# Directional Statistics and Shape Analysis

by K.V. Mardia

Department of Statistics
University of Leeds
Leeds LS2 9JT, UK

## Abstract

There have been various developments in shape analysis in the last decade. We describe here some relationships of shape analysis with directional statistics. For shape, rotations are to be integrated out or to be optimized over whilst they are the basis for directional statistics. However, various concepts are connected. In particular, certain distributions of directional statistics have emerged in shape analysis, such a distribution is Complex Bingham Distribution. This paper first gives some background to shape analysis and then it goes on to directional distributions and their applications to shape analysis. Note that the idea of using tangent space for analysis is common to both manifold as well.

## 1 Introduction

Consider shapes of configurations of points in Euclidean space. There are various contexts in which $k$ labelled points (or "landmarks") $\mathbf{x}_1, ..., \mathbf{x}_k$ in $\mathbb{R}^m$ are given and interest is in the shape of $(\mathbf{x}_1, ..., \mathbf{x}_k)$.

Example 1 The microscopic fossil *Globorotalia truncatulinoides* is shaped rather like a snail. For each of 21 such fossils three landmarks are given in Fig.1. Do the triangles given by these landmarks tend to be almost equilateral?

Insert Fig.1

Example 2 Given a set of a mouse vertebrae bones (T2) with six landmarks (see Fig.2), the question arises for comparison purposes, what is the sample mean shape? How to measure its variation?

Insert Fig.2

Example 3 Given unlabelled triangles arising in a Central Place Theory (see Example 6), how to test the evidence of the equilateral triangle?

There are many other practical examples. For example, object recognition in images through shape, testing whether or not some prehistoric sites tend to lie on ley-lines, whether or not a given set of quasars lies in a straight line.

The above examples all involve "shape" which we can interpret as "what remains when location, scale, and rotational effects are filtered out" (Kendall, 1977). In order to make progress, we need to express this concept more rigorously.

**Shape**

Let $\mathbf{x}_1, ..., \mathbf{x}_k$ and $\mathbf{y}_1, ..., \mathbf{y}_k$ be two (not totally coincident) labelled sets of $k$ points in $\mathbb{R}^m$. We regard these two configurations as having the same shape if

$$\mathbf{y}_i = c\mathbf{R}\mathbf{x}_i + \mathbf{b} \qquad i = 1, ..., k$$

for some scale factor $c > 0$, rotation matrix $\mathbf{R}(m \times m)$, and vector $\mathbf{b}(k \times 1)$, i.e. we can transform $\mathbf{x}_1, ..., \mathbf{x}_k$ into $\mathbf{y}_1, ..., \mathbf{y}_k$ by translation, rotation and scaling. Note that we do not allow $\mathbf{R}$ to be a reflection. Following Kendall (1984), we denote by $\Sigma_m^k$ the shape space of $k$ labelled points in $\mathbb{R}^m$. To measure distances between shapes we require a metric on $\Sigma_m^k$; this gives Kendall's shape space the structure of a Riemannian manifold with Procrustes distance as its metric. We will not use this explicitly.

For a description of shape in coordinate terms, consider a configuration of $k$ points in $\mathbb{R}^m$, represented by a $k \times m$ matrix $\mathbf{X}$. Location and scale effects are easy to eliminate directly. Let us recall that the standard Helmert matrix has its first row with elements equal to $k^{-\frac{1}{2}}$ and the remaining rows are orthogonal to this first row. Let this Helmert matrix minus its first row be $H(k-1 \times k)$ which we will call the Helmert sub-matrix. Explicitly, the $j^{th}$ row of $H$ is a vector with $k$-elements

$$(H)_j = (d_j, ..., d_j, -jd_j, 0, ..., 0), d_j = \{j(j+1)\}^{-\frac{1}{2}} \ , \ j = 2, ..., k, \tag{1.1}$$

and $d_j$ is repeated $j$ times. The rows of $H$ are orthonormal and $\mathbf{H}\mathbf{1}_k = \mathbf{0}, \mathbf{1}_k$ a $k$-vector of ones. Then $\mathbf{Y} = \mathbf{H}\mathbf{X}(k-1 \times m)$ is invariant under location shifts of the configuration $\mathbf{X} \mapsto \mathbf{X} + \mathbf{1}_k\mathbf{w}^T$, for $\mathbf{w} \in \mathbb{R}^m$. The $k$ rows of $\mathbf{X}$ are known as the raw landmarks and the $k-1$ rows of $\mathbf{Y}$ as the "derived landmarks".

Scale effects can be eliminated by the standardization $\mathbf{U} = \mathbf{Y}/ \parallel \mathbf{Y} \parallel$, where $\parallel \mathbf{Y} \parallel^2 = \Sigma\Sigma y_{ij}^2$. Then $\mathbf{U}$ can be regarded as a point on the unit sphere in $\mathbb{R}^{(k-1)m}$, which can be described as the pre-shape matrix.

One route to obtaining shape space, is first to remove the effect of location and scale, giving **"preshape"**. Regard $\mathbf{U}$ as point of $S^{m(k-1)-1}$. In coordinate terms, the preshape of $\mathbf{X}$ is $\mathbf{U}$ defined above. Similarly, removing effect of location and rotation gives **"size-and-shape"** (= "form").

# 2 Shape Spaces and Coordinates

## 2.1 The Two Dimensional Case

Not only is the case $m = 2$ of considerable practical importance but the fact that we can identify $\mathbb{R}^2$ with $\mathbb{C}$ means that the algebra and geometry of complex numbers can be used to give a neat description of the shape space $\Sigma_2^k$. Let $\mathbf{z} = (z_1, ..., z_k)$ denote $k$ complex numbers

which represent $k$ points in the plane i.e. $\mathbf{z}$ denotes the raw landmarks. Then we can remove the effect of location by considering just $(z_1, ..., z_{k-1})$, or equivalently moving the sample mean $\bar{z}$ to the origin by a Helmert transformation. These new $z$'s will be called centered landmarks. Next observe that the effects of rotation and scale are removed by regarding $(z_1, ..., z_{k-1})$ as equivalent to $(cz_1, ..., cz_{k-1})$ for any non-zero complex number $c$. Thus we can represent the shape of $z$ by the equivalence class $[z_1, ..., z_{k-1}]$ of $(z_1, ..., z_{k-1})$ under the above equivalence relation. The set of such equivalence classes forms the $(k-2)$-dimensional complex projective space $\mathbb{C}P^{k-2}$. Thus we have the identification

