NONPARAMETRIC STATISTICS ON MANIFOLDS WITH APPLICATIONS TO SHAPE SPACES

by

Abhishek Bhattacharya

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF MATHEMATICS

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2 0 0 8

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Abhishek Bhattacharya entitled NONPARAMETRIC STATISTICS ON MANIFOLDS WITH APPLICATIONS TO SHAPE SPACES and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

____ Date: 11/24/08
Rabi Bhattacharya

____ Date: 11/24/08
David Glickenstein

____ Date: 11/24/08
Thomas G Kennedy

____ Date: 11/24/08
Douglas M Pickrell

____ Date: 11/24/08
Moshe Shaked

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.
I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

____ Date: 11/24/08
Dissertation Director: Rabi Bhattacharya

Statement by Author

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

Signed: ______________________________
ABHISHEK BHATTACHARYA

ACKNOWLEDGEMENTS

I am very grateful to my advisor Professor Rabi Bhattacharya for guiding me in my research and for helping me write my dissertation. In my PhD career, I have published five articles, three of them being jointly with him. A lot of the material in my dissertation has been taken from those articles. Without my advisor's active support, this thesis would never have been possible.

I am also grateful to the professors who chose to be on my PhD committee, namely Professors David Glickenstein, Thomas G Kennedy, Douglas M Pickrell and Moshe Shaked. I thank them for their kind and helpful suggestions and comments towards writing my thesis.

The different courses that I took at University of Arizona, especially in statistics and differential geometry, helped me a lot in comprehending the geometry of the shape spaces and how to carry out statistical inference on them, which in turn enabled me to derive new results towards statistical analysis of shapes. Their guidance is gratefully acknowledged.

I would also like to acknowledge the use of computer and other facilities at the University of Arizona.

The numerous conferences and workshops that I attended and made presentations throughout my PhD career helped me acquire new ideas towards my research. I would like to thank the organizers for inviting me and providing me financial support to attend those events.

Last, but not the least, I would like to thank my family members, friends and relatives for their active support in all my endeavors.

DEDICATION

I dedicate my work to divine Mother Ma, who is behind all this.

Table of Contents

Table of Contents—*Continued*

Table of Contents—*Continued*

List of Tables

List of Figures

Abstract

This thesis presents certain recent methodologies and some new results for the statistical analysis of probability distributions on non-Euclidean manifolds. The notions of Fréchet mean and variation as measures of center and spread are introduced and their properties are discussed. The sample estimates from a random sample are shown to be consistent under fairly broad conditions. Depending on the choice of distance on the manifold, intrinsic and extrinsic statistical analyses are carried out. In both cases, sufficient conditions are derived for the uniqueness of the population means and for the asymptotic normality of the sample estimates. Analytic expressions for the parameters in the asymptotic distributions are derived. The manifolds of particular interest in this thesis are the shape spaces of k-ads. The statistical analysis tools developed on general manifolds are applied to the spaces of direct similarity shapes, planar shapes, reflection similarity shapes, affine shapes and projective shapes. Two-sample nonparametric tests are constructed to compare the mean shapes and variation in shapes for two random samples. The samples in consideration can be either independent of each other or be the outcome of a matched pair experiment. The testing procedures are based on the asymptotic distribution of the test statistics, or on nonparametric bootstrap methods suitably constructed. Real life examples are included to illustrate the theory.

Chapter 1
Introduction

The statistical analysis of shapes has diverse applications in biology, engineering sciences, computer sciences and many other fields of study. In a wide variety of problems, it is of great practical importance to measure, describe and compare the shapes of objects. Although it is easy to describe 'shape', it is not a trivial matter to come up with a formal mathematical definition of shape. Perhaps the earliest and the most well known notion of shape is the similarity shape, that is all characteristics of the object of study which remains after location, rotation and scale information is removed. In his research, the author is mostly concerned with landmark based shape analysis where a certain number of points or landmarks, say k, are picked from the object whose shape we want to measure, usually with expert help. Such a set of k points is called a k-ad or configuration of k points. One approach to study the shape of a k-ad would be to extract features like ratio of distances, angles etc which are invariant under similarity transformations and perform analysis on them. However it is more efficient to work with the full geometry of the k-ad after removing the similarity transformations. That way one has more information on the shape of the object and the results obtained are easier to interpret. Hence in his dissertation, the author works with the full shape of a k-ad. Among the pioneers of such landmark based shape analysis, the author would like to mention Kendall (1977, 1984) and Bookstein (1978). Bookstein's approach is primarily registration based requiring two or three landmarks to be brought into a standard position by translation, rotation and scaling of the k-ad. For these shapes, the author has preferred Kendall's more invariant view of a shape identified with the orbit under rotation of the k-ad centered at the origin and scaled to have unit size. The set of all shapes then is the quotient of the unit sphere under the group of rotations (in m-dimensions) and hence a quotient manifold.

In many applications of object recognition based on its shape, the notion of 'similarity shape' is not found to be adequate. For example two triangles with labelled vertices may not have the same similarity shape but are similar when one is reflected about an axis. That is when one says that the triangles have the same reflection similarity shape. Further, different photographs of the same object taken from different angles may be related by a projective transformation which may not be a similarity. Then the images have the same projective shape. Hence depending on the application, the appropriate notion of shape of a k-ad has to be considered.

Statistical shape analysis is concerned with methodology for analysing shapes in the presence of randomness. All the shape spaces studied in this thesis are quotient manifolds, often the quotient of the unit sphere. Hence to carry out statistical shape analysis, appropriate estimation and inference tools need to be developed on general non-Euclidean manifolds which are to be applied to the shape space of interest. To do nonparametric inference on manifolds, one uses parameters like mean and variation which are defined for a general probability distribution on the manifold. Then from a random sample on the manifold, the sample estimates are obtained and their properties like consistency and asymptotic distribution are examined. Using the estimates, one carries out inference on the unknown population parameters and hence on the underlying probability distribution either by asymptotic tests or by bootstrap methods. Bootstrap methods are particularly useful when the sample sizes are not too large for the asymptotic distribution approximation to hold good. This is true for shape data, especially when there are too many landmarks. Effron's bootstrap (Effron (1982)) is one of the important tools proposed here for setting critical points for two sample tests and for confidence regions for the parameters. However when the number of landmarks on the sample k-ads is too high compared to the sample size, it may not be possible to construct a pivotal bootstrap confidence region. Then one may set confidence regions for only the first few principle scores or use nonpivotal bootstrap methods. In all the examples studied in this thesis, various bootstrap methods and asymptotic tests yield consistent results (see Sections 7.10, 8.6 and 9.5).

One may also carry out parametric inference where one assumes a specific form for the underlying probability distribution with unknown parameters and carries out inference on the parameters by tests such as likelihood ratio tests. A fairly comprehensive account of parametric models on the planar shape space, with many references to the literature, can be found in Dryden and Mardia (1998). However, the existing models do not seem to fit the shape data that occur in applications very well and examples in Bhattacharya and Bhattacharya (2008a) suggest that appropriate nonparametric methods detect differences in shape distributions better than their parametric counterparts in the literature. Hence in his dissertation, the author has focussed only on nonparametric inference methods.

Apart from comparing the means and variations, one may use more elaborate nonparamteric inference procedures such as nonparametric density estimation. However application of such procedures to the high dimensional shape spaces suffer from the 'curse of dimensionality' problem and hence have not been used in this thesis. Further, in all the examples studied in this thesis and in articles like Bhattacharya and Patrangenaru (2005), Bhattacharya and Bhattacharya (2008 a,c) and Bhattacharya (2008a, b), the means and variations are found to be adequate for discrimination.

This thesis deals with landmark based shape analysis. Apart from that, there is now a vast literature, especially in computer science and electrical engineering on what is called continuous shape analysis. The continuous shape of an object or image is defined by the actual boundary of the object or image, either as a closed contour (in 2D) or by a compact surface (in 3D), modulo a group of transformations. This view was pioneered by Grerenander (1993). In dealing with notions of continuous shape, the infinite dimensional nature of these manifolds and their general lack of local compactness conditions causes problems. In contrast, the landmark based shape spaces are finite dimensional locally compact manifolds. If an adequate number of landmarks are picked at important locations, they can capture most of the continuous shape information. For the current state of research on continuous shapes, see Krim and Yezzi (2006) and Joshi et. al. (2006).

The general nonparametric statistical approach based on Fréchet means is due to Bhattacharya and Patrangenaru (2002, 2003, 2005). Some earlier work on this may also be found in Patrangenaru (1998) and Hendricks and Landsman (1998). Among the goals of the author in the present dissertation are (1) the resolution of a number of geometric and mathematical issues that arise in the application of the above paradigm to manifolds of interest in biology, medical imaging and machine vision, and (2) the construction of proper asymptotic inference procedures such as two-sample tests for discrimination among different distributions on manifolds of interest. Here is a brief outline of the different chapters in this dissertation.

In chapter 2, the concepts of Fréchet mean and variation are introduced. The idea is to define the Fréchet function of a probability distribution Q on a metric space M as the integral (with respect to Q) of a positive convex function of the distance metric on M and define the Frechet mean as the minimizer (if unique) of this Fréchet function and the Fréchet variation as the minimum value (if finite). Among all possible convex functions, general powers of the distance are considered. This generalizes the definition of the Fréchet function and hence Fréchet mean and variation from Bhattacharya and Patrangenaru (2003). Conditions are derived for the consistency and asymptotic normality of the sample estimates. It is to be noted that for the asymptotic normality of the sample variation to hold, it is necessary to assume that there is a unique population mean even though that is not required for the sample variation to be consistent. Confidence regions for the population parameters are constructed both by using the asymptotic distribution of the sample estimates and by pivotal bootstrap methods. Using such confidence regions, one can identify the underlying probability distribution from a random sample on the manifold M. This is called Fréchet analysis on M.

Chapter 3 performs Fréchet analysis on Riemannian manifolds by using the geodesic distance as the distance metric in the definition of the Fréchet function. The resulting Fréchet parameters are called the intrinsic parameters and the corresponding statistical analysis is called intrinsic analysis on a manifold M. In this chapter, sufficient conditions are derived for the existence of a unique intrinsic mean and the asymptotic normality of the sample intrinsic mean. Two-sample nonparametric tests are constructed to compare the sample intrinsic means and variations which can be used to distinguish between the underlying probability distributions. Two different cases are considered, firstly when the two samples are mutually independent and next for matched pair samples, which arise for example when the X and Y observations correspond to the same subjects.

Even though intrinsic analysis would seem to be the natural approach for statistical analysis on manifolds, broad conditions for the existence of the intrinsic mean are not known. The known conditions require that the probability Q has support in a convex geodesic ball. To avoid that, one may perform extrinsic analysis on manifolds which is the subject of Chapter 4. This is also pursued when a natural Riemannian structure on M is not in sight. For extrinsic analysis on M, one embeds it into some higher dimensional Euclidean space and uses the distance induced by this embedding. Then, as the results from Chapter 7 show, the corresponding analysis becomes a lot simpler both mathematically and computationally than its intrinsic counterpart. For example, the extrinsic mean is known to exist under fairly broad conditions and it has a closed form analytic expression in most cases (see Chapters 7, 8, 9 and 10). Also the asymptotic normality of the sample extrinsic mean and variation hold under less restrictive assumptions and analytic expressions for the parameters in the distribution are relatively easy to get. All this is presented in this chapter. In particular, the extrinsic mean set comprises of all projections of the Euclidean mean of the image of Q on the image manifold under the embedding. Hence there is a unique mean iff there is a unique projection. For asymptotic normality of the sample mean and variation, one requires that this projection map is smooth in a neighborhood of the population mean (which is assumed to exist). The chapter concludes with two-sample nonparametric tests to distinguish between two probability distributions by comparing the sample extrinsic means and variations. As in Chapter 3, appropriate tests are constructed both for mutually independent and matched pair samples. For the extrinsic and intrinsic approaches to yield consistent results, it is important to choose a 'nice' embedding. For that, an equivariant embedding is desirable, which preserves a lot of the geometry of the manifold. The numerical examples in Chapter 7 show that then the extrinsic and intrinsic means are very close to each other and the two sample extrinsic and intrinsic tests yield similar results.

Since most of the shape spaces are quotients of the unit sphere, it is important to understand how to carry out nonparamteric statistical analysis on this space. Hence in both Chapters 3 and 4, the results are applied to the sphere.

In Chapter 5, the different notions of shapes that occur in this thesis are introduced. They include direct similarity shapes, reflection similarity shapes, affine shapes and projective shapes. For problems in biology like classification of species, disease detection etc, similarity shape analysis has many uses, while for problems in machine vision and image analysis, affine and projective shape analysis is more appropriate. All these shape spaces can be represented as the quotients of Riemannian manifolds, often the Riemannian sphere. After removing some "singular shapes", the shape space has the natural structure of a Riemannian manifold, which can be used to carry out an intrinsic analysis of shapes. If the shape space is embedded into some Euclidean space via a preferably equivariant embedding, one can also carry out an extrinsic analysis. For all the shape spaces considered in this thesis, the extrinsic mean set and variation are found to depend on the eigenvalues and eigenvectors of the Euclidean mean of the embedded probability distribution. Perturbation theory arguments for eigenvalues and eigenvectors are used to get necessary and sufficient conditions for the existence of a unique extrinsic mean and to prove the asymptotic normality of the sample estimates.

In Chapters 6-10, the geometry of the shape spaces is discussed in more detail and statistical analysis is performed on them using the methods of Chapters 3 and 4.

In particular, Chapter 6 provides an exposition of the geometry of the (direct) similarity shape space of k-ads in m dimensions or Σ_m^k. The cases of interest include $m = 2$ or 3. This space can be represented as the quotient of the unit sphere with respect to all rotations (in m dimensions). It is shown that after removing some singularities, Σ_m^k is a Riemannian manifold. There are no such singularities when $m = 2$. This chapter identifies the tangent space of Σ_m^k, exponential map and geodesic distance on Σ_m^k. Most of the results are taken from Kendall et al. (1999).

Chapter 7 considers in detail the similarity shape space when $m = 2$, which is also called the planar shape space. This is a compact, connected Riemannian manifold which can be identified with the complex projective space $\mathbb{C}P^{k-2}$, a well studied manifold in differential geometry. After identifying Σ_2^k with $\mathbb{C}P^{k-2}$, this chapter presents the geometry of this space, namely its tangent space, exponential and inverse exponential maps, geodesic distance, cut locus, sectional curvature etc. The methods of Chapter 3 are applied to carry out an intrinsic analysis on this space. In particular, a probability distribution Q has a unique intrinsic mean if its support is contained in a geodesic ball of radius $\frac{\pi}{4}$. This result was shown in Kendall (1990). If the support is also contained in a ball of radius $\frac{\pi}{3}$ around the intrinsic mean, then as shown in Bhattacharya and Bhattacharya (2008b), the sample intrinsic mean has an asymptotic Normal distribution. This result is used to construct a nonparametric confidence region for the population mean shape from a random sample of planar shapes. It is shown in this chapter that when the Fréchet function is defined as the integrated geodesic distance cubed, then there is a unique Fréchet mean of Q and the sample Fréchet mean is asymptotically Normally distributed whenever the support of Q lies in a geodesic ball of radius $\frac{\pi}{4}$. To carry out an extrinsic analysis, the planar shape space is embedded into $S(k, \mathbb{C})$, the space of all $k \times k$ complex Hermitian matrices, via the Veronese-Whitney embedding. Using this embedding, expressions for the extrinsic mean and variation are derived in this chapter. No support restriction on Q is necessary for it to have a unique extrinsic mean. Analytic expressions for the parameters in the asymptotic distribution of the sample extrinsic mean are derived. This enables one to perform two-sample tests to compare the extrinsic means and variations for two random sample of shapes and hence discriminate between the underlying probability distributions. The results of extrinsic and intrinsic analyses on Σ_2^k are applied to two examples. In the first example, there are two samples of 2D images of 8 landmarks picked from male and female gorilla skulls. The sample means and variations are computed for the planar shapes of both the sample images. Two-sample tests are carried out to compare the means and variations and hence distinguish between the shapes of skulls for the two sexes. The second example is an application of schizophrenia detection. In this example, one considers the planar shapes of k-ads obtained from the brain scan of 14 schizophrenic and 14 normal children. As in example 1, the mean shapes and shape variations for the two group of children are compared by bootstrap methods and asymptotic tests to detect any difference in brain shape due to schizophrenia.

When $m > 2$, the similarity shape space Σ_m^k fails to be a complete manifold. It has unbounded sectional curvatures. As a result, the results from Chapter 3 cannot be applied to derive sufficient conditions for the existence of a unique intrinsic mean which is necessary to carry out intrinsic analysis. If instead, one considers the reflection (similarity) shape, then it is possible to embed the resulting shape space into the space of symmetric matrices $S(k, \mathbb{R})$ and carry out an extrinsic analysis. This is discussed in Chapter 8. Such an embedding of the reflection (similarity) shape space $R\Sigma_m^k$ was first obtained by Bandulasiri and Patrangenaru (2005) and later independently by Dryden et al. (2007). However the correct computation of the extrinsic mean eluded earlier authors until it was derived in Bhattacharya (2008). In this chapter, that derivation is reproduced along with the condition for its uniqueness. An expression for the extrinsic variation is also obtained. Two-sample tests are carried out to distinguish between two probability distributions by comparing the sample extrinsic means and variations. This enables one to extend nonparametric inference on Kendall type shape spaces from 2D to higher dimensions. Other appropriate distances including the geodesic distance on $R\Sigma_m^k$ are described which can be used to carry out a Fréchet analysis on this space. The chapter concludes with an application to Glaucoma detection. In this example, 3D k-ads are obtained from both the eyes of 12 rhesus monkeys. One of their eyes is infected with Glaucoma while the other one is kept normal. The 3D reflection shapes of the k-ads are considered. Using the matched pair two-sample tests described in this chapter, the mean reflection shapes and the shape variations for the two eyes are compared. The tests for comparing the mean shapes, both the asymptotic chi-squared test and the pivotal bootstrap method yield very small p-values, leading to the conclusion that the mean shapes for the two eyes are very different. Hence Glaucoma changes the shape of one's eye, which can be used for Glaucoma detection.

Chapter 9 focuses on methodologies for nonparametric inference on the affine shape spaces $A\Sigma_m^k$. The affine shape of a k-ad x with landmarks in $\mathbb{R}^m$ is defined as its orbit under all affine transformations, $x \mapsto Ax + b$ ($A \in GL(m, \mathbb{R})$, $b \in \mathbb{R}^m$). The space of affine shapes of all centered k-ads whose columns span $\mathbb{R}^m$ is $A\Sigma_m^k$. This shape space can be identified with the Grassmannian $G_m(k-1)$ of all m dimensional subspaces of $\mathbb{R}^{k-1}$, a result of Sparr (1992). For extrinsic analysis on $A\Sigma_m^k$, one embeds it into $S(k, \mathbb{R})$ via an equivariant embedding, as obtained in Dimitric (1996). Using this embedding, expression for the extrinsic mean and condition for its uniqueness are derived, which was first proved in Sughatadasa (2006). The results from Chapter 4 are used to derive the asymptotic distribution for the sample extrinsic mean and variation which is used to construct two sample nonparametric tests to compare two probability distributions. Expressions for the parameters in the asymptotic distribution, as obtained in Bhattacharya (2008a), are stated. Apart from affine shapes, another notion of shape called the special affine shape is also introduced in this chapter. One gets this shape by removing the effect of all affine transformations $Ax + b$ where $\det(A) > 0$. This does not remove the effects of reflections (orthogonal transformations with determinant = -1) and hence looks at the 'oriented' affine shape. The affine shape space $A\Sigma_m^k$ is the quotient of the special affine shape space $SA\Sigma_m^k$ with respect to reflection. The Riemannian geometry of $SA\Sigma_m^k$ and $A\Sigma_m^k$ are described in this chapter which can enable one to perform intrinsic analysis on these shape spaces. Towards the end of this chapter, an application to handwritten character recognition is described. In this example, a random sample of 30 handwritten digit '3' are collected so as to devise a scheme to automatically classify handwritten characters. The sample extrinsic mean and variation are obtained for the affine shapes of the sample k-ads. After removing some outliers, 95% confidence regions for the mean shape and variation in shape are obtained both by asymptotic and bootstrap methods.

Finally Chapter 10 presents methodologies for statistical analysis of projective shapes. The chapter starts by explaining the connection between the projective shape space $P\Sigma_m^k$ and the real projective space $\mathbb{R}P^m$. To get the projective shape of a k-ad with landmarks in $\mathbb{R}^m$, one views it as a set of axes passing through a common origin (for example the center of a camera) in $\mathbb{R}^{m+1}$. In other words, the k-ad is assumed to lie in $(\mathbb{R}P^m)^k$. Such a k-ad looks like

$$[x] = ([x_1], \ldots, [x_k]),\ [x_j] \in \mathbb{R}P^m,$$

x_j being a representative point in $\mathbb{R}^{m+1}$. To get the projective shape of $[x]$, one first obtains its affine shape, namely the equivalence class $(Ax_1, \ldots, Ax_k)$ for $A \in GL(m+1, \mathbb{R})$, and then takes the corresponding equivalence classes of axes in $\mathbb{R}P^m$. This defines a projective transformation α as

$$\alpha[x] = ([Ax_1], \ldots, [Ax_k]),\ A \in GL(m+1, \mathbb{R}).$$

Hence the projective shape space $P\Sigma_m^k$ is the quotient of $(\mathbb{R}P^m)^k$ with respect to all projective transformations. To give it a manifold structure, only those k-ads are considered for which there exists a set of $m+2$ axes, say $\{[x_{i_1}], \ldots, [x_{i_{m+2}}]\}$, such that the linear span of any $m+1$ points from the set is $\mathbb{R}P^m$, i.e. the linear span of the corresponding representative points is $\mathbb{R}^{m+1}$. Such a k-ad is said to be in general position. The space of projective shapes of k-ads in general position is denoted as $P_0\Sigma_m^k$. In Mardia and Patrangenaru (2005), a diffeomorphism is obtained between $(\mathbb{R}P^m)^{k-m-2}$ and a dense open subset of $P_0\Sigma_m^k$. The diffeomorphism is described in this chapter. It enables one to develop statistical inference tools on projective shapes by developing corresponding inference tools on $\mathbb{R}P^m$.

Apart from application in projective shape analysis, the real projective space $\mathbb{R}P^m$ has many other applications such as in observations on galaxies, on axes of crystals or on the line of geological fissure. This chapter describes the Riemannian structure of $\mathbb{R}P^m$ when identified as S^m modulo antipodal points. Expressions for geodesic distance, exponential and inverse exponential maps, cut-locus, sectional curvature and injectivity radius are obtained. The Riemannian structure is used to carry out an intrinsic analysis on this space. In particular, it follows from the analysis of Chapter 3, that a probability distribution Q on $\mathbb{R}P^m$ with support in an open geodesic ball of radius $\frac{\pi}{4}$ has a unique intrinsic mean. Using results from Chapter 3, expression for the parameters in the asymptotic distribution of the sample intrinsic mean are obtained. To carry out an extrinsic analysis on $\mathbb{R}P^m$, it is embedded into $S(m+1, \mathbb{R})$ via a Veronese-Whitney embedding, which was obtained by Watson (1983). Then the expression for the extrinsic mean and conditions for its uniqueness are stated in this chapter. The results from Chapter 4 are applied to derive expressions for the parameters in the asymptotic distribution of the sample extrinsic mean. Using the identification of the projective shape space as $k - m - 2$ copies of $\mathbb{R}P^m$, one can now carry out extrinsic and intrinsic analysis on this shape space.

Real life examples have been included in various chapters to illustrate the applications of statistical shape analysis. These examples from biology, medical science, image analysis and other fields of study have been taken from various sources such as Anderson (1997), Bhattacharya and Patrangenaru (2005), Bookstein (1991), Derado et al. (2004) and Dryden and Mardia (1998). The numerical calculations involved in these examples have been carried out using Matlab software.

Since statistical analysis of landmark based shapes is still an emerging field of science, many of the terms used in this thesis may be new to the readers. Hence important concepts are highlighted in bold when introduced for the first time in different chapters.

Many of the results presented in this dissertation may also be found in the author's recent articles.

Chapter 2

FRÉCHET ANALYSIS ON METRIC SPACES

2.1 Introduction

To do nonparametric inference on manifolds, we need to introduce parameters which can be defined for a general probability distribution on the manifold, and study their properties. One approach is to extend the usual notion of mean and total variation from Euclidean spaces to arbitrary manifolds. In this chapter, we introduce the concepts of Fréchet mean and variation of probability distributions on manifolds and study their properties. In fact in order to define the above parameters, we do not need our space to be a manifold, the concepts of Fréchet mean and variation can be defined on any metric space. A general notion of a mean of a probability distribution on a metric space was first defined by Fréchet (1948). We generalize that definition in Section 2.2.

2.2 Fréchet Mean and Variation

Let (M, ρ) be a metric space, ρ being the distance, and let $\alpha \geq 1$ be a given real number. For a given probability measure Q on the Borel sigmafield of M, we define the **Fréchet function** of Q as

$$F(p) = \int_M \rho^\alpha(p, x) Q(dx),\ p \in M. \tag{2.2.1}$$

Definition 2.2.1. Suppose $F(p) < \infty$ for some $p \in M$. Then the set of all p for which $F(p)$ is the minimum value of F on M is called the **Fréchet mean set** of Q, denoted by C_Q. If this set is a singleton, say $\{\mu_F\}$, then μ_F is called the **Fréchet mean** of Q. If $X_1, X_2, \ldots, X_n$ are independent and identically distributed (iid) M-valued random variables defined on some probability space $(\Omega, \mathcal{F}, P)$ with common distribution Q, and $Q_n \doteq \frac{1}{n}\sum_{j=1}^{n}\delta_{X_j}$ is the corresponding empirical distribution, then the Fréchet mean set of Q_n is called the **sample Fréchet mean set**, denoted by C_{Q_n}.

Definition 2.2.2. The infimum of F on M is called the **Fréchet variation** of Q, denoted by V and the Fréchet variation of Q_n is called the **sample Fréchet variation**, denoted by V_n.

Proposition 2.2.1 below shows that under mild assumptions, the infimum of F on M is attained, thereby proving that the Fréchet mean set is nonempty and the Fréchet variation is finite.

Proposition 2.2.1. *Suppose every closed and bounded subset of M is compact. If the Fréchet function F of Q is finite for some $p \in M$, then C_Q is nonempty and compact.*

Proof. By triangular inequality on ρ and by convexity of the function $u \mapsto u^\alpha$, $u \geq 0$, we get that

$$\rho^\alpha(q,x) \leq [\rho(p,q)+\rho(p,x)]^\alpha \leq 2^{\alpha-1}[\rho^\alpha(p,q)+\rho^\alpha(p,x)] \tag{2.2.2}$$

which implies that

$$F(q) \leq 2^{\alpha-1}\rho^\alpha(p,q) + 2^{\alpha-1}F(p). \tag{2.2.3}$$

Hence if $F(p) < \infty$ for some p, then $F(q) < \infty\ \forall\ q \in M$. Also

$$|F(p)-F(q)| \leq \int_M |\rho^\alpha(p,x)-\rho^\alpha(q,x)|Q(dx). \tag{2.2.4}$$

As $q \to p$, $|\rho^\alpha(p,x)-\rho^\alpha(q,x)| \to 0$ and from equation (2.2.2), we get that

$$|\rho^\alpha(p,x)-\rho^\alpha(q,x)| \leq (2^{\alpha-1}+1)\rho^\alpha(p,x) + 2^{\alpha-1}\rho^\alpha(p,q).$$

Hence by Lebesgue DCT, we conclude that as $q \to p$, $F(q) \to F(p)$ and hence F is continuous.

For any $p_1, p_2 \in M$, by the triangular inequality, we get that

$$\rho(p_1,p_2) \leq \rho(p_1,x)+\rho(p_2,x)\ \forall x \in M$$

which implies that

$$\rho(p_1, p_2) \leq \int_M \rho(p_1, x)Q(dx) + \int_M \rho(p_2, x)Q(dx). \tag{2.2.5}$$

Apply Jensen's inequality to the Fréchet function, to get

$$[\int_M \rho(q, x)Q(dx)]^\alpha \leq \int_M \rho^\alpha(q, x)Q(dx) \ \forall\ q \in M$$

which implies that

$$\int_M \rho(q, x)Q(dx) \leq [F(q)]^{1/\alpha} \ \forall\ q \in M. \tag{2.2.6}$$

From equations (2.2.5) and (2.2.6), we get that

$$\rho(p_1, p_2) \leq [F(p_1)]^{1/\alpha} + [F(p_2)]^{1/\alpha}. \tag{2.2.7}$$

Since F is finite,

$$V = \inf\{F(q) : q \in M\} < \infty.$$

Get a sequence $\{p_n\}$ such that $F(p_n) \to V$. Then from equation (2.2.7), we get that

$$\rho(p_n, p_1) \leq [F(p_n)]^{1/\alpha} + [F(p_1)]^{1/\alpha}.$$

Hence the sequence $\{p_n\}$ is bounded. Let $\bar{B}$ denote its closure, which is closed and bounded, and hence compact by the proposition hypothesis. Therefore there exists a convergent subsequence $\{p_{n_k}\}$ such that $p_{n_k} \to p$ for some $p \in M$. By continuity of F, $F(p_{n_k}) \to F(p)$. Therefore $F(p) = V$ and hence C_Q is nonempty. From equation (2.2.7), it is easy to show that C_Q is compact. □

We define the **sample Fréchet mean** as a measurable selection from the sample Fréchet mean set C_{Q_n}. Proposition 2.2.2 below proves the strong consistency of the sample Fréchet mean as an estimator of the Fréchet mean of Q if it exists.

Proposition 2.2.2. *Suppose every closed and bounded subset of M is compact and the Fréchet function F of a probability measure Q is finite. Given any $\epsilon > 0$, there exists an integer-valued random variable $N = N(\omega, \epsilon)$ and a P-null set $A(\epsilon)$ such that*

$$C_{Q_n} \subseteq C_Q^\epsilon \doteq \{p \in M : \rho(p, C_Q) < \epsilon\} \ \forall n \geq N \tag{2.2.8}$$

outside of $A(\epsilon)$. *In particular, if* $C_Q = \{\mu_F\}$, *then every measurable selection* μ_{F_n} *from* C_{Q_n} *is a strongly consistent estimator of* μ_F.

Proof. For simplicity of notation, we write $C = C_Q$, $C_n = C_{Q_n}$, $\mu = \mu_F$ and $\mu_n = \mu_{F_n}$. Choose $\epsilon > 0$ arbitrarily. From the proof of Proposition 2.2.1, it follows that F is continuous. Replace Q by Q_n to conclude that F_n is also continuous.

Consider first the case when M is compact. If $C^\epsilon = M$, then equation (2.2.8) holds with $N = 1$. If $D = M \setminus C^\epsilon$ is nonempty, write

$$V = \inf\{F(p) : p \in M\} = F(q)\ \forall q \in C,$$
$$V + \delta(\epsilon) = \inf\{F(p) : p \in D\},\ \delta(\epsilon) > 0.$$

It is enough to show that

$$\sup\{|F_n(p) - F(p)| : p \in M\} \longrightarrow 0 \text{ a.s., as } n \to \infty. \tag{2.2.9}$$

For if equation (2.2.9) holds, then there exists $N \geq 1$ such that, outside a P-null set $A(\epsilon)$,

$$\inf\{F_n(p) : p \in C\} \leq V + \frac{\delta(\epsilon)}{3}, \tag{2.2.10}$$
$$\inf\{F_n(p) : p \in D\} \geq V + \frac{\delta(\epsilon)}{2},\ \forall n \geq N. \tag{2.2.11}$$

Clearly equations (2.2.10) and (2.2.11) imply (2.2.8). To prove equation (2.2.9), choose and fix $\epsilon' > 0$, however small. Since F and F_n are continuous and M is compact, hence they are uniformly continuous. Therefore there exists $\delta(\epsilon', n) > 0$ such that

$$|F(p) - F(p')| \leq \frac{\epsilon'}{4},\ |F_n(p) - F_n(p')| \leq \frac{\epsilon'}{4} \tag{2.2.12}$$

if $\rho(p, p') < \delta(\epsilon', n)$. Let $\{q_1, \ldots, q_k\}$ be a $\delta(\epsilon', n)$-net of M, that is, $\forall\ p \in M$ there exists $q(p) \in \{q_1, \ldots, q_k\}$ such that $\rho(p, q(p)) < \delta(\epsilon', n)$. By the strong law of large numbers, there exists an integer-valued random variable $N(\omega, \epsilon')$ such that outside of a P-null set $A(\epsilon')$, one has

$$|F_n(q_i) - F(q_i)| \leq \frac{\epsilon'}{4} \; \forall i = 1, 2, \ldots, k; \text{ if } n \geq N(\omega, \epsilon'). \tag{2.2.13}$$

From equations (2.2.12) and (2.2.13), we get

$$\begin{aligned} |F(p) - F_n(p)| &\leq |F(p) - F(q(p))| + |F(q(p)) - F_n(q(p))| + |F_n(q(p)) - F_n(p)| \\ &\leq \frac{3\epsilon'}{4} < \epsilon', \; \forall p \in M, \end{aligned}$$

if $n \geq N(\omega, \epsilon')$ outside of $A(\epsilon')$. This proves equation (2.2.9).

Next consider the case when M is non compact and hence unbounded. Let us find a compact set M_1 containing C and $N_1(\omega) > 1$ such that on $M \setminus M_1$,

$$F(p) \geq V + \delta(\epsilon), \; F_n(p) \geq V + \delta(\epsilon) \text{ a.s. } \forall \; n \geq N_1(\omega).$$

Then we can show as in case of compact M, that

$$\sup\{|F_n(p) - F(p)| : p \in M_1\} \longrightarrow 0 \text{ a.s., as } n \to \infty$$

and conclude that equation (2.2.8) holds. To get such a M_1, note that from equation (2.2.7) it follows that for any $p_1 \in C$ and $p \in M$

$$F(p) \geq [\rho(p, p_1) - V^{1/\alpha}]^\alpha. \tag{2.2.14}$$

Let

$$M_1 = \{p \colon \rho(p, C) \leq 2(V + \delta(\epsilon))^{1/\alpha} + V^{1/\alpha}\}.$$

Then from equation (2.2.14), one can check that $F(p) \geq 2(V + \delta(\epsilon)) \; \forall \; p \in M \setminus M_1$. Also from equation (2.2.7), we get for any $p \in M \setminus M_1$

$$F_n(p) \geq [\rho(p, p_1) - (F_n(p_1))^{1/\alpha}]^\alpha. \tag{2.2.15}$$

Since

$$\rho(p, p_1) - (F_n(p_1))^{1/\alpha} \xrightarrow{\text{a.s.}} \rho(p, p_1) - V^{1/\alpha} \geq 2(V + \delta(\epsilon))^{1/\alpha}$$

therefore there exists $N_1(\omega) > 1$ such that $\forall \; n \geq N_1(\omega)$, $F_n(p) \geq V + \delta(\epsilon)$ almost surely. This completes the proof. □

Remark 2.2.1. Proposition 2.2.2 generalizes Theorem 2.3 in Bhattacharya & Patrenganeru (2003) where α is taken to be 2 in the definition of the Fréchet function. For the case $\alpha = 2$, the consistency of the sample Fréchet mean for compact metric spaces, also follows from Ziezold (1977). We will mostly focus on this case but we will consider other Fréchet functions as well (see Section 7.5). In Bhattacharya and Bhattacharya (2008c), more general Fréchet functions are considered. However there the consistency is proved only for compact metric spaces.

Proposition 2.2.3 below proves the consistency of the sample Fréchet variation as an estimator of the Fréchet variation of Q.

Proposition 2.2.3. *Suppose every closed and bounded subset of M is compact and F is finite on M. Then the sample Fréchet variation V_n is a strongly consistent estimator of the Fréchet variation V of Q.*

Proof. In view of Proposition 2.2.2, for any $\epsilon > 0$, there exists $N = N(\omega, \epsilon)$ such that

$$|V_n - V| = |\inf_{p \in M} F_n(p) - \inf_{p \in M} F(p)| \leq \sup_{p \in \overline{C_Q^\epsilon}} |F_n(p) - F(p)| \tag{2.2.16}$$

for all $n \geq N$ almost surely. Also from the proof of Proposition 2.2.2, it follows that for any compact set $K \subseteq M$,

$$\sup_{p \in K} |F_n(p) - F(p)| \longrightarrow 0 \text{ a.s. as } n \to \infty.$$

Since $\overline{C_Q^\epsilon}$ is compact, it follows from equation (2.2.16) that

$$|V_n - V| \longrightarrow 0 \text{ a.s. as } n \to \infty.$$

□

Remark 2.2.2. In view of Proposition 2.2.3, the sample variation is a consistent estimator of the population variation even when the Fréchet function of Q does not have a unique minimizer.

Remark 2.2.3. From a generalization of the Hopf-Rinow theorem (see Hopf and Rinow (1931)), it follows that a complete and locally compact metric space (M, ρ) satisfies the topological hypothesis of Propositions 2.2.1, 2.2.2 and 2.2.3, that every closed and bounded subset of M is compact. Most of our spaces of interest in this thesis are compact metric spaces for which the hypothesis holds.

2.3 Asymptotic Distribution of the Sample Fréchet Mean

In this section we consider the asymptotic distribution of the sample Fréchet mean μ_{F_n}. From now on, we assume M to be a differentiable manifold of dimension d. Let ρ be a distance metrizing the topology of M. Theorem 2.3.1 below proves that under appropriate assumptions, the coordinates of μ_{F_n} are asymptotically normal. Here we denote by D_r the partial derivative with respect to the r^{th} coordinate $(r = 1, \ldots, d)$ and by D the vector of partial derivatives.

Theorem 2.3.1. *Suppose the following assumptions hold:*

A1 *Q has support in a single coordinate patch, (U, ϕ). ($\phi : U \longrightarrow \mathbb{R}^d$ smooth.) Let $\tilde{X}_j = \phi(X_j)$, $j = 1, \ldots, n$.*

A2 *Fréchet mean μ_F of Q is unique.*

A3 *For all x, $y \mapsto h(x, y) = \rho^{\alpha}\left(\phi^{-1}(x), \phi^{-1}(y)\right)$ is twice continuously differentiable in a neighborhood of $\phi(\mu_F) = \mu$.*

A4 $\mathrm{E}\{\mathrm{D}_r h(\tilde{X}_1, \mu)\}^2 < \infty \ \forall r = 1, \ldots, d$.

A5 $\mathrm{E}\{\sup_{|u-v|\le\epsilon} |\mathrm{D}_s\mathrm{D}_r h(\tilde{X}_1, v) - \mathrm{D}_s\mathrm{D}_r h(\tilde{X}_1, u)|\} \to 0$ *as* $\epsilon \to 0 \ \forall\ r, s$.

A6 $\Lambda = ((\ \mathrm{E}\{\mathrm{D}_s\mathrm{D}_r h(\tilde{X}_1, \mu)\}\ ))$ *is nonsingular.*

Let μ_{F_n} be a measurable selection from the sample Fréchet mean set and define $\mu_n = \phi(\mu_{F_n})$. Then under the assumptions A1-A6,

$$\sqrt{n}(\mu_n - \mu) \xrightarrow{\mathcal{L}} N(0, \Lambda^{-1}\Sigma(\Lambda')^{-1}) \tag{2.3.1}$$

where $\Sigma = \mathrm{Cov}[\mathrm{D}h(\tilde{X}_1, \mu)]$.

Proof. Write $\psi^{(r)}(x,y) = \mathrm{D}_r h(x,y) \equiv \frac{\partial}{\partial y_r} h(x,y)$ for $x, y \in \mathbb{R}^d$. Let $Q^\phi = Q \circ \phi^{-1}$. Denote

$$\tilde{F}(y) = \int_{\mathbb{R}^d} \rho^\alpha(\phi^{-1}(x), \phi^{-1}(y)) Q^\phi(dx),\ \tilde{F}_n(y) = \frac{1}{n}\sum_{j=1}^n \rho^\alpha(\phi^{-1}(\tilde{X}_j), \phi^{-1}(y))$$

for $y \in \mathbb{R}^d$. Then $\tilde{F}$ has unique minimizer μ while $\tilde{F}_n$ has minimizer μ_n. Therefore

$$\begin{aligned} 0 = \frac{1}{\sqrt{n}}\sum_{j=1}^n \psi^{(r)}(\tilde{X}_j, \mu_n) &= \frac{1}{\sqrt{n}}\sum_{j=1}^n \psi^{(r)}(\tilde{X}_j, \mu) \\ &+ \sum_{s=1}^d \sqrt{n}(\mu_n - \mu)_s \frac{1}{n}\sum_{j=1}^n \mathrm{D}_s \psi^{(r)}(\tilde{X}_j, \mu) \\ &+ \sum_{s=1}^d \sqrt{n}(\mu_n - \mu)_s (\epsilon_n)_{rs},\ 1 \le r \le d, \qquad (2.3.2) \\ \text{where } (\epsilon_n)_{rs} = \frac{1}{n}\sum_{j=1}^n &[\mathrm{D}_s\psi^{(r)}(\tilde{X}_j, \theta_n) - \mathrm{D}_s\psi^{(r)}(\tilde{X}_j, \mu)] \end{aligned}$$

for some θ_n lying on the line segment joining μ and μ_n. Equation (2.3.2) implies that

$$\left[\left(\left(\frac{1}{n}\sum_{j=1}^n \mathrm{D}_s \mathrm{D}_r h(\tilde{X}_j, \mu) + \epsilon_n\right)\right)\right] \sqrt{n}(\mu_n - \mu) = -\frac{1}{\sqrt{n}}\sum_{j=1}^n \mathrm{D}h(\tilde{X}_j, \mu).$$

In view of assumptions A5 and A6, it follows that

$$\sqrt{n}(\mu_n - \mu) = -\Lambda^{-1}\left(\frac{1}{\sqrt{n}}\sum_{j=1}^n \mathrm{D}h(\tilde{X}_j, \mu)\right) + o_P(1)$$

which implies

$$\sqrt{n}(\mu_n - \mu) \xrightarrow{\mathcal{L}} -\Lambda^{-1}N(0, \Sigma) = N(0, \Lambda^{-1}\Sigma(\Lambda')^{-1}).$$

□

From Theorem 2.3.1, it follows that under assumptions A1-A6 and assuming Σ to be nonsingular,

$$n(\mu_n - \mu)'\Lambda'\Sigma^{-1}\Lambda(\mu_n - \mu) \xrightarrow{\mathcal{L}} \mathcal{X}_d^2 \text{ as } n \to \infty.$$

Here $\mathcal{X}_d^2$ denotes the chi-squared distribution with d degrees of freedom. This can be used to construct an asymptotic confidence set for μ_F, namely

$$\{\mu_F : n(\mu_n - \mu)'\hat{\Lambda}'\hat{\Sigma}^{-1}\hat{\Lambda}(\mu_n - \mu) \leq \mathcal{X}_d^2(1-\theta)\}. \tag{2.3.3}$$

Here $\hat{\Lambda}$ and $\hat{\Sigma}$ are the sample estimates of Λ and Σ respectively and $\mathcal{X}_d^2(1-\theta)$ is the upper $(1-\theta)$-quantile of $\mathcal{X}_d^2$ distribution. The corresponding pivotal bootstrapped confidence region is given by

$$\{\mu_F : n(\mu_n - \mu)'\hat{\Lambda}'\hat{\Sigma}^{-1}\hat{\Lambda}(\mu_n - \mu) \leq c^*(1-\theta)\} \tag{2.3.4}$$

where $c^*(1-\theta)$ is the upper $(1-\theta)$ quantile of the bootstrapped values of the statistic in equation (2.3.3).

