extrinsic mean exists only if V_A has dimension one over \mathbb{C} . Therefore if $\dim_{\mathbb{C}} V_A \geq 2$, even if the Procrustes estimate (extrinsic sample mean) exists, it is inconsistent.

In general, if $[z_r] = [(z_r^1, \dots, z_r^{k-1})]$, $\|z_r\| = 1$, $r = 1, \dots, n$, are independent observations from a random variable on $\mathbb{C}P^{k-2}$, the extrinsic sample mean $\overline{[z]}_E$ is $[m]$, where m is a highest unit eigenvector of

$$(4.6) \quad K := \frac{1}{n} \sum_{r=1}^n z_r z_r^*.$$

Note that $\overline{[z]}_E$ is the full Procrustes estimate for parametric families such as Dryden–Mardia distributions or complex Bingham distribution for planar shapes [Kent (1992)]. Unlike other authors [Kent (1994) and Kendall (1984)], in our computations we do not make any use of the so-called Helmert transform. We simply center the raw landmark data $u_r = (u_r^1, \dots, u_r^k)$, $r = 1, \dots, n$, get $w_r = (w_r^1, \dots, w_r^k)$ with $w_r^j = u_r^j - \bar{u}_r$ and rescale the first $k - 1$ transformed coordinates by taking $z_r = \|\tilde{w}_r\|^{-1} \tilde{w}_r$, where $w_r = (\tilde{w}_r^1, w_r^k)$. Then we evaluate K in (4.6) and take the highest eigenvector of K as a representative of $\overline{[z]}_E$. As noted before, given that our identification of Σ_2^k to $\mathbb{C}P^{k-2}$ differs from Kendall’s identification, the values of $\overline{[z]}_E$ using the two identification methods may differ. This difference will be small in the case of a highly concentrated distribution, as in the example below.

REMARK 4.1. The extrinsic sample mean can be used to determine missing coordinates when most of the landmarks in a new observation are known. We consider the case of one missing landmark although for more missing landmarks the same principle works. Assume $\overline{[z]}_E = [\zeta]$, $\zeta \in \mathbb{C}S^{k-2}$, is the sample mean of a number of complete observations and $o = (z^1, \dots, z^{k-1}, z)$ are the raw coordinates of a new observation, with z unknown (we may assume w.l.o.g. that the missing landmark is the last one). After centering and rescaling we get

$$(4.7) \quad w^j = \left(z^j - \frac{1}{k} \left(z + \sum_{s=1}^{k-1} z^s \right) \right) / \left(\sum_{j=1}^k \left| z^j - \frac{1}{k} \left(z + \sum_{s=1}^{k-1} z^s \right) \right|^2 \right)^{1/2}.$$

Minimizing $d_0([o], [\zeta])$ amounts to maximizing

$$(4.8) \quad h(z) = \frac{(\sum_{j=1}^{k-1} |w^j \bar{\zeta}^j|^2)}{\sum_{j=1}^{k-1} |w^j|^2},$$

where w^j are given in (4.7), and the solution gives the missing landmark z conditionally on the sample data and the other landmarks in o . We will call this the *XM method* of retrieval for a single missing landmark.

If $z, w \in \mathbb{C}S^{k-2}$, the Fubini–Study distance $d_g([z], [w])$ is proportional to $\arccos |z^t \bar{w}|$. Therefore the intrinsic mean $\overline{[z]}_I$ of the sample $[z_r]$, $\|z_r\| = 1$, $r = 1, \dots, n$, is a minimizer of

$$(4.9) \quad g([\zeta]) = \sum_{r=1}^n \arccos^2(|z_r^t \bar{\zeta}|), \quad \|\zeta\| = 1.$$

The minimizer can be determined by selecting $\zeta = (\zeta^1, \dots, \zeta^{k-1})$ with $\zeta^{k-1} > 0$. If for $r = 1, \dots, n$ and $j = 1, \dots, k - 1$, we have $\zeta^j = \xi^j + i\eta^j$ and $z_r^j = x_r^j + iy_r^j$,

such a minimizer is obtained by using numerical methods for the objective function of $2(k - 2)$ real variables