$$\Sigma_2^k = \mathbb{C}P^{k-2}. \tag{2.1}$$

The $(k-2)$ dimensional complex projective space $\mathbb{C}P^{k-2}$ can be considered as the unit sphere in $\mathbb{C}^{k-1}$, with $\mathbf{z}$ equivalent to $\mathbf{z}e^{i\theta}$ for all $\theta$. In the case $k = 3$, the mapping $(z_1, z_2) \mapsto z_2/z_1$ identifies the complex projective line $\mathbb{C}P^1$ with the sphere $S^2$ obtained by adding a point at infinity to the complex plane. Thus we can identify the space $\Sigma_2^3$ of shapes of triangles in the plane with the sphere $S^2$.

*Coordinate systems*

In order to carry out calculations on shapes it is useful to have suitable coordinate systems on the shape spaces. On the shape spaces $\Sigma_2^k$ various coordinate systems are in use. There are two popular sets of coordinates.

**Kendall's coordinates** Kendall's (1984) coordinates for $k$ points in $\mathbb{R}^2$ are in terms of the centered landmarks

$$z_j/z_1, \qquad j = 2, ..., k-1. \tag{2.2}$$

**Bookstein's coordinates** Alternatively we have Bookstein's (1986, 1991) $k-2$ complex shape coordinates

$$w_j = (z_j - z_1)/(z_2 - z_1), \qquad j = 1, 2, 3, ..., k, \tag{2.3}$$

provided $z_1 \neq z_2$ and have $z_i i = 1..k$ denote the raw landmarks. Note that these two representations are very similar. Since $w_1 = 0, w_2 = 1$, this is effectively a $(k-2)$-dimensional complex representation. Landmarks 1 and 2 are called the "base" with respect to which the object is "registered". Note that this is not a symmetrical method of "registration". Working directly on $\mathbb{C}P^{k-2}$ provides a symmetrical way.

## 2.2 Shapes in higher dimensions

The QR decomposition, often used in multivariate analysis, provides an alternative way of obtaining the shape of a configuration $\mathbf{X}$. This decomposition gives shape coordinates in any dimension. Namely, let $\mathbf{X}(k \times m)$ contain the raw landmarks and $\mathbf{Y} = \mathbf{HX}$ the derived landmarks. To remove orientation consider the QR decomposition of $\mathbf{Y}$, namely,

$$\mathbf{Y} = \mathbf{T?} \quad , \quad \mathbf{?} \in V_{n,m}, \| \mathbf{T} \| > 0 \tag{2.4}$$

where $n = \min(k-1, m)$, and $\mathbf{T}(k-1 \times n)$ is lower triangular with non-negative diagonal elements $T_{ii} \geq 0, i = 1, ..., n$, and $\| \mathbf{T} \|^2 = \text{tr } \mathbf{T}^T\mathbf{T}$.

The most important case in practice is when $k > m$. In that case, $\mathbf{?} \in O(m)$ and $|\mathbf{?}| = \pm 1$ so that the QR decomposition removes orientation and reflection. To remove orientation only, we require $\mathbf{?} \in SO(m)$ so $|\mathbf{?}| = +1$ and $T_{mm}$ is unrestricted. Then $\mathbf{T} = \{T_{ij} : 1 \leq i \leq j \leq m\}$ is the *size-and-shape* (or *form*) if $\mathbf{?} \in SO(m)$ and the *reflection size-and-shape* if $\mathbf{?} \in O(m)$.

To remove scale we divide by the Euclidean norm $\| \mathbf{T} \|$ of $\mathbf{T}$, i.e. define $\mathbf{W}$ by

$$\mathbf{W} = \mathbf{T} / \| \mathbf{T} \| \quad , \quad \| \mathbf{T} \| > 0. \tag{2.5}$$

We call $\mathbf{W}$ the *shape* of our configuration if $\mathbf{?} \in SO(m)$ and the *reflection shape* if $\mathbf{?} \in O(m)$. (If $k \leq m$ then these distinctions are irrelevant.) If $\| \mathbf{T} \| = 0$ then the landmarks are coincident and shape is not defined. For practical examples for $m = 3$, see Goodall and Mardia (1993) and Dryden and Mardia (1993). Most of our discussion will concentrate on the 2 dimensional case.

# 3 Procrustes Averages in Various Spaces

Let $\mathbf{u} \in \mathbb{C}^k$ represent a configuration of $k$ points in the plane, $k \geq 3$. Then the two configurations $\mathbf{u}_1$ and $\mathbf{u}_2$ have the same shape if $\mathbf{u}_1 = c\mathbf{1}_k + \alpha e^{i\theta}\mathbf{u}_2$ for some $c \in \mathbb{C}, \theta \in \mathbb{R}$ and $\alpha > 0$, where $\mathbf{1}_k = (1, ..., 1)^T$.

Our aim is to estimate shape $\mu$ where $\mathbf{u}_1, ..., \mathbf{u}_n$ are given by

$$\mathbf{u}_j = c_j \mathbf{1}_k + \alpha_j e^{i\theta_j} \boldsymbol{\mu} \quad , \quad j = 1, ..., n \ , \tag{3.1}$$

where $\boldsymbol{\mu}$ is a fixed configuration, $c_j \in \mathbb{C}$ and $\alpha_j > 0$ are unknown nuisance parameters. The objective is to estimate $\boldsymbol{\mu}$ given $\mathbf{u}_1, ..., \mathbf{u}_n$. Of course only the shape of $\boldsymbol{\mu}$ can be estimated.