2.4 Asymptotic Distribution of the Sample Fréchet Variation

Next we derive the asymptotic distribution of V_n when there is a unique population Fréchet mean.

Theorem 2.4.1. *Let M be a differentiable manifold. Using the notation of Theorem 2.3.1, under assumptions* A1-A6 *and assuming* $\mathrm{E}[\rho^{2\alpha}(X_1, \mu_F)] < \infty$*, one has*

$$\sqrt{n}(V_n - V) \xrightarrow{\mathcal{L}} N\left(0, \mathrm{var}(\rho^{\alpha}(X_1, \mu_F))\right). \tag{2.4.1}$$

Proof. Let

$$\tilde{F}(x) = \int_M \rho^{\alpha}(\phi^{-1}(x), m)Q(dm),\ \tilde{F}_n(x) = \frac{1}{n}\sum_{j=1}^{n}\rho^{\alpha}(\phi^{-1}(x), X_j)$$

for $x \in \mathbb{R}^d$. Let μ_{F_n} be a measurable selection from the sample Fréchet mean set and

$\mu_n = \phi(\mu_{F_n})$. Then

$$\sqrt{n}(V_n - V) = \sqrt{n}(\tilde{F}_n(\mu_n) - \tilde{F}(\mu))$$
$$= \sqrt{n}(\tilde{F}_n(\mu_n) - \tilde{F}_n(\mu)) + \sqrt{n}(\tilde{F}_n(\mu) - \tilde{F}(\mu)), \tag{2.4.2}$$

$$\sqrt{n}(\tilde{F}_n(\mu_n) - \tilde{F}_n(\mu)) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \sum_{r=1}^{d} (\mu_n - \mu)_r \mathrm{D}_r h(\tilde{X}_j, \mu)$$
$$+ \frac{1}{2\sqrt{n}} \sum_{j=1}^{n} \sum_{r=1}^{d} \sum_{s=1}^{d} (\mu_n - \mu)_r (\mu_n - \mu)_s \mathrm{D}_s \mathrm{D}_r h(\tilde{X}_j, \theta_n) \tag{2.4.3}$$

for some θ_n in the line segment joining μ and μ_n. By assumption A5 of Theorem 2.3.1 and because $\sqrt{n}(\mu_n - \mu)$ is asymptotically normal, the second term on the right of equation (2.4.3) converges to 0 in probability. Also

$$\frac{1}{n} \sum_{j=1}^{n} \mathrm{D}h(\tilde{X}_j, \mu) \xrightarrow{P} \mathrm{E}\left(\mathrm{D}h(\tilde{X}_1, \mu)\right) = 0,$$

so that the first term on the right of equation (2.4.3) converges to 0 in probability. Hence (2.4.2) becomes

$$\sqrt{n}(V_n - V) = \sqrt{n}(\tilde{F}_n(\mu) - \tilde{F}(\mu)) + o_P(1)$$
$$= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} (\rho^\alpha(X_j, \mu_F) - \mathrm{E}[\rho^\alpha(X_1, \mu_F)]) + o_P(1). \tag{2.4.4}$$

By the C.L.T. for the iid sequence $\{\rho^\alpha(X_j, \mu_F)\}$, $\sqrt{n}(V_n - V)$ converges in distribution to $N(0, \mathrm{var}(\rho^\alpha(X_1, \mu_F)))$. □

Remark 2.4.1. Although Proposition 2.2.2 does not require the uniqueness of the Fréchet mean of Q for V_n to be a consistent estimator of V, Theorem 2.4.1 requires the Fréchet mean of Q to exist for the sample variation to be asymptotically Normal. It may be shown by examples (see Section 4.5) that it fails to give the correct distribution when there is not a unique mean.

Using Theorem 2.4.1, we can construct the following confidence interval I for V:

$$I = \{V \in [V_n - \frac{s}{\sqrt{n}} Z(1 - \frac{\theta}{2}), V_n + \frac{s}{\sqrt{n}} Z(1 - \frac{\theta}{2})]\}. \tag{2.4.5}$$

The interval I has asymptotic confidence level of $(1-\theta)$. Here s^2 is the sample variance of $\rho^{\alpha}(X_j, \mu_{F_n})$, $j = 1, \ldots, n$ and $Z(1-\theta/2)$ denotes the upper $(1-\frac{\theta}{2})$-quantile of standard Normal distribution. From the confidence interval I, we can also construct a pivotal bootstrap confidence interval for V, the details of which are left to the reader.

Chapter 3

INTRINSIC ANALYSIS ON MANIFOLDS

3.1 Introduction

Let (M, g) be a complete connected Riemannian manifold of dimension d with metric tensor g. Then the natural choice for the distance metric ρ in Chapter 2 is the geodesic distance d_g on M. The statistical analysis on M using this distance is called **intrinsic analysis**. Unless otherwise stated, we consider $\alpha = 2$ in the definition of the Fréchet function in equation (2.2.1). However we will consider other Fréchet functions as well for suitable $\alpha \geq 1$ (see Section 7.5).

3.2 Intrinsic Mean and Variation

Let Q be a probability distribution on M with finite Fréchet function

$$F(p) = \int_M d_g^2(p, m) Q(dm). \tag{3.2.1}$$

Let $X_1, \ldots, X_n$ be an independent and identically distributed (iid) sample from Q.

Definition 3.2.1. The Fréchet mean set of Q under $\rho = d_g$ and $a = 2$ is called the **intrinsic mean set** of Q and the Fréchet variation of Q is called the **intrinsic variation** of Q. The Fréchet mean set of the empirical distribution Q_n is called the **sample intrinsic mean set** and the sample Fréchet variation is called the **sample intrinsic variation**.

Before proceeding further, let us define a few technical terms related to Riemannian manifolds which we will use extensively in this chapter. For details on Riemannian Manifolds, see DoCarmo (1992), Gallot et al. (1993) or Lee (1997).

1. **Geodesic**: These are curves γ on the manifold with zero acceleration. They are locally length minimizing curves. For example, consider great circles on the sphere or straight lines in $\mathbb{R}^d$.

2. **Exponential map**: For $p \in M$, $v \in T_pM$, we define $\exp_p(v) = \gamma(1)$, where γ is a geodesic with $\gamma(0) = p$ and $\dot{\gamma}(0) = v$.

3. **Cut locus**: For a point $p \in M$, we define the cut locus $C(p)$ of p as the set of points of the form $\gamma(t_0)$, where γ is a unit speed geodesic starting at p and t_0 is the supremum of all $t > 0$ such that γ is distance minimizing from p to $\gamma(t)$. For example, $C(p) = \{-p\}$ on the sphere.

4. **Sectional Curvature**: Recall the notion of Gaussian curvature of two dimensional surfaces. On a Riemannian manifold M, choose a pair of linearly independent vectors $u, v \in T_pM$. A two dimensional submanifold of M is swept out by the set of all geodesics starting at p and with initial velocities lying in the two-dimensional section π spanned be u, v. The Gaussian curvature of this submanifold is called the sectional curvature at p of the section π.

5. **Injectivity Radius**: We define the injectivity radius of M as

$$\text{inj}(M) = \inf\{d_g(p, C(p)) : p \in M\}.$$

   For example the sphere of radius 1 has injectivity radius equal to π.

6. **Convex Set**: A subset S of M is said to be convex if for any $x, y \in S$, there exists a unique shortest geodesic in M joining x and y which lies entirely in S.

Also let $r_* = \min\{\text{inj}(M), \frac{\pi}{\sqrt{\overline{C}}}\}$, where $\overline{C}$ is the least upper bound of sectional curvatures of M if this upper bound is positive, and $\overline{C} = 0$ otherwise. The exponential map at p is injective on $\{v \in T_p(M) : |v| < r_*\}$. By $B(p, r)$ we will denote an open ball with center $p \in M$ and radius r, and $\bar{B}(p, r)$ will denote its closure. It is known that $B(p,r)$ is convex whenever $r \leq \frac{r_*}{2}$.

In case Q has a unique intrinsic mean μ_I, it follows from Proposition 2.2.2 and Remark 1.2.3 that the **sample intrinsic mean** μ_{nI}, that is, a measurable selection from the sample intrinsic mean set is a consistent estimator of μ_I. Broad conditions for the existence of a unique intrinsic mean are not known. From results due to Karchar (1977) and Le (2001), it follows that if the support of Q is in a geodesic ball of radius $\frac{r_*}{4}$, i.e. $\text{supp}(Q) \subseteq B(p, \frac{r_*}{4})$, then Q has a unique intrinsic mean. This result has been substantially extended by Kendall (1990) which shows that if $\text{supp}(Q) \subseteq B(p, \frac{r_*}{2})$, then there is a unique local minimum of the Fréchet function F in that ball. Then we redefine the (local) intrinsic mean of Q as that unique minimizer in the ball. In that case one can show that the (local) sample intrinsic mean is a consistent estimator of the intrinsic mean of Q. This is stated in Proposition 3.2.1 below.

Proposition 3.2.1. *Let Q have support in $B(p, \frac{r_*}{2})$ for some $p \in M$. Then (a) Q has a unique (local) intrinsic mean μ_I in $B(p, \frac{r_*}{2})$ and (b) the sample intrinsic mean μ_{nI} in $B(p, \frac{r_*}{2})$ is a strongly consistent estimator of μ_I.*

Proof. (a) Follows from Kendall (1990).
(b) Since $\text{supp}(Q)$ is compact, $\text{supp}(Q) \subseteq B(p,r)$ for some $r < \frac{r_*}{2}$. It is shown in Karchar (1977) that

$$\text{grad}F(q) = -2 \int \exp_q^{-1}(m) Q(dm), \; q \in \bar{B}(p, \frac{r_*}{2}).$$

It can be shown that if $q \in B(p, \frac{r_*}{2}) \setminus B(p,r)$, then there exists a hyperplane in T_qM such that $\exp_q^{-1}(m)$ lies on one side of that hyperplane for all $m \in B(p,r)$ (see Lemma 1, Le (2001)). Hence $\text{grad}F(q)$ cannot be equal to zero. Therefore $\mu_I \in B(p,r)$ and it is the unique intrinsic mean of Q restricted to $\bar{B}(p,r)$. Now consistency follows by applying Proposition 2.2.2 to the compact metric space $\bar{B}(p,r)$. $\square$

3.3 Asymptotic Distribution of the Sample Intrinsic Mean

For the asymptotic distribution of the sample intrinsic mean, we may use Theorem 2.3.1. For that we need to verify assumptions A1-A6. Theorem 3.3.1 which, as proved in Bhattacharya and Bhattacharya (2008b), gives sufficient conditions for those assumptions to hold. In the statement of the theorem, the usual partial order $A \geq B$ between $d \times d$ symmetric matrices A, B, means that $A - B$ is nonnegative definite.

Theorem 3.3.1. *Suppose* $\text{supp}(Q) \subseteq B(p, \frac{r_*}{2})$ *for some* $p \in M$. *Let* $\phi = \exp_{\mu_I}^{-1} : B(p, \frac{r_*}{2}) \longrightarrow T_{\mu_I}M(\approx \mathbb{R}^d)$. *Then the map* $y \mapsto h(x, y) = d_g^2(\phi^{-1}x, \phi^{-1}y)$ *is twice continuously differentiable in a neighborhood of* 0 *and in terms of normal coordinates with respect to a chosen orthonormal basis for* $T_{\mu_I}M$,

$$\mathrm{D}_r h(x, 0) = -2x^r, \ 1 \leq r \leq d, \tag{3.3.1}$$

$$[\mathrm{D}_r \mathrm{D}_s h(x, 0)] \geq \left[2\{\left(\frac{1 - f(|x|)}{|x|^2}\right) x^r x^s + f(|x|)\delta_{rs}\}\right]_{1 \leq r,s \leq d}. \tag{3.3.2}$$

Here $x = (x^1, \ldots, x^d)'$, $|x| = \sqrt{(x^1)^2 + (x^2)^2 + \ldots (x^d)^2}$ *and*

$$f(y) = \begin{cases} 1 \ \text{if } \overline{C} = 0 \\ \sqrt{\overline{C}}y\frac{\cos(\sqrt{\overline{C}}y)}{\sin(\sqrt{\overline{C}}y)} \ \text{if } \overline{C} > 0 \\ \sqrt{-\overline{C}}y\frac{\cosh(\sqrt{-\overline{C}}y)}{\sinh(\sqrt{-\overline{C}}y)} \ \text{if } \overline{C} < 0. \end{cases} \tag{3.3.3}$$

There is equality in equation (3.3.2) *when* M *has constant sectional curvature* $\overline{C}$, *and in this case* Λ *has the expression:*

$$\Lambda_{rs} = 2\mathrm{E}\{\left(\frac{1 - f(|\tilde{X}_1|)}{|\tilde{X}_1|^2}\right) \tilde{X}_1^r \tilde{X}_1^s + f(|\tilde{X}_1|)\delta_{rs}\}, \ \ 1 \leq r, s \leq d, \tag{3.3.4}$$

Λ *being positive definite if* $\text{supp}(Q) \subseteq B(\mu_I, \frac{r_*}{2})$.

Proof. Let $\gamma(s)$ be a geodesic, $\gamma(0) = \mu_I$. Define $c(s, t) = exp_m(texp_m^{-1}\gamma(s))$, $s \in [0, \epsilon]$, $t \in [0, 1]$, as a smooth variation of γ through geodesics lying entirely in $B(p, \frac{r_*}{2})$.

Let $T = \frac{\partial}{\partial t}c(s,t)$, $S = \frac{\partial}{\partial s}c(s,t)$. Since $c(s,0) = m$, $S(s,0) = 0$; and since $c(s,1) = \gamma(s)$, $S(s,1) = \dot{\gamma}(s)$. Also $\langle T, T\rangle = d_g^2(\gamma(s), m)$ is independent of t, and the covariant derivative $D_t T$ vanishes because $t \mapsto c(s,t)$ is a geodesic (for each s). Then

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

Hence $d_g^2(\gamma(s), m)$ is C^∞ smooth, and using the symmetry of the connection on a parametrized surface (see Lemma 3.4, Do Carmo (19992)), we get

$$\begin{aligned} \frac{d}{ds} d_g^2(\gamma(s), m) &= 2\int_0^1 \langle D_s T, T\rangle dt = 2\int_0^1 \frac{d}{dt}\langle T, S\rangle dt \\ &= 2\langle T(s,1), S(s,1)\rangle = -2\langle exp_{\gamma(s)}^{-1} m, \dot{\gamma}(s)\rangle. \end{aligned} \tag{3.3.5}$$

Substituting $s = 0$ in equation (3.3.5), we get expressions for $D_r h(x,0)$ as in equation (3.3.1). Also

$$\begin{aligned} \frac{d^2}{ds^2} d_g^2(\gamma(s), m) &= 2\langle D_s T(s,1), S(s,1)\rangle \\ &= 2\langle D_t S(s,1), S(s,1)\rangle = 2\langle D_t J_s(1), J_s(1)\rangle \end{aligned} \tag{3.3.6}$$

where $J_s(t) = S(s,t)$. Note that J_s is a Jacobi field along $c(s,.)$ with $J_s(0) = 0$, $J_s(1) = \dot{\gamma}(s)$. Let $J_s^\perp$ and J_s^- be the normal and tangential components of J_s. Let η be a unit speed geodesic in M and J a normal Jacobi field along η, $J(0) = 0$. Define

$$u(t) = \begin{cases} t & \text{if } \bar{C} = 0 \\ \frac{\sin(\sqrt{\bar{C}}t)}{\sqrt{\bar{C}}} & \text{if } \bar{C} > 0 \\ \frac{\sinh(\sqrt{-\bar{C}}t)}{\sqrt{-\bar{C}}} & \text{if } \bar{C} < 0. \end{cases}$$

Then $u''(t) = -\bar{C}u(t)$ and

$$(|J|'u - |J|u')'(t) = (|J|'' + \bar{C}|J|)u(t).$$

By exact differentiation and Schwartz inequality, it is easy to show that $|J|'' + \bar{C}|J| \geq 0$, hence $(|J|'u - |J|u')'(t) \geq 0$ whenever $u(t) \geq 0$. This implies that $|J|'u - |J|u' \geq 0$

if $t \leq t_0$, where u is positive on $(0, t_0)$. Also $|J|' = \frac{\langle J', J \rangle}{|J|}$. Therefore $\langle J(t), D_t J(t) \rangle \geq \frac{u'(t)}{u(t)} |J(t)|^2 \; \forall \; t < t_0$. If we drop the unit speed assumption on η, we get

$$\langle J(1), D_t J(1) \rangle \geq |\dot{\eta}| \frac{u^{'}(|\dot{\eta}|)}{u(|\dot{\eta}|)} |J(1)|^2 \text{ if } |\dot{\eta}| < t_0. \tag{3.3.7}$$

Here $t_0 = \infty$ if $\bar{C} \leq 0$ and equals $\frac{\pi}{\sqrt{\bar{C}}}$ if $\bar{C} > 0$. When M has constant sectional curvature $\bar{C}$, $J(t) = u(t)E(t)$ where E is a parallel normal vector field along η. Hence

$$\langle J(t), D_t J(t) \rangle = u(t) u^{'}(t) |E(t)|^2 = \frac{u^{'}(t)}{u(t)} |J(t)|^2.$$

If we drop the unit speed assumption, we get

$$\langle J(t), D_t J(t) \rangle = |\dot{\eta}| \frac{u^{'}(|\dot{\eta}| t)}{u(|\dot{\eta}| t)} |J(t)|^2. \tag{3.3.8}$$

Since $J_s^{\perp}$ is a normal Jacobi field along the geodesic $c(s, .)$, from equations (3.3.7) and (3.3.8), it follows that

$$\langle J_s^{\perp}(1), D_t J_s^{\perp}(1) \rangle \geq f(d(\gamma(s), m)) |J_s^{\perp}(1)|^2 \tag{3.3.9}$$

with equality in equation (3.3.9) when M has constant sectional curvature $\bar{C}$, f being defined in equation (3.3.3).

Next suppose J is a Jacobi field along a geodesic η, $J(0) = 0$ and let $J^-(t)$ be its tangential component. Then $J^-(t) = \lambda t \dot{\eta}(t)$ where $\lambda t = \frac{\langle J(t), \dot{\eta}(t) \rangle}{|\dot{\eta}|^2}$, λ being independent of t. Hence

$$\begin{aligned} (D_t J)^-(t) &= \frac{\langle D_t J(t), \dot{\eta}(t) \rangle}{|\dot{\eta}|^2} \dot{\eta}(t) \\ &= \frac{d}{dt} \left(\frac{\langle J(t), \dot{\eta}(t) \rangle}{|\eta|^2} \right) \dot{\eta}(t) = \lambda \dot{\eta}(t) = D_t (J^-)(t) \end{aligned} \tag{3.3.10}$$

and

$$\begin{aligned} D_t |J^-|^2(1) &= 2\lambda^2 |\dot{\eta}|^2 = 2 \frac{\langle J(1), \dot{\eta}(1) \rangle^2}{|\dot{\eta}(1)|^2} \\ &= D_t \langle J, J^- \rangle (1) = \langle D_t J(1), J^-(1) \rangle + |J^-(1)|^2 \end{aligned}$$

which implies that

$$\langle D_t J(1), J^-(1)\rangle = 2\frac{\langle J(1), \dot{\eta}(1)\rangle^2}{|\dot{\eta}(1)|^2} - |J^-(1)|^2 = \frac{\langle J(1), \dot{\eta}(1)\rangle^2}{|\dot{\eta}(1)|^2}. \tag{3.3.11}$$

Apply (3.3.10) and (3.3.11) to the Jacobi field J_s to get

$$D_t(J_s^-)(1) = (D_t J_s)^-(1) = J_s^-(1) = \frac{\langle J_s(1), T(s,1)\rangle}{|T(s,1)|^2} T(s,1), \tag{3.3.12}$$

$$\langle D_t J_s(1), J_s^-(1)\rangle = \frac{\langle J_s(1), T(s,1)\rangle^2}{|T(s,1)|^2}. \tag{3.3.13}$$

Using (3.3.9), (3.3.12) and (3.3.13), equation (3.3.6) becomes

$$\begin{aligned}
\frac{d^2}{ds^2} d_g^2(\gamma(s), m) &= 2\langle D_t J_s(1), J_s(1)\rangle \\
&= 2\langle D_t J_s(1), J_s^-(1)\rangle + 2\langle D_t J_s(1), J_s^\perp(1)\rangle \\
&= 2\langle D_t J_s(1), J_s^-(1)\rangle + 2\langle D_t (J_s^\perp)(1), J_s^\perp(1)\rangle \\
&\geq 2\frac{< J_s(1), T(s,1) >^2}{|T(s,1)|^2} + 2f(|T(s,1)|)|J_s^\perp(1)|^2 \qquad (3.3.14) \\
&= 2\frac{\langle J_s(1), T(s,1)\rangle^2}{|T(s,1)|^2} + 2f(|T(s,1)|)|J_s(1)|^2 \\
&\quad - 2f(|T(s,1)|)\frac{\langle J_s(1), T(s,1)\rangle^2}{|T(s,1)|^2} \\
&= 2f(d_g(\gamma(s), m))|\dot{\gamma}(s)|^2 \\
&\quad + 2(1 - f(d_g(\gamma(s), m)))\frac{\langle \dot{\gamma}(s), exp_{\gamma(s)}^{-1} m\rangle^2}{d_g^2(\gamma(s), m)} \qquad (3.3.15)
\end{aligned}$$

with equality in (3.3.14) when M has constant sectional curvature $\bar{C}$. Substituting $s = 0$ in equation (3.3.15), we get a lower bound for $[\mathrm{D}_r \mathrm{D}_s h(x, 0)]$ as in equation (3.3.2) and an exact expression for $\mathrm{D}_r \mathrm{D}_s h(x, 0)$ when M has constant sectional curvature. To see this, let $\dot{\gamma}(0) = v$. Then writing $m = \phi^{-1}(x)$, $\gamma(s) = \phi^{-1}(sv)$, one has

$$\begin{aligned}
&\frac{d^2}{ds^2} d_g^2(\gamma(s), m)\Big|_{s=0} = \frac{d^2}{ds^2} d_g^2(\phi^{-1}(x), \phi^{-1}(sv))\Big|_{s=0} \\
=&\frac{d^2}{ds^2} h(x, sv)\Big|_{s=0} = \sum_{r,s=1}^{d} v_r v_s D_r D_s h(x, 0).
\end{aligned}$$

Since $d^2(\gamma(s), m)$ is twice continuously differentiable and Q has compact support, using the Lebesgue DCT, we get,

$$\frac{d^2}{ds^2}F(\gamma(s))|_{s=0} = \int \frac{d^2}{ds^2}d^2(\gamma(s), m)|_{s=0}Q(dm). \tag{3.3.16}$$

Then (3.3.4) follows from (3.3.15). If $\text{supp}(Q) \subseteq B(\mu_I, \frac{r_*}{2})$, then the expression in equation (3.3.15) is strictly positive at $s = 0$ for all $m \in \text{supp}(Q)$, hence Λ is positive definite. This completes the proof. □

Corollary 3.3.2. *Suppose* $\text{supp}(Q) \subseteq B(\mu_I, \frac{r_*}{2})$, μ_I *being the intrinsic mean of* Q*. Let* $X_1, \ldots, X_n$ *be an iid sample from* Q *and* $\tilde{X}_j = \phi(X_j)$, $j = 1, \ldots, n$ *be the normal coordinates of the sample with* ϕ *as in Theorem 3.3.1. Let* μ_{nI} *be the sample intrinsic mean in* $B(\mu_I, \frac{r_*}{2})$*. Then* $\text{E}[\tilde{X}_1] = 0$ *and*

$$\sqrt{n}\phi(\mu_{nI}) \xrightarrow{\mathcal{L}} N(0, \Lambda^{-1}\Sigma\Lambda^{-1})$$

where $\Sigma = 4\text{E}(\tilde{X}_1\tilde{X}_1')$ *and* Λ *being derived in Theorem 3.3.1.*

Proof. Follows from Theorem 3.3.1 and Theorem 2.3.1. □

As in Section 2.3, if Σ is nonsingular, we can construct asymptotic chi-squared and pivotal bootstrapped confidence regions for μ_I. Σ is nonsingular if $Q \circ \phi^{-1}$ has support in no smaller dimensional subspace of $\mathbb{R}^d$. That holds if for example Q has a density with respect to the volume measure on M.

Alternatively one may consider the statistic

$$T_n = d_g^2(\mu_{nI}, \mu_I).$$

Then $T_n = \|\phi(\mu_{nI})\|^2$, hence from Corollary 3.3.2, it follows that

$$nT_n \xrightarrow{\mathcal{L}} \sum_{i=1}^{d} \lambda_i Z_i^2$$

where $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_d$ are the eigenvalues of $\Lambda^{-1}\Sigma\Lambda^{-1}$ and $Z_1, \ldots, Z_d$ are iid $N(0, 1)$.

Using this statistic, an asymptotic level $(1-\alpha)$ confidence set for μ_I can be given by

$$\{\mu_I : nT_n \leq \hat{c}_{1-\alpha}\} \tag{3.3.17}$$

where $\hat{c}_{1-\alpha}$ is the estimated upper $(1-\alpha)$ quantile of the distribution of $\sum_{i=1}^{d} \hat{\lambda}_i Z_i^2$, $\hat{\lambda}_i$ being the sample estimate of λ_i, $i = 1, 2, \ldots, d$ and $(Z_1, Z_2, \ldots)$ is a sample of iid $N(0, 1)$ random variables independent of the original sample $X_1, \ldots, X_n$.

The corresponding bootstrapped confidence region is given by

$$\{\mu_I : nT_n \leq c^*_{1-\alpha}\} \tag{3.3.18}$$

where $c^*(1-\alpha)$ is the upper $(1-\alpha)$ quantile of the bootstrapped values of the statistic nT_n. The advantage of using the confidence set in equation (3.3.18) over that in (2.3.4) is that it is easier to compute and does not require Σ to be nonsingular. However unlike (2.3.4), it is not a pivotal confidence region.

3.4 Example: Directional Space S^d

Consider the space of all directions in $\mathbb{R}^{d+1}$. Since any direction has a unique point of intersection with the unit sphere S^d in $\mathbb{R}^{d+1}$, this space can be identified with S^d which is

$$S^d = \{p \in \mathbb{R}^{d+1} \colon \|p\| = 1\}.$$

At each $p \in S^d$, we endow the tangent space

$$T_pS^d = \{v \in \mathbb{R}^{d+1} \colon v'p = 0\}$$

with the metric tensor $g_p : T_pS^d \times T_pS^d \to \mathbb{R}$ as the restriction of the scaler product at p of the tangent space of $\mathbb{R}^d$: $g_p(v_1, v_2) = v_1'v_2$. Then g is a smooth metric tensor on the tangent bundle

$$TS^d = \{(p, v) : p \in S^d, v \in \mathbb{R}^{d+1} : v'p = 0\}.$$

The geodesics are the great circles,

$$\gamma_{p,v}(t) = \cos(t)p + \sin(t)v, \ -\pi < t \leq \pi$$

Here $\gamma_{p,v}(.)$ is the great circle starting at p at $t = 0$ in the direction of the unit vector v. The exponential map, $\exp : T_pS^d \to S^d$ is given by

$$\begin{aligned} \exp_p(0) &= p, \\ \exp_p(v) &= \cos(\|v\|)p + \sin(\|v\|)\frac{v}{\|v\|}, \ v \neq 0. \end{aligned}$$

The inverse of the exponential map on $S^d \setminus \{-p\}$ into T_pS^d has the expression

$$\begin{aligned} \exp_p^{-1}(q) &= \frac{\arccos(p'q)}{\sqrt{1-(p'q)^2}}[q - (p'q)p] \ (q \neq p, -p), \\ \exp_p^{-1}(p) &= 0. \end{aligned}$$

The geodesic distance between p and q is given by

$$d_g(p, q) = |\arccos(p'q)|$$

which lies in $[0, \pi]$. Hence S^d has a injectivity radius of π. Also it has a constant sectional curvature of 1, therefore $r_* = \pi$.

Let Q be a probability distribution on S^d. It follows from Proposition 3.2.1 that if supp(Q) lies in an open geodesic ball of radius $\frac{\pi}{2}$, then it has a unique intrinsic mean μ_I in that ball. If $X_1, \ldots, X_n$ is an iid random sample from Q, then the sample intrinsic mean μ_{nI} in that ball is a strongly consistent estimator of μ_I. From Corollary 3.3.2 it follows that

$$\sqrt{n}\phi(\mu_{nI}) \xrightarrow{\mathcal{L}} N(0, \Lambda^{-1}\Sigma\Lambda^{-1})$$

where $\Sigma = 4\mathrm{E}[\phi(X_1)\phi(X_1)']$. To get expression for ϕ, pick an orthonormal basis $\{v_1,\ldots,v_d\}$ for $T_{\mu_I}S^d$. For $x \in S^d$, $|x'\mu_I| < 1$, we have

$$\exp_{\mu_I}^{-1}(x) = \frac{\arccos(x'\mu_I)}{\sqrt{1-(x'\mu_I)^2}}[x - (x'\mu_I)\mu_I].$$

Then

$$\phi(x) = y \equiv (y^1,\ldots,y^d)'$$

where $\exp_{\mu_I}^{-1}(x) = \sum_{r=1}^d y^r v_r$, so that

$$y^r = \frac{\arccos(x'\mu_I)}{\sqrt{1-(x'\mu_I)^2}}(x'v_r), \quad r = 1,2,\ldots,d.$$

From Theorem 3.3.1, we get the expression for Λ as

$$\Lambda_{rs} = 2\mathrm{E}[\frac{1}{[1-(X_1'\mu_I)^2]}\left(1 - \frac{\arccos(X_1'\mu_I)}{\sqrt{1-(X_1'\mu_I)^2}}(X_1'\mu_I)\right)(X_1'v_r)(X_1'v_s)$$
$$+ \frac{\arccos(X_1'\mu_I)}{\sqrt{1-(X_1'\mu_I)^2}}(X_1'\mu_I)\delta_{rs}], \quad 1 \le r \le s \le d,$$

Λ being positive definite if $\mathrm{supp}(Q) \subseteq B(\mu_I, \frac{\pi}{2})$.

3.5 Two Sample Intrinsic Tests

In this section, we will construct nonparametric tests to compare the intrinsic means and variations of two probability distributions Q_1 and Q_2 on M. This can be used to distinguish between the two distributions.

3.5.1 Independent Samples

Let $X_1,\ldots,X_{n_1}$ and $Y_1,\ldots,Y_{n_2}$ be two iid samples from Q_1 and Q_2 respectively that are mutually independent. Let μ_i and V_i denote the intrinsic means and variations of Q_i, $i = 1,2$ respectively. Similarly denote by $\hat{\mu}_i$ and $\hat{V}_i$ the sample intrinsic means and variations.

First we test the hypothesis, $H_0 : \mu_1 = \mu_2 = \mu$, say, against $H_1 : \mu_1 \neq \mu_2$. We assume that under H_0, both Q_1 and Q_2 have support in $B(\mu, \frac{r_*}{2})$, so that the normal coordinates of the sample intrinsic means have asymptotic Normal distribution. Let $\phi(\hat{\mu}_i)$, $i = 1, 2$ where $\phi = exp_\mu^{-1}$ be the normal coordinates of the sample means in $T_\mu M$ ($\approx \mathbb{R}^d$). It follows from Corollary 3.3.2 that

$$\sqrt{n_i}\phi(\hat{\mu}_i) \xrightarrow{\mathcal{L}} N(0, \Lambda_i^{-1}\Sigma_i\Lambda_i^{-1}),\ i = 1, 2 \tag{3.5.1}$$

as $n_i \to \infty$. Let $n = n_1 + n_2$ be the pooled sample size. Then if $n_1/n \to \theta$, $0 < \theta < 1$, it follows from (3.5.1) assuming H_0 to be true that,

$$\sqrt{n}(\phi(\hat{\mu}_1) - \phi(\hat{\mu}_2)) \xrightarrow{\mathcal{L}} N\left(0, \frac{1}{\theta}\Lambda_1^{-1}\Sigma_1\Lambda_1^{-1} + \frac{1}{1-\theta}\Lambda_2^{-1}\Sigma_2\Lambda_2^{-1}\right). \tag{3.5.2}$$

Estimate μ by the pooled sample intrinsic mean $\hat{\mu}$, coordinates ϕ by $\hat{\phi} \equiv exp_{\hat{\mu}}^{-1}$, Λ_i and Σ_i be their sample analogs $\hat{\Lambda}_i$ and $\hat{\Sigma}_i$ respectively. Denote by μ_{ni} the coordinates $\hat{\phi}(\hat{\mu}_i)$, $i = 1, 2$, of the two sample intrinsic means. Since under H_0, $\hat{\mu}$ is a consistent estimator of μ, it follows from equation (3.5.2) that the statistic

$$T_{n1} = n(\mu_{n1} - \mu_{n2})'\hat{\Sigma}^{-1}(\mu_{n1} - \mu_{n2}) \tag{3.5.3}$$

where

$$\hat{\Sigma} = n\left(\frac{1}{n_1}\hat{\Lambda}_1^{-1}\hat{\Sigma}_1\hat{\Lambda}_1^{-1} + \frac{1}{n_2}\hat{\Lambda}_2^{-1}\hat{\Sigma}_2\hat{\Lambda}_2^{-1}\right)$$

converges in distribution to chi-squared distribution with d degrees of freedom, d being the dimension of M, i.e.,

$$T_{n1} \xrightarrow{\mathcal{L}} \mathcal{X}_d^2.$$

Hence we reject H_0 at asymptotic level α if $T_{n1} > \mathcal{X}_d^2(1-\alpha)$.

Next we test the hypothesis $H_0 : V_1 = V_2 = V$, say, against $H_1 : V_1 \neq V_2$. We assume that the hypothesis of Theorem 2.4.1 hold so that the sample intrinsic variations have asymptotic Normal distribution. Then under H_0, as $n_i \to \infty$,

$$\sqrt{n_i}(\hat{V}_i - V) \xrightarrow{\mathcal{L}} N(0, \sigma_i^2) \tag{3.5.4}$$

where $\sigma_i^2 = \int_M (d_g^2(x,\mu_i) - V)^2 Q_i(dx)$, $i = 1, 2$. Suppose $n_1/n \to \theta$, $0 < \theta < 1$. Then it follows from (3.5.4) assuming H_0 to be true that,

$$\sqrt{n}(\hat{V}_1 - \hat{V}_2) \xrightarrow{\mathcal{L}} N(0, \left(\frac{\sigma_1^2}{\theta} + \frac{\sigma_2^2}{1-\theta}\right))$$

so that

$$T_{n2} = \frac{\hat{V}_1 - \hat{V}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \xrightarrow{\mathcal{L}} N(0, 1)$$

as $n \to \infty$. Here $s_1^2 = \frac{1}{n_1}\sum_{j=1}^{n_1}(d_g^2(X_j, \mu_1) - \hat{V}_1)^2$ and $s_2^2 = \frac{1}{n_2}\sum_{j=1}^{n_2}(d_g^2(Y_j, \mu_2) - \hat{V}_2)^2$ are the sample estimates of σ_1^2 and σ_2^2 respectively. For a test of asymptotic size α, we reject H_0 if $|T_{n2}| > Z(1 - \frac{\alpha}{2})$ where $Z(1 - \frac{\alpha}{2})$ is the upper $\left(1 - \frac{\alpha}{2}\right)$-quantile of standard Normal distribution.

3.5.2 Matched Pair Samples

Next consider the case when (X_1, Y_1), ..., (X_n, Y_n) is an iid sample from some distribution Q on $\bar{M} = M \times M$. Such a sample is called a matched pair sample and arises when, for example, two different treatments are applied to each subject in the sample. An example of a matched pair sample of shapes is considered in Chapter 8.

Let X_j's come from some distribution Q_1 while Y_j's come from some distribution Q_2 on M. Our objective is to distinguish between Q_1 and Q_2 by comparing the sample intrinsic means and variations. Since the X and Y samples are not independent, we cannot apply the methods of Section 3.5.1. Instead we do our analyses on the Riemannian manifold $\bar{M}$. As in Section 3.5.1, we will denote by μ_i and V_i the intrinsic means and variations of Q_i, $i = 1, 2$ respectively and by $\hat{\mu}_i$ and $\hat{V}_i$ the sample intrinsic means and variations.

First we test the hypothesis $H_0 : \mu_1 = \mu_2 = \mu$, say, against $H_1 : \mu_1 \neq \mu_2$. We assume that under H_0, both Q_1 and Q_2 have support in $B(\mu, \frac{r_*}{2})$. Consider the coordinate map Φ on $\bar{M}$ given by

$$\Phi(m_1, m_2) = (\phi(m_1), \phi(m_2)),\ m_1, m_2 \in M$$

where $\phi = exp_\mu^{-1}$. It follows from Corollary 3.3.2 that under H_0,

$$\sqrt{n}\begin{pmatrix} \phi(\hat{\mu}_1) \\ \phi(\hat{\mu}_2) \end{pmatrix} \xrightarrow{\mathcal{L}} N(0, \Gamma) \tag{3.5.5}$$

where $\Gamma = \Lambda^{-1}\Sigma\Lambda^{-1}$ and Σ, Λ are obtained from Theorem 3.3.1 as follows. For $x = (x_1, x_2)'$, $y = (y_1, y_2)'$, $x_1, x_2, y_1, y_2 \in \mathbb{R}^d$, define

$$\begin{aligned} H(x, y) &= d_g^2(\Phi^{-1}(x), \Phi^{-1}(y)) \\ &= d_g^2(\phi^{-1}(x_1), \phi^{-1}(y_1)) + d_g^2(\phi^{-1}(x_2), \phi^{-1}(y_2)) \\ &= h(x_1, y_1) + h(x_2, y_2). \end{aligned}$$

Then

$$\Lambda = \mathrm{E}[(\mathrm{D}_r\mathrm{D}_s H(\Phi(X_1, Y_1), 0))] = \begin{pmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{pmatrix}$$

and

$$\Sigma = \mathrm{Cov}[(\mathrm{D}_r H(\Phi(X_1, Y_1), 0))] = \begin{pmatrix} \Sigma_1 & \Sigma_{12} \\ \Sigma_{21} & \Sigma_2 \end{pmatrix}.$$

Note that $\Lambda_1, \Lambda_2, \Sigma_1, \Sigma_2$ are as in Section 3.5.1 and

$$\Sigma_{12} = \Sigma_{21}' = \mathrm{Cov}[(\mathrm{D}_r h(\phi(X_1), 0)), \mathrm{D}_r h(\phi(Y_1), 0))].$$

Therefore

$$\begin{aligned} \Gamma = \Lambda^{-1}\Sigma\Lambda^{-1} &= \begin{pmatrix} \Lambda_1^{-1} & 0 \\ 0 & \Lambda_2^{-1} \end{pmatrix} \begin{pmatrix} \Sigma_1 & \Sigma_{12} \\ \Sigma_{21} & \Sigma_2 \end{pmatrix} \begin{pmatrix} \Lambda_1^{-1} & 0 \\ 0 & \Lambda_2^{-1} \end{pmatrix} \\ &= \begin{pmatrix} \Lambda_1^{-1}\Sigma_1\Lambda_1^{-1} & \Lambda_1^{-1}\Sigma_{12}\Lambda_2^{-1} \\ \Lambda_2^{-1}\Sigma_{21}\Lambda_1^{-1} & \Lambda_2^{-1}\Sigma_2\Lambda_2^{-1} \end{pmatrix}. \end{aligned}$$

It follows from equation (3.5.5) that if H_0 is true, then,

$$\sqrt{n}(\phi(\hat{\mu}_1) - \phi(\hat{\mu}_1)) \xrightarrow{\mathcal{L}} N(0, \tilde{\Sigma})$$

where

$$\tilde{\Sigma} = \Lambda_1^{-1}\Sigma_1\Lambda_1^{-1} + \Lambda_2^{-1}\Sigma_2\Lambda_2^{-1} - (\Lambda_1^{-1}\Sigma_{12}\Lambda_2^{-1} + \Lambda_2^{-1}\Sigma_{21}\Lambda_1^{-1}).$$

Estimate $\phi(\hat{\mu}_i)$ by μ_{ni}, $i = 1, 2$, as in Section 3.5.1 and $\tilde{\Sigma}$ by its sample analog $\hat{\tilde{\Sigma}}$. Then, under H_0, the test statistic

$$T_{n3} = n(\mu_{n1} - \mu_{n2})'\hat{\tilde{\Sigma}}^{-1}(\mu_{n1} - \mu_{n2})$$

converges in distribution to chi-squared distribution with d degrees of freedom, i.e. $T_{n3} \xrightarrow{\mathcal{L}} \mathcal{X}_d^2$. Therefore one rejects H_0 at asymptotic level α if $T_{n3} > \mathcal{X}_d^2(1-\alpha)$.