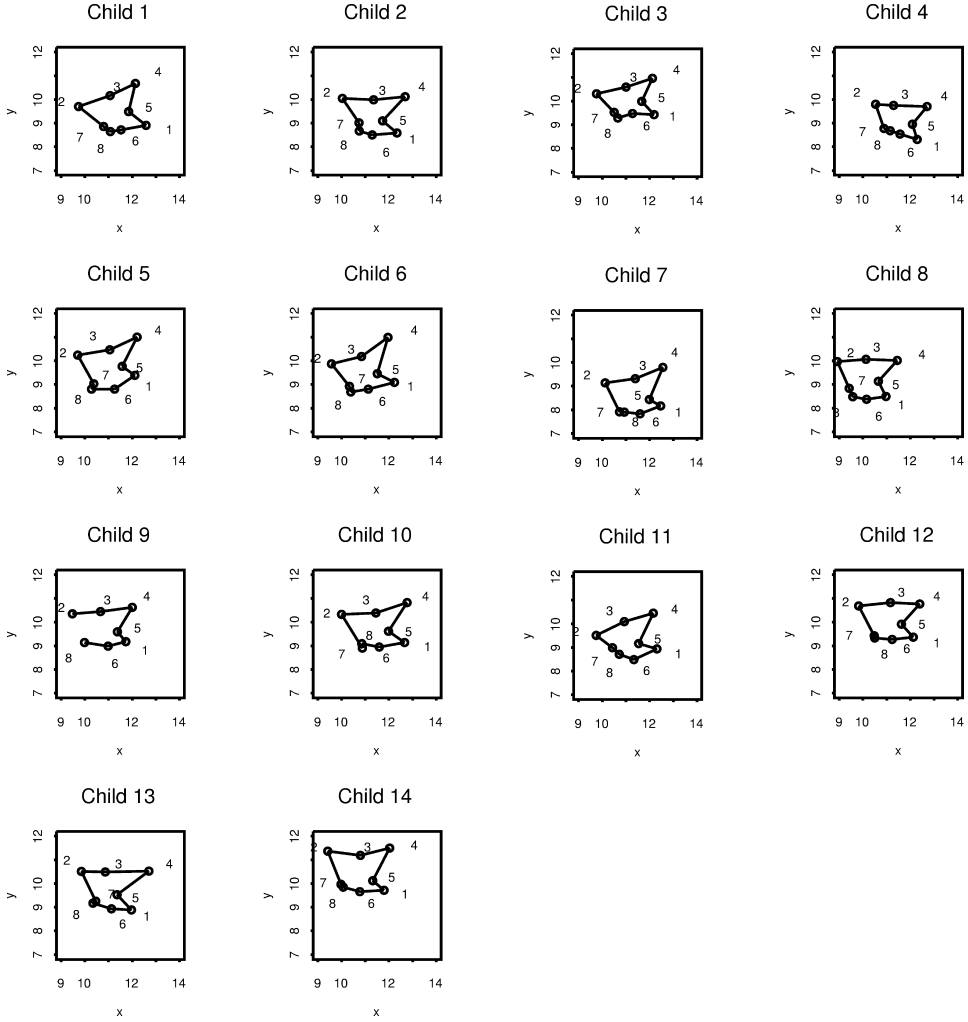
$$(4.10) \quad g(\xi^1, \xi^2, \dots, \xi^{k-2}, \eta^1, \eta^2, \dots, \eta^{k-2}) \\ = \sum_{r=1}^n \arccos^2 \left| z_r^{k-1} \sqrt{1 - \sum_{j=1}^{k-2} ((\xi^j)^2 + (\eta^j)^2) + \sum_{j=1}^{k-2} z_r^j \bar{\zeta}^j} \right|.$$

Because the intrinsic distance is larger than the extrinsic distance (chord $<$ arc) outliers have more influence on the intrinsic sample mean, which makes the use of the extrinsic mean preferable in practice if strong outliers are present. For concentrated data the two means are very close to each other.

We close with an example to compute the mean *Apert syndrome upper midface*, and to use it to estimate a missing landmark. In our example, based on data from Bookstein (1991), we determine the extrinsic mean of a group of 8 landmarks on the Apert syndrome upper midface. The data set represents coordinates of the following landmarks: the Anterior nasal spine, Sella, Spheno-ethmoid registration, Nasion, Orbitale, Inferior zygoma, Pterygomaxillary fissure and Posterior nasal spine taken from lateral X-rays of a group of 14 children suffering from the Apert syndrome. The data are displayed in Figure 1. Note that the coordinates of landmark 7 from “child 9” are missing. The shape variable (in our case, shape of the 8 landmarks on the upper face) is valued in a planar shape space $\mathbb{C}P^6$ (real dimension = 12). The usual statistical methods fail when applied to $\mathbb{C}P^{k-2}$ because, as a Riemannian manifold, $\mathbb{C}P^{k-2}$ is not locally Euclidean, whatever the metric we consider on it. As a special case of Theorem 3.3(b), one may show that if the i.i.d. $\mathbb{C}P^{k-2}$ -valued observations have a Dryden–Mardia distribution, then the intrinsic and extrinsic means are the same. Thus the extrinsic sample mean is a consistent estimator of the intrinsic mean of Q , by Theorem 2.3 or by Ziezold (1977). This result is due to Kent and Mardia (1997) and Le (1998).