One way to estimate $\boldsymbol{\mu}$ is through Procrustes Analysis (Goodall, 1991). The Procrustes solution is a least square solution given by

$$\hat{\boldsymbol{\mu}} = \underset{\boldsymbol{\mu} : \| \boldsymbol{\mu} \| = 1}{\arg\min} \ \underset{\substack{c_j, \theta_j, \alpha_j, \\ j = 1, ..., n}}{\min} \ \sum_{j=1}^n \| \boldsymbol{\mu} - c_j \mathbf{1}_k - \alpha_j \, e^{i\theta_j} \mathbf{u}_j \|^2 \ . \tag{3.2}$$

Some simplification is possible. Location effects can be eliminated by an orthogonal transformation. Let $\mathbf{H}$ be the Helmert sub-matrix above given by (1.1). We can write

$$\boldsymbol{\nu} = \mathbf{H}\boldsymbol{\mu}, \quad \mathbf{v} = \mathbf{H}\mathbf{u},$$

as

$$\mathbf{v}_j = \alpha_j e^{i\theta_j} \boldsymbol{\nu}. \tag{3.3}$$

Thus the vector $\boldsymbol{\nu}$ denotes the new shape mean direction. The estimates $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\nu}}$ are related by $\hat{\boldsymbol{\nu}} = \mathbf{H}\hat{\boldsymbol{\mu}}$, $\hat{\boldsymbol{\mu}} = \mathbf{H}^T \hat{\boldsymbol{\nu}} + c\mathbf{1}_k$ where $c \in \mathbb{C}$ is arbitrary. Further, if $\hat{\boldsymbol{\mu}}^T \mathbf{1}_k = 0$, then $\| \hat{\boldsymbol{\mu}} \| = \| \hat{\boldsymbol{\nu}} \|$. On minimizing over $\alpha_j > 0$, the $\hat{\boldsymbol{\nu}}$ can be rewritten in the form

$$\hat{\boldsymbol{\nu}} = \underset{\boldsymbol{\nu} : \| \boldsymbol{\nu} \| = 1}{\arg\min} \ \sum_{j=1}^n (1 - \boldsymbol{\nu}^* \mathbf{z}_j \mathbf{z}_j^* \boldsymbol{\nu}), \tag{3.4}$$

where $\boldsymbol{\nu}^*$ denotes the complex conjugate transpose of $\boldsymbol{\nu}$. The Procrustes solution $\hat{\boldsymbol{\nu}}$ can be identified with the dominant eigenvector of the complex sum of squares and products matrix

$$\mathbf{S} = \Sigma \mathbf{z}_j \mathbf{z}_j^*. \tag{3.5}$$

Let $l_1, ..., l_{k-1}$ be the eigenvalues of $\mathbf{S}$. Note these are all real and non-negative. We have

$$l_1 + ... + l_{k-1} = n. \tag{3.6}$$

If $l_i$'s are equal, we have uniformity, the shapes are 'diffused' i.e. $z_1, ..., z_k$ of (3.4) are the highly dispersed on shape sphere. If $0 < l_1 < ... < l_{k-1}$, then the dominant eigenvector corresponds to $l_{k-1}$. Let it be $\mathbf{g}_{k-1}$ which is the shape mean direction $\hat{\nu}$. If

$$\mathbf{g}_{k-1} = \mathbf{x}_{k-1} + i\mathbf{y}_{k-1} \tag{3.7}$$

then $||\mathbf{g}_{k-1}|| = 1$, which is defined only up to multiplication by a unit complex number. Hence for visualization purposes we can select a suitable rotation.

If $l_{k-1}$ is very large in proportion to the rest, i.e. $l_{k-1}$ is nearly $n$, then the data set is highly concentrated.

Example 4 As an example consider the data analysed by Dryden and Mardia (1991) and Kent (1994). This analysis follows Kent (1994). The data consist of $k = 6$ landmarks of the $T2$ vertebra from $n = 23$ (small) mice. The landmarks are labelled 1,...,6 as in Figure 2. We find that the eigenvalues are

$$l_1 = 0.004, \quad l_2 = 0.005, \quad l_3 = 0.012, \quad l_4 = 0.072, \quad l_5 = 22.905$$

Insert Fig.3

Thus there is an extremely high concentration around the mean shape as $l_5 = 22.905$ is very close to 23. The mean shape is the eigenvector corresponding to $l_5$ and is given by

$$\mathbf{g}_5 = (0.041 + 0.246i, 0.404 + 0.072i, -0.495 - 0.716i, -0.056 + 0.037i, 0.089)^T, \mathbf{g}_5^* \mathbf{g}_5 = 1.$$

To plot the mean shape, we work on

$$\mathbf{H}^T \mathbf{g}_5$$

so that there are six points, one point for each landmark. Note that $\mathbf{1}^T(\mathbf{H}^T\mathbf{g}_5) = 0$ so that the configuration is centred in addition to being of unit size as $\mathbf{g}_5^* \mathbf{H}\mathbf{H}^T \mathbf{g}_5 = 1$. The eigenvector is defined only up to a rotation so we can plot

$$e^{i\alpha}\mathbf{H}^T\mathbf{g}_1$$

Here we rotate the mean shape by $\alpha$ so that the line joining the two furthest apart landmarks is horizontal. This mean shape is given by

$$e^{i\alpha}\mathbf{H}^T\mathbf{g}_5 =$$

$$(-0.507 - 0.143i, 0.506 - 0.143i, 0.085 + 0.154i, 0.012 + 0.424i, -0.070 + 0.160i, -0.026 - 0.451i)^T$$

and it is plotted in Figure 4. Fig.3 plot the data as well as the mean direction. As expected the data set is highly concentrated around the mean.

Insert Fig.4

To understand detailed variation around landmarks, it is useful to look into the principal components at the tangent space to the shape space at the mean direction (see Kent, 1994). See also Example 7.

Directional Averages The Procrustes procedure used above can be applied to directions (Kent, 1992). That is, consider an equivalence class of vectors in $\mathbb{R}^k$ in which a non-zero vector $\mathbf{x}$ is identified with the ray $\{r\mathbf{x} : r > 0\}$. Given data $\mathbf{u}_1, ..., \mathbf{u}_n$ (up to sign) on $S^{k-1}$, we can define the average $\hat{\boldsymbol{\mu}} \in S^{k-1}$ to minimize

$$\Sigma \parallel r_i \mathbf{u}_i - \boldsymbol{\mu} \parallel^2 = \Sigma(r_i^2 + 1 - 2r_i \boldsymbol{\mu}^T \mathbf{u}_i)$$

over $r_i$ and $\boldsymbol{\mu}$ ; $r_i$ being nuisance parameters here. Minimizing first over $r_i = (\boldsymbol{\mu}^T \mathbf{u}_i)_+$ where $\alpha_+ = \alpha$ if $\alpha > 0$ and $\alpha_+ = 0$ if $\alpha < 0$, the objective function becomes

$$\Sigma[1 - \{(\boldsymbol{\mu}^T \mathbf{u}_i)_+\}^2] \; . \tag{3.8}$$

Here $\hat{\boldsymbol{\mu}}$ is the dominant eigenvector of $\Sigma_+ \mathbf{u}_i \mathbf{u}_i^T$ where $\Sigma_+$ indicates that the sum is taken over data values for which $\hat{\boldsymbol{\mu}}^T \mathbf{u}_i > 0$. The selection of terms in this sum depends on $\hat{\boldsymbol{\mu}}$ and the sign of $\hat{\boldsymbol{\mu}}$ matters here.