Next we test the null hypothesis $H_0 : V_1 = V_2$ against the alternative $H_1 : V_1 \neq V_2$. From equation (2.4.4), it follows that

$$\begin{aligned} \begin{pmatrix} \sqrt{n}(\hat{V}_1 - V_1) \\ \sqrt{n}(\hat{V}_2 - V_2) \end{pmatrix} &= \frac{1}{\sqrt{n}} \begin{pmatrix} \sum_{j=1}^n [d_g^2(X_j, \mu_1) - V_1] \\ \sum_{j=1}^n [d_g^2(Y_j, \mu_2) - V_2] \end{pmatrix} + o_P(1) \\ &\xrightarrow{\mathcal{L}} N\left(0, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}\right) \end{aligned}$$

where $\sigma_1^2 = \mathrm{Var}(d_g^2(X_1, \mu_1))$, $\sigma_2^2 = \mathrm{Var}(d_g^2(Y_1, \mu_2))$ and $\sigma_{12} = \mathrm{Cov}(d_g^2(X_1, \mu_1), d_g^2(Y_1, \mu_2))$. Hence if H_0 is true, then

$$\sqrt{n}(\hat{V}_1 - \hat{V}_2) \xrightarrow{\mathcal{L}} N(0, \sigma_1^2 + \sigma_2^2 - 2\sigma_{12})$$

which implies that the statistic

$$T_{n4} = \frac{\sqrt{n}(\hat{V}_1 - \hat{V}_2)}{\sqrt{s_1^2 + s_2^2 - 2s_{12}}}$$

has asymptotic standard Normal distribution. Here s_1^2, s_2^2 and s_{12} are sample estimates of σ_1^2, σ_2^2 and σ_{12} respectively. Therefore we reject H_0 at asymptotic level α if $|T_{n4}| > Z(1 - \frac{\alpha}{2})$.

CHAPTER 4

EXTRINSIC ANALYSIS ON MANIFOLDS

4.1 Introduction

In Chapter 2, we studied the properties of the Fréchet mean and variation of a probability Q on a Riemannian manifold M, using the geodesic distance in the expression for the Fréchet function. To carry out nonparametric inference using the sample mean and variation, it is necessary to assume that there is a unique population mean. However as we saw (Theorem 3.3.1, Corollary 3.3.2), broad conditions are not known for the existence of the intrinsic mean and for asymptotic normality of the sample intrinsic mean. Also analytic expressions for the intrinsic mean are not available. There are algorithms to compute the intrinsic mean numerically (see Le (2001)), however sufficient conditions for these algorithms to converge put further restrictions on the support of Q and they are computationally slow. In this chapter, we will get introduced to another distance on M obtained from embedding it into some Euclidean space and compute the Fréchet mean and variation. The corresponding statistical analysis is called **extrinsic analysis** on M. As we shall see in the following sections that it is often simpler both mathematically and computationally to carry out an extrinsic analysis on M. The notion of extrinsic mean on a manifold was introduced independently by Hendricks and Landsman (1998) and Patrangenaru (1998), and later generalized in Bhattacharya and Patrangenaru (2003, 2005).

4.2 Extrinsic Mean and Variation

In this chapter, we assume that M is a differentiable manifold of dimension d. Consider an embedding of M into some Euclidean space $E^N \approx \mathbb{R}^N$ via some map $J : M \to E^N$ such that both J and its derivative are injective. Then J induces the distance

$$\rho(x,y) = \|J(x) - J(y)\| \tag{4.2.1}$$

on M, where $\|.\|$ denotes the Euclidean norm

$$(\|u\|^2 = \sum_{i=1}^{N} {u_i}^2 \ \forall\ u = (u_1, u_2, .., u_N)').$$

The distance ρ is called the **extrinsic distance** on M. Given a probability distribution Q on M, we consider the Fréchet function

$$F(x) = \int_M \rho^2(x,y) Q(dy) \tag{4.2.2}$$

with ρ as in equation (4.2.1). This choice of Fréchet function makes the Fréchet mean and variation computable in a number of important examples using Proposition 4.2.1.

Definition 4.2.1. Let (M,ρ), J be as above. Let Q be a probability distribution with finite Fréchet function F. The Fréchet mean set of Q is called the **extrinsic mean set** of Q and the Fréchet variation of Q is called the **extrinsic variation** of Q. If X_i, $i = 1,\ldots,n$ are iid observations from Q and $Q_n = \frac{1}{n}\sum_{i=1}^{n}\delta_{X_i}$ is the corresponding empirical distribution, then the Fréchet mean set of Q_n is called the **sample extrinsic mean set** and the Fréchet variation of Q is called the **sample extrinsic variation**.

Among the possible embeddings, we seek out **equivariant embeddings** which preserve many of the geometric features of M.

Definition 4.2.2. For a Lie group H acting on a manifold M, an embedding $J : M \to \mathbb{R}^N$ is H-**equivariant** if there exists a group homomorphism $\phi : H \to GL(N,\mathbb{R})$ such that

$$J(hp) = \phi(h)J(p)\ \forall p \in M,\ \forall h \in H.$$

Here $GL(N,\mathbb{R})$ is the **general linear group** of all $N \times N$ non-singular matrices.

We say that Q has an **extrinsic mean** μ_E if the extrinsic mean set of Q is a singleton. Proposition 4.2.1 as proved in Bhattacharya and Patrangenaru (2003), gives a necessary and sufficient condition for Q to have an extrinsic mean. It also provides an analytic expression for the extrinsic mean set and extrinsic variation of Q. In the statement of the proposition, we assume that $J(M) = \tilde{M}$ is a closed subset of E^N. Then for every $u \in E^N$, there exists a compact set of points in $\tilde{M}$ whose distance from u is the smallest among all points in $\tilde{M}$. We call this set the **projection set** of u and denote it by

$$P_{\tilde{M}}u = \{x \in \tilde{M} : \|x - u\| \leq \|y - u\| \ \forall y \in \tilde{M}\}. \tag{4.2.3}$$

If this set is a singleton, u is said to be a **nonfocal point** of E^N (w.r.t. $\tilde{M}$), otherwise it is said to be a **focal point** of E^N.

Proposition 4.2.1. *Let $\tilde{Q} = Q \circ J^{-1}$ be the image of Q in E^N. (a) If $\tilde{\mu} = \int_{\mathbb{E}^N} u\tilde{Q}(du)$ is the mean of $\tilde{Q}$, then the extrinsic mean set of Q is given by $J^{-1}(P_{\tilde{M}}\tilde{\mu})$. (b) The extrinsic variation of Q equals*

$$V = \int_{E^N} \|x - \tilde{\mu}\|^2 \tilde{Q}(dx) + \|\tilde{\mu} - \mu\|^2$$

where $\mu \in P_{\tilde{M}}\tilde{\mu}$. (c) If $\tilde{\mu}$ is a nonfocal point of $\mathbb{E}^N$, then the extrinsic mean of Q exists.

We define the **sample extrinsic mean** μ_{nE} to be any measurable selection from the sample extrinsic mean set. In case $\tilde{\mu}$ is a nonfocal point, it follows from Proposition 2.2.2 that μ_{nE} is a strongly consistent estimator of μ_E.

4.3 Asymptotic Distribution of the Sample Extrinsic Mean

From now on we assume that the extrinsic mean μ_E of Q is uniquely defined. It follows from Theorem 2.3.1 in Chapter 2 that under suitable assumptions, the coordinates of the sample extrinsic mean μ_{nE} have asymptotic Gaussian distribution.

However apart from other assumptions, the theorem requires Q to have support in a single coordinate patch and the expression of the asymptotic dispersion depends on what coordinates we choose. In this section, we derive the asymptotic Normality of μ_{nE} in a different way, via Proposition 4.3.1, which is proved in Bhattacharya and Patrangenaru (2005). This proposition makes lesser assumptions on Q and the expression for the asymptotic dispersion is easier to get, as we shall see in subsequent chapters.

When the mean $\tilde{\mu}$ of $\tilde{Q}$ is a nonfocal point of E^N, the projection set in (4.2.3) is a singleton and we can define a projection map

$$P : E^N \to \tilde{M},\ \|\tilde{\mu} - P(\tilde{\mu})\| = \min_{p \in \tilde{M}} \|\tilde{\mu} - p\| \tag{4.3.1}$$

in a neighborhood of $\tilde{\mu}$. Also in a neighborhood of a nonfocal point such as $\tilde{\mu}$, P is smooth. Let $\overline{\tilde{X}} = \frac{1}{n}\sum_{j=1}^{n} \tilde{X}_j$ be the sample mean of $\tilde{X}_j = J(X_j)$, $j = 1, 2, \ldots, n$. Since $\overline{\tilde{X}}$ converges to $\tilde{\mu}$ almost surely, for sample size large enough $\overline{\tilde{X}}$ is nonfocal, and it can be shown by a second order Taylor expansion for $P : E^N \to E^N$, that

$$\sqrt{n}[P(\overline{\tilde{X}}) - P(\tilde{\mu})] = \sqrt{n}(d_{\tilde{\mu}}P)(\overline{\tilde{X}} - \tilde{\mu}) + o_P(1) \tag{4.3.2}$$

where $d_{\tilde{\mu}}P$ is the differential (map) of the projection P, which takes vectors in the tangent space of E^N at $\tilde{\mu}$ to tangent vectors of $\tilde{M}$ at $P(\tilde{\mu})$. Since $\sqrt{n}(\overline{\tilde{X}} - \tilde{\mu})$ has an asymptotic Gaussian distribution and $d_{\tilde{\mu}}P$ is a linear map, from (4.3.2) it follows that $\sqrt{n}[P(\overline{\tilde{X}}) - P(\tilde{\mu})]$ has an asymptotic mean zero Gaussian distribution on the tangent space of $J(M)$ at $P(\tilde{\mu})$. This is stated in Proposition 4.3.1 below.

Proposition 4.3.1. *Suppose $\tilde{\mu}$ is a nonfocal point of E^N and P is continuously differentiable in a neighborhood of $\tilde{\mu}$. Denote by T_j, the vector of coordinates of $(d_{\tilde{\mu}}P)(\tilde{X}_j - \tilde{\mu})$, $j = 1, 2, \ldots, n$, with respect to some orthonormal basis for $T_{P(\tilde{\mu})}\tilde{M}$. Then if $Q \circ J^{-1}$ has finite second moments,*

$$\sqrt{n}\bar{T} \xrightarrow{\mathcal{L}} N(0, \Sigma)$$

where Σ denotes the covariance matrix of T_1.

Let $L_{\tilde{\mu}} : E^N \to T_{P(\tilde{\mu})}\tilde{M}$ denote the linear projection on to $T_{P(\tilde{\mu})}\tilde{M}$. Then from equation (4.3.2) and Proposition 4.3.1, it follows that

$$nL_{\tilde{\mu}}[P(\overline{\tilde{X}}) - P(\tilde{\mu})]'\Sigma^{-1}L_{\tilde{\mu}}[P(\overline{\tilde{X}}) - P(\tilde{\mu})] \xrightarrow{\mathcal{L}} \mathcal{X}_d^2.$$

Using this we can construct the following confidence region for μ_E:

$$\{\mu_E = J^{-1}[P(\tilde{\mu})] : nL[P(\tilde{\mu}) - P(\overline{\tilde{X}})]'\hat{\Sigma}^{-1}L[P(\tilde{\mu}) - P(\overline{\tilde{X}})] \leq \mathcal{X}_d^2(1-\alpha)\} \quad (4.3.3)$$

with asymptotic confidence level $(1-\alpha)$. Here $L : E^N \to T_{P(\overline{\tilde{X}})}\tilde{M}$ denotes the linear projection on to $T_{P(\overline{\tilde{X}})}\tilde{M}$, $\hat{\Sigma}$ is the sample estimate of Σ and $\mathcal{X}_d^2(1-\alpha)$ is the upper $(1-\alpha)$-quantile of the chi-squared distribution with d degrees of freedom ($\mathcal{X}_d^2$). The corresponding pivotal bootstrap confidence region for μ_E is given by

$$\{\mu_E : nL[P(\tilde{\mu}) - P(\overline{\tilde{X}})]'\hat{\Sigma}^{-1}L[P(\tilde{\mu}) - P(\overline{\tilde{X}})] \leq c^*(1-\alpha)\}. \quad (4.3.4)$$

Here $c^*(1-\alpha)$ denotes the upper $(1-\alpha)$-quantile of the bootstrapped values of $nL[P(\tilde{\mu}) - P(\overline{\tilde{X}})]'\hat{\Sigma}^{-1}L[P(\tilde{\mu}) - P(\overline{\tilde{X}})]$.

4.4 Asymptotic Distribution of the Sample Extrinsic Variation

Let V and V_n denote the extrinsic variations of Q and Q_n respectively. We can deduce the asymptotic distribution of V_n from Theorem 2.4.1 in Chapter 2. However for the hypothesis of that theorem to hold, we need to make a number of assumptions including that Q has support in a single coordinate patch. Theorem 4.4.1 proves the asymptotic normality of V_n under less restrictive assumptions. It was first proved in Bhattacharya (2008a). In the statement of the theorem, ρ denotes the extrinsic distance as defined in equation (4.2.1).

Theorem 4.4.1. *If Q has extrinsic mean μ_E and if* $\mathrm{E}\rho^4(X_1, \mu_E) < \infty$, *then*

$$\sqrt{n}(V_n - V) \xrightarrow{\mathcal{L}} N(0, \mathrm{Var}(\rho^2(X_1, \mu_E))).$$

Proof. From definition of V_n and V, it follows that

$$\begin{aligned} V_n - V &= \frac{1}{n}\sum_{j=1}^{n} \rho^2(X_j, \mu_{nE}) - \int_M \rho^2(x, \mu_E) Q(dx) \\ &= \frac{1}{n}\sum_{j=1}^{n} \rho^2(X_j, \mu_{nE}) - \frac{1}{n}\sum_{j=1}^{n} \rho^2(X_j, \mu_E) \\ &+ \frac{1}{n}\sum_{j=1}^{n} \rho^2(X_j, \mu_E) - \mathrm{E}[\rho^2(X_1, \mu_E)] \end{aligned} \tag{4.4.1}$$

where μ_{nE} is the sample extrinsic mean, i.e. some measurable selection from the sample extrinsic mean set. Now,

$$\begin{aligned} &\frac{1}{n}\sum_{j=1}^{n} \rho^2(X_j, \mu_{nE}) = \frac{1}{n}\sum_{j=1}^{n} \|Y_j - P(\bar{Y})\|^2 \\ =&\frac{1}{n}\sum_{j=1}^{n} \|Y_j - P(\tilde{\mu})\|^2 + \|P(\tilde{\mu}) - P(\bar{Y})\|^2 - 2\langle \bar{Y} - P(\tilde{\mu}), P(\bar{Y}) - P(\tilde{\mu})\rangle \end{aligned} \tag{4.4.2}$$

Substitute (4.4.2) in (4.4.1) to get

$$\begin{aligned} V_n - V = \|P(\bar{Y}) - P(\tilde{\mu})\|^2 - 2\langle \bar{Y} - P(\tilde{\mu}), P(\bar{Y}) - P(\tilde{\mu})\rangle \\ + \frac{1}{n}\sum_{j=1}^{n} \rho^2(X_j, \mu_E) - \mathrm{E}[\rho^2(X_1, \mu_E)] \end{aligned}$$

which implies that

$$\sqrt{n}(V_n - V) = T_1 + T_2 \tag{4.4.3}$$

where

$$T_1 = \sqrt{n}\|P(\bar{Y}) - P(\tilde{\mu})\|^2 - 2\sqrt{n}\langle \bar{Y} - P(\tilde{\mu}), P(\bar{Y}) - P(\tilde{\mu})\rangle, \tag{4.4.4}$$

$$T_2 = \sqrt{n}\left(\frac{1}{n}\sum_{j=1}^{n} \rho^2(X_j, \mu_E) - \mathrm{E}\rho^2(X_1, \mu_E)\right) \tag{4.4.5}$$

From the classical CLT, if $\mathrm{E}\rho^4(X_1, \mu_E) < \infty$, then

$$T_2 \xrightarrow{\mathcal{L}} N(0, \mathrm{Var}(\rho^2(X_1, \mu_E))). \tag{4.4.6}$$

Compare the expression of T_1 with equation (4.3.2) to get

$$T_1 = -2\langle d_{\tilde{\mu}} P(\bar{Y} - \tilde{\mu}), \tilde{\mu} - P(\tilde{\mu})\rangle + o_P(1). \tag{4.4.7}$$

From the definition of P, $P(\tilde{\mu}) = \mathrm{argmin}_{p \in \tilde{M}} \|\tilde{\mu} - p\|^2$. Hence the Euclidean derivative of $\|\tilde{\mu} - p\|^2$ at $p = P(\tilde{\mu})$ must be orthogonal to $T_{P(\tilde{\mu})}\tilde{M}$, or

$$\tilde{\mu} - P(\tilde{\mu}) \in (T_{P(\tilde{\mu})}\tilde{M})^{\perp}.$$

Since $d_{\tilde{\mu}} P(\bar{Y} - \tilde{\mu}) \in T_{P(\tilde{\mu})}\tilde{M}$, the first term in the expression of T_1 in (4.4.7) is 0, and hence $T_1 = o_P(1)$. From equations (4.4.3) and (4.4.6), we conclude that

$$\begin{aligned} \sqrt{n}(V_n - V) &= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} [\rho^2(X_j, \mu_E) - \mathrm{E}\rho^2(X_1, \mu_E)] + o_P(1) \qquad (4.4.8) \\ &\xrightarrow{\mathcal{L}} N(0, \mathrm{Var}(\rho^2(X_1, \mu_E))). \end{aligned}$$

This completes the proof. □

Remark 4.4.1. Although Proposition 2.2.3 does not require the uniqueness of the extrinsic mean of Q for V_n to be a consistent estimator of V, Theorem 4.4.1 breaks down in the case of non-uniqueness. (see Section 4.5).

Using Theorem 4.4.1, one can construct the following confidence interval I for V:

$$I = \{V \in [V_n - \frac{s}{\sqrt{n}} Z(1 - \frac{\alpha}{2}), V_n + \frac{s}{\sqrt{n}} Z(1 - \frac{\alpha}{2})]\}. \tag{4.4.9}$$

The interval I has asymptotic confidence level of $(1 - \alpha)$. Here s^2 is the sample variance of $\rho^2(X_j, \mu_{nE})$, $j = 1, \ldots, n$ and $Z(1 - \frac{\alpha}{2})$ denotes the upper $(1 - \frac{\alpha}{2})$-quantile of $N(0, 1)$ distribution. From equation (4.4.9), one can also construct a pivotal bootstrap confidence interval for V, the details of which are left to the reader.

4.5 Extrinsic Analysis on the unit sphere S^d

Consider the inclusion map

$$i : S^d \to \mathbb{R}^{d+1},\ i(x) = x.$$

as the embedding of S^d in $\mathbb{R}^{d+1}$. The extrinsic mean set of a probability measure Q on S^d is then the projection set of $\tilde{\mu} = \int_{\mathbb{R}^d} x\tilde{Q}(dx)$ on S^d, where $\tilde{Q}$ is Q regarded as a probability measure on $\mathbb{R}^{d+1}$. Note that $\tilde{\mu}$ is nonfocal iff $\tilde{\mu} \neq 0$. Then

$$P(\tilde{\mu}) = \frac{\tilde{\mu}}{\|\tilde{\mu}\|}$$

and if $\tilde{\mu} = 0$, then $P_{S^d}(0) = S^d$. The extrinsic variation of Q is

$$\begin{aligned} V &= \int_{\mathbb{R}^{d+1}} \|x - \tilde{\mu}\|^2 \tilde{Q}(dx) + (\|\tilde{\mu}\| - 1)^2 \\ &= 2(1 - \|\tilde{\mu}\|). \end{aligned}$$

If V_n denotes the sample extrinsic variation, it is easy to check that $\sqrt{n}(V_n - V)$ is asymptotically Normal iff $\tilde{\mu} \neq 0$.

4.6 Two Sample Extrinsic Tests

In this section, we will use the asymptotic distribution of the sample extrinsic mean and variation to construct nonparametric tests to compare two probability distributions Q_1 and Q_2 on M.

4.6.1 Independent Samples

Let $X_1, \ldots, X_{n_1}$ and $Y_1, \ldots, Y_{n_2}$ be two iid samples from Q_1 and Q_2 respectively that are mutually independent. Let μ_{iE} and V_i denote the extrinsic means and variations of Q_i, $i = 1, 2$ respectively. Similarly denote by $\hat{\mu}_{iE}$ and $\hat{V}_i$ the sample extrinsic means and variations. We want to test the hypothesis, $H_0 : Q_1 = Q_2$.

We start by comparing the sample extrinsic means. Let $\tilde{X}_j = J(X_j)$, $\tilde{Y}_j = J(Y_j)$ be the embeddings of the sample points into E^N. Let μ_i be the mean of $\tilde{Q}_i = Q_i \circ J^{-1}$. Then under H_0, $\mu_1 = \mu_2 = \mu$ (say). Let $\hat{\mu}_i$, $i = 1, 2$ be the sample means of $\{\tilde{X}_j\}$ and $\{\tilde{Y}_j\}$ respectively. Then from equation (4.3.2), it follows that

$$\sqrt{n_i}[P(\hat{\mu}_i) - P(\mu)] = \sqrt{n_i}(d_\mu P)(\hat{\mu}_i - \mu) + o_P(1) \; i = 1, 2. \tag{4.6.1}$$

Hence, if $n_i \to \infty$ such that $\frac{n_i}{n_1+n_2} \to p_i$, $0 < p_i < 1$, $p_1 + p_2 = 1$, then

$$\begin{aligned}\sqrt{n}(P(\hat{\mu}_1) - P(\hat{\mu}_2)) &= \sqrt{n} d_\mu P(\hat{\mu}_1 - \mu) - \sqrt{n} d_\mu P(\hat{\mu}_2 - \mu) + o_P(1) \\ &\xrightarrow{\mathcal{L}} N(0, \frac{\Sigma^1}{p_1} + \frac{\Sigma^2}{p_2}).\end{aligned} \tag{4.6.2}$$

Here $n = n_1 + n_2$ is the pooled sample size, Σ^i, $i = 1, 2$ are the covariance matrices of the coordinates of $d_\mu P(\tilde{X}_1 - \mu)$ and $d_\mu P(\tilde{Y}_1 - \mu)$ with respect to some chosen basis for $T_\mu \tilde{M}$. We estimate μ by the pooled sample mean $\hat{\mu} = \frac{1}{n}(n_1\hat{\mu}_1 + n_2\hat{\mu}_2)$. Denote the coordinates of $\{d_{\hat{\mu}} P(\tilde{X}_j - \hat{\mu})\}_{j=1}^{n_1}$ and $\{d_{\hat{\mu}} P(\tilde{Y}_j - \hat{\mu})\}_{j=1}^{n_2}$ by $\{S_j^1\}_{j=1}^{n_1}$ and $\{S_j^2\}_{j=1}^{n_2}$ respectively. Let $\hat{\Sigma}^i$ $i = 1, 2$ denote the sample covariance matrices of $\{S_j^i\}_{j=1}^{n_i}$ $i = 1, 2$ respectively. Then if H_0 is true, the statistic

$$T_1 = (\bar{S^1} - \bar{S^2})' \left(\frac{1}{n_1}\hat{\Sigma}^1 + \frac{1}{n_2}\hat{\Sigma}^2 \right)^{-1} (\bar{S^1} - \bar{S^2}) \tag{4.6.3}$$

converges in distribution to $\mathcal{X}_d^2$ distribution, where d is the dimension of M. Hence we reject H_0 at asymptotic level α if $T_1 > \mathcal{X}_d^2(1 - \alpha)$.

Next we test the null hypothesis $H_0 : \mu_{1E} = \mu_{2E}$ against the alternative $H_a : \mu_{1E} \neq \mu_{2E}$. From equation (4.3.2), it follows that

$$\begin{aligned}&\sqrt{n}[\{P(\hat{\mu}_1) - P(\mu_1)\} - \{P(\hat{\mu}_2) - P(\mu_2)\}] \\ =&\sqrt{n} d_{\mu_1} P(\hat{\mu}_1 - \mu_1) - \sqrt{n} d_{\mu_2} P(\hat{\mu}_2 - \mu_2) + o_P(1).\end{aligned} \tag{4.6.4}$$

Since the samples are independent, equation (4.6.4) implies that, for $\tilde{\mu} \in M$,

$$\begin{aligned}
&L_{\tilde{\mu}}\left[\sqrt{n}[\{P(\hat{\mu}_1) - P(\mu_1)\} - \{P(\hat{\mu}_2) - P(\mu_2)\}]\right] \\
=&L_{1\tilde{\mu}}[\sqrt{n} d_{\mu_1} P(\hat{\mu}_1 - \mu_1)] - L_{2\tilde{\mu}}[\sqrt{n} d_{\mu_2} P(\hat{\mu}_2 - \mu_2)] + o_P(1) \\
\xrightarrow{\mathcal{L}}&N(0, \frac{1}{p_1} L_{1\tilde{\mu}} \Sigma_1 L'_{1\tilde{\mu}} + \frac{1}{p_2} L_{2\tilde{\mu}} \Sigma_2 L'_{2\tilde{\mu}}).
\end{aligned} \tag{4.6.5}$$

In equation (4.6.5), $L_{\tilde{\mu}}$ denotes the linear projection from E^N on to $T_{\tilde{\mu}}\tilde{M}$, while $L_{i\tilde{\mu}}$, $i = 1, 2$ denote the linear projections from $T_{\mu_i}\tilde{M}$ on to $T_{\tilde{\mu}}\tilde{M}$, and Σ_i, $i = 1, 2$, are the covariance matrices of the coordinates of $d_{\mu_1}P(\tilde{X}_1 - \mu_1)$ and $d_{\mu_2}P(\tilde{Y}_1 - \mu_2)$ with respect to some chosen bases for $T_{\mu_1}\tilde{M}$ and $T_{\mu_2}\tilde{M}$ respectively. Hence if H_0 is true, then $P(\mu_1) = P(\mu_2)$, and hence from equation (4.6.5), it follows that

$$L_{\tilde{\mu}}[\sqrt{n}\{P(\hat{\mu}_1) - P(\hat{\mu}_2)\}] \xrightarrow{\mathcal{L}} N(0, \frac{1}{p_1} L_{1\tilde{\mu}} \Sigma_1 L'_{1\tilde{\mu}} + \frac{1}{p_2} L_{2\tilde{\mu}} \Sigma_2 L'_{2\tilde{\mu}}). \tag{4.6.6}$$

Note that $\tilde{\mu}$ can be any point on $\tilde{M}$ for (4.6.6) to hold. Using this one can construct the test statistic

$$T_2 = L[P(\hat{\mu}_1) - P(\hat{\mu}_2)]' \left(\frac{1}{n_1} L_1 \hat{\Sigma}_1 L'_1 + \frac{1}{n_2} L_2 \hat{\Sigma}_2 L'_2\right)^{-1} L[P(\hat{\mu}_1) - P(\hat{\mu}_2)] \tag{4.6.7}$$

to test if H_0 is true. In the statistic T_2, L is the linear projection from E^N on to $T_p\tilde{M}$, where $p \in \tilde{M}$. L_i, $i = 1, 2$ are the matrices of linear projection on to $T_p\tilde{M}$ from $T_{P(\hat{\mu}_i)}\tilde{M}$ respectively. Again, p can be any point on $\tilde{M}$, but tangent space analysis is expected to provide better approximation to the asymptotic limit if we choose $p = P(\hat{\mu})$, $\hat{\mu}$ being the pooled sample mean. Here $\hat{\Sigma}_i$, $i = 1, 2$ denote the sample covariance matrices of the coordinates of $\{d_{\hat{\mu}_1}P(\tilde{X}_j - \hat{\mu}_1)\}_{j=1}^{n_1}$ and $\{d_{\hat{\mu}_2}P(\tilde{Y}_j - \hat{\mu}_2)\}_{j=1}^{n_2}$ respectively with respect to some chosen basis for $T_{\hat{\mu}_1}\tilde{M}$ and $T_{\hat{\mu}_2}\tilde{M}$. Under H_0, $T_2 \xrightarrow{\mathcal{L}} \mathcal{X}_d^2$. Hence we reject H_0 at asymptotic level α if $T_2 > \mathcal{X}_d^2(1 - \alpha)$. In our examples from Chapter 7, the two statistics T_1 and T_2 yield values that are quite close to each other.

When the sample sizes are not too large, it is more efficient to construct a bootstrap confidence region for $P(\mu_1) - P(\mu_2)$ using the test statistic T_2 in (4.6.7), and use that to test if H_0 is true. Let $\{X_j^*, j = 1, \ldots, n_1\}$ and $\{Y_j^*, j = 1, \ldots, n_2\}$ denote the bootstrap resamples from the original samples in $\tilde{M}$. Denote by μ_i^*, $i = 1, 2$ the bootstrap sample means, and by μ^* the pooled sample mean. Let L_i^*, $i = 1, 2$, be the matrices of projections from $T_{P(\mu_i^*)}\tilde{M}$ on to $T_{P(\mu^*)}\tilde{M}$, L^* be the linear projection from E^N on to $T_{P(\mu^*)}\tilde{M}$, Σ_i^*, $i = 1, 2$, be the bootstrap sample covariance matrices of the coordinates of $\{d_{\mu_1^*}P(X_j^* - \mu_1^*)\}$ and $\{d_{\mu_2^*}P(Y_j^* - \mu_2^*)\}$ respectively. Then the bootstrap version of T_2 in (4.6.7) is

$$T_2^* = v^{*\prime}\Sigma^{*-1}v^* \tag{4.6.8}$$

where

$$v^* = L^*[\{P(\mu_1^*) - P(\hat{\mu}_1)\} - \{P(\mu_2^*) - P(\hat{\mu}_2)\}] \tag{4.6.9}$$

and

$$\Sigma^* = \frac{1}{n_1}L_1^*\hat{\Sigma}_1^*L_1^{*\prime} + \frac{1}{n_2}L_2^*\hat{\Sigma}_2^*L_2^{*\prime}. \tag{4.6.10}$$

We reject H_0 at level α if $T_2 > c^*(1-\alpha)$, where $c^*(1-\alpha)$ is the upper $(1-\alpha)$ quantile of the bootstrap distribution of T_2^*.

Next we test if Q_1 and Q_2 have the same extrinsic variations, i.e. $H_0 : V_1 = V_2$. From Theorem 4.4.1 and using the fact that the samples are independent, we get, under H_0,

$$\sqrt{n}(\hat{V}_1 - \hat{V}_2) \xrightarrow{\mathcal{L}} N\left(0, \frac{\sigma_1^2}{p_1} + \frac{\sigma_2^2}{p_2}\right) \tag{4.6.11}$$

$$\Rightarrow \frac{\hat{V}_1 - \hat{V}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \xrightarrow{\mathcal{L}} N(0, 1)$$

where $\sigma_1^2 = \mathrm{Var}\,(\rho^2(X_1, \mu_{1E}))$, $\sigma_2^2 = \mathrm{Var}\rho^2(Y_1, \mu_{2E})$ and s_1^2, s_2^2 are their sample estimates. Hence to test if H_0 is true, we can use the test statistic

$$T_3 = \frac{\hat{V}_1 - \hat{V}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}. \tag{4.6.12}$$

For a test of asymptotic size α, we reject H_0 if $|T_3| > Z\left(1 - \frac{\alpha}{2}\right)$, where $Z\left(1 - \frac{\alpha}{2}\right)$ is the upper $\left(1 - \frac{\alpha}{2}\right)$-quantile of $N(0,1)$ distribution. We can also construct a bootstrap confidence interval for $V_1 - V_2$ and use that to test if $V_1 - V_2 = 0$. The details of that are left to the reader.

4.6.2 Matched Pair Samples

Next consider the case when (X_1, Y_1), ..., (X_n, Y_n) is an iid sample from some distribution Q on $\bar{M} = M \times M$. Such samples arise, when for example we have two different observations from each subject (see Section 8.6).

Let X_j's have distribution Q_1 while Y_j's come from some distribution Q_2 on M. Our objective is to distinguish Q_1 and Q_2 by comparing the sample extrinsic means and variations. Since the X and Y samples are not independent, we cannot apply the methods of the earlier section. Instead we do our analyses on $\bar{M}$. Note that $\bar{M}$ is a differentiable manifold which can be embedded into $E^N \times E^N$ via the map

$$\bar{J} : \bar{M} \to E^N \times E^N, \;\; \bar{J}(x, y) = (J(x), J(y)).$$

Let $\tilde{Q} = Q \circ \bar{J}^{-1}$. Then if $\tilde{Q}_i$ has mean μ_i, $i = 1, 2$, then $\tilde{Q}$ has mean $\bar{\mu} = (\mu_1, \mu_2)$. The projection of $\bar{\mu}$ on $\tilde{\bar{M}} \equiv \bar{J}(\bar{M})$ is given by $\bar{P}(\mu) = (P(\mu_1), P(\mu_2))$. Hence if Q_i has extrinsic mean μ_{iE}, $i = 1, 2$, then Q has extrinsic mean $\bar{\mu}_E = (\mu_{1E}, \mu_{2E})$. Denote the paired sample as $Z_j \equiv (X_j, Y_j)$, $j = 1, \ldots, n$ and let $\hat{\bar{\mu}} = (\hat{\mu}_1, \hat{\mu}_2)$, $\hat{\bar{\mu}}_E = (\hat{\mu}_{1E}, \hat{\mu}_{2E})$ be the sample estimates of $\bar{\mu}$ and $\bar{\mu}_E$ respectively. From equation (4.3.2), it follows that

$$\sqrt{n}(\bar{P}(\hat{\bar{\mu}}) - \bar{P}(\bar{\mu})) = \sqrt{n} d_{\bar{\mu}} \bar{P}(\hat{\bar{\mu}} - \bar{\mu}) + o_P(1)$$

which can be written as

$$\sqrt{n}\begin{pmatrix} P(\hat{\mu}_1) - P(\mu_1) \\ P(\hat{\mu}_2) - P(\mu_2) \end{pmatrix} = \sqrt{n}\begin{pmatrix} d_{\mu_1}P(\hat{\mu}_1 - \mu_1) \\ d_{\mu_2}P(\hat{\mu}_2 - \mu_2) \end{pmatrix} + o_P(1) \tag{4.6.13}$$

Hence if H_0: $\mu_1 = \mu_2 = \mu$, then under H_0,

$$\sqrt{n}\begin{pmatrix} P(\hat{\mu}_1) - P(\mu) \\ P(\hat{\mu}_2) - P(\mu) \end{pmatrix} = \sqrt{n}\begin{pmatrix} d_{\mu}P(\hat{\mu}_1 - \mu) \\ d_{\mu}P(\hat{\mu}_2 - \mu) \end{pmatrix} + o_P(1)$$
$$\xrightarrow{\mathcal{L}} N(0, \Sigma = \begin{pmatrix} \Sigma^1 & \Sigma^{12} \\ \Sigma^{21} & \Sigma^2 \end{pmatrix}). \tag{4.6.14}$$

In (4.6.14), Σ^i, $i = 1, 2$ are the same as in (4.6.2) and $\Sigma^{12} = (\Sigma^{21})'$ is the covariance between the coordinates of $d_\mu P(\tilde{X}_1 - \mu)$ and $d_\mu P(\tilde{Y}_1 - \mu)$. From (4.6.14), it follows that

$$\sqrt{n} d_\mu P(\hat{\mu}_1 - \hat{\mu}_2) \xrightarrow{\mathcal{L}} N(0, \Sigma^1 + \Sigma^2 - \Sigma^{12} - \Sigma^{21}).$$

This gives rise to the test statistic

$$T_{1p} = n(\bar{S^1} - \bar{S^2})'(\hat{\Sigma}^1 + \hat{\Sigma}^2 - \hat{\Sigma}^{12} - \hat{\Sigma}^{21})^{-1}(\bar{S^1} - \bar{S^2}) \tag{4.6.15}$$

where $\bar{S^1}$, $\bar{S^2}$, $\hat{\Sigma}^1$ and $\hat{\Sigma}^2$ are as in (4.6.3) and $\hat{\Sigma}^{12} = (\hat{\Sigma}^{21})'$ is the sample covariance between $\{S_j^1\}_{j=1}^n$ and $\{S_j^2\}_{j=1}^n$. If H_0 is true, T_{1p} converges in distribution to $\mathcal{X}_d^2$ distribution. Hence we reject H_0 at asymptotic level α if $T_{1p} > \mathcal{X}_d^2(1-\alpha)$.

If we are testing $H_0 : \mu_{1E} = \mu_{2E}$, then from (4.6.13), it follows that, under H_0,

$$\sqrt{n}[P(\hat{\mu}_1) - P(\hat{\mu}_2)] = \sqrt{n} d_{\mu_1} P(\hat{\mu}_1 - \mu_1) - \sqrt{n} d_{\mu_2} P(\hat{\mu}_2 - \mu_2) + o_P(1) \tag{4.6.16}$$

which implies that, for any $\tilde{\mu} \in M$,

$$\begin{aligned} &L_{\tilde{\mu}}[\sqrt{n}\{P(\hat{\mu}_1) - P(\hat{\mu}_2)\}] \\ =&L_{1\tilde{\mu}}[\sqrt{n} d_{\mu_1} P(\hat{\mu}_1 - \mu_1)] - L_{2\tilde{\mu}}[\sqrt{n} d_{\mu_2} P(\hat{\mu}_2 - \mu_2)] + o_P(1) \\ \xrightarrow{\mathcal{L}}&N(0, \Sigma) \end{aligned} \tag{4.6.17}$$

where

$$\Sigma = L_{1\tilde{\mu}}\Sigma_1 L'_{1\tilde{\mu}} + L_{2\tilde{\mu}}\Sigma_2 L'_{2\tilde{\mu}} - L_{1\tilde{\mu}}\Sigma_{12} L'_{2\tilde{\mu}} - L_{2\tilde{\mu}}\Sigma_{21} L'_{1\tilde{\mu}}.$$

In (4.6.17), $L_{\tilde{\mu}}$, $L_{i\tilde{\mu}}$ and Σ_i, $i = 1, 2$ are the same as in (4.6.6), and $\Sigma_{12} = \Sigma_{21}'$ denotes the covariance between the coordinates of $d_{\mu_1}P(\tilde{X}_1 - \mu_1)$ and $d_{\mu_2}P(\tilde{Y}_1 - \mu_2)$. Hence to test if H_0 is true, one can use the test statistic

$$T_{2p} = nL[P(\hat{\mu}_1) - P(\hat{\mu}_2)]'\hat{\Sigma}^{-1}L[P(\hat{\mu}_1) - P(\hat{\mu}_2)] \text{ where} \tag{4.6.18}$$

$$\hat{\Sigma} = L_1\hat{\Sigma}_1L_1' + L_2\hat{\Sigma}_2L_2' - L_1\hat{\Sigma}_{12}L_2' - L_2\hat{\Sigma}_{21}L_1'. \tag{4.6.19}$$

In the statistic T_{2p}; L, L_i and $\hat{\Sigma}_i$, $i = 1, 2$ are as in (4.6.7) and $\hat{\Sigma}_{12} = (\hat{\Sigma}_{12})'$ denotes the sample covariance between the coordinates of $\{d_{\hat{\mu}_1}P(\tilde{X}_j - \hat{\mu}_1)\}_{j=1}^n$ and $\{d_{\hat{\mu}_2}P(\tilde{Y}_j - \hat{\mu}_2)\}_{j=1}^n$. Under H_0, $T_{2p} \xrightarrow{\mathcal{L}} \mathcal{X}_d^2$. Hence we reject H_0 at asymptotic level α if $T_{2p} > \mathcal{X}_d^2(1-\alpha)$. In the application considered in Section 8.6, the values for the two statistics T_{1p} and T_{2p} are very close to each other.

One can also find a bootstrap confidence region for $P(\mu_1) - P(\mu_2)$ and use that to test if H_0 is true. The details are left to the reader.

Let V_1 and V_2 denote the extrinsic variations of Q_1 and Q_2 and let $\hat{V}_1$, $\hat{V}_2$ be their sample analogues. Suppose we want to test the hypothesis $H_0 : V_1 = V_2$. From (4.4.8), we get that,

$$\begin{aligned}\begin{pmatrix} \sqrt{n}(\hat{V}_1 - V_1) \\ \sqrt{n}(\hat{V}_2 - V_2) \end{pmatrix} &= \frac{1}{\sqrt{n}}\begin{pmatrix} \sum_{j=1}^n[\rho^2(X_j, \mu_{1E}) - \mathrm{E}\rho^2(X_1, \mu_{1E})] \\ \sum_{j=1}^n[\rho^2(Y_j, \mu_{2E}) - \mathrm{E}\rho^2(Y_1, \mu_{2E})] \end{pmatrix} + o_P(1) \\ &\xrightarrow{\mathcal{L}} N\left(0, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}\right)\end{aligned} \tag{4.6.20}$$

where $\sigma_{12} = \mathrm{Cov}(\rho^2(X_1, \mu_{1E}), \rho^2(Y_1, \mu_{2E}))$, σ_1^2 and σ_2^2 are as in (4.6.11). Hence if H_0 is true,

$$\sqrt{n}(\hat{V}_1 - \hat{V}_2) \xrightarrow{\mathcal{L}} N(0, \sigma_1^2 + \sigma_2^2 - 2\sigma_{12}).$$

This gives rise to the test statistic,

$$T_{3p} = \frac{\sqrt{n}(\hat{V}_1 - \hat{V}_2)}{\sqrt{s_1^2 + s_2^2 - 2s_{12}}} \tag{4.6.21}$$

where s_1^2, s_2^2, s_{12} are sample estimates of $\sigma_1^2, \sigma_2^2, \sigma_{12}$ respectively. We reject H_0 at asymptotic level α if $|T_{3p}| > Z(1 - \frac{\alpha}{2})$. We can also get a $(1 - \alpha)$ level confidence interval for $V_1 - V_2$ using bootstrap simulations and use that to test if H_0 is true.