Using MINITAB, from the 13 complete observations, after rescaling, we found the following representative for the extrinsic sample mean shape corresponding to $[\bar{z}]_E = [z^1 : z^2 : z^3 : z^4 : z^5 : z^6 : z^7] \in \mathbb{C}P^6$:

$$\begin{aligned} z^1 &= -0.174205 + 0.351359i, & z^2 &= 0.258564 - 0.431477i, \\ z^3 &= -0.112506 - 0.233028i, & z^4 &= -0.527492 - 0.069521i, \\ z^5 &= -0.117264 + 0.109873i, & z^6 &= 0.113351 + 0.209546i, \\ z^7 &= 0.279319 + 0.000000i, & z^8 &= 0.280233 + 0.063249i. \end{aligned}$$

FIG. 1. *Apert data.*

Using MATHEMATICA for the function g in (4.10), we obtained a representative for the intrinsic sample mean shape corresponding to $[\bar{z}]_I = [z^1 : z^2 : z^3 : z^4 : z^5 : z^6 : z^7] \in \mathbb{C}P^6$, which after rescaling, is given by

$$\begin{aligned}
 z^1 &= -0.174180 + 0.351085i, & z^2 &= 0.258289 - 0.431400i, \\
 z^3 &= -0.112757 - 0.232802i, & z^4 &= -0.527627 - 0.069441i, \\
 z^5 &= -0.117257 + 0.109764i, & z^6 &= 0.113449 + 0.209658i, \\
 z^7 &= 0.279347 + 0.000000i, & z^8 &= 0.280736 + 0.063136i.
 \end{aligned}$$

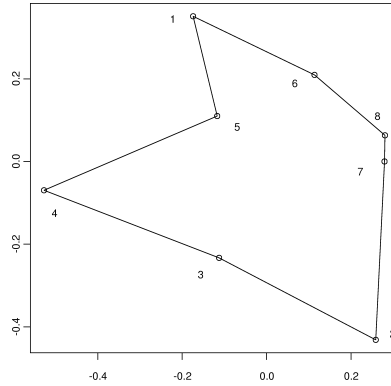


FIG. 2. *Extrinsic sample mean of the 13 complete Apert observations.*

The representative of the intrinsic sample mean, including the resulting coordinate $z^8 = -\sum_{j=1}^7 z^j$, is displayed in Figure 3; it cannot be distinguished from the extrinsic sample mean (cf. Figure 2) since the coordinates of corresponding landmarks are identical to the third decimal.

As explained above, we used a different method of identification of a shape with a point in $\mathbb{C}P^6$. For this reason our result slightly differs from the extrinsic mean obtained using Kendall's method of identification. We also include a representative of the extrinsic sample mean, using Kendall's method of identification [for details on Kendall's method see Kendall (1984)], provided by one of the referees:

$$\begin{aligned} z^1 &= -0.1764454 + 0.3503738i, & z^2 &= 0.2619642 - 0.4296601i, \\ z^3 &= -0.1109860 - 0.2335313i, & z^4 &= -0.5270207 - 0.0731476i, \\ z^5 &= -0.1178114 + 0.1094137i, & z^6 &= 0.1123381 + 0.2108623i, \\ z^7 &= 0.2771263 + 0.0000000i, & z^8 &= 0.2808349 + 0.0656893i. \end{aligned}$$