Note that $\hat{\boldsymbol{\mu}}$ is in some sense a resistant estimator of average direction because data values more than $90^o$ away from $\hat{\boldsymbol{\mu}}$ are given weight 0. However, $\hat{\boldsymbol{\mu}}$ is not necessarily unique. Consider a data set on the circle consisting of 3 tight clusters of data approximately $120^o$ apart from one another. Then $\hat{\boldsymbol{\mu}}$ can take 3 values; in each case $\hat{\boldsymbol{\mu}}$ points towards the centre of one of the clusters and ignores the other two.

Note that this construction of $\hat{\boldsymbol{\mu}}$ coincides with the Procrustes analysis for shapes in $m = 1$ dimension. Thus, Procrustes methods do not necessarily lead to a uniquely defined average in this case.

A similar type of argument can be applied in finding an average for axial data. However, the solution is not as explicit. In fact, the average (see Kent, 1992) is

$$\hat{\boldsymbol{\mu}} \; \propto \; \Sigma \text{ sign } (\hat{\boldsymbol{\mu}}^T \mathbf{u}_i) \mathbf{u}_i \tag{3.9}$$

obtained by minimizing

$$\Sigma \parallel r_i \sigma_i \mathbf{u}_i - \boldsymbol{\mu} \parallel^2$$

over $r_i > 0, \sigma_i = \pm 1$ and $\boldsymbol{\mu}$. Thus $\hat{\boldsymbol{\mu}}$ is proportional to the resultant vector of the $\mathbf{u}_i$ after the signs of the $\mathbf{u}_i$ have been switched so that they all lie in the same hemisphere. Again, $\hat{\boldsymbol{\mu}}$ is no longer unique.

Form Average Note that we can define the procedure for Procrustes size-and-shape average (see, e.g. Kent, 1992), i.e. two configurations $\mathbf{u}_1$ and $\mathbf{u}_2$ have the same "form" if $\mathbf{u}_1 = c\mathbf{1}_k + e^{i\theta}\mathbf{u}_2$ for some $c \in \mathbb{C}$.

Following the same method described earlier for the Procrustes shape average, we find that the Procrustes form average is

$$\hat{\boldsymbol{\nu}} = \underset{\boldsymbol{\nu}}{\arg\min} \quad \sum_{j=1}^{n} \{\boldsymbol{\nu}^* \boldsymbol{\nu} - 2|\boldsymbol{\nu}^* \mathbf{v}_j| + \mathbf{v}_j^* \mathbf{v}_j\} \tag{3.10}$$

where $\mathbf{v}_j = \mathbf{H}\mathbf{u}_j$ , $\boldsymbol{\nu} = \mathbf{H}\boldsymbol{\mu}$. This solution cannot be simplified further in contrast to the shape case.

Higher Dimensions We have treated the two-dimensional case, but the approach can be extended for higher dimension as follows. Given the data matrices $\mathbf{U}_i(k \times m)$, $i = 1, ..., n$ (Helmertised for convenience), define the average shape $\hat{\boldsymbol{\mu}}(k \times m)$ (Helmertised $\hat{\boldsymbol{\mu}}$ is defined up to a rotation $\hat{\boldsymbol{\mu}}\mathbf{R}$) to minimize

$$\Sigma \parallel r_i \mathbf{U}_i \mathbf{R}_i - \boldsymbol{\mu} \parallel^2 \tag{3.11}$$

over $\mathbf{R}_i \in SO(m), r_i > 0$ and $\boldsymbol{\mu}$. For the Procrustes form average minimize

$$\Sigma \parallel \mathbf{U}_i \mathbf{R}_i - \boldsymbol{\mu} \parallel^2 \tag{3.12}$$

over $\mathbf{R}_i \in SO(m)$ and $\boldsymbol{\mu}$. In practice, except for the 2-dimensional case above of shape, the solution can be obtained iteratively.

# 4 Distributions

We have seen in directional statistics that there are two main approaches which have produced directional distributions from the multivariate normal distribution

**(i)** the marginal approach - where we integrate out the non-directional variables (as in the derivation of the off-set normal distribution, Mardia, 1972).

**(ii)** the conditional approach - where the non-directional variables are held constant (as for the von-Mises density, we fix the length in a suitable bivariate normal distribution, Mardia, 1972).

Recently both approaches have produced useful distributions starting with Mardia and Dryden's distribution (Mardia and Dryden, 1989a,b) following the marginal approach. Kent (1994) adopted the conditional approach and introduced the complex Bingham distribution. We now describe these approaches.

## 4.1 The Marginal Approach

Let $\mathbf{X}_1, ..., \mathbf{X}_k$ be independently distributed points in $\mathbb{R}^2$. Then the shape of $\mathbf{X}_1, ..., \mathbf{X}_k$ is a random point in $\Sigma_2^k$. A case of particular interest is that in which $\mathbf{X}_i \sim N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$ for $i = 1, ..., k$. For this case Mardia & Dryden (1989), Mardia (1989) and Dryden & Mardia (1991) have given expressions for the probability density function of the corresponding shape distribution on $\Sigma_2^k$. Dryden & Mardia (1991) also allows correlations between landmarks.