Chapter 5

INTRODUCTION TO SHAPE SPACES

5.1 Landmark Based Shape Spaces

The statistical analysis of shape distributions based on random samples is important in many areas such as morphometrics (discrimination and classification of biological shapes), medical diagnostics (detection of change or deformation of shapes in some organs due to some disease, for example), machine vision (e.g., digital recording and analysis based on planar views of 3-D objects) and robotics (for robots to visually recognize a scene). Among the pioneers on foundational studies leading to such applications, we mention Kendall (1977, 1984) and Bookstein (1991). In this chapter and the chapters that follow, we will be mostly interested in the analysis of shapes of landmark based data, in which each observation consists of $k > m$ points in m-dimension, representing k locations on an object, called a k-ad. The choice of landmarks is generally made with expert help in the particular field of application. Depending on the way the data are collected or recorded, the appropriate shape of a k-ad is the maximal invariant specified by the space of orbits under a group of transformations.

For example, one may look at k-ads modulo size and Euclidean rigid body motions of translation and rotation. The analysis of shapes under this invariance was pioneered by Kendall (1977, 1984) and Bookstein (1978). Bookstein's approach is primarily registration-based requiring two or three landmarks to be brought into a standard position by translation, rotation and scaling of the k-ad. For these shapes, we would prefer Kendall's more invariant view of a shape identified with the orbit under rotation (in m-dimension) of the k-ad centered at the origin and scaled to have unit size. The resulting shape spaces are called the similarity shape spaces. A fairly comprehensive account of parametric inference on these spaces, with many references to the literature, may be found in Dryden and Mardia (1998). Once we consider the orbits under all orthogonal transformations, we get the reflection shape spaces.

Recently there has been much emphasis on the statistical analysis of other notions of shapes of k-ads, namely, affine shapes invariant under affine transformations, and projective shapes invariant under projective transformations. Reconstruction of a scene from two (or more) aerial photographs taken from a plane is one of the research problems in affine shape analysis. Potential applications of projective shape analysis include face recognition and robotics.

In this chapter, we will briefly describe the geometry of the above shape spaces and return to them one by one in the subsequent chapters.

5.2 Geometry of Shape Manifolds

Many differentiable manifolds M naturally occur as submanifolds, or surfaces or hypersurfaces, of an Euclidean space. One example of this is the sphere $S^d = \{p \in \mathbb{R}^{d+1}\colon \|p\| = 1\}$. The shape spaces of interest here are not of this type. They are quotients of a Riemannian manifold N under the action of a transformation group G, i.e., $M = N/G$. A number of them are quotient spaces of $N = S^d$ under the action of a compact group G, i.e., the elements of the space are orbits in S^d traced out by the application of G. Among important examples of this kind are the Kendall's shape spaces and reflection shape spaces. In some cases the **action of the group is free**, i.e., $gp = p$ only holds for the identity element $g = e$. Then the elements of the orbit $O_p = \{gp\colon g \in G\}$ are in one-one correspondence with elements of G, and one can identify the orbit with the group. The orbit inherits the differential structure of the Lie group G. The tangent space T_pN at a point p may then be decomposed into a **ver-**

tical subspace V_p of dimension that of the group G along the orbit space to which p belongs, and a **horizontal subspace** H_p which is orthogonal to it. The vertical subspace is isomorphic to the tangent space of G and the horizontal one can be identified with the tangent space of M at the orbit O_p. With this identification, M is a differentiable manifold of dimension that of N minus the dimension of G. Further if G acts as isometries of N, then the projection π, $\pi(p) = O_p$ is a **Riemannian submersion** of N onto the quotient space $M = N/G$. In other words, $\langle d\pi(v), d\pi(w)\rangle_{\pi(p)} = \langle v, w\rangle_p$ for horizontal vectors $v, w \in T_pN$, where $d\pi : T_pN \to T_{\pi(p)}M$ denotes the differential, or Jacobian, of the projection π. With this metric tensor, M has the natural structure of a Riemannian manifold. This provides the framework for carrying out an intrinsic analysis on M.

To carry out an extrinsic analysis on M, we use a smooth map J from N into some Euclidean space E which is an embedding of M into that Euclidean space. Then the image $J(M)$ is a differentiable submanifold of E. The tangent space at $J(\pi(p))$ is $dJ(H_p)$ where dJ is the differential of the map $J : N \to E$. Among all possible embeddings, we choose J to be equivariant under the action of a large group H on M. In most cases, H is compact.

5.2.1 (Direct) Similarity Shape Spaces Σ_m^k

Consider a k-ad in 2D or 3D with not all landmarks identical. Its **(direct) similarity shape** is what remains after removing the effects of translation, one dimensional scaling and rotation. The space of all similarity shapes forms the **(Direct) Similarity Shape Space** Σ_m^k, with m being the dimension of the Euclidean space where the landmarks lie, which is usually 2 or 3. Similarity shape analysis finds many applications in morphometrics - classification of biological species based on their shapes, medical diagnostics - disease detection based on change in shape of an organ due to disease or deformation, evolution studies - studying the change in shape of an organ or organism with time, age etc, and many more. Some such applications will be considered in subsequent chapters.

5.2.2 Reflection Similarity Shape Spaces $R\Sigma_m^k$

When one considers the **reflection similarity shape** of a k-ad, that is, features invariant under translation, scaling and all orthogonal transformations, then it is possible to embed the resulting shape space in some higher dimensional Euclidean space and carry out an extrinsic analysis. The embedding was first considered by Bandulasiri and Patrangenaru (2005) and later independently by Dryden et al. (2008). Such an embedding which is equivariant under a large group action is known for the similarity shape spaces only when $m = 2$. Thus considering the reflection shape makes it possible to extend the results of nonparametric inference on shapes from 2 to m (in particular 3) dimensions. The correct computation of the mean reflection shape was first carried out in Bhattacharya (2008a). For details, see Chapter 8 of this book.

5.2.3 Affine Shape Spaces $A\Sigma_m^k$

An application in bioinformatics consists in matching two marked electrophoresis gels. Proteins are subjected to stretches in two directions. Due to their molecular mass and electrical charge, the amount of stretching depends on the strength and duration of the electrical fields applied. For this reason, the same tissue analyzed by different laboratories may yield different constellations of protein spots. The two configurations differ by a change of coordinates that can be approximately given by an **affine transformation** which may not be similarity.

Another application of **affine shape** analysis is in scene recognition: to reconstruct a larger image from partial views in a number of ariel images of that scene. For a remote scene, the image acquisition process will involve a parallel projection, which in general is not orthogonal. Two common parts of the same scene seen in different images will essentially differ by an affine transformation but not a similarity.

5.2.4 Projective Shape Spaces $P\Sigma_m^k$

In machine vision, if images are taken from a great distance, affine shape analysis is appropriate. Otherwise, **projective shape** is a more appropriate choice. If images are obtained through a central projection, a ray is received as a point on the image plane. Since axes in 3D comprise the **projective space** $\mathbb{R}P^2$, k-ads in this view are valued in $\mathbb{R}P^2$. To have invariance with regard to camera angles, one may first look at the original 3D k-ad and achieve affine invariance by its affine shape and finally take the corresponding equivalence class of axes in $\mathbb{R}P^2$, to define the projective shape of the k-ad invariant with respect to all projective transformations on $\mathbb{R}P^2$. Potential applications of projective shape analysis arise in robotics, particularly in machine vision for robots to visually recognize a scene, avoid an obstacle, etc.

For a remote view, the rays falling on the image plane are more or less parallel, and then a projective transformation can be approximated by an affine transformation. Further, if it is assumed that the rays fall perpendicular to the image plane, then similarity or reflection similarity shape space analysis becomes appropriate.

Chapter 6

KENDALL'S (DIRECT) SIMILARITY SHAPE SPACES Σ_m^k.

6.1 Introduction

Kendall's shape spaces are quotient spaces S^d/G, under the action of the special orthogonal group $G = SO(m)$ of $m \times m$ orthogonal matrices with determinant $+1$. Important cases include $m = 2, 3$.

For the case $m = 2$, consider the space of all planar k-ads $(z_1, z_2, \ldots, z_k)$ $(z_j = (x_j, y_j))$, $k > 2$, excluding those with k identical points. The set of all centered and normed k-ads, say $u = (u_1, u_2, \ldots, u_k)$ comprise a unit sphere in a $(2k-2)$ dimensional vector space and is, therefore, a $(2k-3)$ dimensional sphere S^{2k-3}, called the **preshape sphere**. The group $G = SO(2)$ acts on the sphere by rotating each landmark by the same angle. The orbit under G of a point u in the preshape sphere can thus be seen to be a circle S^1, so that Kendall's **planar shape space** Σ_2^k can be viewed as the quotient space $S^{2k-3}/G \sim S^{2k-3}/S^1$, a $(2k-4)$ dimensional compact manifold. An algebraically simpler representation of Σ_2^k is given by the complex projective space $\mathbb{C}P^{k-2}$. For nonparametric inference on Σ_2^k, see Bhattacharya and Patrangenaru (2003, 2005), Bhattacharya and Bhattacharya (2008a, 2008b). For many applications in archaeology, astronomy, morphometrics, medical diagnostics etc, see Bookstein (1986, 1997), Kendall (1989), Dryden and Mardia (1998), Bhattacharya and Patrangenaru (2003, 2005), Bhattacharya and Bhattacharya (2008a, c) and Small (1996). We will return back to this shape space in the next chapter.

When $m > 2$, consider a set of k points in $\mathbb{R}^m$, not all points being the same.

Such a set is called a k-ad or a configuration of k landmarks. We will denote a k-ad by the $m \times k$ matrix, $x = (x_1, \ldots, x_k)$ where $x_i,\ i = 1, \ldots, k$ are the k landmarks from the object of interest. Assume $k > m$. The **direct similarity shape** of the k-ad is what remains after we remove the effects of translation, rotation and scaling. To remove translation, we subtract the mean $\bar{x} = \frac{1}{k}\sum_{i=1}^{k} x_i$ from each landmark to get the centered k-ad $u = (x_1 - \bar{x}, \ldots, x_k - \bar{x})$. We remove the effect of scaling by dividing u by its Euclidean norm to get

$$z = (\frac{x_1 - \bar{x}}{\|u\|}, \ldots, \frac{x_k - \bar{x}}{\|u\|}) = (z_1, z_2, \ldots, z_k). \tag{6.1.1}$$

This z is called the **preshape** of the k-ad x and it lies in the unit sphere S_m^k in the hyperplane $H_m^k = \{z \in \mathbb{R}^{m \times k} \colon z\mathbf{1}_k = 0\}$. Hence

$$S_m^k = \{z \in \mathbb{R}^{m \times k} \colon \mathrm{Trace(zz')} = 1,\ \mathrm{z}\mathbf{1}_\mathrm{k} = 0\} \tag{6.1.2}$$

Here $\mathbf{1}_k$ denotes the $k \times 1$ vector of all ones. Thus the **preshape sphere** S_m^k may be identified with the sphere S^{km-m-1}. Then the shape of the k-ad x is the orbit of z under left multiplication by $m \times m$ rotation matrices. In other words $\Sigma_m^k = S_m^k / SO(m)$. One can also remove the effect of translation from the original k-ad x by postmultiplying the centered k-ad u by a **Helmert** matrix H which is a $k \times (k-1)$ matrix satisfying $H'H = I_{k-1}$ and $\mathbf{1}_k' H = 0$. The resulting k-ad $\tilde{u} = uH$ lies in $\mathbb{R}^{m \times (k-1)}$ and is called the **Helmertized k-ad**. Then the preshape of x or $\tilde{u}$ is $\tilde{z} = \tilde{u}/\|\tilde{u}\|$ and the preshape sphere is

$$S_m^k = \{z \in \mathbb{R}^{m \times (k-1)} \colon \mathrm{Trace(zz')} = 1\}. \tag{6.1.3}$$

The advantage of using this representation of S_m^k is that there is no linear constraint on the coordinates of z and hence analysis becomes simpler. However, now the choice of the preshape depends on the choice of H which can vary. In most cases, including applications, we will represent the preshape of x as in equation (6.1.1) and the preshape sphere as in (6.1.2).

6.2 Geometry of Similarity Shape Spaces

In this section, we study the topological and geometrical properties of Σ_m^k represented as $S_m^k/SO(m)$. We are interested in the case when $m > 2$. The case $m = 2$ is studied in Chapter 7.

For $m > 2$, the direct similarity shape space Σ_m^k fails to be a manifold. That is because the action of $SO(m)$ is not in general free. Indeed, the orbits of preshapes under $SO(m)$ have different dimensions in different regions (see, e.g., Kendall et al. (1999) and Small (1996)). To avoid that, one may consider the shape of only those k-ads whose preshapes have rank at least $m - 1$. Define

$$NS_m^k = \{z \in S_m^k : \text{ rank}(z) \geq m - 1\}$$

as the **nonsingular part** of S_m^k and $\Sigma_{0m}^k = NS_m^k/SO(m)$. Then, since the action of $SO(m)$ on NS_m^k is free, Σ_{0m}^k is a differentiable manifold of dimension $km - m - 1 - \frac{m(m-1)}{2}$. Also since $SO(m)$ acts as isometries of the sphere, Σ_{0m}^k inherits the Riemannian metric tensor of the sphere and hence is a Riemannnian manifold. However it is not complete because of the 'holes' created by removing the singular part.

Consider the projection map

$$\pi : NS_m^k \longrightarrow \Sigma_{0m}^k, \ \pi(z) = \{Az : \ A \in SO(m)\}.$$

This map is shown to be a **Riemannian submersion** (see Kendall et. al. (1999)). This means that if we write $T_z S_m^k$ as the direct sum of the horizontal subspace H_z and vertical subspace V_z, then $d\pi$ is a isometry from H_z into $T_{\pi(z)}\Sigma_{0m}^k$. The tangent space $T_z S_m^k$ is

$$T_z S_m^k = \{v \in H_m^k : \text{ Trace}(vz') = 0\}.$$

The vertical subspace V_z consists of initial velocity vectors of curves in S_m^k starting at z and remaining in the orbit $\pi(z)$. Such a curve will have the form $\gamma(t) = \tilde{\gamma}(t)z$

where $\tilde{\gamma}(t)$ is a curve in $SO(m)$ starting at the identity matrix I_m. Geodesics in $SO(m)$ starting at I_m have the form $\tilde{\gamma}(t) = \exp(tA)$ where

$$\exp(A) = I + A + \frac{A^2}{2} + \frac{A^3}{3!} + \ldots$$

and A is skew-symmetric ($A + A' = 0$). For such a curve, $\dot{\tilde{\gamma}}(0) = A$, therefore $\dot{\gamma}(0) = Az$ which implies that

$$V_z = \{Az : \ A + A' = 0\}.$$

The horizontal subspace is its ortho-complement, which is

$$H_z = \{v \in H_m^k : \ \text{Trace}(vz') = 0, \ vz' = zv'\}.$$

Since π is a Riemannian submersion, $T_{\pi(z)}\Sigma_{0m}^k$ is isometric to H_z.

The geodesic distance between two shapes $\pi(x)$ and $\pi(y)$, where $x, y \in S_m^k$, is given by

$$d_g(\pi(x), \pi(y)) = \min_{T \in SO(m)} d_{gs}(x, Ty).$$

Here $d_{gs}(.,.)$ is the geodesic distance on S_m^k which is

$$d_{gs}(x, y) = \arccos(\text{Trace}(yx')).$$

Therefore

$$d_g(\pi(x), \pi(y)) = \arccos(\max_{T \in SO(m)} \text{Trace}(Tyx')). \tag{6.2.1}$$

Consider the **pseudo-singular value decomposition** of yx' which is

$$yx' = U\Lambda V; \ U, V \in SO(m),$$

$$\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_m), \ \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{m-1} \geq |\lambda_m|, \ \text{sign}(\lambda_m) = \text{sign}(\det(yx')).$$

Then the value of T for which $\text{Trace}(Tyx')$ in equation (6.2.1) is maximized is $T = V'U'$ and then

$$d_g(\pi(x), \pi(y)) = \arccos(\text{Trace}(\Lambda)) = \arccos(\sum_{j=1}^{m} \lambda_j)$$

which lies between 0 and $\frac{\pi}{2}$.

Define the singular part D_{m-2} of S_m^k as the set of all preshapes with rank less than $m-1$. Then it is shown in Kendall et. al. (1999) that for $x \in S_m^k \setminus D_{m-2} \equiv NS_m^k$, the cut-locus of $\pi(x)$ in Σ_{0m}^k is given by

$$C(\pi(x)) = \pi(D_{m-2}) \cup C_0(\pi(X))$$

where $C_0(\pi(X))$ is defined to be the set of all shapes $\pi(y) \in \Sigma_{0m}^k$ such that there exists more than one length minimizing geodesic joining $\pi(x)$ and $\pi(y)$. It is also shown that the least upper bound on all sectional curvatures of Σ_{0m}^k is $+\infty$. Hence we cannot apply the results of Chapter 3 to carry out intrinsic analysis on this space.

Once we remove the effects of reflections along with rotations from the preshapes, we can embed the shape space into a higher dimensional Euclidean space and carry out extrinsic analysis of shapes. This is done in Chapter 8.

Chapter 7

THE PLANAR SHAPE SPACE Σ_2^k

7.1 Introduction

Consider a set of k points on the plane, not all points being the same. We will assume $k > 2$ and refer to such a set as a k-ad or a set of k landmarks. For convenience we will denote a k-ad by k complex numbers ($z_j = x_j + iy_j, 1 \leq j \leq k$), i.e., we will represent k-ads on a complex plane. Then the similarity shape of a k-ad $z = (z_1, z_2, \ldots, z_k)'$ represents the equivalence class, or orbit of z under translation, one dimensional scaling and rotation. To remove translation, one subtracts

$$\langle z \rangle \equiv (\bar{z}, \bar{z}, \ldots, \bar{z})' \;\; (\bar{z} = \frac{1}{k}\sum_{j=1}^{k} z_j)$$

from z to get $z - \langle z \rangle$. Rotation of the k-ad by an angle θ and scaling by a factor $r > 0$ are achieved by multiplying $z - \langle z \rangle$ by the complex number $\lambda = re^{i\theta}$. Hence one may represent the shape of the k-ad as the complex line passing through $z - \langle z \rangle$, namely, $\{\lambda(z - \langle z \rangle) \colon \lambda \in \mathbb{C} \setminus \{0\}\}$. Thus the space of similarity shapes of k-ads is the set of all complex lines on the (complex $(k-1)$-dimensional) hyperplane, $H^{k-1} = \{w \in \mathbb{C}^k \setminus \{0\} \colon \sum_1^k w_j = 0\}$. Therefore the similarity shape space Σ_2^k of planer k-ads has the structure of the **complex projective space** $\mathbb{C}P^{k-2}$- the space of all complex lines through the origin in $\mathbb{C}^{k-1}$.

7.2 Geometry of the Planar Shape Space

When identified with $\mathbb{C}P^{k-2}$, Σ_2^k is a compact connected Riemannian manifold of (real) dimension $2k-4$. As in the case of $\mathbb{C}P^{k-2}$, it is convenient to represent the shape $\sigma(z)$ of a k-ad z by the curve

$$\sigma(z) = \pi(u) = \{e^{i\theta}u\colon\ -\pi < \theta \le \pi\},\ u = \frac{z - \langle z \rangle}{\|z - \langle z \rangle\|}$$

on the unit sphere $\mathbb{C}S^{k-1}$ in H^{k-1}. The quantity u is called the **preshape** of the shape of the original k-ad z and it lies on $\mathbb{C}S^{k-1}$ which is

$$\mathbb{C}S^{k-1} = \{u \in \mathbb{C}^k : \sum_{j=1}^{k} u_j = 0,\ \|u\| = 1\}.$$

The map $\pi : \mathbb{C}S^{k-1} \to \Sigma_2^k$ is a Riemannian submersion. Hence its derivative $d\pi$ is an isometry from H_u into $T_{\pi(u)}\Sigma_2^k$, where H_u is the horizontal subspace of the tangent space $T_u\mathbb{C}S^{k-1}$ of $\mathbb{C}S^{k-1}$ at u, which is

$$H_u = \{v \in \mathbb{C}^k\ : z^{'}\bar{v} = 0,\ v'\mathbf{1}_k = 0\}.$$

The preshape sphere $\mathbb{C}S^{k-1}$ can be identified with the real sphere of dimension $2k-3$, namely S^{2k-3}. Hence if exp denotes the exponential map of $\mathbb{C}S^{k-1}$ as derived in Chapter 3, then the exponential map of Σ_2^k is given by

$$\mathrm{Exp}_{\pi(u)} : T_{\pi(u)}\Sigma_2^k \to \Sigma_2^k,\ \mathrm{Exp}_{\pi(u)} = \pi \circ \exp_u \circ\, d\pi_u^{-1}.$$

The geodesic distance between two shapes $\sigma(x)$ and $\sigma(y)$ is given by

$$d_g(\sigma(x), \sigma(y)) = d_g(\pi(z), \pi(w)) = \inf_{\theta \in (-\pi, \pi]} d_{gs}(z, e^{i\theta}w)$$

where x and y are two k-ads in $\mathbb{C}^k$, z and w are their preshapes in $\mathbb{C}S^{k-1}$, and $d_{gs}(.,.)$ denotes the geodesic distance on $\mathbb{C}S^{k-1}$, which is given by $d_{gs}(z, w) = \arccos(\mathrm{Re}(\bar{w}'z))$ as mentioned in Section 3.4. Hence the geodesic distance on Σ_2^k has the following expression:

$$\begin{aligned} d_g(\pi(z), \pi(w)) &= \inf_{\theta \in (-\pi, \pi]} \arccos(\mathrm{Re}(e^{-i\theta}\bar{w}'z)) \\ &= \arccos \sup_{\theta \in (-\pi, \pi]} \mathrm{Re}(e^{-i\theta}\bar{w}'z) = \arccos(|\bar{w}'z|). \end{aligned}$$

Hence the geodesic distance between any pair of planar shapes lies between 0 and $\frac{\pi}{2}$ which means that Σ_2^k has an injectivity radius of $\frac{\pi}{2}$. The cut-locus $C(\pi(z))$ of $z \in \mathbb{C}S^{k-1}$ is given by

$$C(\pi(z)) = \{\pi(w) : w \in \mathbb{C}S^{k-1},\ d_g(\pi(z), \pi(w)) = \frac{\pi}{2}\} = \{\pi(w) : \bar{w}'z = 0\}.$$

The exponential map $\mathrm{Exp}_{\pi(z)}$ is invertible outside the cut-locus of z and its inverse is given by

$$\mathrm{Exp}_{\pi(z)}^{-1} : \Sigma_2^k \setminus C(\pi(z)) \to T_{\pi(z)}\Sigma_2^k,\ \pi(w) \mapsto d\pi_{\pi(z)} \left\{ \frac{r}{\sin(r)}(-\cos(r)z + e^{i\theta}w) \right\} \tag{7.2.1}$$

where $r = d_g(\pi(z), \pi(w))$ and $e^{i\theta} = \frac{\bar{w}'z}{|\bar{w}'z|}$. It has been shown in Kendall (1984) that Σ_2^k has constant holomorphic sectional curvature of 4.

Given two preshapes u and v, the **Procrustes coordinates** of v with respect to u is defined as

$$v^P = e^{i\theta} v$$

where $\theta \in (-\pi, \pi]$ is chosen so as to minimize the Euclidean distance between u and $e^{i\theta}v$, namely $d_P(\theta) = \|u - e^{i\theta}v\|$. In other words, one trys to rotate the preshape v so as to bring it closest to u. Then

$$d_P^2(\theta) = 2 - 2\mathrm{Re}(e^{i\theta}\bar{u}'v)$$

which is minimized when $e^{i\theta} = \frac{\bar{v}'u}{|\bar{v}'u|}$ and then the minimum value of the Euclidean distance turns out to be

$$d_P = \min_{\theta \in (-\pi, \pi]} d_P(\theta) = \sqrt{2(1 - |\bar{v}'u|)}.$$

This d_P is a distance metric on Σ_2^k, called the **Procrustes distance** (see Dryden and Mardia (1998) for details). The Procrustes coordinates can be particularly useful for plotting shapes as we shall see in the next section.

7.3 Examples

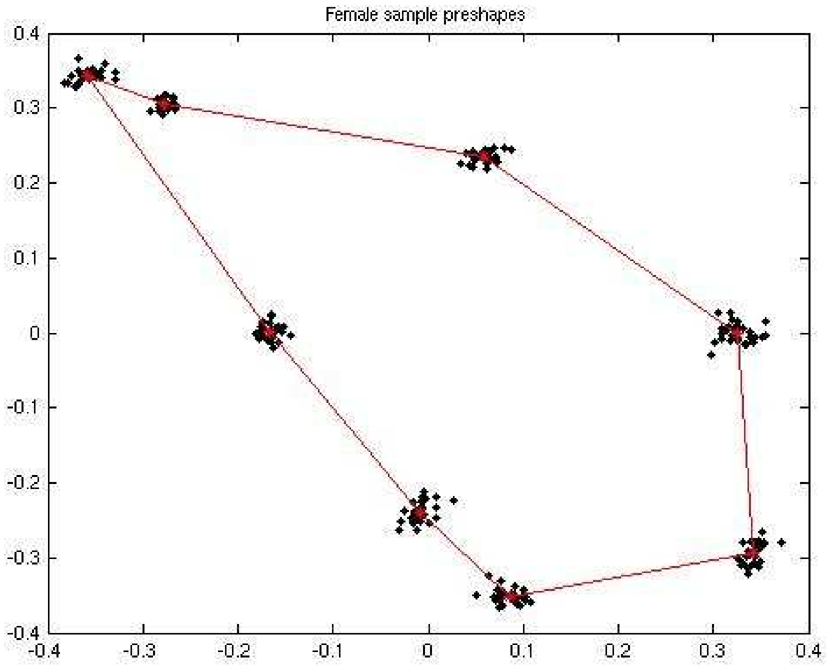
In this section, we discuss two applications of planar shape analysis. We will return back to these examples in Section 7.10.

7.3.1 Gorilla Skulls

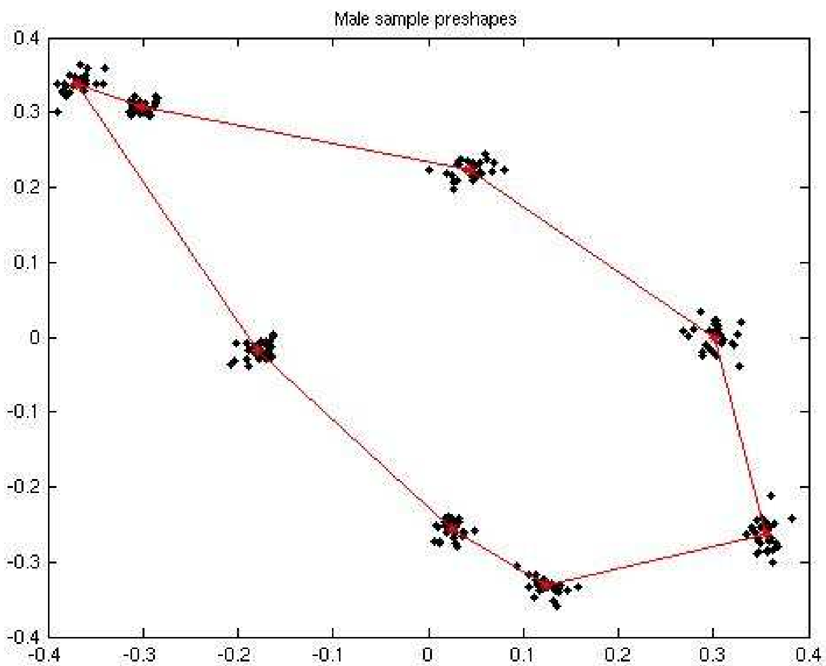
Consider eight locations on a gorilla skull projected on a plane. There are 29 male and 30 female gorillas and the eight landmarks are chosen on the midline plane of the 2D image of the skulls. The data can be found in Dryden and Mardia (1998). It is of interest to study the shapes of the skulls and use that to detect difference in shapes between the sexes. This finds application in morphometrics and other biological sciences. To analyze the planar shapes of the k-ads, the observations lie in Σ_2^k, $k = 8$. Figure 7.1(a) shows the Procrustes coordinates of the shapes of the female gorilla skulls. The coordinates are obtained with respect to a preshape of the sample extrinsic mean, which is defined in Section 7.7. Figure 7.1(b) shows the Procrustes coordinates of the shapes of the male gorilla skulls with respect to a preshape of the male sample extrinsic mean.

7.3.2 Schizophrenic Children

In this example from Bookstein (1991), 13 landmarks are recorded on a midsagittal two-dimensional slice from a Magnetic Resonance brain scan of each of 14 schizophrenic children and 14 normal children. It is of interest to study differences in shapes of brains between the two groups which can be used to detect schizophrenia. This is an application of disease detection. The shapes of the sample k-ads lie in Σ_2^k, $k = 13$. Figure 7.2(a) shows the Procrustes coordinates of the shapes of the schizophrenic children while Figure 7.2(b) shows the coordinates for the normal children. As in Section 7.3.1, the coordinates are obtained with respect to the preshapes of the respective sample extrinsic means.

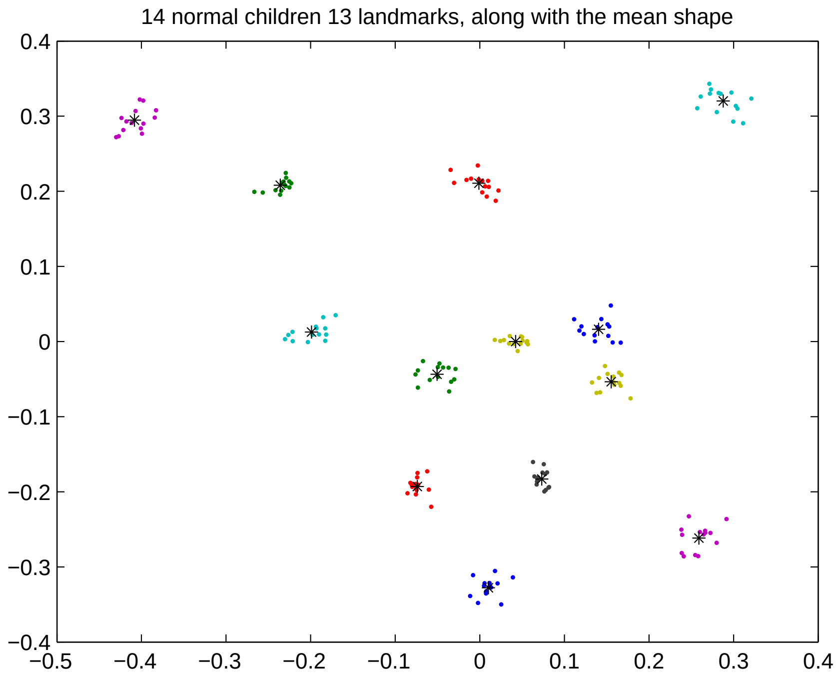


(a)

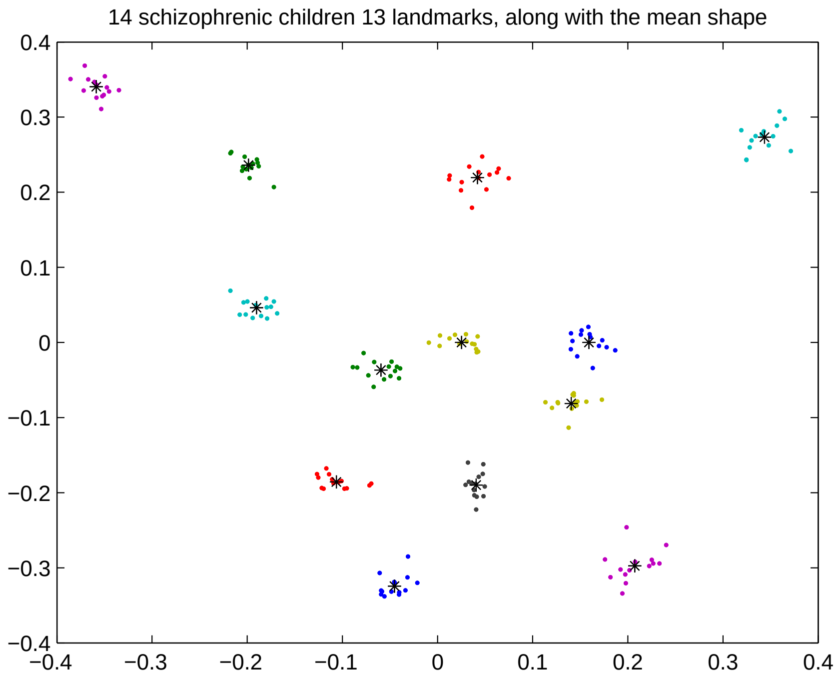


(b)

Figure 7.1. (a) and (b) show 8 landmarks from skulls of 30 female and 29 male gorillas respectively along with the respective sample mean shapes. * correspond to the mean shapes' landmarks.



(a)



(b)

FIGURE 7.2. (a) and (b) show 13 landmarks for 14 normal and 14 schizophrenic children respectively along with the respective mean shapes. * correspond to the mean shapes' landmarks.

7.4 Intrinsic Analysis on the Planar Shape Space

Let Q be a probability distribution on Σ_2^k. From Proposition 3.2.1, it follows that if the support of Q is contained in a geodesic ball of radius $\frac{\pi}{4}$ then it has a unique intrinsic mean in that ball. In this section we assume that, that is $\text{supp}(Q) \subseteq B(p, \frac{\pi}{4})$ for some $p \in \Sigma_2^k$. Let $\mu_I = \pi(\mu)$ be the (local) intrinsic mean of Q in $B(p, \frac{\pi}{4})$, with μ being one of its preshapes. Let $X_1, \ldots, X_n$ be an iid sample from Q on Σ_2^k and let μ_{nI} be the (local) sample intrinsic mean in $B(p, \frac{\pi}{4})$. From Proposition 3.2.1, it follows that μ_{nI} is a consistent estimator of μ_I. Furthermore if we assume that $\text{supp}(Q) \subseteq B(\mu_I, \frac{\pi}{4})$, then Theorem 3.3.1 implies that the coordinates of μ_{nI} have asymptotic Normal distribution. However this theorem does not give expression for the asymptotic parameter Λ because Σ_2^k does not have constant sectional curvature. Theorem 7.4.1 below shows us how to get the analytic expression for Λ and relaxes the support condition for its positive definiteness. This theorem is stated in Bhattacharya and Bhattacharya (2008b) and the following proof is taken from there.

Theorem 7.4.1. *Let $\phi : B(p, \frac{\pi}{4}) \rightarrow \mathbb{C}^{k-2} (\approx \mathbb{R}^{2k-4})$ be the coordinates of $d\pi_\mu^{-1} \circ Exp_{\mu_I}^{-1} : B(p, \frac{\pi}{4}) \rightarrow H_\mu$ with respect to some orthonormal basis $\{v_1, \ldots, v_{k-2}, iv_1, \ldots, iv_{k-2}\}$ for H_μ (over $\mathbb{R}$). Define $h(x, y) = d_g^2(\phi^{-1}(x), \phi^{-1}(y))$. Let $((D_r h))_{r=1}^{2k-4}$ and $((D_r D_s h))_{r,s=1}^{2k-4}$ be the matrix of first and second order derivatives of $y \mapsto h(x, y)$. Let $\tilde{X}_j = \phi(X_j) = (\tilde{X}_j^1, \ldots, \tilde{X}_j^{k-2})$; $j = 1, \ldots, n$ be the coordinates of the sample observations. Define $\Lambda = E((D_r D_s h(\tilde{X}_1, 0)))_{r,s=1}^{2k-4}$. Then Λ is positive definite if the support of Q is contained in $B(\mu_I, R)$ where R is the unique solution of $\tan(x) = 2x$, $x \in (0, \frac{\pi}{2})$.*

Proof. For a geodesic γ starting at μ_I, write $\gamma = \pi \circ \tilde{\gamma}$, where $\tilde{\gamma}$ is a geodesic in $\mathbb{C}S^{k-1}$ starting at μ. From the proof of Theorem 3.3.1, for $m = \pi(z) \in B(p, \frac{\pi}{4})$,

$$\frac{d}{ds} d_g^2(\gamma(s), m) = 2\langle T(s, 1), \dot{\gamma}(s) \rangle = 2\langle \tilde{T}(s, 1), \dot{\tilde{\gamma}}(s) \rangle \tag{7.4.1}$$

$$\frac{d^2}{ds^2} d_g^2(\gamma(s), m) = 2\langle D_s T(s, 1), \dot{\gamma}(s) \rangle = 2\langle D_s \tilde{T}(s, 1), \dot{\tilde{\gamma}}(s) \rangle \tag{7.4.2}$$

where $\tilde{T}(s,1) = d\pi^{-1}_{\gamma(s)}(T(s,1))$. From equation (7.2.1), this has the expression

$$\tilde{T}(s,1) = -\frac{\rho(s)}{\sin(\rho(s))}\left[-\cos(\rho(s))\tilde{\gamma}(s) + e^{i\theta(s)}z\right] \tag{7.4.3}$$
$$\text{where } e^{i\theta(s)} = \frac{\bar{z}'\tilde{\gamma}(s)}{\cos(\rho(s))},\ \rho(s) = d_g(\gamma(s), m).$$

The inner product in equations (7.4.1) and (7.4.2) is the Riemannian metric on $T\mathbb{C}S^{k-1}$ which is $\langle v, w\rangle = \mathrm{Re}(v'\bar{w})$. Observe that $D_s\tilde{T}(s,1)$ is $\frac{d}{ds}\tilde{T}(s,1)$ projected onto $H_{\tilde{\gamma}(s)}$. Since $\langle \mu, \dot{\tilde{\gamma}}(0)\rangle = 0$, we get

$$\frac{d^2}{ds^2}d_g^2(\gamma(s), m)|_{s=0} = 2\langle\frac{d}{ds}\tilde{T}(s,1)|_{s=0}, \dot{\tilde{\gamma}}(0)\rangle.$$

From equation (7.4.3) we have,

$$\begin{aligned}\frac{d}{ds}\tilde{T}(s,1)|_{s=0} = {} & \left(\frac{d}{ds}(\frac{\rho(s)\cos(\rho(s))}{\sin(\rho(s))})|_{s=0}\right)\mu + \left(\frac{\rho(s)\cos(\rho(s))}{\sin\rho(s)}|_{s=0}\right)\dot{\tilde{\gamma}}(0)\\ & - \left(\frac{d}{ds}(\frac{\rho(s)}{\sin(\rho(s))\cos(\rho(s))})|_{s=0}\right)(\bar{z}'\mu)z\\ & - \left(\frac{\rho(s)}{\sin(\rho(s))\cos(\rho(s))}|_{s=0}\right)(\bar{z}'\dot{\tilde{\gamma}}(0))z\end{aligned}$$

and along with equation (7.4.1), we get

$$\frac{d}{ds}\rho(s)|_{s=0} = \frac{-1}{\sin(r)}\langle\dot{\tilde{\gamma}}(0), \frac{\bar{z}'\mu}{\cos(r)}z\rangle$$

where $r = d_g(m, \mu_I)$. Hence

$$\begin{aligned}\langle\frac{d}{ds}\tilde{T}(s,1)|_{s=0}, \dot{\tilde{\gamma}}(0)\rangle = {} & r\frac{\cos(r)}{\sin(r)}\|\dot{\tilde{\gamma}}(0)\|^2 - \left(\frac{1}{\sin^2 r} - r\frac{\cos(r)}{\sin^3(r)}\right)(\mathrm{Re}(x))^2\\ & + \frac{r}{\sin(r)\cos(r)}(\mathrm{Im}(x))^2 \end{aligned} \tag{7.4.4}$$

where

$$x = e^{i\theta}z'\overline{\dot{\tilde{\gamma}}(0)},\ e^{i\theta} = \frac{\bar{z}'\mu}{\cos(r)}. \tag{7.4.5}$$

The value of x in equation (7.4.5) and hence the expression in equation (7.4.4) depend on z only through $m = \pi(z)$. Also if $\gamma = \pi(\gamma_1) = \pi(\gamma_2)$, γ_1 and γ_2 being two geodesics on $\mathbb{C}S^{k-1}$ starting at μ_1 and μ_2 respectively, with $\pi(\mu_1) = \pi(\mu_2) = \pi(\mu)$, then $\gamma_1(t) = \lambda\gamma_2(t)$, where $\mu_2 = \lambda\mu_1$, $\lambda \in \mathbb{C}$. Now it is easy to check that the expression in (7.4.4) depends on μ only through $\pi(\mu) = \mu_I$. Note that

$$|x|^2 < 1 - \cos^2(r).$$

So when $|\dot{\gamma}(0)| = 1$, (7.4.4) becomes

$$\begin{aligned} & r\frac{\cos(r)}{\sin(r)} - \left(\frac{1}{\sin^2(r)} - r\frac{\cos(r)}{\sin^3(r)}\right)(\mathrm{Re}(x))^2 + \frac{r}{\sin(r)\cos(r)}(\mathrm{Im}(x))^2 \\ > & r\frac{\cos(r)}{\sin(r)} - \left(\frac{1}{\sin^2(r)} - r\frac{\cos(r)}{\sin^3(r)}\right)\sin^2(r) \\ = & \frac{2r - \tan(r)}{\tan(r)} \end{aligned} \tag{7.4.6}$$

which is strictly positive if $r \leq R$ where

$$\tan(R) = 2R,\ R \in (0, \frac{\pi}{2}).$$

Therefore if $\mathrm{supp}(Q) \subseteq B(\mu_I, R)$, then $\frac{d^2}{ds^2} d^2(\gamma(s), m)|_{s=0} > 0$ and hence Λ is positive definite. □

Remark 7.4.1. It can be shown that $R \in (\frac{\pi}{3}, \frac{2\pi}{5})$. It is approximately 0.37101π.