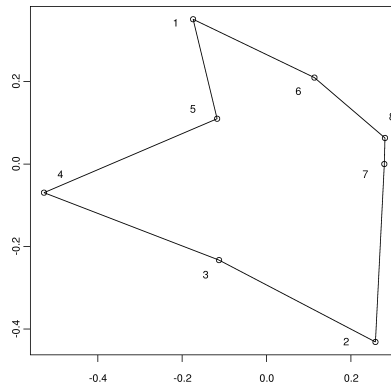


FIG. 3. *Intrinsic sample mean of the 13 complete Apert observations.*

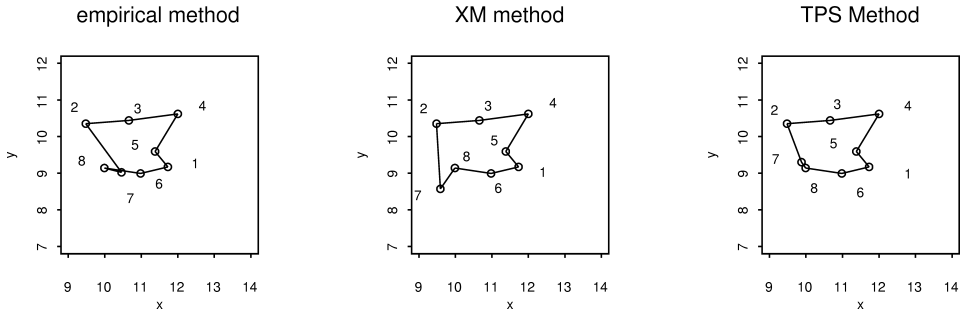


FIG. 4. Retrieval of landmark 7 in observation 9.

Finally, in Figure 4 we display observation 9 completed by various methods. Our XM method, from formulas (4.7)–(4.8), yields for the missing landmark the coordinate $z_7 = 9.59 + 8.57i$. Unlike the empirical method which yields $z_7 = 10.46 + 9.02i$, the XM method places landmark 7 to the left of landmark 8, in agreement with 9 out of the complete 13 observations.

However, when applied to the Apert data, both the empirical method and the XM method perform worse than the TPS (thin-plate spline) method [see Dryden and Mardia (1998), page 206]. We owe to one of the referees of this paper the value of $z_7 = 9.88 + 9.307i$ based on the TPS method, thus putting landmark 7 above landmark 8, which is the case for 12 out of 13 observations.

For work on missing landmark data see Bookstein and Mardia (2000).

REMARK 4.2. While the computations for Apert data are only illustrative, we believe that similar computations for random samples of clinically normal children from various groups of populations, may be useful in reconstructive plastic surgery.

Acknowledgments. The authors are very grateful to three anonymous referees for many valuable suggestions. We also thank Gordana Derado and Ray Pruett for help with MATHEMATICA and with S-Plus, and Fred Bookstein for providing us with the data on Apert syndrome.

REFERENCES

- AMARI, S.-I. (1985). *Differential-Geometrical Methods in Statistics. Lecture Notes in Statist.* **28**. Springer, New York.
- BARNDORFF-NIELSEN, O. E. and COX, D. R. (1994). *Inference and Asymptotics*. Chapman and Hall, London.
- BERAN, R. J. (1979). Exponential models for directional data. *Ann. Statist.* **7** 1162–1178.
- BERAN, R. and FISHER, N. I. (1998). Nonparametric comparison of mean directions or mean axes. *Ann. Statist.* **26** 472–493.
- BOOKSTEIN, F. L. (1991). *Morphometric Tools for Landmark Data: Geometry and Biology*. Cambridge Univ. Press.