For the case $k = 3$, the identification of $\Sigma_2^3$ with the sphere $S^2$ with $\mathbf{X}_i \sim N(\boldsymbol{\mu}_i, \sigma^2\mathbf{I}), i = 1, 2, 3,$, Mardia (1989) showed that the probability density function of the shape distribution on $\mathcal{S}^2$ is

$$f(\mathbf{x}; \boldsymbol{\lambda}, \kappa) = \{1 + \kappa(\boldsymbol{\lambda}'\mathbf{x} + 1)\} \exp \kappa(\boldsymbol{\lambda}'\mathbf{x} - \mathbf{1}), \mathbf{x} \in S^2, |\boldsymbol{\lambda}| = 1, \tag{4.1}$$

and $\boldsymbol{\lambda}$ is the "mean shape". For all $\kappa$ and $k = 3$, this density is quite close to a Fisher density with the same mean (see, Mardia, 1989) and hence Mardia (1989) recommended the use of the Fisher distribution for shapes. Further, Goodall & Mardia (1993) have obtained shape distributions in higher dimensions following the marginal approach.

## 4.2 The Fisher Distribution for Shapes of Triangles

Let $w_1, w_2$ and $w_3$ be the three complex variables corresponding to vertices $A, B, C$ of a triangle. Define the variables

$$\begin{aligned} z_0 &= (w_1 + w_2 + w_3)/\sqrt{3} \\ z_1 &= (w_2 - w_1)/(\sqrt{2} \parallel \mathbf{Hw} \parallel) \\ z_2 &= (2w_3 - w_1 - w_2)/(\sqrt{6} \parallel \mathbf{Hw} \parallel) \end{aligned} \tag{4.2}$$

where

$$\parallel \mathbf{Hw} \parallel^2 = \frac{1}{2}(w_2 - w_1)^2 + \frac{1}{6}(2w_3 - w_1 - w_2)^2.$$

We have $|z_1|^2 + |z_2|^2 = 1$, i.e. $z_1$ and $z_2$ are preshape variables. Now define the spherical shape variables

$$t_1 = |z_1|^2 - |z_2|^2,\ t_2 = \bar{z}_1 z_2 + z_1 \bar{z}_2,\ t_3 = i(z_1 \bar{z}_2 - \bar{z}_1 z_2) \tag{4.3}$$

so that

$$t_1^2 + t_2^2 + t_3^2 = 1,$$

and $(t_1, t_2, t_3)$ lie on a sphere. Let

$$t_1 = \sin\theta\cos\phi, t_2 = \sin\theta\sin\phi, t_3 = \cos\theta\ , 0 \le \theta \le \pi, 0 \le \phi \le 2\pi. \tag{4.4}$$

These variables are Kendall's spherical shape variables. The above transformation of $(e^{i\alpha}z_1, e^{i\alpha}z_2)$ for any $\alpha$ to $(t_1, t_2, t_3)$ gives an explicit identification of $\mathbb{C}P^1$ to $S^2$ as described earlier after (2.1). Fig.5 shows triangle shapes located on the spherical shape space. If we map $A$ to $(-1/\sqrt{3}, 0)$ and $B$ to $(1/\sqrt{3}, 0)$ then the Cartesian coordinates of $C$ are $(U, V)$:

$$U = (\sin\theta\sin\phi)/(1 + \sin\theta\cos\phi), V = (\cos\theta)/(1 + \sin\theta\cos\phi). \tag{4.5}$$

Note that the equilateral triangles correspond to the 'North pole' and the 'South pole' and the flat triangles (the triangles which are almost collinear) lie around the equator. The relabelling of the triangles divides the sphere into six lunes of equal area e.g. one of the lunes is $0 \le \theta \le \pi, 0 \le \phi \le \pi/3$.

Insert Fig.5

Further, reflecting a triangle in one of its sides reflects the corresponding point on the shape sphere in the equator. Thus, we can reduce the distinct triangles to half a lune. One of the half-lunes is given by $0 < \theta < \pi/2, 0 < \phi < \pi/3$. Fig.6 shows an equal area projection of the half-lune using a Schmidt net. Thus if invariance under labelling is required, we would be restricted to a half-lune.

Because the space of all shapes of triangles is a shape sphere, an appropriate distribution for shapes $(t_1, t_2, t_3)$ of triangles can be taken as a Fisher distribution $F(\boldsymbol{\mu}, k)$. This fact was exploited by Mardia (1989).

Various hypothesis related to triangle shape are of interest.

**(i)** The hypothesis that the triangles are 'uniformly' distributed corresponds to $\kappa = 0$.

**(ii)** The hypothesis that the triangles are equilateral, implies that $\boldsymbol{\mu} = (0, 0, 1)^T$. Since the triangles are labelled, we have taken without any loss of generality, $\boldsymbol{\mu} = (0, 0, 1)^T$ instead of $\boldsymbol{\mu} = (0, 0, -1)^T$.

**(iii)** Similarly, if $\boldsymbol{\mu}$ corresponds to an isosceles triangle, we can relabel the triangles so that $\phi = 0$ and then $\boldsymbol{\mu} = (\sin\theta, 0, \cos\theta)$ for some $\theta$.

**(iv)** The hypothesis that the triangles are flat, implies $\mu^2 + \nu^2 = 1$ or $\eta = 0$ where $\boldsymbol{\mu}^T = (\mu, \nu, \eta)$.

Example 5 Bookstein (1991, p.406) gives landmarks for 20 triangles for the microscopic fossil data of Lohmann (actually each observation is a mean value). Table 1 gives the Bookstein shape variables $(u_i, v_i), i = 1, ..., 20$ extracted from the figure and after shifting the base line to $(-\frac{1}{2}, 0), (\frac{1}{2}, 0)$. Using the relationship

$$t_1 = (1 - r^2)/(1 + r^2), t_2 = (4/\sqrt{3})u/(1 + r^2), t_3 = (4/\sqrt{3})v/(1 + r^2), r^2 = 4(u^2 + v^2)/\sqrt{3}$$

between the Bookstein shape variables $(u, v)$ and the Kendall spherical shape variables $(t_1, t_2, t_3)$ given at (13.4.3), we obtain corresponding $(\theta, \phi)$. These values are given in Table 1.