From Theorems 2.3.1 and 7.4.1, we conclude that if $\mathrm{supp}(Q) \subseteq B(p, \frac{\pi}{4}) \cap B(\mu_I, R)$ and if Σ is nonsingular (e.g., if Q is absolutely continuous), then the coordinates of the sample mean shape from an iid sample have an asymptotically Normal distribution with nonsingular dispersion. Note that the coordinate map ϕ in Theorem 7.4.1 has the form

$$\phi(m) = (\tilde{m}^1, \ldots, \tilde{m}^{k-2})',\ \tilde{m}^j = \frac{r}{\sin(r)} e^{i\theta} \bar{v}_j{}' z$$

where $m = \pi(z)$, $\mu_I = \pi(\mu)$, $r = \arccos(|\bar{z}'\mu|)$ and $e^{i\theta} = \frac{\bar{z}'\mu}{|\bar{z}'\mu|}$. Corollary 7.4.2 below derives expressions for Λ and Σ in terms of ϕ. It is proved in Bhattacharya and Bhattacharya (2008b).

Corollary 7.4.2. *Consider the same set up as in Theorem 7.4.1. If Q has support in a geodesic ball of radius $\frac{\pi}{4}$, then Λ has the following expression:*

$$\Lambda = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{12}' & \Lambda_{22} \end{bmatrix} \tag{7.4.7}$$

where for $1 \le r, s \le k-2$,

$$\begin{aligned}
(\Lambda_{11})_{rs} &= 2\mathrm{E}\Big[\ d_1 \cot(d_1)\delta_{rs} - \frac{(1 - d_1 \cot(d_1))}{d_1^2}(\mathrm{Re}(\tilde{X}_1^r))(\mathrm{Re}(\tilde{X}_1^s)) \\
&\qquad + \frac{\tan(d_1)}{d_1}(\mathrm{Im}(\tilde{X}_1^r))(\mathrm{Im}(\tilde{X}_1^s))\ \Big], \\
(\Lambda_{22})_{rs} &= 2\mathrm{E}\Big[\ d_1 cot(d_1)\delta_{rs} - \frac{(1 - d_1 cot(d_1))}{d_1^2}(Im\tilde{X}_1^r)(Im\tilde{X}_1^s) \\
&\qquad + \frac{tan(d_1)}{d_1}(Re\tilde{X}_1^r)(Re\tilde{X}_1^s)\ \Big], \\
(\Lambda_{12})_{rs} &= -2\mathrm{E}\Big[\ \frac{(1 - d_1 \cot(d_1))}{d_1^2}(\mathrm{Re}(\tilde{X}_1^r))(\mathrm{Im}(\tilde{X}_1^s)) \\
&\qquad + \frac{\tan(d_1)}{d_1}(\mathrm{Im}(\tilde{X}_1^r))(\mathrm{Re}(\tilde{X}_1^s))\ \Big]
\end{aligned}$$

with $d_1 = d_g(X_1, \mu_I)$. If we define $\Sigma = \mathrm{Cov}((D_r h(\tilde{X}_1, 0))_{r=1}^{2k-4}$, then it can be expressed as

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{bmatrix} \tag{7.4.8}$$

where for $1 \le r, s \le k-2$,

$$\begin{aligned}
(\Sigma_{11})_{rs} &= 4\mathrm{E}(\mathrm{Re}(\tilde{X}_1^r)\mathrm{Re}(\tilde{X}_1^s)), \\
(\Sigma_{12})_{rs} &= 4\mathrm{E}(\mathrm{Re}(\tilde{X}_1^r)\mathrm{Im}(\tilde{X}_1^s)), \\
(\Sigma_{22})_{rs} &= 4\mathrm{E}(\mathrm{Im}(\tilde{X}_1^r)\mathrm{Im}(\tilde{X}_1^s)).
\end{aligned}$$

Proof. With respect to the orthonormal basis $\{v_1, \ldots, v_{k-2},\ iv_1, \ldots, iv_{k-2}\}$ for H_μ, $\tilde{X}_j$ has coordinates

$$(\mathrm{Re}(\tilde{X}_j^1), \ldots, \mathrm{Re}(\tilde{X}_j^{k-2}), \mathrm{Im}(\tilde{X}_j^1), \ldots, \mathrm{Im}(\tilde{X}_j^{k-2})).$$

in $\mathbb{R}^{2k-4}$. Now the expression for Σ follows from Corollary 3.3.2. If one writes Λ as in (7.4.7) and if $\dot{\tilde{\gamma}}(0) = \sum_{j=1}^{k-2} x^j v_j + \sum_{j=1}^{k-2} y^j (iv_j)$, then

$$\mathrm{E}\left(\frac{d^2}{ds^2} d_g^2(\gamma(s), X_1)\right)\Big|_{s=0} = x'\Lambda_{11}x + y'\Lambda_{22}y + 2x'\Lambda_{12}y$$

where $x = (x^1, \dots, x^{k-2})'$ and $y = (y^1, \dots, y^{k-2})'$. Now expressions for Λ_{11}, Λ_{12} and Λ_{22} follow from the proof of Theorem 7.4.1. □

Using the expressions for Λ and Σ from Corollary 7.4.2, one can construct confidence regions for the population intrinsic mean as in Sections 2.3 and 3.3. Also one may carry out two sample tests as in Section 3.5 to distinguish between two probability distributions on Σ_2^k by comparing the sample intrinsic means.

7.5 Other Fréchet Functions

Consider the general definition of Fréchet function as in equation (2.2.1) with ρ being the geodesic distance on Σ_2^k, that is

$$F(p) = \int_{\Sigma_2^k} d_g^\alpha(p, m) Q(dm).$$

In this section we investigate conditions for existence of a unique Fréchet mean.

Suppose the support of Q is contained in a convex geodesic ball $B(p, \frac{\pi}{4})$. Let $m \in B(p, \frac{\pi}{4})$. Let $\gamma(s)$ be a geodesic in $\bar{B}(p, \frac{\pi}{4})$. Then it is is easy to show that

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}s} d_g^\alpha(\gamma(s), m) &= \frac{\alpha}{2} d_g^{\alpha-2}(\gamma(s), m) \frac{\mathrm{d}}{\mathrm{d}s} d_g^2(\gamma(s), m), \\
\frac{\mathrm{d}^2}{\mathrm{d}s^2} d_g^\alpha(\gamma(s), m) &= \frac{\alpha}{2}(\frac{\alpha}{2} - 1) d_g^{\alpha-4}(\gamma(s), m) \frac{\mathrm{d}}{\mathrm{d}s} d_g^2(\gamma(s), m) \\
&+ \frac{\alpha}{2} d_g^{\alpha-2}(\gamma(s), m) \frac{\mathrm{d}^2}{\mathrm{d}s^2} d_g^2(\gamma(s), m).
\end{aligned}$$

We can get expressions for $\frac{\mathrm{d}}{\mathrm{d}s}d_g^2(\gamma(s), m)$ and $\frac{\mathrm{d}^2}{\mathrm{d}s^2}d_g^2(\gamma(s), m)$ from equations (7.4.1) and (7.4.2). For example when $\alpha = 3$,

$$\frac{\mathrm{d}}{\mathrm{d}s}d_g^3(\gamma(s), m) = -3d_g(\gamma(s), m)\langle Exp_{\gamma(s)}^{-1}m, \dot{\gamma}(s)\rangle$$
$$\frac{\mathrm{d}^2}{\mathrm{d}s^2}d_g^3(\gamma(s), m) = 3d^2\frac{\cos(d)}{\sin(d)}|\dot{\gamma}(s)|^2 + 3d^2\frac{\cos(d)}{\sin^3(d)}(\mathrm{Re}(z))^2 + \frac{3d^2}{\sin(d)\cos(d)}(\mathrm{Im}(z))^2$$

where $d = d_g(\gamma(s), m)$, $z = e^{i\theta}\tilde{m}'\overline{\dot{\tilde{\gamma}}(s)}$, $e^{i\theta} = \frac{\bar{\tilde{m}}'\tilde{\gamma}(s)}{\cos(d)}$, $m = \pi(\tilde{m})$ and $\gamma(s) = \pi(\tilde{\gamma}(s))$. The expression for $\frac{\mathrm{d}^2}{\mathrm{d}s^2}d_g^3(\gamma(s), m)$ is strictly positive if $m \neq \gamma(s)$. Hence the Fréchet function of Q is strictly convex in $B(p, \frac{\pi}{4})$ and hence has a unique minimizer which is called the (local) Fréchet mean of Q and denoted by μ_F. Replace Q by the empirical distribution Q_n to get the (local) sample Fréchet mean μ_{nF}. This proves the following theorem.

Theorem 7.5.1. *Suppose* $\mathrm{supp}(Q) \subseteq B(p, \frac{\pi}{4})$. *Consider the Fréchet function of* Q

$$F(q) = \int_{\Sigma_2^k} d_g^3(q, m)Q(dm).$$

Then (a) Q *has a unique (local) Fréchet mean* μ_F *in* $B(p, \frac{\pi}{4})$ *and if* μ_{nF} *denotes the (local) sample Fréchet mean from an iid random sample from* Q*, then (b)* $\sqrt{n}\phi(\mu_{nF})$ *has a asymptotic mean zero Normal distribution,* ϕ *being defined in Theorem 7.4.1.*

In Theorems 7.4.1 and 7.5.1, we differentiate the Fréchet function pointwise by constructing a geodesic variation. To construct this smooth geodesic variation, we required that the support of Q is contained in some convex ball. In case we differentiate the Fréchet function with respect to some coordinate chart, then it may be possible to extend Theorem 7.5.1 to show that there is a unique Fréchet mean even when Q has full support. Such an extension will be considered in a later article.

7.6 Extrinsic Analysis on the Planar Shape Space

For extrinsic analysis on the planar shape space, we embed it into the space $S(k, \mathbb{C})$ of all $k \times k$ complex Hermitian matrices. Here $S(k, \mathbb{C})$ is viewed as a (real) vector space with respect to the scaler field $\mathbb{R}$. The embedding is called the **Veronese-Whitney embedding** and is given by

$$\begin{aligned} J &: \Sigma_2^k \to S(k, \mathbb{C}), \\ J(\sigma(z)) &= J(\pi(u)) = uu^* \ (u = (u_1, \ldots, u_k)' \in \mathbb{C}S^{k-1}) \\ &= ((u_i \bar{u}_j))_{1 \le i,j \le k} \end{aligned}$$

where $u = \frac{z - \langle z \rangle}{\|\mathbf{z} - \langle z \rangle\|}$ is the preshape of the planar k-ad z. Define the extrinsic distance ρ on Σ_2^k by that induced from this embedding, namely,

$$\rho^2(\sigma(z), \sigma(w)) = \|uu^* - vv^*\|^2, \ \ u \doteq \frac{z - \langle z \rangle}{\|z - \langle z \rangle\|}\ , v \doteq \frac{w - \langle w \rangle}{\|w - \langle w \rangle\|}$$

where for arbitrary $k \times k$ complex matrices A and B,

$$\|A - B\|^2 = \sum_{j,j'} \|a_{jj'} - b_{jj'}\|^2 = \text{Trace}[(A - B)(A - B)^*]$$

is just the squared Euclidean distance between A and B regarded as elements of $\mathbb{C}^{k^2}$ (or $\mathbb{R}^{2k^2}$). Hence we get

$$\rho^2(\sigma(z), \sigma(w)) = 2(1 - |u^* v|^2).$$

The image of Σ_2^k under the Veronese-Whitney embedding is given by

$$J(\Sigma_2^k) = \{A \in S^+(k, \mathbb{C}) : \text{rank}(A) = 1, \ \text{Trace}(A) = 1, A\mathbf{1}_k = 0\}.$$

Here $S^+(k, \mathbb{C})$ is the space of all complex positive semidefinite matrices, "rank" denotes the complex rank and $\mathbf{1}_k$ is the k dimensional vector of all ones. Thus the image is a compact submanifold of $S(k, \mathbb{C})$ of (real) dimension $2k - 4$. Kendall (1984) shows that the embedding J is equivariant under the action of the special unitary group

$$SU(k) = \{A \in GL(k, \mathbb{C}) : AA^* = I, \det(A) = 1\}$$

which acts on the left: $A\pi(u) = \pi(Au)$. Indeed, then

$$J(A\pi(u)) = Auu^*A^* = \phi(A)J(\pi(u))$$

where

$$\phi(A) : S(k, \mathbb{C}) \to S(k, \mathbb{C}),\ \phi(A)B = ABA^*$$

is an isometry.

7.7 Extrinsic Mean and Variation

Let Q be a probability measure on the shape space Σ_2^k, let $X_1, X_2, \ldots, X_n$ be an iid sample from Q and let $\tilde{\mu}$ denote the mean vector of $\tilde{Q} \doteq Q \circ J^{-1}$, regarded as a probability measure on $\mathbb{C}^{k^2}$ (or $\mathbb{R}^{2k^2}$). Note that $\tilde{\mu}$ belongs to the convex hull of $\tilde{M} = J(\Sigma_2^k)$ and therefore is positive semidefinite and satisfies

$$\tilde{\mu}\mathbf{1}_k = 0,\ \text{Trace}(\tilde{\mu}) = 1,\ \text{rank}(\tilde{\mu}) \geq 1.$$

Let T be a matrix in $SU(k)$ such that

$$T\tilde{\mu}T^* = D = \text{Diag}(\lambda_1, \lambda_2, \ldots, \lambda_k),$$

where $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_k$ are the eigenvalues of $\tilde{\mu}$ in ascending order. Then, writing $v = Tu$ with $u \in \mathbb{C}S^{k-1}$, we get

$$\begin{aligned}
\|uu^* - \tilde{\mu}\|^2 = \|vv^* - D\|^2 &= \sum_{j=1}^{k}(|v_j|^2 - \lambda_j)^2 + \sum_{j \neq j'} |v_j \bar{v}_{j'}|^2 \\
&= \sum {\lambda_j}^2 + \sum_{j=1}^{k} |v_j|^4 - 2\sum_{j=1}^{k} \lambda_j |v_j|^2 + \sum_{j=1}^{k} |v_j|^2 . \sum_{j'=1}^{k} |v_{j'}|^2 - \sum_{j=1}^{k} |v_j|^4 \\
&= \sum {\lambda_j}^2 + 1 - 2\sum_{j=1}^{k} \lambda_j |v_j|^2
\end{aligned}$$

which is minimized (on $J(\Sigma_2^k)$) by taking $v = e_k = (0, \ldots, 0, 1)'$, i.e., $u = T^* e_k$- a unit eigenvector of $\tilde{\mu}$ having the largest eigenvalue λ_k. This implies that the projection set of $\tilde{\mu}$ on $\tilde{M}$, as defined in Section 4.2, consists of all $\mu\mu^*$ where μ is a unit eigenvector of $\tilde{\mu}$ corresponding to λ_k. The projection set is a singleton, in other words $\tilde{\mu}$ is a nonfocal point of $S(k, \mathbb{C})$, if and only if the eigenspace for the largest eigenvalue of $\tilde{\mu}$

is (complex) one dimensional, that is when $\lambda_k > \lambda_{k+1}$, or λ_k is a simple eigenvalue. Then Q has a unique extrinsic mean μ_E, say, which is given by $\mu_E = \pi(\mu)$. This is proved in Bhattacharya and Patrangenaru (2003).

If one writes $X_j = \pi(Z_j)$, $j = 1, 2, \ldots, n$ where Z_j is a preshape of X_j in $\mathbb{C}S^{k-1}$, then from Proposition 4.2.1 it follows that the extrinsic variation of Q has the expression

$$\begin{aligned} V &= \mathrm{E}\left[\|Z_1 Z_1^* - \tilde{\mu}\|^2\right] + \|\tilde{\mu} - \mu\mu^*\|^2 \\ &= 2(1 - \lambda_k). \end{aligned}$$

Therefore, we have the following consequence of Propositions 2.2.2 and 2.2.3.

Corollary 7.7.1. *Let μ_n denote a unit eigenvector of $\frac{1}{n}\sum_{j=1}^n Z_j Z_j^*$ having the largest eigenvalue λ_{kn}. (a) If the largest eigenvalue λ_k of $\tilde{\mu}$ is simple, then the sample extrinsic mean $\pi(\mu_n)$ is a strongly consistent estimator of the extrinsic mean $\pi(\mu)$ of Q. (b) The sample extrinsic variation $V_n = 2(1 - \lambda_{kn})$ is a strongly consistent estimator of the extrinsic variation $V = 2(1 - \lambda_k)$ of Q.*

7.8 Asymptotic Distribution of the Sample Extrinsic Mean

In this section, we assume that Q has a unique extrinsic mean $\mu_E = \pi(\mu)$ where μ is a unit eigenvector corresponding to the largest eigenvalue of the mean $\tilde{\mu}$ of $Q \circ J^{-1}$. To get the asymptotic distribution of the sample extrinsic mean μ_{nE} using Proposition 4.3.1, we need to differentiate the projection map

$$P : S(k, \mathbb{C}) \to J(\Sigma_2^k),\ \ P(\tilde{\mu}) = \mu\mu^*$$

in a neighborhood of a nonfocal point such as $\tilde{\mu}$. We consider $S(k, \mathbb{C})$ as a linear subspace of $\mathbb{C}^{k^2}$ (over $\mathbb{R}$) and as such a regular submanifold of $\mathbb{C}^{k^2}$ embedded by the inclusion map, and inheriting the metric tensor

$$\langle A, B\rangle = \mathrm{Re}\left(\mathrm{Trace}(AB^*)\right).$$

The (real) dimension of $S(k,\mathbb{C})$ is k^2. An orthonormal basis for $S(k,\mathbb{C})$ is given by $\{v_b^a : 1 \le a \le b \le k\}$ and $\{w_b^a : 1 \le a < b \le k\}$, defined as

$$v_b^a = \begin{cases} \frac{1}{\sqrt{2}}(e_a e_b^t + e_b e_a^t), & a < b \\ e_a e_a^t, & a = b \end{cases}$$

$$w_b^a = +\frac{i}{\sqrt{2}}(e_a e_b^t - e_b e_a^t),\ a < b$$

where $\{e_a : 1 \le a \le k\}$ is the standard canonical basis for $\mathbb{R}^k$. One can also take $\{v_b^a : 1 \le a \le b \le k\}$ and $\{w_b^a : 1 \le a < b \le k\}$ as the (constant) orthonormal frame for $S(k,\mathbb{C})$. For any $U \in SU(k)$ ($UU^* = U^*U = I$, $\det(U)$=+1), $\{Uv_b^aU^* : 1 \le a \le b \le k\}$, $\{Uw_b^aU^* : 1 \le a < b \le k\}$ is also an orthonormal frame for $S(k,\mathbb{C})$. We view $\mathrm{d}_{\tilde{\mu}}P : S(k,\mathbb{C}) \to T_{P(\tilde{\mu})}J(\Sigma_2^k)$. Choose $U \in SU(k)$ such that $U^*\tilde{\mu}U = D$,

$$U = (U_1, \ldots, U_k) \text{ and } D = \mathrm{Diag}(\lambda_1, \ldots, \lambda_k).$$

Here $\lambda_1 \le \ldots \le \lambda_{k-1} < \lambda_k$ are the eigenvalues of $\tilde{\mu}$ and $U_1, \ldots, U_k$ are corresponding eigenvectors. Choose the orthonormal basis frame $\{Uv_b^aU^*, Uw_b^aU^*\}$ for $S(k,\mathbb{C})$. Then it can be shown that

$$\mathrm{d}_{\tilde{\mu}}P(Uv_b^aU^*) = \begin{cases} 0 & \text{if } 1 \le a \le b < k,\ \text{a = b = k}, \\ (\lambda_k - \lambda_a)^{-1} Uv_k^aU^* & \text{if } 1 \le a < k, b = k. \end{cases}$$

$$\mathrm{d}_{\tilde{\mu}}P(Uw_b^aU^*) = \begin{cases} 0 & \text{if } 1 \le a < b < k \\ (\lambda_k - \lambda_a)^{-1} Uw_k^aU^* & \text{if } 1 \le a < k, b = k. \end{cases} \tag{7.8.1}$$

The proof is similar to that for the real projective shape which is considered in Section 10.6. Let $\tilde{X}_j = J(X_j)$, $j = 1, 2, \ldots, n$, where $X_1, \ldots, X_n$ is an iid random sample from Q. Write

$$\begin{aligned} \tilde{X}_j - \tilde{\mu} &= \sum\sum_{1 \le a \le b \le k} \langle(\tilde{X}_j - \tilde{\mu}), Uv_b^aU^*\rangle Uv_b^aU^* \\ &+ \sum\sum_{1 \le a < b \le k} \langle(\tilde{X}_j - \tilde{\mu}), Uw_b^aU^*\rangle Uw_b^aU^*. \end{aligned} \tag{7.8.2}$$

Since $\tilde{X}_j \mathbf{1}_k = \tilde{\mu}\mathbf{1}_k = 0$, hence $\lambda_1 = 0$ and one can choose $U_1 = \alpha \mathbf{1}_k$ where $|\alpha| = 1/\sqrt{k}$. Therefore

$$\langle(\tilde{X}_j - \tilde{\mu}), Uv_b^1 U^*\rangle = \langle(\tilde{X}_j - \tilde{\mu}), Uw_b^1 U^*\rangle = 0,\ 1 \leq b \leq k.$$

Then from equations (7.8.1) and (7.8.2), it follows that

$$\begin{aligned}
&\mathrm{d}_{\tilde{\mu}} P(\tilde{X}_j - \tilde{\mu}) \\
&= \sum_{a=2}^{k-1} \langle(\tilde{X}_j - \tilde{\mu}), Uv_k^a U^*\rangle (\lambda_k - \lambda_a)^{-1} Uv_k^a U^* \\
&+ \sum_{a=2}^{k-1} \langle(\tilde{X}_j - \tilde{\mu}), Uw_k^a U^*\rangle (\lambda_k - \lambda_a)^{-1} Uw_k^a U^*. \\
&= \sum_{a=2}^{k-1} \sqrt{2}\mathrm{Re}(U_a^* \tilde{X}_j U_k)(\lambda_k - \lambda_a)^{-1} Uv_k^a U^* \\
&+ \sum_{a=2}^{k-1} \sqrt{2}\mathrm{Im}(U_a^* \tilde{X}_j U_k)(\lambda_k - \lambda_a)^{-1} Uw_k^a U^*.
\end{aligned} \tag{7.8.3}$$

From equation (7.8.3), it is easy to check that the vectors

$$\{Uv_k^a U^*, Uw_k^a U^* :\ a = 2, \ldots, k-1\} \tag{7.8.4}$$

form an orthonormal basis for $T_{P(\tilde{\mu})}\tilde{M}$. Further $\mathrm{d}_{\tilde{\mu}} P(\tilde{X}_j - \tilde{\mu})$ has coordinates

$$T_j(\tilde{\mu}) \equiv (T_j^1(\tilde{\mu}), \ldots, T_j^{2k-4}(\tilde{\mu}))'$$

with respect to this orthonormal basis, where

$$T_j^a(\tilde{\mu}) = \begin{cases} \sqrt{2}(\lambda_k - \lambda_a)^{-1}\mathrm{Re}(U_{a+1}^* \tilde{X}_j U_k) & \text{if } 1 \leq a \leq k-2, \\ \sqrt{2}(\lambda_k - \lambda_a)^{-1}\mathrm{Im}(U_{a-k+3}^* \tilde{X}_j U_k) & \text{if } k-1 \leq a \leq 2k-4. \end{cases} \tag{7.8.5}$$

It follows from Proposition 4.3.1 that

$$\sqrt{n}\bar{T} \xrightarrow{\mathcal{L}} N(0, \Sigma)$$

where $\Sigma = \mathrm{Cov}(T_1)$.

7.9 Two Sample Extrinsic Tests on the Planar Shape Space

Suppose Q_1 and Q_2 are two probability distributions on the planar shape space. Let $X_1, \ldots, X_{n1}$ and $Y_1, \ldots, Y_{n2}$ be two iid samples from Q_1 and Q_2 respectively that are mutually independent. One many detect differences between Q_1 and Q_2 by comparing the sample extrinsic mean shapes or the sample extrinsic variations. This puts us in the same set up as in Section 4.6.1.

To compare the extrinsic means, one may use the statistics T_1 or T_2 defined through equations (4.6.3) and (4.6.7) respectively. To get the expression for T_1, one needs to find the coordinates of $\mathrm{d}_{\hat{\mu}}P(\tilde{X}_j - \hat{\mu})$ and $\mathrm{d}_{\hat{\mu}}P(\tilde{Y}_j - \hat{\mu})$ which are obtained from equation (7.8.5) by replacing $\tilde{\mu}$ by $\hat{\mu}$. For the statistic T_2, which is

$$T_2 = L[P(\hat{\mu}_1) - P(\hat{\mu}_2)]'\hat{\Sigma}^{-1}L[P(\hat{\mu}_1) - P(\hat{\mu}_2)]$$

where

$$\hat{\Sigma} = \frac{1}{n_1}L_1\hat{\Sigma}_1L_1' + \frac{1}{n_2}L_2\hat{\Sigma}_2L_2', \tag{7.9.1}$$

we need expressions for the linear projections L, L_1 and L_2. With respect to the orthonormal basis in equation (7.8.4) for $T_{P(\tilde{\mu})}J(\Sigma_2^k)$, the linear projection $L(A)$ of a matrix $A \in S(k, \mathbb{R})$ on to $T_{P(\tilde{\mu})}J(\Sigma_2^k)$ has coordinates

$$\begin{aligned} L(A) &= \{\langle A, Uv_k^aU^*\rangle, \langle A, Uw_k^aU^*\rangle : \ a = 2, \ldots, k-1\} \\ &= \sqrt{2}\{\mathrm{Re}(U_a^*AU_k), \mathrm{Im}(U_a^*AU_k) : \ a = 2, \ldots, k-1\}. \end{aligned}$$

For $A_1, A_2 \in S(k, \mathbb{R})$, if we label the bases for $T_{P(A_i)}J(\Sigma_2^k)$ as $\{v_1^i, \ldots, v_d^i\}$, $i = 1, 2$, then it is easy to check that the linear projection matrix L_1 from $T_{P(A_1)}J(\Sigma_2^k)$ on to $T_{P(A_2)}J(\Sigma_2^k)$ is the $d \times d$ matrix with coordinates

$$(L_1)_{ab} = \langle v_a^2, v_b^1\rangle \ 1 \leq a, b \leq d.$$

When the sample sizes are smaller than the dimension d (see Section 7.10.2), the standard error $\hat{\Sigma}$ in equation (7.9.1) may be singular or close to singular. Then it becomes more effective to estimate it from bootstrap simulations. When the sample sizes are small, we can also perform a bootstrap test using the test statistic T_2^* defined in equation (4.6.8), which is

$$T_2^* = v^{*\prime}\Sigma^{*-1}v^*.$$

However due to not enough observations, Σ^* may be singular or close to singular in most simulations. Then we may compare only the first few principal scores of the coordinates of the means. If $d_1 < d$ is the number of principal scores that we want to compare, then the appropriate test statistic to be used is given by

$$T_{21} = L[P(\hat{\mu}_1) - P(\hat{\mu}_2)]'\hat{\Sigma}_{11}^{-1}L[P(\hat{\mu}_1) - P(\hat{\mu}_2)] \tag{7.9.2}$$

where $\hat{\Sigma} = U\Lambda U'$, $U = (U_1, \ldots, U_d) \in SO(d)$, $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_d)$, $\lambda_1 \geq \ldots \geq \lambda_d$ is a s.v.d. for $\hat{\Sigma}$ and

$$\hat{\Sigma}_{11}^{-1} \doteq \sum_{j=1}^{d_1} \lambda_j^{-1} U_j U_j'.$$

Then T_{21} has an asymptotic $\mathcal{X}_{d_1}^2$ distribution. We can construct its bootstrap analogue, say T_{21}^* and compare the first d_1 principal scores by a pivotal bootstrap test. Alternatively, we may use a nonpivotal bootstrap test statistic

$$T_2^{**} = w^{*\prime}\Sigma^{**-1}w^* \tag{7.9.3}$$

for comparing the mean shapes, where

$$w^* = L[\{P(\mu_1^*) - P(\hat{\mu}_1)\} - \{P(\mu_2^*) - P(\hat{\mu}_2)\}]$$

and Σ^{**} is the sample covariance of w^* values, estimated from the bootstrap resamples.

To compare the sample extrinsic variations, one may use the statistic T_3 defined through equation (4.6.12). If $\hat{\lambda}_i$ denotes the largest eigenvalue of $\hat{\mu}_i$, $i = 1, 2$, then

$$T_3 = 2\frac{\hat{\lambda}_2 - \hat{\lambda}_1}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}. \tag{7.9.4}$$

The bootstrap version of T_3 is given by

$$T_3^* = 2\frac{(\lambda_2^* - \hat{\lambda}_2) - (\lambda_1^* - \hat{\lambda}_1)}{\sqrt{\frac{s_1^{*2}}{n_1} + \frac{s_2^{*2}}{n_2}}}$$

where λ_i^* and s_i^* are the bootstrap analogues of $\hat{\lambda}_i$ and s_i, $i = 1, 2$, respectively.

7.10 Applications

In this section, we record the results of two sample tests carried out in the two examples from Section 7.3.

7.10.1 Gorilla Skulls

Consider the data on gorilla skull images from Section 7.3.1. There are 30 female and 29 male gorillas giving rise to two independent samples of sizes 30 and 29 respectively on Σ_2^k, $k = 8$. To detect difference in the shapes of skulls between the two sexes, one may compare the sample mean shapes or variations in shape.

Figure 7.3 shows the plots of the sample extrinsic means for the two sexes along with the pooled sample extrinsic mean. In fact, the Procrustes coordinates for the two means with respect to a preshape of the pooled sample extrinsic mean have been plotted. The coordinates are

$$\hat{\mu}_1 = \begin{array}{llllllll}(-0.37, & -0.33; & 0.35, & 0.28; & 0.09, & 0.35; & -0.00, & 0.24; \\ -0.17, & 0.00; & -0.28, & -0.30; & 0.05, & -0.24; & 0.32, & -0.01)\end{array}$$

$$\hat{\mu}_2 = \begin{array}{llllllll}(-0.36, & -0.35; & 0.35, & 0.27; & 0.11, & 0.34; & 0.02, & 0.26; \\ -0.18, & 0.01; & -0.29, & -0.32; & 0.05, & -0.22; & 0.30, & 0.01)\end{array}$$

$$\hat{\mu} = \begin{array}{llllllll}(-0.36, & -0.34; & 0.35, & 0.28; & 0.10, & 0.34; & 0.01, & 0.25; \\ -0.17, & 0.01; & -0.29, & -0.31; & 0.05, & -0.23; & 0.31, & 0.00)\end{array}$$

where $\hat{\mu}_i$, $i = 1, 2$, denotes the Procrustes coordinates of the extrinsic mean shapes for the female and male samples respectively, and $\hat{\mu}$ is a preshape of the pooled sample extrinsic mean. The x and y coordinates for each landmark are separated by comma, while the different landmarks are separated by semicolons. The sample intrinsic means are very close to their extrinsic counterparts, the geodesic distance between the intrinsic and extrinsic means being 5.54×10^{-7} for the female sample and 1.96×10^{-6} for the male sample.

The value of the two sample test statistic defined through equation (3.5.3) for comparing the intrinsic mean shapes and the asymptotic p-value for the chi-squared test are

$$T_{n1} = 391.63, \text{ p-value } = P(\mathcal{X}^2_{12} > 391.63) < 10^{-16}.$$

Hence we reject the null hypothesis that the two sexes have the same intrinsic mean shape.
The two sample test statistics defined through equations (4.6.3) and (4.6.7) for comparing the extrinsic mean shapes and the corresponding asymptotic p-values are

$$T_1 = 392.6, \text{ p-value } = P(\mathcal{X}^2_{12} > 392.6) < 10^{-16},$$
$$T_2 = 392.0585, \text{ p-value } < 10^{-16}.$$

Hence we reject the null hypothesis that the two sexes have the same extrinsic mean shape. We can also compare the mean shapes by pivotal bootstrap method using the test statistic T_2^* defined in equation (4.6.8). The p-value for the bootstrap test using 10^5 simulations turns out to be 0.

The sample extrinsic variations for the female and male samples are 0.0038 and 0.005 respectively. The value of the two sample test statistic in (7.9.4) for testing equality of extrinsic variations is 0.923, and the asymptotic p-value is

$$P(|Z| > 0.923) = 0.356 \text{ where } Z \sim N(0,1).$$

Hence we accept the null hypothesis that the two underlying distributions have the

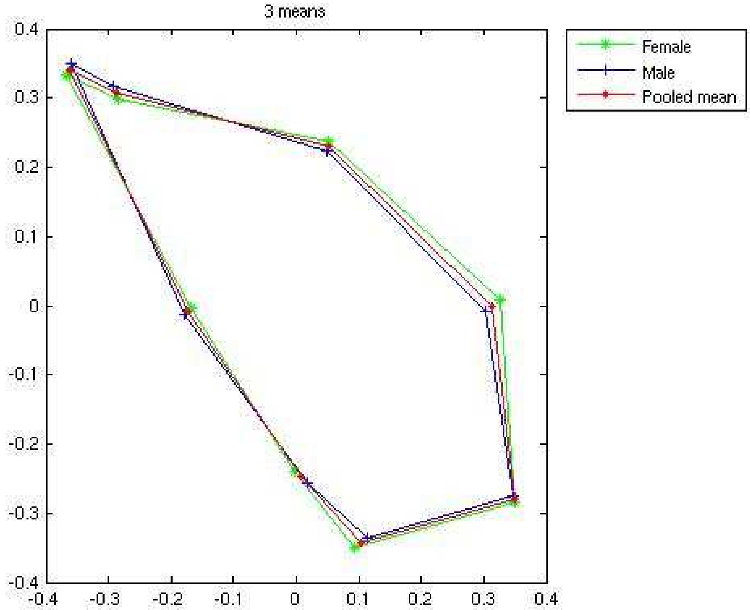


FIGURE 7.3. The sample extrinsic means for the 2 groups along with the pooled sample mean, corresponding to Figure 7.1.

same extrinsic variation. However since the mean shapes are different, it is possible to distinguish between the distribution of shapes for the two sexes.

7.10.2 Schizophrenia Detection

In this example from Section 7.3.2, we have two independent random samples of size 14 each on Σ_2^k, $k = 13$. To distinguish between the underlying distributions, we compare the mean shapes and shape variations.

Figure 7.4 shows the Procrustes coordinates of the sample extrinsic means for the two group of children along with a preshape for the pooled sample extrinsic mean. The coordinates for the two sample means have been obtained with respect to the pooled sample mean's preshape. The coordinates for the three means are

$$\hat{\mu}_1 = (0.14, 0.01; -0.22, 0.22; 0.01, 0.21; 0.31, 0.30; 0.24, -0.28; 0.15, -0.06; 0.06, -0.19; \\ -0.01, -0.33; -0.05, -0.04; -0.09, -0.19; -0.20, 0.02; -0.39, 0.32; 0.04, -0.00)$$

$$\hat{\mu}_2 = (0.16, 0.02; -0.22, 0.22; 0.02, 0.22; 0.31, 0.31; 0.24, -0.28; 0.15, -0.07; 0.06, -0.18; \\ -0.01, -0.33; -0.06, -0.04; -0.09, -0.20; -0.19, 0.03; -0.39, 0.30; 0.03, 0.00)$$

$$\hat{\mu} = (0.15, 0.01; -0.22, 0.22; 0.02, 0.22; 0.31, 0.30; 0.24, -0.28; 0.15, -0.06; 0.06, -0.19; \\ -0.01, -0.33; -0.05 - 0.04; -0.09, -0.19; -0.20, 0.03; -0.39, 0.310.03, 0.00)$$

Here $\hat{\mu}_i$, $i = 1, 2$, denotes the Procrustes coordinates of the extrinsic mean shape for the sample of normal and schizophrenic children respectively, and $\hat{\mu}$ is the preshape of the pooled sample extrinsic mean.

As in case of the gorilla skull images from the last section, the sample intrinsic means are very close to their extrinsic counterparts, the geodesic distance between the intrinsic and extrinsic means being 1.65×10^{-5} for the normal children sample and 4.29×10^{-5} for the sample of schizophrenic children.

The values of the two sample test statistic in equation (4.6.3) for testing equality of the population intrinsic mean shapes, along with the asymptotic p-values are

$$T_{n1} = 95.4587, \text{ p-value} = P(\mathcal{X}_{22}^2 > 95.4587) = 3.97 \times 10^{-11}.$$

The values of the two sample test statistics defined through equations (4.6.3) and (4.6.7) for comparing the extrinsic mean shapes and the corresponding asymptotic p-values are

$$T_1 = 95.5476, \text{ p-value } = P(\mathcal{X}_{22}^2 > 95.5476) = 3.8 \times 10^{-11},$$
$$T_2 = 95.2549, \text{ p-value } = 4.3048 \times 10^{-11}.$$

Hence we reject the null hypothesis that the two groups have the same mean shape (both extrinsic and intrinsic) at asymptotic levels greater than or equal to 10^{-10}.

TABLE 7.1. Percent of variation (P.V.) explained by different Principal Components (P.C.) of $\hat{\Sigma}$

P.C.	1	2	3	4	5	6	7	8	9	10	11
P.V.	21.6	18.4	12.1	10.0	9.9	6.3	5.3	3.6	3.0	2.5	2.1
P.C.	12	13	14	15	16	17	18	19	20	21	22
P.V.	1.5	1.0	0.7	0.5	0.5	0.3	0.2	0.2	0.1	0.1	0.0

Next we compare the extrinsic means by bootstrap methods. Since the dimension 22 of the underlying shape space is much higher than the sample sizes, it becomes difficult to construct a bootstrap test statistic as in the earlier section. That is because, the bootstrap estimate of the standard error $\hat{\Sigma}$ defined in equation (7.9.1) tends to be singular in most simulations. Hence we only compare the first few principal scores of the coordinates of the sample extrinsic means. Table 7.1 displays the percentage of variation explained by each principal component of $\hat{\Sigma}$. The value of T_{21} from equation (7.9.2) for comparing the first five principal scores of $L[P(\hat{\mu}_1) - P(\hat{\mu}_2)]$ with $\mathbf{0}$ and the asymptotic p-value are

$$T_{21} = 12.1872, \text{ p-value } = P(\mathcal{X}_5^2 > 12.1872) = 0.0323.$$

The bootstrap p-value from 10^4 simulations equals 0.0168 which is fairly small. When we use the nonpivotal bootstrap test statistic T_2^{**} from equation (7.9.3), the p-value for testing equality of the extrinsic mean shapes from 10^4 simulations equals 0. The value of T_2 with $\hat{\Sigma}$ replaced by its bootstrap estimate Σ^{**} equals 105.955 and the asymptotic p-value using $\mathcal{X}_{22}^2$ approximation is 5.7798×10^{-13}. Hence we again reject H_0 and conclude the extrinsic mean shapes are different.

Next we test equality of extrinsic variations for the two group of children. The sample extrinsic variations for patient and normal samples turn out to be 0.0107 and 0.0093 respectively. The value of the two sample test statistic in equation (7.9.4) for testing equality of population extrinsic variations is 0.9461 and the asymptotic p-value using standard Normal approximation is 0.3441. The bootstrap p-value with

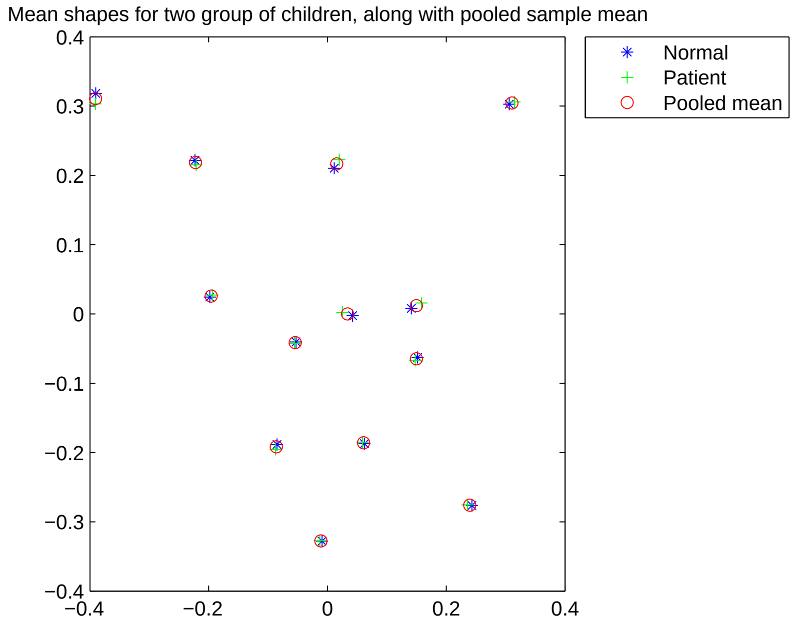


FIGURE 7.4. The sample extrinsic means for the 2 groups along with the pooled sample mean, corresponding to Figure 7.2.

10^4 simulations equals 0.3564. Hence we conclude at levels of significance less than or equal to 0.3 that the extrinsic variations in shapes for the two distributions are equal.

Since the mean shapes are different, we conclude that the probability distributions of the shapes of brain scans of normal and schizophrenic children are distinct.

Chapter 8

REFLECTION (SIMILARITY) SHAPE SPACES $R\Sigma_m^k$

8.1 Introduction

The **reflection (similarity) shape** of a k-ad as defined in Section 5.2.2, is its orbit under translation, scaling and all orthogonal transformations. Let $x = (x_1, \ldots, x_k)$ be a configuration of k points in $\mathbb{R}^m$, and let z denote its preshape in S_m^k as defined in equation (6.1.1). Then the reflection (similarity) shape of the k-ad x is given by the orbit

$$\sigma(x) = \sigma(z) = \{Az\colon A \in O(m)\} \tag{8.1.1}$$

where $O(m)$ is the group of all $m \times m$ orthogonal matrices (with determinants either +1 or -1). For the action of $O(m)$ on S_m^k to be free and the reflection shape space to be a Riemannian manifold, we consider only those shapes where the columns of z span $\mathbb{R}^m$. The set of all such z is called the nonsingular part of S_m^k and denoted by NS_m^k. Then the **reflection (similarity) shape space** is

$$R\Sigma_m^k = \{\sigma(z)\colon z \in S_m^k,\ \text{rank}(z) = m\} = NS_m^k/O(m) \tag{8.1.2}$$

which is a Riemannian manifold of dimension $km - m - 1 - m(m-1)/2$. Note that $R\Sigma_m^k = \Sigma_{0m}^k/G$ where $\Sigma_{0m}^k = NS_m^k/SO(m)$- a dense open subset of the similarity shape space Σ_m^k (see Section 6.2) and G is the group of reflections which maps a similarity shape to the shape of its reflected configuration. Since G is generated by a single element in $O(m)$ with determinant -1, therefore $R\Sigma_m^k$ is locally like Σ_{0m}^k with the same tangent space and Riemannian metric.