- BOOKSTEIN, F. L. and MARDIA, K. V. (2000). A family of EM-type algorithms for missing morphometric data. In *Abstracts of the 19th L.A.S.R Workshop* (J. T. Kent and R. G. Aykroyd, eds.).
- BURBEA, J. and RAO, C. R. (1982). Entropy differential metric, distance and divergence measures in probability spaces: A unified approach. *J. Multivariate Anal.* **12** 575–596.
- CHAVEL, I. (1993). *Riemannian Geometry: A Modern Introduction*. Cambridge Univ. Press.
- DO CARMO, M. P. (1992). *Riemannian Geometry*. Birkhäuser, Boston.
- DRYDEN, I. L. and MARDIA, K. V. (1993). Multivariate shape analysis. *Sankhyā Ser. A* **55** 460–480.
- DRYDEN, I. L. and MARDIA, K. V. (1998). *Statistical Shape Analysis*. Wiley, New York.
- EFRON, B. (1975). Defining the curvature of a statistical problem (with applications to second order efficiency) (with discussion). *Ann. Statist.* **3** 1189–1242.
- EFRON, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans*. SIAM, Philadelphia.
- ÉMERY, M. and MOKOBODZKI, G. (1991). Sur le barycentre d'une probabilité dans une variété. *Séminaire de Probabilités XXV. Lecture Notes in Math.* **1485** 220–233. Springer, Berlin.
- FISHER, N. I. (1993). *Statistical Analysis of Circular Data*. Cambridge Univ. Press.
- FISHER, N. I. and HALL, P. (1992). Bootstrap methods for directional data. In *The Art of Statistical Science: A Tribute to G. S. Watson* (K. V. Mardia, ed.) 47–63. Wiley, New York.
- FISHER, N. I., HALL, P., JING, B.-Y. and WOOD, A. T. A. (1996). Improved pivotal methods for constructing confidence regions with directional data. *J. Amer. Statist. Assoc.* **91** 1062–1070.
- HELGASON, S. (1978). *Differential Geometry, Lie Groups, and Symmetric Spaces*. Academic Press, New York.
- KARCHER, H. (1977). Riemannian center of mass and mollifier smoothing. *Comm. Pure Appl. Math.* **30** 509–541.
- KENDALL, D. G. (1984). Shape manifolds, Procrustean metrics, and complex projective spaces. *Bull. London Math. Soc.* **16** 81–121.
- KENDALL, D. G., BARDEN, D., CARNE, T. K. and LE, H. (1999). *Shape and Shape Theory*. Wiley, New York.
- KENDALL, W. S. (1990). Probability, convexity, and harmonic maps with small image. I. Uniqueness and fine existence. *Proc. London Math. Soc.* **61** 371–406.
- KENT, J. T. (1992). New directions in shape analysis. In *The Art of Statistical Science: A Tribute to G. S. Watson* (K. V. Mardia, ed.) 115–128. Wiley, New York.
- KENT, J. T. (1994). The complex Bingham distribution and shape analysis. *J. Roy. Statist. Soc. Ser. B* **56** 285–299.
- KENT, J. T. and MARDIA, K. V. (1997). Consistency of Procrustes estimators. *J. Roy. Statist. Soc. Ser. B* **59** 281–290.
- KOBAYASHI, S. (1968). Isometric imbeddings of compact symmetric spaces. *Tôhoku Math. J.* **20** 21–25.
- KOBAYASHI, S. (1972). *Transformation Groups in Differential Geometry*. Springer, New York.
- KOBAYASHI, S. and NOMIZU, K. (1996). *Foundations of Differential Geometry 2*. (Reprint of the 1969 original.) Wiley, New York.
- LE, H. (1998). On the consistency of Procrustean mean shapes. *Adv. in Appl. Probab.* **30** 53–63.
- LE, H. and KUME, A. (2000). The Fréchet mean shape and the shape of means. *Adv. in Appl. Probab.* **32** 101–113.
- MARDIA, K. V. and JUPP, P. E. (1999). *Directional Statistics*. Wiley, New York.
- MILNOR, J. (1963). *Morse Theory*. Princeton Univ. Press.
- OLLER, J. M. and CORCUERA, J. M. (1995). Intrinsic analysis of statistical estimation. *Ann. Statist.* **23** 1562–1581.
- PRENTICE, M. J. (1984). A distribution-free method of interval estimation for unsigned directional data. *Biometrika* **71** 147–154.

- PRENTICE, M. J. and MARDIA, K. V. (1995). Shape changes in the plane for landmark data. *Ann. Statist.* **23** 1960–1974.
- RAO, C. R. (1945). Information and the accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.* **37** 81–91.
- SIBSON, R. (1978). Studies in the robustness of multidimensional scaling: Procrustes analysis. *J. Roy. Statist. Soc. Ser. B* **40** 234–238.
- WANG, H.-C. (1952). Two-point homogeneous spaces. *Ann. Math.* **55** 177–191.
- WARNER, F. W. (1965). The conjugate locus of a Riemannian manifold. *Amer. J. Math.* **87** 575–604.
- WATSON, G. S. (1983). *Statistics on Spheres*. Wiley, New York.
- ZIEZOLD, H. (1977). On expected figures and a strong law of large numbers for random elements in quasi-metric spaces. In *Transactions of the Seventh Prague Conference on Information Theory, Statistical Decision Functions, Random Processes and of the Eighth European Meeting of Statisticians* **A** 591–602. Reidel, Dordrecht.

DEPARTMENT OF MATHEMATICS
INDIANA UNIVERSITY
BLOOMINGTON, INDIANA 47405

DEPARTMENT OF MATHEMATICS
AND STATISTICS
GEORGIA STATE UNIVERSITY
ATLANTA, GEORGIA 30303
E-MAIL: matvnp@geomstat.cs.gsu.edu