Insert Table 1

We find that

$$\bar{R}_x = 0.297, \bar{R}_y = -0.153, \bar{R}_z = 0.932 .$$

Thus the mean resultant length is

$$\bar{R} = 0.99$$

The mean directional vector is

$$(0.30, -0.15, 0.94) .$$

Note $\bar{R}$ is quite large and the hypothesis of uniformity is rejected. Indeed $\hat{\kappa} = 100$ so that the data set is highly concentrated. We can test the hypothesis that the shape is equilateral, i.e.

$$H_0 : \boldsymbol{\mu} = (0, 0, 1)^T.$$

We have

$$(n - 1)(\bar{R} - \bar{R}_z)/(1 - \bar{R}_z) \sim F_{2,2n-2}.$$

Here $F_{2,40} = 16.2$, the 0.1% of $F_{2,40}$ is 8.3. Hence this hypothesis is clearly rejected. ■

Goodall and Lange (1989) give growth curve models for correlated triangle shapes using the shape sphere with local linear Gaussion approximation.

Unlabelled triangles

We now consider labelling triangles so that $AB \geq AC \geq BC$. Thus unlabelled triangles are restricted to just 1/12 of the sphere, a half-lune with "vertices" at

$$L(1,0,0), M(1/2, \sqrt{3}/2, 0), N(0,0,1)$$

i.e. at $L$, the vertices $A = C$, at $M, B = C$ and $N$ corresponds to the equilateral triangle.

Kendall (1983, 1984) uses Lambert's equivalent cylindrical projection to obtain a 'spherical blackboard'. Consider the point $P$ such that $LP = MP = NP$. Basically points on the half lune are mapped to the cylinder touching the sphere at $P$, and this cylinder is unwrapped, with $P$ at the origin to give the blackboard, a bell shaped region of $\mathbb{R}^2$ suitable for the representation of triangles as points in the bell. Coordinates on the bell (X,Y) can be shown explicitly to be (Mardia & Walder, 1988):

$$X = \tan^{-1}\left\{\frac{\sqrt{7}}{2}\frac{-x+\sqrt{3}y}{2\sqrt{3}x+y+\sqrt{3}z}\right\},$$

$$Y = \frac{1}{2\sqrt{7}}\left\{-3x-\sqrt{3}y+4z\right\}.$$

Figure 7 is analogous to Figure 6 giving the triangle shapes on the bell; it is somewhat a matter of personal preference which representation is used.

Insert Fig.6

Example 6 Central place theory is concerned with the pattern of human settlement (see, for example, Okabe et al, 1993). It states that towns tend to occur on a hexagonal lattice.

Mardia, Edwards and Puri (1977) and Mardia (1989) consider the shapes of Delaunay triangles formed by the towns as the points in $\mathbb{R}^2$. They point out that if the points are almost on a regular lattice then the proportion of nearly equilateral triangles will be higher than it would if the points were generated by a Poisson process, i.e. "randomly". Fig.9 shows the data (63 triangles) from 44 towns in Iowa State. In Fig.8, it is also plotted on a Schmidt net.

Insert Fig.8

Insert Fig.9

For $l = (0,0,1)$, for the unlabelled triangle is

$$6f(\theta,\phi;\kappa)d\theta d\phi,\ 0 < \theta < \pi, 0 < \phi < \pi/3$$

where $f(\theta,\phi;\kappa)$ is the p.d.f. of the Fisher distribution.

If we identify the unlabelled triangle with its reflection, then the unlabelled triangle can be restricted to a half-lune, say $0 < \theta < \pi/2, 0 < \phi < \pi/3$. The 'folded' p.d.f. is

$$f(\theta,\phi) = 6c(\kappa)\{f(\theta,\phi;\kappa) + f(\theta,\phi;-\kappa)\}, 0 < \theta < \pi/2,\ 0 < \phi < \pi/3.$$

For totally unlabelled triangles, it may be appropriate to consider a Dimroth-Watson distribution on $0 < \theta < \pi/2, 0 < \phi < \pi/3$ as the null distribution. Thus to test $H_0 : \mu = (0, 0, 1)$, we should use various statistics for the data on a half-lune are given by

$$\bar{R}_x = 0.467, \bar{R}_y = 0.299, \bar{R}_2 = 0.765, \bar{R} = 0.945.$$

The value of $\bar{R}$ indicates that the data is highly concentrated but note that Mardia (1989) in large simulation studies show that for the $(\theta, \phi)$ uniformly distributed on the half-lune $\bar{R} = 0.901$, and under a Poisson configuration $\bar{R} = 0.927$.

The entries for the matrix of sum of squares and products are given by

$$\Sigma l_x^2 = 15.92, \Sigma l_y^2 = 8.34, \Sigma l_z^2 = 38.74$$

$$\Sigma l_x l_y = 9.61, \Sigma l_x l_z = 20.84, \Sigma l_y l_z = 12.74.$$

The eigenvalues and the corresponding eigenvectors are

$$l_3 = 56.34,\ l_2 = 5.03,\ l_1 = 1.63$$

and

$$(0.49, 0.31, 0.81), (-0.54, -0.62, 0.57), (0.68, -0.72, -0.14)$$

Now from (Mardia, 1972)

$$(n-2)(l_3 - \boldsymbol{\mu}_0^T \mathbf{T} \boldsymbol{\mu}_0)/2(n - l_3) \sim F_{2,n-2}.$$

We have

$$l_3 = 56.34,\ \boldsymbol{\mu}_0^T \mathbf{T} \boldsymbol{\mu}_0 = 38.74$$

so that the above statistic is 80.6, indicating no evidence of "equilateralness".

There is very little evidence for the Central Place Theory. Indeed, the configuration of Delaunay triangles in Fig.8 do reinforce this conclusion. This data set was first analysed by Mardia et al (1977) using some intractable von Mises distributions. Also note that these analyses assume that the triangles are independent which is not a realistic assumption. However, from results of Dryden et al (1995), variances are higher if we allow for dependence and therefore the tests under independence are conservative.

Mardia (1989) provides some further details, eg. on the effect of "triangles on the boundary". Kendall (1989) argues that the size of triangles is also important and provides a new method of analysis.