8.2 Extrinsic Analysis on the Reflection Shape Space

It has been shown that the map

$$J : R\Sigma_m^k \to S(k, \mathbb{R}),\ J(\sigma(z)) = z'z \tag{8.2.1}$$

is an embedding of the reflection shape space into $S(k, \mathbb{R})$ (see Bandulasiri and Patrangenaru (2005), Bandulasiri et al. (2008), and Dryden et al. (2008)). It induces the extrinsic distance

$$\begin{aligned}\rho_E^2(\sigma(z_1), \sigma(z_2)) &= \|J(\sigma(z_1)) - J(\sigma(z_2))\|^2 = \text{Trace}(z_1'z_1 - z_2'z_2)^2 \\ &= \text{Trace}(z_1z_1')^2 + \text{Trace}(z_2z_2')^2 - 2\text{Trace}(z_1z_2'z_2z_1'),\ z_1, z_2 \in S_m^k\end{aligned}$$

on $R\Sigma_m^k$. The embedding J is H-equivariant where $H = O(k)$ acts on the right: $A\sigma(z) \doteq \sigma(zA')$, $A \in O(k)$. Indeed, then

$$J(A\sigma(z)) = Azz'A' = \phi(A)J(\sigma(z))$$

where

$$\phi(A) : S(k, \mathbb{R}) \to S(k, \mathbb{R}),\ \phi(A)B = ABA'$$

is an isometry.

Define M_m^k as the set of all $k \times k$ positive semi-definite matrices of rank m and trace 1. Then the image of $R\Sigma_m^k$ under the embedding J in (8.2.1) is

$$J(R\Sigma_m^k) = \{A \in M_m^k : A\mathbf{1}_k = 0\}. \tag{8.2.2}$$

If we represent the preshape sphere S_m^k as in (6.1.3), then $M_m^k = J(R\Sigma_m^{k+1})$. Hence M_m^k is a submanifold (not complete) of $S(k, \mathbb{R})$ of dimension $km - 1 - m(m-1)/2$. Proposition 8.2.1 below identifies the tangent and normal spaces of M_m^k. The proof is taken from Bhattacharya (2008a).

Proposition 8.2.1. *Let $A \in M_m^k$. (a) The tangent space of M_m^k at A is given by*

$$T_A(M_m^k) = \{U \begin{pmatrix} T & S \\ S' & 0 \end{pmatrix} U' \colon T \in S(m, \mathbb{R}),\ \mathrm{Trace}(T) = 0\} \tag{8.2.3}$$

where $A = UDU'$ is a singular value decomposition (s.v.d.) of A, $U \in SO(m)$ and $D = \mathrm{Diag}(\lambda_1, \ldots, \lambda_k)$. (b) The orthocomplement of the tangent space in $S(k, \mathbb{R})$ or the normal space is given by

$$T_A(M_m^k)^\perp = \{U \begin{pmatrix} \lambda I_m & 0 \\ 0 & T \end{pmatrix} U' \colon \lambda \in \mathbb{R},\ T \in S(k-m, \mathbb{R})\} \tag{8.2.4}$$

Proof. Represent the preshape of a $(k+1)-$ad x by the $m \times k$ matrix z where $\|z\|^2 = \mathrm{Trace(zz')} = 1$ and let S_m^{k+1} be the preshape sphere,

$$S_m^{k+1} = \{z \in \mathbb{R}^{m \times k} \colon \|z\| = 1\}.$$

Let NS_m^{k+1} be the nonsingular part of S_m^{k+1}, i.e.,

$$NS_m^{k+1} = \{z \in S_m^{k+1} \colon \mathrm{rank(z) = m}\}.$$

Then $R\Sigma_m^{k+1} = NS_m^{k+1}/O(m)$ and $M_m^k = J(R\Sigma_m^{k+1})$. The map

$$J : R\Sigma_m^{k+1} \longrightarrow S(k, \mathbb{R}),\ J(\sigma(z)) = z'z = A$$

is a embedding. Hence

$$T_A(M_m^k) = dJ_{\sigma(z)}(T_{\sigma(z)} R\Sigma_m^{k+1}). \tag{8.2.5}$$

Since $R\Sigma_m^{k+1}$ is locally like Σ_{0m}^{k+1}, $T_{\sigma(z)} R\Sigma_m^{k+1}$ can be identified with the horizontal subspace H_z of $T_z S_m^{k+1}$ obtained in Section 6.2, which is

$$H_z = \{v \in R^{m \times k} \colon\ \mathrm{trace}(zv') = 0,\ zv' = vz'\}. \tag{8.2.6}$$

Consider the map

$$\tilde{J} : NS_m^{k+1} \to S(k, \mathbb{R}),\ \tilde{J}(z) = z'z. \tag{8.2.7}$$

Its derivative is a isomorphism between the horizontal subspace of $TNS_m^{k+1} \equiv TS_m^{k+1}$ and TM_m^k. The derivative is given by

$$d\tilde{J} : TS_m^{k+1} \to S(k, \mathbb{R}),\ d\tilde{J}_z(v) = z'v + v'z. \tag{8.2.8}$$

Hence

$$T_A M_m^k = d\tilde{J}_z(H_z) = \{z'v + v'z \colon v \in H_z\}. \tag{8.2.9}$$

From the description of H_z in equation (8.2.6), and using the fact that z has full row rank, it follows that

$$H_z = \{zv \colon v \in \mathbb{R}^{k\times k},\ \text{trace}(z'zv) = 0, zvz' \in S(m, \mathbb{R})\}. \tag{8.2.10}$$

From equations (8.2.9) and (8.2.10), we get that

$$T_A M_m^k = \{Av + v'A \colon AvA \in S(k, \mathbb{R}),\ \text{trace}(Av) = 0\}. \tag{8.2.11}$$

Let $A = UDU'$ be a s.v.d. of A as in the statement of the proposition. Using the fact that A has rank m, (8.2.11) can be written as

$$\begin{aligned} T_A M_m^k &= \{U(Dv + v'D)U' : DvD \in S(k, \mathbb{R}),\ \text{trace}(Dv) = 0\} \\ &= \{U \begin{pmatrix} T & S \\ S' & 0 \end{pmatrix} U' \colon T \in S(m, \mathbb{R}),\ \text{Trace}(T) = 0\}. \end{aligned} \tag{8.2.12}$$

This proves part (a). From the definition of orthocomplement and (8.2.12), we get that

$$\begin{aligned} T_A {M_m^k}^{\perp} &= \{v \in S(k, \mathbb{R}) \colon \text{trace}(v'w) = 0\ \forall\ w \in T_A M_m^k\} \\ &= \{U \begin{pmatrix} \lambda I_m & 0 \\ 0 & R \end{pmatrix} U' \colon \lambda \in \mathbb{R},\ R \in S(k - m, \mathbb{R})\} \end{aligned} \tag{8.2.13}$$

where I_m is the $m \times m$ identity matrix. This proves (b) and completes the proof. $\square$

For a $k \times k$ positive semi definite matrix μ with rank at least m, its projection on to M_m^k is defined as

$$P(\mu) = \{A \in M_m^k \colon \|\mu - A\|^2 = \underset{x \in M_m^k}{\text{argmin}} \|\mu - x\|^2\} \tag{8.2.14}$$

if this set is non empty. The following theorem, as proved in Bhattacharya (2008a), shows that the projection set is nonempty and derives formula for the projection matrices.

Theorem 8.2.2. *$P(\mu)$ is non empty and consists of*

$$A = \sum_{j=1}^{m} (\lambda_j - \bar{\lambda} + \frac{1}{m}) U_j U_j' \tag{8.2.15}$$

where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k$ are the ordered eigenvalues of μ; $U_1, U_2, \ldots, U_k$ are some corresponding orthonormal eigenvectors and $\bar{\lambda} = \frac{1}{m}\sum_{j=1}^{m} \lambda_j$.

Proof. Let

$$f(x) = \|\mu - x\|^2,\ x \in S(k, \mathbb{R}). \tag{8.2.16}$$

If f has a minimizer A in M_m^k then $(\text{grad } f)(A) \in T_A(M_m^k)^{\perp}$ where grad denotes the Euclidean derivative operator. But $(\text{grad } f)(A) = 2(A - \mu)$. Hence if A minimizes f, then

$$A - \mu = U^A \begin{pmatrix} \lambda I_m & 0 \\ 0 & T \end{pmatrix} U^{A\prime} \tag{8.2.17}$$

where $U^A = (U_1^A, U_2^A, \ldots, U_k^A)$ is a $k \times k$ matrix consisting of an orthonormal basis of eigenvectors of A corresponding to its ordered eigenvalues $\lambda_1^A \geq \lambda_2^A \geq \ldots \geq \lambda_m^A > 0 = \ldots = 0$. From (8.2.17) it follows that

$$\mu U_j^A = (\lambda_j^A - \lambda) U_j^A;\ j = 1, 2, \ldots, m. \tag{8.2.18}$$

Hence $\{\lambda_j^A - \lambda\}_{j=1}^m$ are eigenvalues of μ with $\{U_j^A\}_{j=1}^m$ as corresponding eigenvectors. Since these eigenvalues are ordered, this implies that there exists a singular value decomposition of μ: $\mu = \sum_{j=1}^k \lambda_j U_j U_j'$, and a set of indices $S = \{i_1, i_2, \ldots, i_m\}$, $1 \leq i_1 < i_2 < \ldots < i_m \leq k$ such that

$$\lambda_j^A - \lambda = \lambda_{i_j} \text{and} \tag{8.2.19}$$

$$U_j^A = U_{i_j},\ j = 1, \ldots, m. \tag{8.2.20}$$

Add the equations in (8.2.19) to get $\lambda = \frac{1}{m} - \bar{\lambda}$ where $\bar{\lambda} = \frac{\sum_{j\in S}\lambda_j}{m}$. Hence

$$A = \sum_{j\in S}(\lambda_j - \bar{\lambda} + \frac{1}{m})U_jU_j'. \tag{8.2.21}$$

Since $\sum_{j=1}^{k}\lambda_j = 1$, hence $\bar{\lambda} \leq 1/m$ and $\lambda_j - \bar{\lambda} + \frac{1}{m} > 0\ \forall j \in S$. So A is positive semi definite of rank m. It is easy to check that trace(A)=1 and hence $A \in M_m^k$. It can be shown that among the matrices A of the form (8.2.21), the function f defined in equation (8.2.16) is minimized when

$$S = \{1, 2, \ldots, m\}. \tag{8.2.22}$$

Define $M_{\leq m}^k$ as the set of all $k \times k$ positive semi-definite matrices of rank $\leq m$ and trace $= 1$. This is a compact subset of $S(k, \mathbb{R})$. Hence f restricted to $M_{\leq m}^k$ attains a minimum value. Let A_0 be a corresponding minimizer. If rank(A_0) $< m$, say $= m_1$, then A_0 minimizes f restricted to $M_{m_1}^k$. $M_{m_1}^k$ is a Riemannian manifold (it is $J(R\Sigma_{m_1}^{k+1})$). Hence A_0 must have the form

$$A_0 = \sum_{j=1}^{m_1}(\lambda_j - \bar{\lambda} + \frac{1}{m_1})U_jU_j' \tag{8.2.23}$$

where $\bar{\lambda} = \frac{\sum_{j=1}^{m_1}\lambda_j}{m_1}$. But if one defines

$$A = \sum_{j=1}^{m}(\lambda_j - \bar{\lambda} + \frac{1}{m})U_jU_j' \tag{8.2.24}$$

with $\bar{\lambda} = \frac{\sum_{j=1}^{m}\lambda_j}{m}$, then it is easy to check that $f(A) < f(A_0)$. Hence A_0 cannot be a minimizer of f over $M_{\leq m}^k$, that is, a minimizer must have rank $= m$. Then it lies in M_m^k and from equations (8.2.21) and (8.2.22), it follows that it has the form as in equation (8.2.24). This completes the proof. $\square$

Let Q be a probability distribution on $R\Sigma_m^k$ and let $\tilde{\mu}$ be the mean of $\tilde{Q} \equiv Q \circ J^{-1}$ in $S(k, \mathbb{R})$. Then $\tilde{\mu}$ is positive semi definite of rank at least m and $\tilde{\mu}\mathbf{1}_k = 0$. Theorem 8.2.2 can be used to get the formula for the extrinsic mean set of Q. This is obtained in Corollary 8.2.3.

Corollary 8.2.3. *(a) The projection of $\tilde{\mu}$ into $J(R\Sigma_m^k)$ is given by*

$$P_{J(R\Sigma_m^k)}(\tilde{\mu}) = \{A \colon A = \sum_{j=1}^{m}(\lambda_j - \bar{\lambda} + \frac{1}{m})U_j{U_j}'\} \tag{8.2.25}$$

where $\lambda_1 \geq \ldots \geq \lambda_k$ are the ordered eigenvalues of $\tilde{\mu}$, $U_1, \ldots, U_k$ are corresponding orthonormal eigenvectors and $\bar{\lambda} = \frac{\sum_{j=1}^m \lambda_j}{m}$. (b) The projection set is a singleton and Q has a unique extrinsic mean μ_E iff $\lambda_m > \lambda_{m+1}$. Then $\mu_E = \sigma(F)$ where $F = (F_1, \ldots, F_m)'$, $F_j = \sqrt{\lambda_j - \bar{\lambda} + \frac{1}{m}}U_j$.

Proof. Since $\tilde{\mu}\mathbf{1}_k = 0$, therefore $U_j'\mathbf{1}_k = 0 \; \forall j \leq m$. Hence any A in (8.2.25) lies in $J(R\Sigma_m^k)$ Now part (a) follows from Theorem 8.2.2 using the fact that $J(R\Sigma_m^k) \subseteq M_m^k$. For simplicity, let us denote $\lambda_j - \bar{\lambda} + \frac{1}{m}$, $j = 1, \ldots, m$ by λ_j^*. To prove part (b), note that if $\lambda_m = \lambda_{m+1}$, clearly $A_1 = \sum_{j=1}^m \lambda_j^* U_j {U_j}'$ and $A_2 = \sum_{j=1}^{m-1} \lambda_j^* U_j {U_j}' + \lambda_m^* U_{m+1} U_{m+1}'$ are two distinct elements in the projection set of (8.2.25). Consider next the case when $\lambda_m > \lambda_{m+1}$. Let $\tilde{\mu} = U\Lambda U' = V\Lambda V'$ be two different s.v.d. of $\tilde{\mu}$. Then $U'V$ consists of orthonormal eigenvectors of $\Lambda = \text{Diag}(\lambda_1, \ldots, \lambda_k)$. The condition $\lambda_m > \lambda_{m+1}$ implies that

$$U'V = \begin{pmatrix} V_{11} & 0 \\ 0 & V_{22} \end{pmatrix} \tag{8.2.26}$$

where $V_{11} \in SO(m)$ and $V_{22} \in SO(k-m)$. Write

$$\Lambda = \begin{pmatrix} \Lambda_{11} & 0 \\ 0 & \Lambda_{22} \end{pmatrix}.$$

Then $\Lambda U'V = U'V\Lambda$ implies $\Lambda_{11}V_{11} = V_{11}\Lambda_{11}$ and $\Lambda_{22}V_{22} = V_{22}\Lambda_{22}$. Hence

$$\begin{aligned}
&\sum_{j=1}^{m} \lambda_j^* V_j V_j' \\
=&U \sum_{j=1}^{m} \begin{pmatrix} \lambda_j^*(V_{11})_j(V_{11})_j' & 0 \\ 0 & 0 \end{pmatrix} U' \\
=&U \begin{pmatrix} \Lambda_{11} + (\frac{1}{m} - \bar{\lambda})I_m & 0 \\ 0 & 0 \end{pmatrix} U' \\
=&\sum_{j=1}^{m} \lambda_j^* U_j U_j'.
\end{aligned}$$

This proves that the projection set in (8.2.25) is a singleton when $\lambda_m > \lambda_{m+1}$. Then for any F in part (b) and A in the projection set of equation (8.2.25), $A = F'F = J(\sigma(F))$. This proves part (b) and completes the proof. □

From Proposition 4.2.1 and Corollary 8.2.3, it follows that the extrinsic variation of Q has the following expression:

$$V = \int_{J(R\Sigma_m^k)} \|x - \tilde{\mu}\|^2 \tilde{Q}(dx) + \|\tilde{\mu} - A\|^2,\ A \in P_{J(R\Sigma_m^k)}(\tilde{\mu}).$$

$$= \int_{J(R\Sigma_m^k)} \|x\|^2 \tilde{Q}(dx) + m(\frac{1}{m} - \bar{\lambda})^2 - \sum_{j=1}^{m} \lambda_j^2. \tag{8.2.27}$$

Remark 8.2.1. From the proof of Theorem 8.2.2 and Corollary 8.2.3, it follows that the extrinsic mean set C_Q of Q is also the extrinsic mean set of Q restricted to $M_{\leq m}^k$. Since $M_{\leq m}^k$ is a compact metric space, from Proposition 2.2.1, it follows that C_Q is compact. Let $X_1, X_2, \ldots, X_n$ be an iid sample from Q and let μ_{nE} and V_n be the sample extrinsic mean and variation. Then from Proposition 2.2.3, it follows that V_n is a consistent estimator of V. From Proposition 2.2.2, it follows that if Q has a unique extrinsic mean μ_E, then μ_{nE} is a consistent estimator of μ_E.

8.3 Asymptotic Distribution of the Sample Extrinsic Mean

Let $X_1, \ldots, X_n$ be an iid sample from some probability distribution Q on $R\Sigma_m^k$ and let μ_{nE} be the sample extrinsic mean (any measurable selection from the sample extrinsic mean set). In the last section, we saw that if Q has a unique extrinsic mean μ_E, that is, if the mean $\tilde{\mu}$ of $\tilde{Q} = Q \circ J^{-1}$ is a nonfocal point of $S(k, \mathbb{R})$, then μ_{nE} converges a.s. to μ_E as $n \to \infty$. Also from Proposition 4.3.1 it follows that if the projection map $P \equiv P_{J(R\Sigma_m^k)}$ is continuously differentiable at $\tilde{\mu}$, then $\sqrt{n}[J(\mu_{nE}) - J(\mu_E)]$ has asymptotic mean zero Gaussian distribution on $T_{J(\mu_E)} J(R\Sigma_m^k)$. To find the asymptotic dispersion, we need to compute the differential of P at $\tilde{\mu}$ (if it exists).

Consider first the map $P : N(\tilde{\mu}) \to S(k,\mathbb{R})$, $P(\mu) = \sum_{j=1}^{m}(\lambda_j(\mu) - \bar{\lambda}(\mu) + 1/m)U_j(\mu)U_j(\mu)'$ as in Theorem 8.2.2. Here $N(\tilde{\mu})$ is an open neighborhood of $\tilde{\mu}$ in $S(k,\mathbb{R})$ where P is defined. Hence for $\mu \in N(\tilde{\mu})$, $\lambda_m(\mu) > \lambda_{m+1}(\mu)$. It can be shown that P is smooth on $N(\tilde{\mu})$ (see Theorem 8.3.1). Let $\gamma(t) = \tilde{\mu} + tv$ be a curve in $N(\tilde{\mu})$ with $\gamma(0) = \tilde{\mu}$ and $\dot{\gamma}(0) = v \in S(k,\mathbb{R})$. Let $\tilde{\mu} = U\Lambda U'$, $U = (U_1,\ldots,U_k)$, $\Lambda = \text{Diag}(\lambda_1,\ldots,\lambda_k)$ be a s.v.d. of $\tilde{\mu}$ as in Corollary 8.2.3. Then

$$\gamma(t) = U(\Lambda + tU'vU)U' = U\tilde{\gamma}(t)U' \tag{8.3.1}$$

where $\tilde{\gamma}(t) = \Lambda + tU'vU$. Then $\tilde{\gamma}(t)$ is a curve in $S(k,\mathbb{R})$ starting at Λ. Say $\tilde{v} = \dot{\tilde{\gamma}}(0) = U'vU$. From equation (8.3.1) and from the definition of P, we get that

$$P[\gamma(t)] = UP[\tilde{\gamma}(t)]U'. \tag{8.3.2}$$

Differentiate equation (8.3.2) at $t = 0$, noting that $\frac{\mathrm{d}}{\mathrm{d}t}P[\gamma(t)]|_{t=0} = d_{\tilde{\mu}}P(v)$ and $\frac{\mathrm{d}}{\mathrm{d}t}P[\tilde{\gamma}(t)]|_{t=0} = d_{\Lambda}P(\tilde{v})$, to get

$$d_{\tilde{\mu}}P(v) = Ud_{\Lambda}P(\tilde{v})U'. \tag{8.3.3}$$

Let us find $\frac{\mathrm{d}}{\mathrm{d}t}P[\tilde{\gamma}(t)]|_{t=0}$. For that without loss of generality, we may assume that $\lambda_1 > \lambda_2 > \ldots > \lambda_k$. That is because, the set of all such matrices forms an open dense set of $S(k,\mathbb{R})$. Then we can choose a s.v.d. for $\tilde{\gamma}(t)$: $\tilde{\gamma}(t) = \sum_{j=1}^{k}\lambda_j(t)e_j(t)e_j(t)'$ such that $\{e_j(t),\lambda_j(t)\}_{j=1}^{k}$ are some smooth functions of t satisfying $e_j(0) = e_j$ and $\lambda_j(0) = \lambda_j$, where $\{e_j\}_{j=1}^{k}$ is the canonical basis for $\mathbb{R}^k$. Since $e_j(t)'e_j(t) = 1$, we get by differentiating,

$$e_j'\dot{e}_j(0) = 0,\ j = 1,\ldots,k. \tag{8.3.4}$$

Also since $\tilde{\gamma}(t)e_j(t) = \lambda_j(t)e_j(t)$, we get that

$$\tilde{v}e_j + \Lambda\dot{e}_j(0) = \lambda_j\dot{e}_j(0) + \dot{\lambda}_j(0)e_j,\ j = 1,\ldots,k. \tag{8.3.5}$$

Consider the orthonormal basis (frame) for $S(k,\mathbb{R})$: $\{E_{ab} : 1 \le a \le b \le k\}$ defined as

$$E_{ab} = \begin{cases} \frac{1}{\sqrt{2}}(e_a e_b^t + e_b e_a^t) & \text{if } a < b \\ e_a e_a^t & \text{if } a = b. \end{cases} \tag{8.3.6}$$

Let $\tilde{v} = E_{ab}$, $1 \le a \le b \le k$. From equations (8.3.4) and (8.3.5), we get that

$$\dot{e}_j(0) = \begin{cases} 0 & \text{if } a = b \text{ or } j \notin \{a, b\} \\ 2^{-1/2}(\lambda_a - \lambda_b)^{-1} e_b & \text{if } j = a < b \\ 2^{-1/2}(\lambda_b - \lambda_a)^{-1} e_a & \text{if } j = b > a. \end{cases} \tag{8.3.7}$$

and

$$\dot{\lambda}_j(0) = \begin{cases} 1 & \text{if } j = a = b \\ 0 & \text{o.w.} \end{cases} \tag{8.3.8}$$

Since

$$P[\tilde{\gamma}(t)] = \sum_{j=1}^{m} [\lambda_j(t) - \bar{\lambda}(t) + \frac{1}{m}] e_j(t) e_j(t)'$$

where $\bar{\lambda}(t) = \frac{1}{m}\sum_{j=1}^{m} \lambda_j(t)$, therefore

$$\begin{aligned} \dot{\bar{\lambda}}(0) &= \frac{1}{m}\sum_{j=1}^{m} \dot{\lambda}_j(0), \\ \frac{\mathrm{d}}{\mathrm{d}t} P[\tilde{\gamma}(t)]|_{t=0} &= \sum_{j=1}^{m} [\dot{\lambda}_j(0) - \dot{\bar{\lambda}}(0)] e_j e_j' \\ &\quad + \sum_{j=1}^{m} [\lambda_j - \bar{\lambda} + \frac{1}{m}][e_j \dot{e}_j(0)' + \dot{e}_j(0) e_j']. \end{aligned} \tag{8.3.9}$$

Take $\dot{\tilde{\gamma}}(0) = \tilde{v} = E_{ab}$, $1 \le a \le b \le k$ in equation (8.3.9). From equations (8.3.7) and (8.3.8), we get that

$$\frac{\mathrm{d}}{\mathrm{d}t} P[\tilde{\gamma}(t)]|_{t=0} = d_\Lambda P(E_{ab}) = \begin{cases} E_{ab} & \text{if } a < b \le m, \\ E_{aa} - \frac{1}{m}\sum_{j=1}^{m} E_{jj} & \text{if } a = b \le m, \\ (\lambda_a - \bar{\lambda} + \frac{1}{m})(\lambda_a - \lambda_b)^{-1} E_{ab} & \text{if } a \le m < b \le k, \\ 0 & \text{if } m < a \le b \le k. \end{cases} \tag{8.3.10}$$

Then from equations (8.3.3) and (8.3.10), we get that

$$d_{\tilde{\mu}} P(U E_{ab} U') = \begin{cases} U E_{ab} U' & \text{if } a < b \le m, \\ U\left(E_{aa} - \frac{1}{m}\sum_{j=1}^{m} E_{jj}\right) U' & \text{if } a = b \le m, \\ (\lambda_a - \bar{\lambda} + \frac{1}{m})(\lambda_a - \lambda_b)^{-1} U E_{ab} U' & \text{if } a \le m < b \le k, \\ 0 & \text{if } m < a \le b \le k. \end{cases} \tag{8.3.11}$$

From the description of the tangent space $T_{P(\tilde{\mu})}M_m^k$ in Proposition 8.2.1, it is clear that

$$d_{\tilde{\mu}}P(UE_{ab}U') \in T_{P(\tilde{\mu})}M_m^k \ \forall a \leq b.$$

Let us denote by

$$F_{ab} = UE_{ab}U',\ 1 \leq a \leq m, a < b \leq k, \tag{8.3.12}$$

$$F_a = UE_{aa}U',\ 1 \leq a \leq m. \tag{8.3.13}$$

Then from equation (8.3.11), we get that

$$d_{\tilde{\mu}}P(UE_{ab}U') = \begin{cases} F_{ab} & \text{if } 1 \leq a < b \leq m, \\ F_a - \bar{F} & \text{if } a = b \leq m, \\ \left(\lambda_a - \bar{\lambda} + \frac{1}{m}\right)(\lambda_a - \lambda_b)^{-1}F_{ab} & \text{if } 1 \leq a \leq m < b \leq k, \\ 0 & \text{o.w.} \end{cases} \tag{8.3.14}$$

where $\bar{F} = \frac{1}{m}\sum_{a=1}^m F_a$. Note that the vectors $\{F_{ab}, F_a\}$ in equations (8.3.12) and (8.3.13) are orthonormal and $\sum_{a=1}^m F_a - \bar{F} = 0$. Hence from equation (8.3.14), we conclude that the subspace spanned by

$$\{d_{\tilde{\mu}}P(UE_{ab}U'):\ 1 \leq a \leq b \leq k\}$$

has dimension

$$\frac{m(m-1)}{2} + m - 1 + m(k-m) = km - m - \frac{m(m-1)}{2}$$

which is the dimension of M_m^k. This proves that

$$T_{P(\tilde{\mu})}M_m^k = \text{Span}\{d_{\tilde{\mu}}P(UE_{ab}U')\}_{a \leq b}.$$

Consider the orthonormal basis $\{UE_{ab}U' : 1 \leq a \leq b \leq k\}$ of $S(k, \mathbb{R})$. Define

$$\tilde{F}_a = \sum_{j=1}^m H_{aj}F_j,\ 1 \leq a \leq m-1 \tag{8.3.15}$$

where H is a $(m-1) \times m$ Helmert matrix, that is $HH' = I_{m-1}$ and $H\mathbf{1}_m = 0$. Then the vectors $\{F_{ab}\}$ defined in equation (8.3.12) and $\{\tilde{F}_a\}$ defined in (8.3.15) together form an orthonormal basis of $T_{P\tilde{\mu}}M_m^k$. This is proved in Theorem 8.3.1. It is taken from Bhattacharya (2008a).

Theorem 8.3.1. *Let $\tilde{\mu}$ be a nonfocal point in $S(k,\mathbb{R})$. Let $\tilde{\mu} = U\Lambda U'$ be a s.v.d. of $\tilde{\mu}$. (a) The projection map $P : N(\tilde{\mu}) \to S(k,\mathbb{R})$ is smooth and its derivative $dP : S(k,\mathbb{R}) \to TM_m^k$ is given by equation (8.3.11). (b) The vectors (matrices) $\{F_{ab} : 1 \le a \le m, a < b \le k\}$ defined in equation (8.3.12) and $\{\tilde{F}_a : 1 \le a \le (m-1)\}$ defined in equation (8.3.15) together form an orthonormal basis of $T_{P(\tilde{\mu})}M_m^k$. (c) Let $A \in S(k,\mathbb{R}) \equiv T_{\tilde{\mu}}S(k,\mathbb{R})$ have coordinates $((a_{ij}))_{1\le i\le j\le k}$ with respect to the orthonormal basis $\{UE_{ij}U'\}$ of $S(k,\mathbb{R})$. That is,*

$$A = \sum\sum_{1\le i\le j\le k} a_{ij} UE_{ij}U',$$

$$a_{ij} = \langle A, UE_{ij}U'\rangle = \begin{cases} \sqrt{2}U_i'AU_j & \text{if } i<j \\ U_i'AU_i & \text{if } i=j. \end{cases}$$

Then $d_{\tilde{\mu}}P(A)$ has coordinates

$$a_{ij},\ 1 \le i < j \le m,$$

$$\tilde{a}_i,\ 1 \le i \le (m-1),$$

$$\left(\lambda_i - \bar{\lambda} + \frac{1}{m}\right)(\lambda_i - \lambda_j)^{-1} a_{ij},\ 1 \le i \le m < j \le k$$

w.r.t. the orthonormal basis $\{F_{ij} : 1 \le i < j \le m\}$, $\{\tilde{F}_i : 1 \le i \le (m-1)\}$ and $\{F_{ij} : 1 \le i \le m < j \le k\}$ of $T_{P(\tilde{\mu})}M_m^k$. Here

$$\mathbf{a} \equiv (a_{11}, a_{22}, \ldots, a_{mm})',$$

$$\tilde{\mathbf{a}} \equiv (\tilde{a}_1, \tilde{a}_2, \ldots, \tilde{a}_{m-1})' = H\mathbf{a}.$$

Proof. Let $\mu \in N(\tilde{\mu})$ have ordered eigenvalues $\lambda_1(\mu) \ge \lambda_2(\mu) \ge \ldots \ge \lambda_k(\mu)$ with corresponding orthonormal eigenvectors $U_1(\mu), U_2(\mu), \ldots, U_k(\mu)$. Then from Perturbation theory, it follows that if $\lambda_m(\mu) > \lambda_{m+1}(\mu)$, then

$$\mu \mapsto \text{Span}(U_1(\mu), \ldots, U_m(\mu)),\ \sum_{i=1}^{m} \lambda_i(\mu)$$

are smooth maps into their respective codomains (see Dunford and Schwartz (1958), p. 598). Write

$$P(\mu) = \sum_{j=1}^{m} \lambda_j(\mu) U_j(\mu) U_j(\mu)' + \left(\frac{1}{m} - \bar{\lambda}(\mu)\right) \sum_{j=1}^{m} U_j(\mu) U_j(\mu)'$$

Then $\sum_{j=1}^{m} U_j(\mu)U_j(\mu)'$ is the projection matrix of the subspace Span$(U_1(\mu), \ldots, U_m(\mu))$, which is a smooth function of μ. $\sum_{j=1}^{m} \lambda_j U_j(\mu)U_j(\mu)'$ is the projection of μ on the subspace Span$(U_1(\mu)U_1(\mu)', \ldots, U_m(\mu)U_m(\mu)')$ and hence is a smooth function of μ. Thus $\mu \mapsto P(\mu)$ is a smooth map on $N(\tilde{\mu})$. This proves part (a).

From equation (8.3.14), we conclude that $\{F_{ab} : 1 \leq a \leq m, a < b \leq k\}$ and $\{F_a - \bar{F} : 1 \leq a \leq m\}$ span $T_{P(\tilde{\mu})}M_m^k$. It is easy to check from the definition of H that Span$\{\tilde{F}_a : 1 \leq a \leq (m-1)\} =$ Span$\{F_a - \bar{F} : 1 \leq a \leq m\}$. Also since $\{F_a\}$ are mutually orthogonal, so are $\{\tilde{F}_a\}$. This proves that $\{F_{ab} : 1 \leq a \leq m, a < b \leq k\}$ and $\{\tilde{F}_a : 1 \leq a \leq (m-1)\}$ together form an orthonormal basis of $T_{P\tilde{\mu}}M_m^k$, which is claimed in part (b).

If $A = \sum\sum_{1 \leq i \leq j \leq k} a_{ij} U E_{ij} U'$, then

$$d_{\tilde{\mu}}P(A) = \sum\sum a_{ij} d_{\tilde{\mu}}P(UE_{ij}U') \tag{8.3.16}$$

$$= \sum\sum_{1 \leq i < j \leq m} a_{ij} F_{ij} + \sum_{i=1}^{m} a_{ii}(F_i - \bar{F}) + \sum_{i=1}^{m} \sum_{j=m+1}^{k} a_{ij}(\lambda_i - \bar{\lambda} + \frac{1}{m})(\lambda_i - \lambda_j)^{-1} F_{ij} \tag{8.3.17}$$

$$= \sum\sum_{1 \leq i < j \leq m} a_{ij} F_{ij} + \sum_{i=1}^{m-1} \tilde{a}_i \tilde{F}_i + \sum_{i=1}^{m} \sum_{j=m+1}^{k} (\lambda_i - \bar{\lambda} + \frac{1}{m})(\lambda_i - \lambda_j)^{-1} a_{ij} F_{ij}. \tag{8.3.18}$$

This proves part (c). To get (8.3.18) from (8.3.17), we use the fact that $\sum_{i=1}^{m} a_{ii}(F_i - \bar{F}) = \sum_{i=1}^{m-1} \tilde{a}_i \tilde{F}_i$. To show that, denote by F the matrix $(F_1, \ldots, F_m)$, by $F - \bar{F}$ the matrix $(F_1 - \bar{F}, \dots, F_m - \bar{F})$ and by $\tilde{F}$ the matrix $(\tilde{F}_1, \dots, \tilde{F}_{m-1})$. Then

$$\sum_{i=1}^{m-1} \tilde{a}_i \tilde{F}_i = \tilde{F}\tilde{\mathbf{a}}$$
$$= \tilde{F} H \mathbf{a} = F(I_m - \mathbf{1}_m \mathbf{1}_m')\mathbf{a}$$
$$= (F - \bar{F})\mathbf{a} = \sum_{i=1}^{m} a_{ii}(F_i - \bar{F}).$$

This completes the proof. □

Corollary 8.3.2. *Consider the projection map restricted to* $S_0(k, \mathbb{R}) \equiv \{A \in S(k, \mathbb{R}) \colon A\mathbf{1}_k = 0\}$. *Then its derivative is given by*

$$dP \colon S_0(k, \mathbb{R}) \to TJ(\mathbb{R}\Sigma_m^k),$$
$$d_{\tilde{\mu}}P(A) = \sum\sum_{1 \le i < j \le m} a_{ij}F_{ij} + \sum_{i=1}^{m-1} \tilde{a}_i \tilde{F}_i + \sum_{i=1}^{m} \sum_{j=m+1}^{k-1} (\lambda_i - \bar{\lambda} + \frac{1}{m})(\lambda_i - \lambda_j)^{-1} a_{ij} F_{ij}. \tag{8.3.19}$$

Hence $d_{\tilde{\mu}}P(A)$ *has coordinates* $\{a_{ij}, 1 \le i < j \le m\}$, $\{\tilde{a}_i, 1 \le i \le (m-1)\}$, $\{(\lambda_i - \bar{\lambda} + \frac{1}{m})(\lambda_i - \lambda_j)^{-1} a_{ij}, 1 \le i \le m < j < k\}$ *w.r.t. the orthonormal basis* $\{F_{ij} : 1 \le i < j \le m\}$, $\{\tilde{F}_i : 1 \le i \le (m-1)\}$ *and* $\{F_{ij} : 1 \le i \le m < j < k\}$ *of* $T_{P\tilde{\mu}}J(\mathbb{R}\Sigma_m^k)$.

Proof. Follows from the fact that

$$T_{P(\tilde{\mu})}J(\mathbb{R}\Sigma_m^k) = \{v \in T_{P(\tilde{\mu})}M_m^k \colon v\mathbf{1}_k = 0\}$$

and $\{F_{ij} \colon j = k\}$ lie in $T_{P(\tilde{\mu})}J(\mathbb{R}\Sigma_m^k)^{\perp}$. □

Consider the same set up as in Section 4.3. Let $\tilde{X}_j = J(X_j)$, $j = 1, \dots, n$ be the embedded sample in $J(R\Sigma_m^k)$. Let d be the dimension of $R\Sigma_m^k$. Let T_j, $j = 1, \dots, n$ be the coordinates of $d_{\tilde{\mu}}P(\tilde{X}_j - \tilde{\mu})$ in $T_{P(\tilde{\mu})}J(R\Sigma_m^k) \approx \mathbb{R}^d$. Then from equation (4.3.2) and Proposition 4.3.1, it follows that

$$\sqrt{n}[P(\overline{\tilde{X}}) - P(\tilde{\mu})] = \sqrt{n}\bar{T} + o_P(1) \xrightarrow{\mathcal{L}} N(0, \mathrm{Cov}(\mathrm{T}_1)).$$

We can get expression for T_j and hence $\bar{T}$ from Corollary 8.3.2 as follows. Define

$$
\begin{aligned}
(Y_j)_{ab} &= \begin{cases} \sqrt{2}U_a' Y_j U_b \text{ if } 1 \leq a < b \leq k, \\ U_a' Y_j U_a - \lambda_a \text{ if } a = b, \end{cases} \\
S_j &= H((Y_j)_{11}, (Y_j)_{22}, \ldots, (Y_j)_{mm})', \\
(T_j)_{ab} &= \begin{cases} (Y_j)_{ab} \text{ if } 1 \leq a < b \leq m, \\ (S_j)_a \text{ if } 1 \leq a = b \leq (m-1), \\ (\lambda_a - \bar{\lambda} + \frac{1}{m})(\lambda_a - \lambda_b)^{-1}(Y_j)_{ab} \text{ if } 1 \leq a \leq m < b < k. \end{cases}
\end{aligned}
\tag{8.3.20}
$$

Then $T_j \equiv ((T_j)_{ab})$ is the vector of coordinates of $d_{\tilde{\mu}}P(\tilde{X}_j - \tilde{\mu})$ in $\mathbb{R}^d$.

8.4 Two Sample Tests on the Reflection Shape Spaces

Now we are in the same set up as in Section 4.6: there are two samples on $R\Sigma_m^k$ and we want to test if they come from the same distribution, by comparing their sample extrinsic means and variations. To use the test statistic T_1 from equation (4.6.3) to compare the extrinsic means, we need the coordinates of $\{d_{\hat{\mu}}P(\tilde{X}_j - \hat{\mu})\}$ and $\{d_{\hat{\mu}}P(\tilde{Y}_j - \hat{\mu})\}$ in $T_{P(\hat{\mu})}J(R\Sigma_m^k)$. We get those from Corollary 8.3.2 as described in equation (8.3.20). To use the test statistic T_2 from equation (4.6.7), we need expressions for $L : S(k, \mathbb{R}) \to T_{P(\hat{\mu})}(JR\Sigma_m^k)$ and $L_i : T_{P(\hat{\mu}_i)}(JR\Sigma_m^k) \to T_{P(\hat{\mu})}(JR\Sigma_m^k)$, $i = 1, 2$. Let $\hat{\mu} = U\Lambda U'$ be a s.v.d. of $\hat{\mu}$. Consider the orthonormal basis $\{UE_{ij}U' : 1 \leq i \leq j \leq k\}$ of $S(k, \mathbb{R})$ and the orthonormal basis of $T_{P(\hat{\mu})}(JR\Sigma_m^k)$ derived in Corollary 8.3.2. Then if $A \in S(k, \mathbb{R})$ has coordinates $\{a_{ij}, 1 \leq i \leq j \leq k\}$, it is easy to show that $L(A)$ has coordinates $\{a_{ij}, 1 \leq i < j \leq m\}$, $\{\tilde{a}_i, 1 \leq i \leq m-1\}$ and $\{a_{ij}, 1 \leq i \leq m < j < k\}$ in $T_{P(\hat{\mu})}(JR\Sigma_m^k)$. If we label the bases of $T_{P(\hat{\mu}_i)}J(R\Sigma_m^k)$ as $\{v_1^i, \ldots, v_d^i\}$, $i = 1, 2$ and that of $T_{P(\hat{\mu})}J(R\Sigma_m^k)$ as $\{v_1, \ldots, v_d\}$, then one can show that L_i is the $d \times d$ matrix with coordinates

$$(L_i)_{ab} = \langle v_a, v_b^i \rangle \; 1 \leq a, b \leq d, \; i = 1, 2.$$

8.5 Other distances on the Reflection Shape Spaces

In this section, we introduce some distances other than the extrinsic distance on $R\Sigma_m^k$ which can be used to construct appropriate Fréchet functions and hence Fréchet mean and variation.

8.5.1 Full Procrustes Distance

Given two k-ads X_1 and X_2 in $\mathbb{R}^{m\times k}$, we define the **full Procrustes distance** between their reflection shapes as

$$d_F(\sigma(X_1), \sigma(X_2)) = \inf_{\Gamma\in O(m), \beta\in\mathbb{R}^+} \|Z_2 - \beta\Gamma Z_1\| \tag{8.5.1}$$

where Z_1 and Z_2 are the preshapes of X_1 and X_2 respectively. By a proof similar to that of Result 4.1 in Dryden and Mardia (1998), it can be shown that

$$d_F(X_1, X_2) = [1 - (\sum_{i=1}^{m} \lambda_i)^2]^{1/2}$$

and the values of Γ and β for which the infimum in equation (8.5.1) is attained are

$$\hat{\Gamma} = VU', \;\; \hat{\beta} = \sum_{i=1}^{m} \lambda_i.$$

Here $Z_1 Z_2' = U\Lambda V'$ is the singular value decomposition of $Z_1 Z_2'$, i.e. $U, V \in O(m)$ and

$$\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_m), \;\; \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m \geq 0.$$

The quantity $\hat{\beta}\hat{\Gamma}Z_1$ is called the **full Procrustes coordinates** of the shape of Z_1 with respect to that of Z_2.

8.5.2 Partial Procrustes Distance

Now define the **partial Procrustes distance** between the shapes of X_1 and X_2 as

$$d_P(\sigma(X_1), \sigma(X_2)) = \inf_{\Gamma\in O(m)} \|Z_2 - \Gamma Z_1\| \tag{8.5.2}$$

which is

$$d_P(X_1, X_2) = \sqrt{2}(1 - \sum_{i=1}^{m} \lambda_i)^{1/2}.$$

The value $\hat{\Gamma}$ of Γ for which the infimum in equation (8.5.2) is attained is the same as in Section 8.5.1. The quantity $\hat{\Gamma} Z_1$ is called the **partial Procrustes coordinates** of the shape of Z_1 with respect to that of Z_2.

8.5.3 Geodesic Distance

We saw in Section 8.1 that $R\Sigma_m^k = NS_m^k/O(m)$. Therefore the geodesic distance between the shapes of two k-ads X_1 and X_2 is given by

$$d_g(\sigma(X_1), \sigma(X_2)) = d_g(\sigma(Z_1), \sigma(Z_2)) = \inf_{\Gamma \in O(m)} d_{gs}(Z_1, \Gamma Z_2). \tag{8.5.3}$$

Here Z_1 and Z_2 are the preshapes of X_1 and X_2 respectively in the unit sphere S_m^k and $d_{gs}(.,.)$ denotes the geodesic distance on S_m^k which is given by

$$d_{gs}(Z_1, Z_2) = \arccos(\text{Trace}(Z_1 Z_2')).$$

Therefore

$$d_g(\sigma(X_1), \sigma(X_2)) = \inf_{\Gamma \in O(m)} \arccos(\text{Trace}(Z_1 Z_2')) = \arccos(\max_{\Gamma \in O(m)} (\text{Trace}(\Gamma Z_1 Z_2'))).$$

Let $Z_1 Z_2' = U \Lambda V$ be the singular value decomposition of $Z_1 Z_2'$, that is, $U, V \in O(m)$ and

$$\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_m),\ \lambda_1 \geq \lambda_2 \geq \ldots \lambda_m \geq 0.$$

Then

$$\begin{aligned} \text{Trace}(\Gamma Z_1 Z_2') &= \text{Trace}(\Gamma U \Lambda V) = \text{Trace}(V \Gamma U \Lambda) \\ &= \sum_{j=1}^{m} \lambda_j (V \Gamma U)_{jj}. \end{aligned}$$

This is maximized when $V\Gamma U = I_m$ or $\Gamma = V'U'$ and then

$$\text{Trace}(\Gamma Z_1 Z_2') = \sum_{j=1}^{m} \lambda_j.$$

Therefore the geodesic distance is

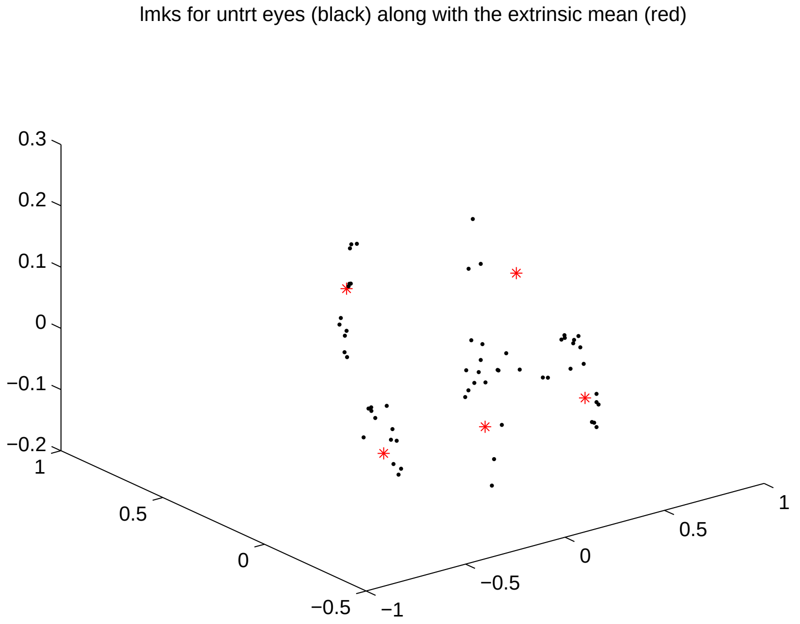
$$d_g(\sigma(X_1), \sigma(X_2)) = \arccos(\sum_{j=1}^{m} \lambda_j).$$

8.6 Application: Glaucoma Detection

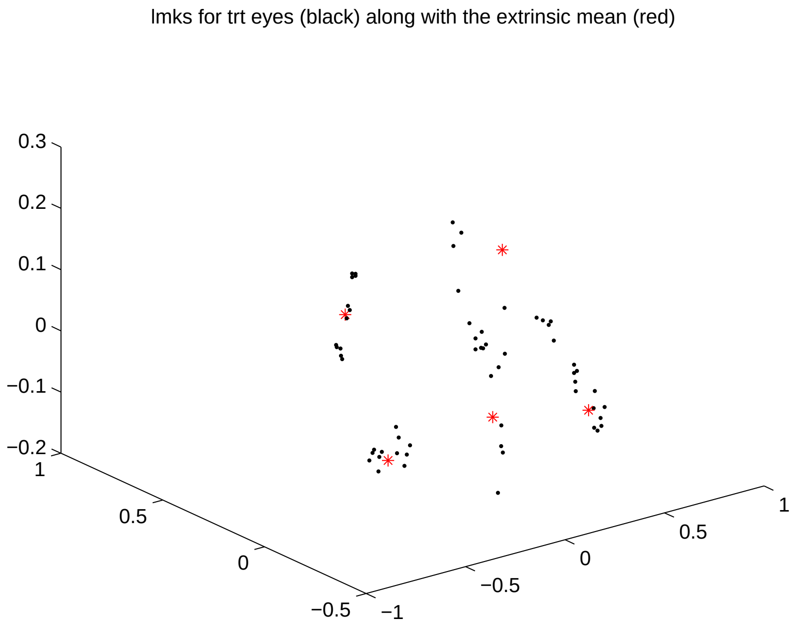
In this section, we see an application of 3D similarity shape analysis in disease detection.

Glaucoma is a leading cause of eye blindness. To detect any shape change due to Glaucoma, 3D images of the Optic Nerve Head (ONH) of both eyes of 12 mature rhesus monkeys were collected. One of the eyes was treated to increase the Intra Ocular Pressure (IOP) which is often the case of glaucoma onset, while the other was left untreated. Five landmarks were recorded on each eye. For details on landmark registration, see Derado et al. (2004). The landmark coordinates can be found in Bhattacharya and Patrangenaru (2005). In this section, we consider the reflection shape of the k-ads in $R\Sigma_3^k$, $k = 5$. We want to test if there is any significant difference between the shapes of the treated and untreated eyes by comparing the extrinsic means and variations. The analysis is carried out in Bhattacharya (2008a).

Figure 8.1(a) shows the partial Procrustes coordinates of the untreated eyes' shapes along with a preshape of the untreated eye sample extrinsic mean. Figure 8.1(b) shows the coordinates for the treated eyes' shapes along with a preshape of the treated eye sample extrinsic mean. In both cases the Procrustes coordinates are obtained with respect to the respective sample means. Figure 8.2 shows the Procrustes



(a)



(b)

FIGURE 8.1. (a) and (b) show 5 landmarks from untreated and treated eyes of 12 monkeys respectively, along with the mean shapes. * correspond to the mean shapes' landmarks.

coordinates of the mean shapes for the two eyes along with a preshape of the pooled sample extrinsic mean. Here the coordinates are with respect to the preshape of the pooled sample extrinsic mean. The sample extrinsic means have coordinates

$$L[P(\hat{\mu}_1) - P(\hat{\mu})] = (0.003, -0.011, -0.04, 0.021, 0.001, -0.001, 0.007, -0.004),$$
$$L[P(\hat{\mu}_2) - P(\hat{\mu})] = (-0.003, 0.011, 0.04, -0.021, -0.001, 0.001, -0.007, 0.005)$$

in the tangent space of $P(\hat{\mu})$. Here $P(\hat{\mu}_1)$ and $P(\hat{\mu}_2)$ are the embeddings of the sample extrinsic mean shapes of the untreated and treated eyes respectively, $P(\hat{\mu})$ is the embedded extrinsic mean shape for the pooled sample and L denotes the linear projection on to $T_{P(\hat{\mu})}J(R\Sigma_3^5)$. The sample extrinsic variations for the untreated and treated eyes are 0.041 and 0.038 respectively.

This is an example of a matched paired sample. To compare the extrinsic means and variations, we use the methodology of Section 4.6.2. The value of the matched pair test statistic T_{1p} in equation (4.6.15) is 36.29 and the asymptotic p-value for testing if the shape distributions for the two eyes are the same is

$$P(\mathcal{X}_8^2 > 36.29) = 1.55 \times 10^{-5}.$$

The value of the test statistic T_{2p} from equation (4.6.18) for testing whether the extrinsic means are the same is 36.56 and the p-value of the chi-squared test turns out to be 1.38×10^{-5}. Hence we conclude at asymptotic level 0.0001 or higher that the mean shapes of the two eyes are significantly different. Because of lack of sufficient data and high dimension, the bootstrap estimates of the covariance matrix $\hat{\Sigma}$ in (4.6.19) turn out to be singular or close to singular in many simulations. To avoid that, we construct a pivotal bootstrap confidence region for the first few principal scores of $L_{\tilde{\mu}}[P(\mu_1) - P(\mu_2)]$ and see if it includes $\mathbf{0}$. Here $P(\mu_i)$ is the embedding of the extrinsic mean of Q_i, $i = 1, 2$ (see Section 4.6.2) and $\tilde{\mu} = (\mu_1 + \mu_2)/2$. The first two principal components of $\hat{\Sigma}$ explain more than 80% of its variation. A bootstrap confidence region for the first two principal scores is given by the set

$$\{nT_n'\hat{\Sigma}_{11}^{-1}T_n \leq c^*(1-\alpha)\} \text{ where} \tag{8.6.1}$$

$$T_n = L[P(\hat{\mu}_1) - P(\hat{\mu}_2) - P(\mu_1) + P(\mu_2)]. \tag{8.6.2}$$

Here $n = 12$ is the sample size and $c^*(1-\alpha)$ is the upper $(1-\alpha)$-quantile of the bootstrap distribution of $nv^*\Sigma_{11}^{*-1}v^*$, v^* being defined in equation (4.6.9). If $\hat{\Sigma} = \sum_{j=1}^{8} \lambda_j U_j U_j'$ is a s.v.d. for $\hat{\Sigma}$, then $\hat{\Sigma}_{11}^{-1} \doteq \sum_{j=1}^{2} \lambda_j^{-1} U_j U_j'$ and Σ_{11}^{*-1} is its bootstrap estimate. The bootstrap p-value with 10^4 simulations turns out to be 0.0098. Hence we again reject $H_0 : P(\mu_1) = P(\mu_2)$. The corresponding p-value using $\mathcal{X}_2^2$ approximation for the distribution of $nT_n'\hat{\Sigma}_{11}^{-1}T_n$ in (8.6.1) turns out to be 0.002. It may be noted that the p-values are much smaller than those obtained by different methods in Bhattacharya and Patrangenaru (2005) and Bandulasiri et al. (2008).

Next we test if the two eye shapes have the same extrinsic variation. The value of the test statistic T_{3p} from equation (4.6.21) equals -0.5572 and the asymptotic p-value equals

$$P(|Z| > 0.5572) = 0.577, \; Z \sim N(0,1).$$

The bootstrap p-value with 10^4 simulations equals 0.59. Hence we accept H_0 and conclude that the extrinsic variations are equal at levels 0.5 or lower.

Since the mean shapes for the two eyes are found to be different, we conclude that the underlying probability distributions are distinct and hence Glaucoma indeed changes the shape of the eyes.

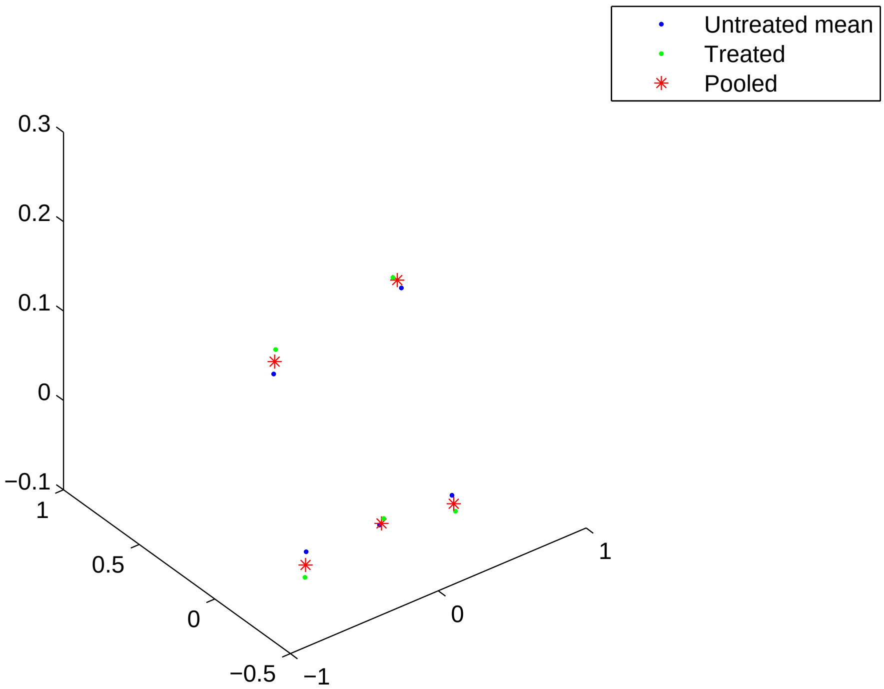


FIGURE 8.2. Extrinsic mean shapes for the 2 eyes along with the pooled sample extrinsic mean.

Chapter 9

AFFINE SHAPE SPACES $A\Sigma_m^k$

9.1 Introduction

The **affine shape** of a k-ad x with landmarks in $\mathbb{R}^m$ may be defined as the orbit of this k-ad under the group of all **affine transformations**

$$x \mapsto F(x) = Ax + b,$$

where A is an arbitrary $m \times m$ nonsingular matrix and b is an arbitrary point in $\mathbb{R}^m$. Then the **affine shape space** $A\Sigma_m^k$ may be defined as the collection of all affine shapes, that is

$$A\Sigma_m^k = \{\sigma(x) : \ x \in \mathbb{R}^{m \times k}\} \text{ where}$$
$$\sigma(x) = \{Ax + b : \ A \in GL(m, \mathbb{R}), \ b \in \mathbb{R}^m\}$$

and $GL(m, \mathbb{R})$ is the general linear group on $\mathbb{R}^m$ of all $m \times m$ nonsingular matrices. Note that two k-ads $x = (x_1, \dots, x_k)$ and $y = (y_1, \dots, y_k)$, $(x_j, \ y_j \in \mathbb{R}^m$ for all $j)$ have the same affine shape if and only if the corresponding centered k-ads $u = (u_1, u_2, \dots, u_k) = (x_1 - \bar{x}, \dots, x_k - \bar{x})$ and $v = (v_1, v_2, \dots, v_k) = (y_1 - \bar{y}, \dots, y_k - \bar{y})$ are related by a transformation

$$Au \doteq (Au_1, \dots, Au_k) = v, \ A \in GL(m, \mathbb{R}).$$

The centered k-ads lie in a linear subspace of $\mathbb{R}^{m \times k}$, call it $H(m, k)$ which is

$$H(m, k) = \{u \in \mathbb{R}^{m \times k} : \ \sum_{j=1}^{k} u_j = 0\}.$$

Hence $A\Sigma_m^k$ can be represented as the quotient of this subspace under all general linear transformations, that is

$$A\Sigma_m^k = H(m, k)/GL(m, \mathbb{R}).$$

The subspace $H(m,k)$ is an Euclidean manifold of dimension $m(k-1)$. The group $GL(m,\mathbb{R})$ has the relative topology (and distance) of $\mathbb{R}^{m^2}$ and hence is a manifold of dimension m^2. Assume $k > m+1$. For the action of $GL(m,\mathbb{R})$ on $H(m,k)$ to be free and the affine shape space to be a manifold, we require that the columns of u ($u \in H(m,k)$) span $\mathbb{R}^m$. Indeed the condition

$$Au = u \Leftrightarrow A = I_m$$

holds if and only if $\mathrm{rank}(u) = m$. Hence we consider only such centered k-ads u, that is

$$u \in H_0(m,k) \doteq \{v \in H(m,k): \ \mathrm{rank}(v) = m\}$$

and redefine the affine shape space as

$$A\Sigma_m^k = H_0(m,k)/GL(m,\mathbb{R}).$$

Then it follows that $A\Sigma_m^k$ is a manifold of dimension $m(k-1)-m^2$. To get rid of the linear constraint $\sum_{j=1}^k u_j = 0$ on $H(m,k)$, one may postmultiply u by a Helmert matrix H and consider the Helmertized k-ad $\tilde{u} = uH$ as in Section 6.1. Then $H(m,k)$ can be identified with $\mathbb{R}^{m(k-1)}$ and $H_0(m,k)$ is an open dense subset of $H(m,k)$.

For $u, v \in H_0(m,k)$, the condition $Au = v$ holds if and only if $u'A' = v'$ and as A varies over $GL(m,\mathbb{R})$, $u'A'$ generates the linear subspace L of $\mathbb{R}^{k-1}$ spanned by the m rows of u. The affine shape of u (or of the original k-ad x) can be identified with this subspace. Thus $A\Sigma_m^k$ may be identified with the set of all m dimensional subspaces of $\mathbb{R}^{k-1}$, namely, the **Grassmannian** $G_m(k-1)$, a result of Sparr (1992) (also see Boothby (1986)). This identification enables us to give a Riemannian structure to $A\Sigma_m^k$ and carry out an intrinsic analysis. This is discussed in Section 9.2.

To carry out an extrinsic analysis on $A\Sigma_m^k$, we embed it into the space of all $k \times k$ symmetric matrices $S(k,\mathbb{R})$ via an equivariant embedding. Then analytic expressions for the extrinsic mean and variation are available. This is the subject of Section 9.3. To get the asymptotic distribution of the sample extrinsic mean and carry out nonparametric inference on affine shapes, we need to differentiate the projection map of Proposition 9.3.1 which requires Perturbation theory arguments for eigenvalues and eigenvectors. This is carried out in Section 9.4.

Affine shape spaces arise in many problems of bioinformatics, cartography, machine vision and pattern recognition (see Berthilsson and Heyden (1999), Berthilsson and Astrom (1999), Sepiashvili et al. (2003), Sparr (1992)). We will see such an application in Section 9.5. The tools developed in Sections 9.3 and 9.4 are applied to this example to carry out an extrinsic analysis.

9.2 Geometry of Affine Shape Spaces

Consider a Helmertized k-ad x in $\mathbb{R}^{m\times(k-1)}$. Define its **special affine shape** as the orbit

$$s\sigma(x) = \{Ax : A \in GL(m, \mathbb{R}),\ \det(\mathrm{A}) > 0\}. \tag{9.2.1}$$

Any $A \in GL(m, \mathbb{R})$ has a **pseudo singular value decomposition** $A = U\Lambda V$ where $U, V \in SO(m)$ and

$$\Lambda = \mathrm{Diag}(\lambda_1, \ldots, \lambda_m),\ \lambda_1 \geq \ldots \geq \lambda_{m-1} \geq |\lambda_m|,\ \mathrm{sign}(\lambda_m) = \mathrm{sign}(\det(A)).$$

Therefore a linear transformation $x \mapsto Ax$ consists of a rotation and different amount of stretching in different directions followed by another rotation or reflection. When $\det(A) > 0$, that is, when we consider the special affine shape, we look at the affine shape without any reflections. We can get the affine shape $\sigma(x)$ of x from its special affine shape $s\sigma(x)$ by identifying $s\sigma(x)$ with $s\sigma(Tx)$ where $T \in O(m)$, $\det(T) = -1$. This T can be chosen to be any reflection matrix. Let the **special affine shape**

space $S A\Sigma_m^k$ be the collection of all special affine shapes, which is

$$S A\Sigma_m^k = \{s\sigma(x) : x \in \mathbb{R}^{m\times(k-1)},\ \text{rank}(x) = m\}$$

We restrict to full rank k-ads so that the group action is free and $S A\Sigma_m^k$ is a manifold. From the expression of $s\sigma(x)$ in equation (9.2.1), it is clear that it is a function of the 'oriented' span of the rows of x, which in turn is a function of an orthogonal m-frame for the rowspace of x. In fact $S A\Sigma_m^k$ can be viewed as the quotient of $SO(k-1)$ as follows. Denote by $St_m(k)$ the **Steifel manifold** of all orthogonal m-frames in $\mathbb{R}^k$. For $V \in SO(k-1)$, write $V = \begin{pmatrix} V_1 \\ V_2 \end{pmatrix}$ where $V_1 \in St_m(k-1)$, $V_2 \in St_{k-m-1}(k-1)$. Then the oriented span of the rows of V_1 which is the special affine shape of V_1 can be identified with the orbit

$$\pi(V) = \{\begin{pmatrix} AV_1 \\ BV_2 \end{pmatrix} : A \in SO(m),\ B \in SO(k-m-1)\} = \{\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} V\}. \quad (9.2.2)$$

This implies that

$$S A\Sigma_m^k = SO(k-1)/SO(m) \times SO(k-m-1).$$

Then $A\Sigma_m^k = S A\Sigma_m^k/G$ where G is a finite group generated by any $T \in SO(k-1)$ which looks like

$$T = \begin{pmatrix} T_1 & 0 \\ 0 & T_2 \end{pmatrix},\ T_1 \in O(m),\ T_2 \in O(k-m-1),\ \det(T_1) = \det(T_2) = -1.$$

This means that two elements V, W in $SO(k-1)$ have the same affine shape iff either $\pi(V) = \pi(W)$ or $\pi(TV) = \pi(W)$. Hence $A\Sigma_m^k$ is locally like $S A\Sigma_m^k$. Since $SO(m) \times SO(k-m-1)$ acts as isometries on $SO(k-1)$, therefore the map $\pi : SO(k-1) \to S A\Sigma_m^k$ in equation (9.2.2) is a Riemannian submersion. Then $S A\Sigma_m^k$ and hence $A\Sigma_m^k$ inherits the Riemannian metric tensor from $SO(k-1)$ making it a Riemannian manifold.

To derive expression for the tangent space of $S A\Sigma_m^k$ (or of $A\Sigma_m^k$), we need to identify the horizontal subspace of the tangent space of $SO(k-1)$. Then $d\pi$ provides an isometry between the horizontal subspace and the tangent space of $SA\Sigma_m^k$. We saw in Section 6.2 that geodesics in $SO(k-1)$ starting at $V \in SO(k-1)$ look like

$$\gamma(t) = \exp(tA)V$$

where $A \in Skew(k-1)$ $(A + A' = 0)$ and

$$\exp(B) = I + B + \frac{B^2}{2!} + \dots$$

This geodesic is vertical if it lies in the orbit $\pi(V)$. That is when

$$\gamma(t) = \begin{pmatrix} \exp(tA) & 0 \\ 0 & \exp(tB) \end{pmatrix} V$$

where $A \in Skew(m)$, $B \in Skew(k-m-1)$. Then

$$\dot{\gamma}(0) = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} V.$$

Therefore the vertical subspace of the tangent space $T_V SO(k-1)$ of $SO(k-1)$ at V has the form

$$Ver_V = \{\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} V \colon A + A' = 0,\ B + B' = 0\}.$$

The horizontal subspace H_V is its orthocomplement in $T_V SO(k-1)$ which is given by

$$\begin{aligned} H_V = \{AV :\ & A \in Skew(k-1),\ \text{Trace}\left(A \begin{pmatrix} B_1 & 0 \\ 0 & B_2 \end{pmatrix}\right) = 0\ \forall\ B_1 \in Skew(m), \\ & B_2 \in Skew(k-m-1)\} \\ = \{AV \colon\ & A = \begin{pmatrix} 0 & B \\ -B' & 0 \end{pmatrix},\ B \in \mathbb{R}^{m\times(k-m-1)}\}. \end{aligned}$$

Then

$$T_{\pi(V)} SA\Sigma_m^k = d\pi_V(H_V).$$

9.3 Extrinsic Analysis on Affine Shape Spaces

Let u be a centered k-ad in $H_0(m,k)$ and let $\sigma(u)$ denote its affine shape, which is the orbit

$$\sigma(u) = \{Au\colon A \in GL(m,\mathbb{R})\}.$$

Consider the map

$$J : A\Sigma_m^k \to S(k,\mathbb{R}),\; J(\sigma(u)) \equiv A = FF' \tag{9.3.1}$$

where $F = (f_1, f_2, \dots, f_m)$ is an orthonormal basis for the row space of u. It has been shown that J is an embedding of $A\Sigma_m^k$ into $S(k,\mathbb{R})$, equivariant under the action of $O(k)$ (see Dimitric (1996)). In (9.3.1), A is the projection (matrix) on to the subspace spanned by the rows of u. Hence through the embedding J, we identify a m-dimensional subspace of $\mathbb{R}^{k-1}$ with the projection map (matrix) on to that subspace. Since A is a projection matrix, it is characterized by

$$A^2 = A,\; A = A' \text{ and } \operatorname{trace}(A) = \operatorname{rank}(A) = m.$$

Also since u is a centered k-ad, that is, the rows of u are orthogonal to $\mathbf{1}_k$, therefore $A\mathbf{1}_k = 0$. Hence the image of $A\Sigma_m^k$ into $S(k,\mathbb{R})$ under the embedding J is given by

$$J(A\Sigma_m^k) = \{A \in S(k,\mathbb{R})\colon A^2 = A,\; \operatorname{trace}(A) = m,\; A\mathbf{1}_k = 0\} \tag{9.3.2}$$

which is a compact Riemannian submanifold of $S(k,\mathbb{R})$ of dimension $mk - m - m^2$. It is easy to show that $A = u'(uu')^{-1}u$.

Let Q be a probability distribution on $A\Sigma_m^k$ and let $\tilde{Q} = Q \circ J^{-1}$ be its image in $J(A\Sigma_m^k)$. Let $\tilde{\mu}$ be the mean of $\tilde{Q}$, that is $\tilde{\mu} = \int_{J(A\Sigma_m^k)} x\tilde{Q}(dx)$. Then $\tilde{\mu}$ is a $k \times k$ positive semi definite matrix satisfying

$$\operatorname{trace}(\tilde{\mu}) = m,\; \operatorname{rank}(\tilde{\mu}) \geq m \text{ and } \tilde{\mu}\mathbf{1}_k = 0.$$

To carry out an extrinsic analysis on $A\Sigma_m^k$, we need to identify the extrinsic mean (set) of Q which is the projection (set) of $\tilde{\mu}$ on $J(A\Sigma_m^k)$. Denote by $P(\tilde{\mu})$ the set of projections of $\tilde{\mu}$ on $J(A\Sigma_m^k)$, as defined in equation (4.2.3). Proposition 9.3.1 below gives an expression for $P(\tilde{\mu})$ and hence finds the extrinsic mean set of Q. It was first proved in Sughatadasa (2006). The proof below is constructed independently and is included here for the sake of completeness.

Proposition 9.3.1. *(a) The projection of $\tilde{\mu}$ into $J(A\Sigma_m^k)$ is given by*

$$P(\tilde{\mu}) = \{\sum_{j=1}^{m} U_j U_j'\} \tag{9.3.3}$$

where $U = (U_1, \ldots, U_k) \in SO(k)$ is such that $\tilde{\mu} = U\Lambda U'$, $\Lambda = \text{Diag}(\lambda_1, \ldots, \lambda_k)$, $\lambda_1 \geq \ldots \geq \lambda_k = 0$. (b) $\tilde{\mu}$ is nonfocal and Q has a unique extrinsic mean μ_E iff $\lambda_m > \lambda_{m+1}$. Then $\mu_E = \sigma(F')$ where $F = (U_1, \ldots, U_m)$.

Proof. From the definition of $P(\tilde{\mu})$, it follows that for any $A_0 \in P(\tilde{\mu})$,

$$\|\tilde{\mu} - A_0\|^2 = \min_{A \in J(A\Sigma_m^k)} \|\tilde{\mu} - A\|^2.$$

Here $\|.\|$ denotes the Euclidean norm which is

$$\|A\|^2 = \text{Trace}(AA'),\ A \in \mathbb{R}^{k\times k}.$$

Then for any $A \in J(A\Sigma_m^k)$,

$$\|\tilde{\mu} - A\|^2 = \text{Trace}(\tilde{\mu} - A)^2 = \sum_{i=1}^{k} \lambda_i^2 + m - 2\text{Trace}(\tilde{\mu}A) \tag{9.3.4}$$

where $\lambda_1, \ldots, \lambda_k$ are the eigenvalues of $\tilde{\mu}$ defined in the statement of the Proposition. Since A is a projection matrix, it can be written as

$$A = FF' \text{ where } F \in \mathbb{R}^{k\times m},\ F'F = I_m.$$

Also write $\tilde{\mu} = U\Lambda U'$ as in the proposition. Then

$$\|\tilde{\mu} - A\|^2 = \sum_{i=1}^{k} \lambda_i^2 + m - 2\text{Trace}(F'U\Lambda U'F)$$
$$= \sum_{i=1}^{k} \lambda_i^2 + m - 2\text{Trace}(E\Lambda E'),\ E = F'U. \tag{9.3.5}$$

To minimize $\|\tilde{\mu} - A\|^2$, we need to maximize $\text{Trace}(E\Lambda E')$ over $E \in \mathbb{R}^{m\times k}$, $EE' = I_m$. Note that

$$\text{Trace}(E\Lambda E') = \sum_{i=1}^{m}\sum_{j=1}^{k} e_{ij}^2 \lambda_j = \sum_{j=1}^{k} w_j \lambda_j$$

where $E = ((e_{ij}))$ and $w_j = \sum_{i=1}^{m} e_{ij}^2$, $j = 1, 2, \ldots, k$. Then $0 \le w_j \le 1$ and $\sum_{j=1}^{k} w_j = m$. Therefore the maximum value of $\text{Trace}(E\Lambda E')$ equals $\sum_{j=1}^{m} \lambda_j$ which is attained iff

$$w_1 = w_2 = \ldots = w_m = 1,\ w_i = 0 \text{ for } i > m.$$

That is when

$$E = (E_{11}, 0)$$

for some $E_{11} \in O(m)$. Then from equation (9.3.5), it follows that $F = UE'$ and the value of A which minimizes (9.3.4) is given by

$$A_0 = FF' = UE'EU' = U \begin{pmatrix} I_m & 0 \\ 0 & 0 \end{pmatrix} U' = \sum_{j=1}^{m} U_j U_j'. \tag{9.3.6}$$

This proves part (a) of the proposition.

To prove part (b), note that $\sum_{j=1}^{m} U_j U_j'$ is the projection matrix of the subspace spanned by $\{U_1, \ldots, U_m\}$ which is unique iff $\lambda_m > \lambda_{m+1}$. Then $\mu_E = \sigma(F')$ for any F satisfying $A_0 = FF'$, A_0 being defined in equation (9.3.6). Hence one can choose $F = (U_1, \ldots, U_m)$. This completes the proof. $\square$

We can use Proposition 9.3.1 and Proposition 4.2.1 to get an expression for the extrinsic variation V of Q as follows:

$$V = \|\tilde{\mu} - \mu\|^2 + \int_{J(A\Sigma_m^k)} \|\tilde{\mu} - x\|^2 \tilde{Q}(dx), \ \mu \in P(\tilde{\mu}).$$
$$= 2(m - \sum_{i=1}^{m} \lambda_i). \qquad (9.3.7)$$

Let $X_1, \ldots, X_n$ be an iid sample from Q and let μ_{nE} be the sample extrinsic mean, which can be any measurable selection from the sample extrinsic mean set. It follows from Proposition 2.2.2 that if Q has a unique extrinsic mean μ_E, that is, if $\tilde{\mu}$ is a nonfocal point of $S(k, \mathbb{R})$, then μ_{nE} is a consistent estimator of μ_E.

9.4 Asymptotic Distribution of the Sample Extrinsic Mean

In this section, we assume that $\tilde{\mu}$ is a nonfocal point of $S(k, \mathbb{R})$. Then the map $P(\tilde{\mu}) = \sum_{j=1}^{m} U_j U_j'$ is well defined and smooth in a neighborhood $N(\tilde{\mu})$ of $\tilde{\mu}$ in $S(k, \mathbb{R})$. That follows from Perturbation theory, because if $\lambda_m > \lambda_{m+1}$, then the subspace spanned by $\{U_1, \ldots, U_m\}$ is a smooth map from $S(k, \mathbb{R})$ into $G_m(k)$ and $P(\tilde{\mu})$ is the matrix of projection onto that subspace. Then it follows from the calculations of Section 4.3 that $\sqrt{n}(J(\mu_{nE}) - J(\mu_E))$ is asymptotically normal in the tangent space of $J(A\Sigma_m^k)$ at $J(\mu_E) \equiv P(\tilde{\mu})$. To get the asymptotic coordinates and the dispersion matrix as in Proposition 4.3.1, we need to find the derivative of P. Define

$$N_m^k = \{A \in S(k, \mathbb{R})\colon A^2 = A, \mathrm{trace}(A) = m\}. \qquad (9.4.1)$$

Then $N_m^k = J(A\Sigma_m^{k+1})$, which is a Riemannian manifold of dimension $km - m^2$. It has been shown in Dimitric (1996) that the tangent and normal spaces to N_m^k are given by

$$T_A N_m^k = \{v \in S(k, \mathbb{R})\colon vA + Av = v\}, \qquad (9.4.2)$$
$$T_A {N_m^k}^{\perp} = \{v \in S(k, \mathbb{R})\colon vA = Av\}. \qquad (9.4.3)$$

Consider the map

$$P : N(\tilde{\mu}) \to N_m^k,\ P(A) = \sum_{j=1}^{m} U_j(A)U_j(A)' \tag{9.4.4}$$

where $A = \sum_{j=1}^{k} \lambda_j(A)U_j(A)U_j(A)'$ is a s.v.d. of A as in Proposition 9.3.1. The expression for the derivative of P is obtained in Bhattacharya (2008a) which is mentioned in Proposition 9.4.1 below.

Proposition 9.4.1. *The derivative of P is given by*

$$dP : S(k, \mathbb{R}) \to TN_m^k,\ d_{\tilde{\mu}}P(A) = \sum_{i=1}^{m} \sum_{j=m+1}^{k} (\lambda_i - \lambda_j)^{-1} a_{ij} UE_{ij}U' \tag{9.4.5}$$

where $A = \sum\sum_{1 \le i \le j \le k} a_{ij}UE_{ij}U'$ and $\{UE_{ij}U' : 1 \le i \le j \le k\}$ is the orthogonal basis (frame) for $S(k, \mathbb{R})$ obtained in Section 8.3.

Proof. Let $\gamma(t) = \tilde{\mu} + tv$ be a curve in $N(\tilde{\mu})$ with $\gamma(0) = \tilde{\mu}$ and $\dot{\gamma}(0) = v \in S(k, \mathbb{R})$. Then

$$\gamma(t) = U(\Lambda + tU'vU)U' = U\tilde{\gamma}(t)U' \tag{9.4.6}$$

where $\tilde{\gamma}(t) = \Lambda + tU'vU$, which is a curve in $S(k, \mathbb{R})$ satisfying $\tilde{\gamma}(0) = \Lambda$ and $\dot{\tilde{\gamma}}(0) = \tilde{v} = U'vU$. From equations (9.4.4) and (9.4.6), we get that

$$P[\gamma(t)] = UP[\tilde{\gamma}(t)]U'. \tag{9.4.7}$$

Differentiate equation (9.4.7) at $t = 0$ to get

$$d_{\tilde{\mu}}P(v) = Ud_{\Lambda}P(\tilde{v})U'. \tag{9.4.8}$$

To find $d_{\Lambda}P(\tilde{v}) \equiv \frac{\mathrm{d}}{\mathrm{d}t}P[\tilde{\gamma}(t)]|_{t=0}$, we may assume without loss of generality that $\lambda_1 > \lambda_2 > \ldots > \lambda_k$. Then we can choose a s.v.d. for $\tilde{\gamma}(t)$ as $\tilde{\gamma}(t) = \sum_{j=1}^{k} \lambda_j(t)e_j(t)e_j(t)'$ such that $\{e_j(t), \lambda_j(t)\}_{j=1}^{k}$ are some smooth functions of t satisfying $e_j(0) = e_j$ and $\lambda_j(0) = \lambda_j$, where $\{e_j\}_{j=1}^{k}$ is the canonical basis for $\mathbb{R}^k$. Let $\tilde{v} = E_{ab}$, $1 \le a \le b \le k$. Then we can get expressions for $\dot{e}_j(0)$ from equation (8.3.7). Since

$$P[\tilde{\gamma}(t)] = \sum_{j=1}^{m} e_j(t)e_j(t)',$$

therefore

$$\frac{\mathrm{d}}{\mathrm{d}t}P[\tilde{\gamma}(t)]|_{t=0} = \sum_{j=1}^{m}[e_j\dot{e}_j(0)' + \dot{e}_j(0)e_j'] \tag{9.4.9}$$

From equations (8.3.7) and (9.4.9), we get that

$$d_\Lambda P(E_{ab}) = \begin{cases}(\lambda_a - \lambda_b)^{-1}E_{ab} & \text{if } a \leq m < b \leq k,\\ 0 & \text{o.w.}\end{cases} \tag{9.4.10}$$

Then from equation (9.4.8), we get that

$$d_{\tilde{\mu}}P(UE_{ab}U') = \begin{cases}(\lambda_a - \lambda_b)^{-1}UE_{ab}U' & \text{if } a \leq m < b \leq k,\\ 0 & \text{o.w.}\end{cases} \tag{9.4.11}$$

Hence if $A = \sum\sum_{1\leq i\leq j\leq k} a_{ij}UE_{ij}U'$, from (9.4.11), we get that

$$d_{\tilde{\mu}}P(A) = \sum_{i=1}^{m}\sum_{j=m+1}^{k}(\lambda_i - \lambda_j)^{-1}a_{ij}UE_{ij}U'. \tag{9.4.12}$$

This completes the proof. □

Corollary 9.4.2. *Consider the projection map of equation* (9.4.4) *restricted to*

$$S_0(k,\mathbb{R}) := \{A \in S(k,\mathbb{R}) : A\mathbf{1}_k = 0\}.$$

It has the derivative

$$dP : S_0(k,\mathbb{R}) \to TJ(A_m^k),\ d_{\tilde{\mu}}P(A) = \sum_{i=1}^{m}\sum_{j=m+1}^{k-1}(\lambda_i - \lambda_j)^{-1}a_{ij}UE_{ij}U'.$$

Proof. Follows from Proposition 9.4.1 and the fact that

$$T_{P(\tilde{\mu})}J(A_m^k) = \{v \in T_{P(\tilde{\mu})}N_m^k : v\mathbf{1}_k = 0\}.$$

□

From Corollary 9.4.2, it follows that

$$\{UE_{ij}U' : 1 \leq i \leq m < j < k\} \tag{9.4.13}$$

forms an orthonormal basis for $T_{P(\tilde{\mu})}J(A_m^k)$ and if $A \in S(k,\mathbb{R})$ has coordinates $\{a_{ij} : 1 \leq i \leq j \leq k\}$ with respect to the orthonormal basis $\{UE_{ij}U' : 1 \leq i \leq j \leq k\}$ of $S(k,\mathbb{R})$, then $d_{\tilde{\mu}}P(A)$ has coordinates $\{(\lambda_i - \lambda_j)^{-1}a_{ij} : 1 \leq i \leq m < j < k\}$ in $T_{P(\tilde{\mu})}J(A_m^k)$. Also it is easy to show that the linear projection $L(A)$ of A into $T_{P(\tilde{\mu})}J(A_m^k)$ has coordinates $\{a_{ij} : 1 \leq i \leq m < j < k\}$. Therefore we have the following corollary to Proposition 4.3.1. In the statement of Corollary 9.4.3, $\tilde{X}_j = J(X_j)$, $j = 1, \ldots, n$ denotes the embedded sample in $J(A\Sigma_m^k)$ and

$$T_j = ((T_j)_{ab} : \ 1 \leq a \leq m < b < k)$$

denotes the coordinates of $d_{\tilde{\mu}}P(\tilde{X}_j - \tilde{\mu})$ in $\mathbb{R}^{km-m-m^2}$ with respect to the orthonormal basis of $T_{P(\tilde{\mu})}J(A_m^k)$ obtained in equation (9.4.13). Then T_j has the following expression:

$$(T_j)_{ab} = \sqrt{2}(\lambda_a - \lambda_b)^{-1}U_a'\tilde{X}_jU_b, \ 1 \leq a \leq m < b < k.$$

Corollary 9.4.3. *If $\tilde{\mu} = \mathrm{E}[\tilde{X}_1]$ is a nonfocal point of $S(k,\mathbb{R})$, then*

$$\sqrt{n}[J(\mu_{nE}) - J(\mu_E)] = \sqrt{n}d_{\tilde{\mu}}P(\overline{\tilde{X}} - \tilde{\mu}) + o_P(1) \xrightarrow{\mathcal{L}} N(0,\Sigma)$$

where Σ denotes the covariance matrix of T_1.