### 4.3 The Complex Bingham Distribution

We have noted above that the pre-shape

$$\mathbf{z} = (z_1, z_2, ..., z_k)$$

lies on a complex sphere $\mathbb{C}S^{k-1}$, i.e. $\mathbf{z}^*\mathbf{z} = 1$, where $\mathbf{z}^*$ is the complex conjugate transpose of $\mathbf{z}$. One way to construct an appropriate distribution is to condition the complex multivariate normal distribution $\mathbb{C}N_k(\mathbf{0}, \mathbf{A})$ where $\mathbf{A}$ is Hermitian, i.e. to condition the pdf

$$C \exp(-\frac{1}{2}\mathbf{z}^* \mathbf{A} \mathbf{z})$$

where $\mathbf{A}$ is Hermitian (i.e. $\mathbf{A} = \mathbf{A}^*$). Conditioning it on $\mathbf{z}^*\mathbf{z} = 1$ gives rise to the complex Bingham distribution. For a similar construction of the real Bingham distribution (see, Mardia, 1990, Ch.8).

Definition The complex Bingham distribution (Kent, 1994) on $\mathbb{C}S^{k-1}$ (or specifically on $\mathbb{C}P^{k-1}$) is

$$f(\mathbf{z}) = C(\mathbf{A})^{-1} \exp(\mathbf{z}^*\mathbf{A}\mathbf{z}) \ , \ \mathbf{z} \in \mathbb{C}S^{k-1}, \tag{4.6}$$

where the matrix $\mathbf{A}$ is $k \times k$ Hermitian and $C(\mathbf{A})$ is the normalizing constant. We write

$$\mathbf{z} \sim \mathbb{C}B_{k-1}(\mathbf{A}).$$

The complex Bingham distribution is analogous to the real Bingham distribution and in fact, as is shown below, it is a particular case. Note that the distribution has the property that

$$f(e^{i\theta}\mathbf{z}) = f(\mathbf{z}) \tag{4.7}$$

and is thus invariant under rotation of the preshape $\mathbf{z}$ i.e. it lies on $\mathbb{C}P^{k-1}$. This property therefore makes the distribution suitable for shape analysis (location and scale are already removed as $\mathbf{z}$ is on the preshape sphere). Kent (1994) proposed this distribution for the analysis of two dimensional shape. He also commented on its limitations for its uses in shape analysis since imposing a complex symmetry assumption might not be realistic. However, it does provide a tractable model for shape data.

Since $\mathbf{z}^*\mathbf{z} = 1$ for $\mathbf{z} \in \mathbb{C}P^{k-1}$, it can be checked that the parameter matrices $\mathbf{A}$ and $\mathbf{A} + \alpha\mathbf{I}$ define the same complex Bingham distribution with $C(\mathbf{A} + \alpha\mathbf{I}) = C(\mathbf{A}) \exp \alpha$ for any complex to number $\alpha$. It is convenient to remove this non-identifiability by setting $\lambda_{\max}(\mathbf{A}) = 0$ where $\lambda_{\max}(\mathbf{A})$ denotes the largest eigenvalue of $\mathbf{A}$. Let $\boldsymbol{\gamma}_1, ..., \boldsymbol{\gamma}_k$ denote the normalized eigenvectors of $\mathbf{A}$ so that

$$\boldsymbol{\gamma}_j^*\boldsymbol{\gamma}_j = 1, \boldsymbol{\gamma}_i^*\boldsymbol{\gamma}_j = 0 (i \neq j), \mathbf{A}\boldsymbol{\gamma}_j = \lambda\boldsymbol{\gamma}_j, j = 1, ..., k.$$

Note that each $\boldsymbol{\gamma}_j$ is defined only up to rotation by a unit complex number. If $\lambda_1 = ... = \lambda_k = 0$, the distribution reduced to the uniform distribution on $\mathbb{C}\mathbf{P}^{k-1}$. Provided $\lambda_{k-1} < 0$, it can be seen that $\mathbf{z} = \boldsymbol{\gamma}_k$ maximizes the density and $\boldsymbol{\gamma}_k$ is unique up to a scalar rotation $\exp(i\theta)\boldsymbol{\gamma}_k$. If $\lambda_1, ..., \lambda_{k-1}$ are far below zero, the distribution is highly concentrated about this modal axis.

**(a)** The normalizing constant (Kent, 1994). Let $\lambda_1, ..., \lambda_k$ be the eigenvalues of $\mathbf{A}$, $\lambda_1 < \lambda_2 < ... < \lambda_{k-1} < \lambda_k = 0$. Then

$$C(\mathbf{A}) = 2\pi^{k-1} \sum_{j=1}^{k-1} a_j \exp(\lambda_j), \quad a_j^{-1} = \Pi_{i \neq j}^{k} (\lambda_j - \lambda_i) \tag{4.8}$$

*Proof:* Since $\mathbf{A}$ is Hermitian, we can write $\mathbf{A} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^*$ where $\mathbf{U}$ is unitary ($\mathbf{U}^*\mathbf{U} = \mathbf{I}_k$) and $\boldsymbol{\Lambda} = \text{ diag}(\lambda_1, .., \lambda_k)$ where $\lambda_1, ..., \lambda_k$ are real.

If $\mathbf{z} \sim CB_{k-1}(\mathbf{A})$, we have $\mathbf{U}^*\mathbf{z} \sim CB_{k-1}(\boldsymbol{\Lambda})$. These two distributions have the same normalization constant $C(\mathbf{A}) = C(\boldsymbol{\Lambda})$. Consider the transformation of $x_1, y_1, ..., x_k, y_k$ on $\mathbb{R}^{2k}$

$$x_j = t_j^{\frac{1}{2}} \cos\theta_j, \ y_j = t_j^{\frac{1}{2}} \sin\theta_j, \ t_j > 0, \ 0 < \theta < 2\pi. \tag{4.9}$$

Hence

$$dx_1 dy_1 ... dx_k dy_k = 2^{-k} dt_1 d\theta_1 ... dt_k d\theta_k. \tag{4.10}$$