Using Corollary 9.4.3, we may construct confidence regions for μ_E as in Section 4.3 or perform two sample tests to compare the extrinsic means from two populations on $A\Sigma_m^k$ as in Section 4.6.

The asymptotic distribution of the sample extrinsic variation follows from Theorem 4.4.1 using which we may construct confidence intervals for the extrinsic variation of Q or compare the extrinsic variations for two populations via two sample tests described in Section 4.6.

9.5 Application to Handwritten Digit Recognition

A random sample of 30 handwritten digit '3' were collected so as to devise a scheme to automatically classify handwritten characters. 13 landmarks were recorded on each image by Anderson (1997). The landmark data can be found in Dryden and Mardia (1998).

We analyse the affine shape of the sample points and estimate the mean shape and variation in shape. This can be used as a prior model for digit recognition from images of handwritten codes. Our observations lie on the affine shape space $A\Sigma_2^k$, $k = 13$. A representative of the sample extrinsic mean shape has coordinates

$$u = (-0.53, -0.32, -0.26, -0.41, 0.14, -0.43, 0.38, -0.29, 0.29, -0.11, 0.06, 0,$$
$$-0.22, 0.06, 0.02, 0.08, 0.19, 0.13, 0.30, 0.21, 0.18, 0.31, -0.13, 0.38, -0.42, 0.38).$$

The coordinates are in pairs, x coordinate followed by y. Figure 9.1 shows the plot of u.

The sample extrinsic variation turns out to be 0.27 which is fairly large. There seems to be a lot of variability in the data. Following are the extrinsic distances squared of the sample points from the mean affine shape:

$$(\rho^2(X_j, \mu_E),\ j = 1, \ldots, n) = (1.64, 0.28, 1.00, 0.14, 0.13, 0.07, 0.20, 0.09, 0.17, 0.15,$$
$$0.26, 0.17, 0.14, 0.20, 0.42, 0.31, 0.14, 0.12, 0.51, 0.10, 0.06, 0.15, 0.05, 0.31, 0.08,$$
$$0.08, 0.11, 0.18, 0.64, 0.12).$$

Here $n = 30$ is the sample size. From these distances, it is clear that observations 1 and 3 are outliers. We remove them and recompute the sample extrinsic mean and variation. The sample variation now turns out to be 0.19.

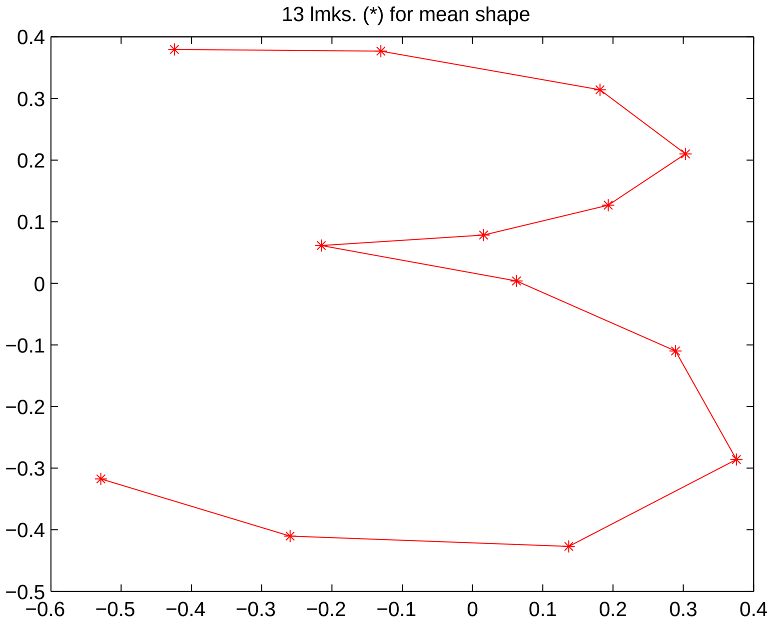


FIGURE 9.1. Extrinsic mean shape for handwritten digit 3 sample.

An asymptotic 95% confidence region for the extrinsic mean μ_E as in equation (4.3.3) is given by

$$\{\mu_E : nL[P(\tilde{\mu}) - P(\overline{\tilde{X}})]'\hat{\Sigma}^{-1}L[P(\tilde{\mu}) - P(\overline{\tilde{X}})] \leq \mathcal{X}^2_{20}(0.95) = 31.4104\}.$$

The dimension 20 of $A\Sigma_2^{13}$ is quite high compared to the sample size of 28. It is difficult to construct a pivotal bootstrap confidence region as in equation (4.3.4) because the bootstrap covariance estimates Σ^* tend to be singular or close to singular in most simulations. Instead, we construct a nonpivotal bootstrap confidence region by considering the linear projection $L[P(\overline{\tilde{X}}) - P(\overline{\tilde{X}^*})]$ into the tangent space of $P(\overline{\tilde{X}})$ and replacing Σ^* by $\hat{\Sigma}$. Then the 95$^{\text{th}}$ bootstrap percentile $c^*(0.95)$ turns out be 1.077 using 10^5 simulations. Hence bootstrap methods yield much smaller confidence region for the true mean shape compared to that obtained from chi-squared approximation.

A 95% confidence interval for the extrinsic variation V by normal approximation as described in equation (4.4.9) is given by $V \in [0.140, 0.243]$ while a pivotal bootstrap confidence interval using 10^5 simulations turns out to be $[0.119, 0.264]$.

In Dryden and Mardia (1998), the 2D similarity shapes (planar shapes) of the sample k-ads are analysed. A multivariate Normal distribution is assumed for the Procrustes coordinates of the planar shapes of the sample points, using which a F test is carried out to test if the population mean shape corresponds to that of an idealized template. The test yields a p-value of 0.0002 (see Example 7.1, Dryden and Mardia (1998)).

CHAPTER 10

REAL PROJECTIVE SPACES AND PROJECTIVE SHAPE SPACES

10.1 Introduction

Consider a k-ad picked on a planar image of an object or scene in 3D. If one thinks of images or photographs obtained through a central projection (a pinhole camera is an example of this), a ray is received as a landmark on the image plane (e.g., the film of the camera). Since axes in 3D comprise the projective space $\mathbb{R}P^2$, the k-ad in this view is valued in $\mathbb{R}P^2$. For a k-ad in 3D to represent a k-ad in $\mathbb{R}P^2$, the corresponding axes must all be distinct. To have invariance with regard to camera angles, one may first look at the original noncollinear 3D k-ad u and achieve affine invariance by its affine shape (i.e., by the equivalence class Au, $A \in GL(3, \mathbb{R})$), and finally take the corresponding equivalence class of axes in $\mathbb{R}P^2$ to define the projective shape of the k-ad as the equivalence class, or orbit, with respect to projective transformations on $\mathbb{R}P^2$. The projective shape of a k-ad is singular if the k axes lie on the vector plane $\mathbb{R}P^1$. For $k > 4$, the space of all non-singular shapes is the **2D projective shape space**, denoted as $P_0\Sigma_2^k$.

In general, the **projective space** $\mathbb{R}P^m$ comprises of axes or lines through the origin in $\mathbb{R}^{m+1}$. Thus elements of $\mathbb{R}P^m$ may be represented as equivalence classes

$$[x] = [x^1 : x^2 : \ldots : x^{m+1}] = \{\lambda x : \lambda \neq 0\},\ x = (x^1, \ldots, x^{m+1})' \in \mathbb{R}^{m+1} \setminus \{0\}.$$

Then a **projective transformation** α on $\mathbb{R}P^m$ is defined in terms of an $(m+1) \times (m+1)$ nonsingular matrix $A \in GL(m+1, \mathbb{R})$ by

$$\alpha([x]) = [Ax].$$

The group of all projective transformations on $\mathbb{R}P^m$ is denoted by $PGL(m)$. Now consider a k-ad $y = (y_1, \ldots, y_k)$ in $(\mathbb{R}P^m)^k$, say $y_j = [x_j]$, $j = 1, \ldots, k$, $x_j \in \mathbb{R}^{m+1} \backslash \{0\}$ and $k > m+2$. The **projective shape** of this k-ad is its orbit under $PGL(m)$, i.e., $\alpha(y) \doteq \{(\alpha y_1, \ldots, \alpha y_k) \colon \alpha \in PGL(m)\}$. To exclude singular shapes, define a k-ad $y = (y_1, \ldots, y_k) = ([x_1], \ldots, [x_k])$ to be in **general position** if there exists a subset of $m+2$ landmarks, say $(y_{i_1}, \ldots, y_{i_{m+2}})$, such that the linear span of any $m+1$ points from this set is $\mathbb{R}P^m$, i.e., if the linear span of their representative points in $\mathbb{R}^{m+1}$ is $\mathbb{R}^{m+1}$. The space of shapes of all k-ads in general position is the **projective shape space** $P_0\Sigma_m^k$.

10.2 Geometry of the Real Projective Space $\mathbb{R}P^m$

Since any line through the origin in $\mathbb{R}^{m+1}$ is uniquely determined by its points of intersection with the unit sphere S^m, one may identify $\mathbb{R}P^m$ with S^m/G, with G comprising the identity map and the antipodal map $p \mapsto -p$. Its structure as a m-dimensional manifold (with quotient topology) and its Riemannian structure both derive from this identification. Among applications are observations on galaxies, on axes of crystals, or on the line of a geological fissure (Watson (1983), Mardia and Jupp (1999), Fisher et al. (1987), Beran and Fisher (1998), Kendall (1989)).

For $u, v \in S^m$ the geodesic distance between the corresponding elements $[u], [v] \in \mathbb{R}P^m$ is given by

$$d_g([u],[v]) = \min\{d_{gs}(u,v), d_{gs}(u,-v)\}$$

where $d_{gs}(u,v) = \arccos(u'v)$ is the geodesic distance on S^m. Therefore

$$d_g([u],[v]) = \min\{\arccos(u'v), \arccos(-u'v)\} = \arccos(|u'v|).$$

The injectivity radius of $\mathbb{R}P^m$ is $\frac{\pi}{2}$. The map

$$\pi : S^m \to \mathbb{R}P^m,\ u \mapsto [u]$$

is a Riemannian submersion. The exponential map of $\mathbb{R}P^m$ at $[u]$ is $Exp_{[u]} = \pi \circ exp_u \circ d\pi_u^{-1}$ where $exp_u : T_uS^m \to S^m$ is the exponential map of the sphere, which is

$$exp_u(v) = \cos(\|v\|)u + \sin(\|v\|)\frac{v}{\|v\|},\ v \in T_uS^m.$$

The cutlocus of $[u]$ is

$$\begin{aligned} C([u]) &= \{[v] \in \mathbb{R}P^m :\ d_g([u],[v]) = \frac{\pi}{2}\} \\ &= \{[v] \in \mathbb{R}P^m :\ u'v = 0\}. \end{aligned}$$

The exponential map $Exp_{[u]}$ is invertible on $\mathbb{R}P^m \setminus C([u])$ and its inverse is given by

$$Exp_{[u]}^{-1}([v]) = \frac{\arccos(|u'v|)}{\sqrt{1-(u'v)^2}} d\pi_u \left(\frac{u'v}{|u'v|}v - |u'v|u\right),\ u'v \neq 0.$$

The projective space has a constant sectional curvature of 4.

10.3 Geometry of the Projective Shape Space $P_0\Sigma_m^k$

Recall that the projective shape of a k-ad $y \in (\mathbb{R}P^m)^k$ is given by the orbit

$$\sigma(y) = \{\alpha y :\ \alpha \in PGL(m)\}.$$

This orbit has full rank if y is in general position. Then we defined the projective shape space $P_0\Sigma_m^k$ to be the set of all shapes of k-ads in general position. Define a **projective frame** in $\mathbb{R}P^m$ to be an ordered system of $m+2$ points in general position, that is, the linear span of any $m+1$ points from this set is $\mathbb{R}P^m$. Let $I = i_1 < \ldots < i_{m+2}$ be an ordered subset of $\{1,\ldots,k\}$. A manifold structure on $P_I\Sigma_m^k$, an open dense subset of $P_0\Sigma_m^k$ of projective shapes of k-ads $(y_1,\ldots,y_k)$, for which $(y_{i1},\ldots,y_{i_{m+2}})$ is a projective frame in $\mathbb{R}P^m$, is derived in Mardia and Patrangenaru (2005) as follows. The **standard frame** is defined to be $([e_1],\ldots,[e_{m+1}],[e_1+e_2+\ldots+e_{m+1}])$, where $e_j \in \mathbb{R}^{m+1}$ has 1 in the j-th coordinate and zeros elsewhere. Given two projective frames $(p_1,\ldots,p_{m+2})$ and $(q_1,\ldots,q_{m+2})$, there exists a unique $\alpha \in PGL(m)$ such that $\alpha(p_j) = q_j$ $(j = 1, \ldots, m+2)$. By ordering the points in a k-ad such that the first $m+2$ points are in general position, one may bring this ordered set, say, $(p_1, \ldots, p_{m+2})$, to the standard form by a unique $\alpha \in PGL(m)$. Then the ordered set of remaining $k-m-2$ points is transformed to a point in $(\mathbb{R}P^m)^{k-m-2}$. This provides a diffeomorphism between $P_I\Sigma_m^k$ and the product of $k-m-2$ copies of the real projective space $\mathbb{R}P^m$. Hence by developing corresponding inference tools on $\mathbb{R}P^m$, one can perform statistical inference in a dense open subset of $P_0\Sigma_m^k$. In the subsequent sections, we develop intrinsic and extrinsic analysis tools on $\mathbb{R}P^m$.

10.4 Intrinsic Analysis on $\mathbb{R}P^m$

Let Q be a probability distribution on $\mathbb{R}P^m$ and let $X_1, \ldots, X_n$ be an iid random sample from Q. The value of r_* on $\mathbb{R}P^m$ as defined in Chapter 2 turns out to be the minimum of its injectivity radius of $\frac{\pi}{2}$ and $\frac{\pi}{4\sqrt{C}}$ where C is its constant sectional curvature of 4. Hence $r_* = \frac{\pi}{2}$ and therefore if the support of Q is contained in an open geodesic ball of radius $\frac{\pi}{4}$, then it has a unique intrinsic mean in that ball. In this section, we assume that $\text{supp}(Q) \subseteq B(p, \frac{\pi}{4})$, $p \in \mathbb{R}P^m$. Let $\mu_I = [\mu]$ $(\mu \in S^m)$ be the intrinsic mean of Q in the ball. Choose an orthonormal basis $v_1, \ldots, v_d$ for $T_\mu S^m$ so that $\{d\pi_\mu(v_j)\}$ forms an orthonormal basis for $T_{\mu_I}\mathbb{R}P^m$. For $[x] \in B(p, \frac{\pi}{4})$ $(x \in S^m)$, let $\phi([x])$ be the coordinates of $Exp_{\mu_I}^{-1}([x])$ with respect to this basis, which are

$$\begin{aligned} \phi([x]) &= (x^1, \ldots, x^m), \\ x^j &= \frac{x'\mu}{|x'\mu|} \frac{\arccos(|x'\mu|)}{\sqrt{1-(x'\mu)^2}} (x'v_j), \ j = 1, 2, \ldots, m. \end{aligned}$$

Let $X_j = [Y_j]$ $(Y_j \in S^m)$ and $\tilde{X}_j = \phi(X_j)$, $j = 1, 2, \ldots, n$. Let μ_{nI} be the sample intrinsic mean in $B(p, \frac{\pi}{4})$ and let $\mu_n = \phi(\mu_{nI})$. Then from Theorem 3.3.1 and Corollary 3.3.2, it follows that if $\text{supp}(Q) \subseteq B(\mu_I, \frac{\pi}{4})$, then

$$\sqrt{n}\mu_n \xrightarrow{\mathcal{L}} N(0, \Lambda^{-1}\Sigma\Lambda^{-1})$$

where $\Sigma = 4\mathrm{E}(\tilde{X}_1\tilde{X}_1')$ and $\Lambda = ((\Lambda_{rs}))_{1\le r,s\le d}$ where

$$\Lambda_{rs} = \Lambda_{sr} = 2\mathrm{E}[\frac{1}{(1-|\tilde{X}_1'\mu|^2)}\left(1-\frac{\arccos(|\tilde{X}_1'\mu|)(2|\tilde{X}_1'\mu|^2-1)}{|\tilde{X}_1'\mu|\sqrt{1-(\tilde{X}_1'\mu)^2}}\right)(\tilde{X}_1'v_r)(\tilde{X}_1'v_s)$$
$$+\frac{\arccos(|\tilde{X}_1'\mu|)(2|\tilde{X}_1'\mu|^2-1)}{|\tilde{X}_1'\mu|\sqrt{1-(\tilde{X}_1'\mu)^2}}\delta_{rs}],\quad 1\le r\le s\le d.$$

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

$$\{\mu_I : n\mu_n'\hat{\Lambda}\hat{\Sigma}^{-1}\hat{\Lambda}\mu_n \le \mathcal{X}_m^2(1-\alpha)\}$$

where $\hat{\Lambda}$ and $\hat{\Sigma}$ are sample estimates of Λ and Σ respectively. We can also construct a pivotal bootstrap confidence region by methods developed in Section 2.3.

To compare the intrinsic means or variations of two probability distribution on $\mathbb{R}P^m$ and hence distinguish between them, we can use the methods developed in Section 3.5.

10.5 Extrinsic Analysis on $\mathbb{R}P^m$

Another representation of $\mathbb{R}P^m$ is via the **Veronese-Whitney embedding** J of $\mathbb{R}P^m$ into the space of all $(m+1)\times(m+1)$ symmetric matrices $S(m+1,\mathbb{R})$ which is a real vector space of dimension $\frac{(m+1)(m+2)}{2}$. This embedding was introduced by Watson (1983) and is given by

$$J([u]) = uu' = ((u_iu_j))_{1\le i,j\le m+1},\ u=(u_1,..,u_{m+1})'\in S^m.$$

It induces the extrinsic distance

$$\rho^2([u],[v]) = \|uu'-vv'\|^2 = \mathrm{Trace}(uu'-vv')^2 = 2(1-(u'v)^2).$$

If one denotes the the space of all $(m+1)\times(m+1)$ positive semi definite matrices as $S^+(m+1,\mathbb{R})$, then

$$J(\mathbb{R}P^m) = \{A\in S^+(m+1,\mathbb{R}) : \mathrm{rank}(A) = \mathrm{Trace}(A) = 1\}$$

which is a compact Riemannian submanifold of $S(m+1, \mathbb{R})$ of dimension m. The embedding J is equivariant under the action of the orthogonal group $O(m+1)$ which acts on $\mathbb{R}P^m$ as $A[u] = [Au]$ (see Kent (1992), Bhattacharya and Patrangenaru (2005)).

Let Q be a probability measure on $\mathbb{R}P^m$, and let $\tilde{\mu}$ be the mean of $\tilde{Q} \doteq Q \circ J^{-1}$ considered as a probability distribution on $S(m+1, \mathbb{R})$. To find the extrinsic mean set of Q, we need to find the projection of $\tilde{\mu}$ on $\tilde{M} \doteq J(\mathbb{R}P^m)$, say $P_{\tilde{M}}(\tilde{\mu})$, as in Proposition 4.2.1. The projection set has been obtained in Bhattacharya and Patrangenaru (2003) but we include the derivation here for the sake of completeness. Since $\tilde{\mu}$ belongs to the convex hull of $\tilde{M}$, it lies in $S^+(m+1, \mathbb{R})$ and satisfies

$$\text{rank}(\tilde{\mu}) \geq 1, \;\; \text{Trace}(\tilde{\mu}) = 1.$$

There exists an orthogonal $(m+1) \times (m+1)$ matrix U such that $\tilde{\mu} = UDU'$, $D \equiv \text{Diag}(\lambda_1, \ldots, \lambda_{m+1})$ where the eigenvalues may be taken to be ordered: $0 \leq \lambda_1 \leq \ldots \leq \lambda_{m+1}$. To find $P_{\tilde{M}}(\tilde{\mu})$, note first that, writing $v = U'u$, we get that

$$\begin{aligned}
\|\tilde{\mu} - uu'\|^2 &= \text{Trace}[(\tilde{\mu} - uu')^2] \\
&= \text{Trace}[\{U'(\tilde{\mu} - uu')U\}\{U'(\tilde{\mu} - uu')U\}] = \text{Trace}[(D - vv')^2].
\end{aligned}$$

Write $v = (v_1, \ldots, v_{m+1})$, so that

$$\begin{aligned}
\|\tilde{\mu} - uu'\|^2 &= \sum_{i=1}^{m+1} (\lambda_i - v_i^2)^2 + \sum_{j \neq j'} (v_i v_j)^2 \\
&= \sum_{i=1}^{m+1} \lambda_i^2 + \sum_{i=1}^{m+1} v_i^4 - 2 \sum_{i=1}^{m+1} \lambda_i v_i^2 + (\sum_j v_j^2)(\sum_{j'} v_{j'}^2) - \sum_{j=1}^{m+1} v_j^4 \\
&= \sum_{i=1}^{m+1} \lambda_i^2 - 2 \sum_{i=1}^{m+1} \lambda_i v_i^2 + 1. \qquad (10.5.1)
\end{aligned}$$

The minimum of equation (10.5.1) is achieved when $v = (0, 0, \ldots, 0, 1)' = e_{m+1}$. That is when $u = Uv = Ue_{m+1}$ is an unit eigenvector of $\tilde{\mu}$ having the eigenvalue λ_{m+1}. Hence the minimum distance between $\tilde{\mu}$ and $\tilde{M}$ is attained by $\mu\mu'$ where μ is a unit vector in the eigenspace of the largest eigenvalue of $\tilde{\mu}$. There is a unique minimizer iff the largest eigenvalue of $\tilde{\mu}$ is simple, i.e., if the eigenspace corresponding to the largest eigenvalue is one dimensional. In that case, one says that $\tilde{\mu}$ is a nonfocal point of $S^+(m+1,\mathbb{R})$ and then from Proposition 4.2.1 it follows that the extrinsic mean μ_E of Q is $[\mu]$. Also the extrinsic variation of Q has the expression

$$V = \mathrm{E}[\|J(X_1) - \tilde{\mu}\|^2] + \|\tilde{\mu} - uu'\|^2 = 2(1 - \lambda_{m+1})$$

where $X_1 \sim Q$. Therefore we have the following corollary to Proposition 4.2.1.

Corollary 10.5.1. *Let Q be a probability distribution on $\mathbb{R}P^m$ and let $\tilde{Q} = Q \circ J^{-1}$ be its image in $S(m+1,\mathbb{R})$. Let $\tilde{\mu} = \int_{S(m+1,\mathbb{R})} x\tilde{Q}(dx)$ denote the mean of $\tilde{Q}$. (a) Then the extrinsic mean set of Q consists of all $[\mu]$, where μ is a unit eigenvector of $\tilde{\mu}$ corresponding to its largest eigenvalue λ_{m+1}. (b) This set is a singleton and Q has a unique extrinsic mean iff $\tilde{\mu}$ is nonfocal, that is λ_{m+1} is a simple eigenvalue. (c) The extrinsic variation of Q has the expression $V = 2(1 - \lambda_{m+1})$.*

Consider a random sample $X_1, \ldots, X_n$ iid Q. Let μ_n denote a measurable unit eigenvector of $\tilde{\mu}_n = \frac{1}{n}\sum_{i=1}^n J(X_i)$ corresponding to its largest eigenvalue $\lambda_{m+1,n}$. Then it follows from Proposition 2.2.2 and Corollary 10.5.1 that if $\tilde{\mu}$ is nonfocal, then the sample extrinsic mean $\mu_{nE} = [\mu_n]$ is a strongly consistent estimator of the extrinsic mean of Q. Proposition 2.2.3 implies that the sample extrinsic variation $2(1-\lambda_{m+1,n})$ is a strongly consistent estimator of the extrinsic variation of Q.

10.6 Asymptotic distribution of the Sample Extrinsic Mean

In this section we assume that $\tilde{\mu}$ is a nonfocal point of $S(m+1,\mathbb{R})$. Let $\tilde{X}_j = J(X_j)$, $j = 1, \ldots, n$ be the image of the sample in $\tilde{M}$ $(= J(\mathbb{R}P^m))$. Then it follows from Proposition 4.3.1 that if the projection map $P : S(m+1,\mathbb{R}) \to J(\mathbb{R}P^m)$, $P(A) = vv'$, v being a unit eigenvector from the eigenspace of the largest eigenvalue of A, is continuously differentiable in a neighborhood of $\tilde{\mu}$, then $\sqrt{n}[J(\mu_{nE}) - J(\mu_E)]$ has an asymptotic mean zero Gaussian distribution on $T_{J(\mu_E)}\tilde{M}$. It has asymptotic coordinates $\sqrt{n}\bar{T}$ where T_j is the coordinates of $d_{\tilde{\mu}}P(\tilde{X}_j - \tilde{\mu})$ with respect to some orthonormal basis for $T_{J(\mu_E)}\tilde{M}$. To get these coordinates and hence derive analytic expression for the parameters in the asymptotic distribution, we need to compute the differential of P at $\tilde{\mu}$ (if it exists). The computations can be found in Prentice (1984) and Bhattacharya and Patrangenaru (2005). We mention the derivation below so as to derive expressions for two-sample test statistics as in Section 4.6 to compare the extrinsic means from two populations.

Let $\gamma(t) = \tilde{\mu} + tv$ be a curve in $S(m+1, \mathbb{R})$ with $\gamma(0) = \tilde{\mu}$ and $\dot{\gamma}(0) = v \in S(m+1, \mathbb{R})$. Let $\tilde{\mu} = UDU'$, $U = (U_1, \ldots, U_{m+1})$, $D = \text{diag}(\lambda_1, \ldots, \lambda_{m+1})$ be a s.v.d. of $\tilde{\mu}$ as in Section 10.5. Then

$$\gamma(t) = U(D + tU'vU)U' = U\tilde{\gamma}(t)U'$$

where $\tilde{\gamma}(t) = D + tU'vU$ is a curve in $S(m+1, \mathbb{R})$ starting at D with initial velocity $\dot{\tilde{\gamma}}(0) = \tilde{v} \equiv U'vU$. Since D has largest eigenvalue simple, for t sufficiently small, $\tilde{\gamma}(t)$ is nonfocal. Choose $e_{m+1}(t)$ be a unit eigenvector corresponding to the largest (simple) eigenvalue $\lambda_{m+1}(t)$ of $\tilde{\gamma}(t)$, such that $t \mapsto e_{m+1}(t)$, $t \mapsto \lambda_{m+1}(t)$ are smooth (near $t = 0$) with $e_{m+1}(0) = e_{m+1}$, $\lambda_{m+1}(0) = \lambda_{m+1}$. Such a choice is possible by Perturbation theory of matrices since $\lambda_{m+1} > \lambda_m$ (see Dunford and Schwartz (1958)). Then

$$\tilde{\gamma}(t)e_{m+1}(t) = \lambda_{m+1}(t)e_{m+1}(t), \tag{10.6.1}$$

$$e'_{m+1}(t)e_{m+1}(t) = 1. \tag{10.6.2}$$

Differentiate equations (10.6.1) and (10.6.2) with respect to t at $t = 0$ to get

$$(\lambda_{m+1}I_{m+1} - D)\dot{e}_{m+1}(0) = -\dot{\lambda}_{m+1}(0)e_{m+1} + \tilde{v}e_{m+1}, \tag{10.6.3}$$

$$e'_{m+1}\dot{e}_{m+1}(0) = 0 \tag{10.6.4}$$

where $\dot{e}_{m+1}(0)$ and $\dot{\lambda}_{m+1}(0)$ refer to $\frac{\mathrm{d}}{\mathrm{d}t}e_{m+1}(t)|_{t=0}$ and $\frac{\mathrm{d}}{\mathrm{d}t}\lambda_{m+1}(t)|_{t=0}$ respectively. Consider the orthonormal basis (frame) $\{E_{ab} : 1 \leq a \leq b \leq m+1\}$ for $S(m+1, \mathbb{R})$ as defined in Section 8.3. Choose $\tilde{v} = E_{ab}$ for $1 \leq a \leq b \leq m+1$. From equations (10.6.3) and (10.6.4), we get that

$$\dot{e}_{m+1}(0) = \begin{cases} 0 & \text{if } 1 \leq a \leq b \leq m \text{ or } a = b = m+1, \\ 2^{-1/2}(\lambda_{m+1} - \lambda_a)^{-1} e_a & \text{if } 1 \leq a < b = m+1. \end{cases} \tag{10.6.5}$$

Since $P(\tilde{\gamma}(t)) = e_{m+1}(t)e'_{m+1}(t)$, therefore

$$\frac{\mathrm{d}}{\mathrm{d}t} P(\tilde{\gamma}(t))|_{t=0} = d_D P(\tilde{v}) = e_{m+1}\dot{e}'_{m+1}(0) + \dot{e}_{m+1}(0)e'_{m+1}. \tag{10.6.6}$$

From equations (10.6.5) and (10.6.6), we get that

$$d_D P(E_{ab}) = \begin{cases} 0 \text{ if } 1 \leq a \leq b \leq m \text{ or } a = b = m+1, \\ (\lambda_{m+1} - \lambda_a)^{-1} E_{ab} \text{ if } 1 \leq a < b = m+1. \end{cases} \tag{10.6.7}$$

Since P commutes with isometries $A \mapsto UAU'$, i.e. $P(UAU') = UP(A)U'$ and $\gamma(t) = U\tilde{\gamma}(t)U'$, therefore

$$\frac{\mathrm{d}}{\mathrm{d}t} P(\gamma(t))|_{t=0} = U \frac{\mathrm{d}}{\mathrm{d}t} P(\tilde{\gamma}(t))|_{t=0} U'$$

or

$$d_{\tilde{\mu}} P(v) = U d_D P(\tilde{v}) U'.$$

Hence from equation (10.6.7), it follows that

$$d_{\tilde{\mu}} P(UE_{ab}U') = \begin{cases} 0 \text{ if } 1 \leq a \leq b \leq m \text{ or } a = b = m+1, \\ (\lambda_{m+1} - \lambda_a)^{-1} UE_{ab}U' \text{ if } 1 \leq a < b = m+1. \end{cases} \tag{10.6.8}$$

Note that for all $U \in SO(m+1)$, $\{UE_{ab}U' : 1 \leq a \leq b \leq m+1\}$ is also an orthonormal frame for $S(m+1, \mathbb{R})$. Further from equation (10.6.8), it is clear that

$$\{UE_{ab}U' : 1 \leq a < b = m+1\} \tag{10.6.9}$$

forms an orthonormal frame for $T_{P(\tilde{\mu})}\tilde{M}$. If $A \in S(m+1, \mathbb{R})$ has coordinates $\{a_{ij} : 1 \le i \le j \le m+1\}$ with respect to the basis $\{UE_{ab}U' : 1 \le a \le b \le m+1\}$, that is,

$$A = \sum\sum_{1 \le i \le j \le m+1} a_{ij} U E_{ij} U',$$

$$a_{ij} = \langle A, UE_{ij}U' \rangle = \begin{cases} \sqrt{2} U_i' A U_j & \text{if } i < j \\ U_i' A U_i & \text{if } i = j, \end{cases}$$

then from equation (10.6.8), it follows that

$$\begin{aligned} d_{\tilde{\mu}} P(A) &= \sum\sum_{1 \le i \le j \le m+1} a_{ij} d_{\tilde{\mu}} P(UE_{ij}U') \\ &= \sum_{i=1}^{m} a_{im+1} (\lambda_{m+1} - \lambda_i)^{-1} U E_{im+1} U'. \end{aligned}$$

Hence $d_{\tilde{\mu}} P(A)$ has coordinates

$$\{\sqrt{2}(\lambda_{m+1} - \lambda_i)^{-1} U_i' A U_{m+1} : 1 \le i \le m\} \tag{10.6.10}$$

with respect to the orthonormal basis in equation (10.6.9) for $T_{P(\tilde{\mu})}\tilde{M}$. This proves the following Proposition.

Proposition 10.6.1. *Let Q be a probability distribution on $\mathbb{R}P^m$ with unique extrinsic mean μ_E. Let $\tilde{\mu}$ be the mean of $\tilde{Q} \doteq Q \circ J^{-1}$ regarded as a probability distribution on $S(m+1, \mathbb{R})$. Let μ_{nE} be the sample extrinsic mean from an iid sample $X_1, \ldots, X_n$. Let $\tilde{X}_j = J(X_j)$, $j = 1, \ldots, n$ and $\overline{\tilde{X}} = \frac{1}{n}\sum_{j=1}^{n} \tilde{X}_j$. (a) The projection map P is twice continuously differentiable in a neighborhood of $\tilde{\mu}$ and*

$$\begin{aligned} \sqrt{n}[J(\mu_{nE}) - J(\mu_E)] &= \sqrt{n} d_{\tilde{\mu}} P(\overline{\tilde{X}} - \tilde{\mu}) + o_P(1) \\ &\xrightarrow{\mathcal{L}} N(0, \Sigma) \end{aligned}$$

where Σ is the covariance of the coordinates of $d_{\tilde{\mu}} P(\tilde{X}_1 - \tilde{\mu})$.
(b) If $T_j = (T_j^1, \ldots, T_j^m)$ denotes the coordinates of $d_{\tilde{\mu}} P(\tilde{X}_j - \tilde{\mu})$ with respect to the orthonormal basis of $T_{P(\tilde{\mu})}\tilde{M}$ as in equation (10.6.9), then

$$T_j^a = \sqrt{2}(\lambda_{m+1} - \lambda_a)^{-1} U_a' \tilde{X}_j U_{m+1}, \ a = 1, \ldots, m.$$

Proof. See Proposition 4.3.1. □

Proposition 10.6.1 can be used to construct an asymptotic or bootstrap confidence region for μ_E as in Section 4.3.

Given two random samples on $\mathbb{R}P^m$, we can distinguish between the underlying probability distributions by comparing the sample extrinsic means and variations by methods developed in Section 4.6.

References

[1] Anderson, C. R. (1997). *Object recognition using statistical shape analysis.* PhD thesis, University of Leeds.

[2] Bandulasiri, A., Bhattacharya, R. N. and Patrangenaru, V. (2008). Nonparametric Inference on Shape Manifolds with Applications in Medical Imaging. *To appear.*

[3] Bandulasiri A. and Patrangenaru, V. (2005). Algorithms for Nonparametric Inference on Shape Manifolds, *Proc. of JSM 2005, Minneapolis, MN* 1617-1622.

[4] Beran, R. and Fisher, N. I. (1998). Nonparametric comparison of mean axes. *Ann. Statist.* **26** 472-493.

[5] Berthilsson, R. and Heyden A. (1999). Recognition of Planar Objects using the Density of Affine Shape. *Computer Vision and Image Understanding* **76** 2 135-145.

[6] Berthilsson, Rikard and Astrom, K. (1999). Extension of affine shape. *J. Math. Imaging Vision* **11**, no. 2 119136.

[7] Bhattacharya, A. (2008a). Statistical Analysis on Manifolds: A Nonparametric Approach for Inference on Shape Spaces. *To appear.*

[8] Bhattacharya, A. (2008b). Nonparametric Inference on Shape Spaces. *To appear in JSM Proceedings.*

[9] Bhattacharya, A. and Bhattacharya, R. (2008a). Nonparametric Statistics on Manifolds with Applications to Shape Spaces. *Pushing the Limits of Contemporary Statistics: Contributions in honor of J.K. Ghosh. IMS Collections* **3** 282-301.

[10] Bhattacharya, A. and Bhattacharya, R. (2008b). Statistics on Riemannian Manifolds: Asymptotic Distribution and Curvature. *Proc. Amer. Math. Soc* **136** 2957-2967.

[11] Bhattacharya, R. and Bhattacharya, A. (2008c). Statistics on Manifolds with Applications to Shape Spaces. *Perspectives in Mathematical Sciences.* Indian Statistical Institute, Bangalore. *In Press.*

[12] Bhattacharya, R. N. and Patrangenaru, V. (2002). Nonparametric estimation of location and dispersion on Riemannian manifolds. *J. Statist. Plann. Inference* **108** 23-35.

[13] Bhattacharya, R. N. and Patrangenaru, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds-I. *Ann. Statist.* **31** 1-29.

[14] Bhattacharya, R. and Patrangenaru, V. (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds-II. *Ann. Statist.* **33** 1225-1259.

[15] Bookstein, F. (1978). *The Measurement of Biological Shape and Shape Change.* Lecture Notes in Biomathematics **24**. Springer, Berlin.

[16] Bookstein, F. L. (1986). Size and shape spaces of landmark data (with discussion). *Statistical Science* **1** 181-242.

[17] Bookstein, F.L. (1991). *Morphometric Tools for Landmark data: Geometry and Biology.* Cambridge Univ. Press, Cambridge, U.K.

[18] Boothby, W. (1986). *An Introduction to Differentiable Manifolds and Riemannian Geometry.* 2d ed. Academic Press, Orlando.

[19] Derado, G., Mardia, K.V., Patrangenaru, V. and Thompson, H.W. (2004). A Shape Based Glaucoma Index for Tomographic Images. *J. Appl. Statist.* **31** 1241-1248.

[20] Dimitrić, Ivko (1996). A note on equivariant embeddings of Grassmannians. *Publ. Inst. Math.* (Beograd) (N.S.) **59**(73) 131-137.

[21] Do Carmo, M. P. (1992). *Riemannian Geometry.* Birkhauser, Boston. English translation by F. Flaherty.

[22] Dryden, I. L., Kume, A., Le, H. and Wood, A. T.A. (2008). A multidimensional scaling approach to shape analysis. *To appear in Biometrika.*

[23] Dryden, I. L. and Mardia, K. V. (1998). *Statistical Shape Analysis.* Wiley N.Y.

[24] Dunford, N. and Schwartz, J. (1958). *Linear Operators - I*. Wiley N.Y.

[25] Efron, Bradley (1982). *The Jackknife, the Bootstrap and Other Resampling Plans.* CBMS-NSF Regional Conference Series in Applied Mathematics, **38**. SIAM, Philadelphia, Pa.

[26] Fisher, N.I., Lewis, T. and Embleton, B. J. (1987). *Statistical Analysis of Spherical Data.* Cambridge U. Press.

[27] Fréchet, M. (1948). Lés élements aléatoires de nature quelconque dans un espace distancié. *Ann. Inst. H. Poincaré* **10** 215-310.

[28] Gallot, S., Hulin, D. and Lafontaine, J. (1993). *Riemannian Geometry.* 2nd ed. Springer, Berlin.

[29] Grenander, U. (1993). *General Pattern Theory.* Oxford University Press, Oxford.

[30] Hendricks, H. and Landsman, Z. (1998). Mean location and sample mean location on manifolds: Asymptotics, tests, confidence regions. *J. Multivariate Anal.* **67** 227-243.

[31] Hendricks, H. and Landsman, Z. (2007). Asymptotic data analysis on manifolds. *Ann. Statist.* **35** 109-131.

[32] Hopf, H. and Rinow, W. (1931). *Über den Begriff der vollständigen differentialgeometrischen Fläche.* Comment. Math. Helv. **3** 209-225.

[33] Joshi, S., Kaziska, D., Srivastava, A. and Mio, W. (2006). Riemannian structures on shape spaces: A framework for statistical inferences. In *Statistics and Analysis of Shapes* (H. Krim and A. Yezzi eds.). Birkhauser, Boston. 313-333.

[34] Karchar, H. (1977). Riemannian center of mass & mollifier smoothing. *Comm. on Pure & Applied Math.* **30** 509-541.

[35] Kendall, D.G. (1977). The diffusion of shape. *Adv. Appl. Probab.* **9** 428-430.

[36] Kendall, D. G. (1984). Shape manifolds, Procrustean metrics, and complex projective spaces. *Bull. London Math. Soc.* **16** 81-121.

[37] Kendall, David G. (1989) A survey of the statistical theory of shape. *Statist. Sci.* **4** 87120.

[38] Kendall, D. G., Barden, D., Carne, T. K. and Le, H. (1999). *Shape & Shape Theory.* Wiley N.Y.

[39] Kendall, W. S. (1990). Probability, convexity, and harmonic maps with small image-I, Uniqueness and the fine existence. *Proc. London Math. Soc.* **61** (1990) 371-406.

[40] 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.

[41] Krim, H. and Yezzi, A. eds. (2006). *Statistics and Analysis of Shapes.* Birkhauser, Boston.

[42] Le, H. (2001). Locating frechet means with application to shape spaces. *Adv. Appl. Prob.* **33** 324-338.

[43] Lee, J. M. (1997). *Riemannian Manifolds: An Introduction to Curvature.* Springer, N.Y.

[44] Mardia, K.V. and Jupp, P. E. (1999). *Statistics of Directional Data.* Wiley, New York.

[45] Mardia, K.V. and Patrangenaru, V. (2005). Directions and projective shapes. *Ann. Statist.* **33** 1666-1699.

[46] Patrangenaru, V. (1998). *Asymptotic Statistics on Manifolds and Their Applications*, PhD. dissertation, Indiana University, Bloomington.

[47] Prentice, Michael J. (1984). A distribution-free method of interval estimation for unsigned directional data. *Biometrika* **71** 147-154.

[48] Sepiashvili, D., Moura, J.M.F. and Ha, V.H.S. (2003). Affine-Permutation Symmetry: Invariance and Shape Space (2003). *Proceedings of the 2003 Workshop on Statistical Signal Processing.* 293-296.

[49] Small, C.G. (1996). *The Statistical Theory of Shape.* Springer, New York.

[50] Sparr, G. (1992). Depth-computations from polihedral images. *Proc. 2nd European Conf. on Computer Vision*(ECCV-2) (G. Sandini, editor). 378-386.

[51] Sugathadasa, M. S. (2006). *Affine and Projective Shape Analysis with Applications.* PhD dissertation, Texas Tech University.

[52] Watson, G. S. (1983). *Statistics on spheres.* University of Arkansas Lecture Notes in the Mathematical Sciences, 6. Wiley, New York.

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