Now change $t_1, ..., t_k$ to $s_1, ..., s_{k-1}, r$ by

$$t_j = r^2 s_j, j = 1, ..., k-1, \quad t_k = r^2(1 - s_1 - ... - s_{k-1})$$

so that

$$t_1 + ... + t_k = r^2.$$

Thus

$$dx_1 dy_1 ... dx_k dy_k = 2^{-k+1} r^{2k-1} d\theta_1 ... d\theta_k ds_1 ... ds_{k-1} dr,$$

which on fixing $r = 1$ yields the volume measure on $CS^{k-1}$ as

$$2^{-k+1} d\theta_1 ... d\theta_k ds_1 ... ds_{k-1}.$$

Also,

$$\mathbf{z}^* \mathbf{A} \mathbf{z} = \Sigma \lambda_j \mathbf{z}_j^* \mathbf{z}_j = \Sigma \lambda_j (x_j^2 + y_j^2) = \sum_{j=1}^{k} \lambda_j s_j,$$

where

$$s_k = 1 - s_1 - ... - s_{k-1}.$$

Note that $(\theta_1, ..., \theta_k)$ and $(s_1, ..., s_k)$ are independently distributed. Further, the $\theta_i$ are independent uniform, whereas $(s_1, ..., s_{k-1})$ has a truncated exponential distribution (see (4.30) below). Denote the $k$-dimensional unit simplex in $\mathbb{R}^k$ by

$$\mathcal{S}_k = \{\mathbf{s} = (s_1, ..., s_k)^T, s_k \geq 0 \text{ and } \sum_{j=1}^{k} s_j \leq 1\}.$$

Hence

$$\begin{aligned} C(\mathbf{\Lambda}) &= 2^{-k+1} \int_{S_k} \int_{[0,2\pi]^k} \exp\{\textstyle\sum_{j=1}^{k} \lambda_j s_j\} d\theta_1 ... d\theta_k ds_1 .... ds_{k-1} \\ &= 2\pi^k \int_{\mathcal{S}_k} \exp\{\textstyle\sum_{j=1}^{k} \lambda_j s_j\} ds_1 ... ds_{k-1}. \end{aligned} \tag{4.11}$$

Let

$$\alpha \mathcal{S}_{k-1} = \{(s_1, ..., s_{k-1})^T; s_j \geq 0 \text{ and } \sum_{j=1}^{k-1} s_j \leq \alpha\}$$

denote the "scaled" unit simplex. We now prove a general result

$$I_k(\lambda_1, ..., \lambda_k; \alpha) = \int_{\alpha \mathcal{S}_{k-1}} \exp\left(\sum_{j=1}^{k} \lambda_j s_j\right) ds_1 ... ds_{k-1} = \sum_{j=1}^{k} a_j \exp(\alpha \lambda_j) \tag{4.12}$$

For our result, $\alpha = 1$.

*Proof:* Proof is by induction. For $k = 2$,

$$\int_0^{\alpha} \exp\{\lambda_1 s_1 + \lambda_2 (1 - s_1)\} ds_1 = \sum_{j=1}^{2} a_j e^{\alpha \lambda_j},$$

where $a_1^{-1} = \lambda_2 - \lambda_1,\ a_2^{-1} = \lambda_1 - \lambda_2$. Hence the result holds.

For $k > 2$, we proceed as follows. Let

$$b_j^{-1} = \Pi_{i \neq j}^{k}(\lambda_j - \lambda_i),\ j = 2, ..., k-1.$$

Then using

$$\sum_{j=1}^{k} \lambda_j s_j = \lambda_1 s_1 + \sum_{j=2}^{k} \lambda_j s_j,\ \alpha \mathcal{S}_{k-1} = \{0 < s_1 < \alpha, (\alpha - s_1)\mathcal{S}_{k-2}\}$$

we have

$$\begin{aligned} I_k(\mathbf{\Lambda}; \alpha) &= \int_0^\alpha \exp(\lambda_1 s_1) I_{k-1}(\lambda_1, ..., \lambda_{k-1}; \alpha - s_1) ds_1 \\ &= \int_0^\alpha \exp(\lambda_1 s_1)[\textstyle\sum_{j=2}^{k} b_j \exp\{\lambda_j(\alpha - s_1)\}] ds_1. \end{aligned}$$

On integrating with respect to $s_1$

$$I_k(\mathbf{\Lambda}; \alpha) = \sum_{j=2}^{k} b_j(\lambda_1 - \lambda_j)^{-1}\{\exp(\alpha\lambda_1) - \exp(\alpha\lambda_j)\}. \tag{4.13}$$

We can see in (4.12) that

$$-\sum_{j=2}^{k} b_j(\lambda_1 - \lambda_j)^{-1} \exp(\alpha\lambda_j) = \sum_{j=2}^{k} a_j \exp(\alpha\lambda_j).$$

To show the following equality in (4.12)

$$\sum_{j=2}^{k} b_j(\lambda_1 - \lambda_j)^{-1} = a_1$$

use the partial fraction expansion as a function of $\lambda_1$. Hence the proof is complete.

If equalities exist between some $\lambda_i$'s then $C(\mathbf{A})$ can be obtained using repeated applications of L'Hôpital's rule. For example, for $\lambda_{k-1} = \lambda_k$ but all other $\lambda$'s distinct, we get (Bingham et al, 1992)

$$C(\mathbf{A}) = (k-1)! \left[ \sum_{i=1}^{k-2} \frac{e^{-\lambda_i}}{\Pi_{j \neq i}(\lambda_j - \lambda_i)} + \frac{e^{-\lambda_k}\{1 - \sum_{i=1}^{k-2}(\lambda_i - \lambda_k)^{-1}\}}{\Pi_{j \leq k-2}(\lambda_j - \lambda_k)} \right]. \tag{4.14}$$

**(b)** The $(k-1)$ dimensional complex Bingham distribution can be regarded as a special case of a $(2k-1)$ dimensional real Bingham distribution as follows. If $\mathbf{z}_j = x_j + iy_j$, define a $2k$-dimensional vector $\mathbf{u} = (x_1, y_1, ..., x_k, y_k)^T = V(\mathbf{z})$, say, by splitting each complex number into its real and imaginary parts. Also, if $\mathbf{A} = (a_{hj})$ has entries $a_{hj} = \alpha_{hj} \exp(i\theta_{hj})$ with $\theta_{jh} = -\theta_{hj}$, define a $2k \times 2k$ matrix $\mathbf{B}$ made up of $k^2$ $2 \times 2$ blocks given by

$$\mathbf{B}_{hj} = \alpha_{hj} \begin{pmatrix} \cos\theta_{hj} & -\sin\theta_{hj} \\ \sin\theta_{hj} & \cos\theta_{hj} \end{pmatrix}, -\pi < \theta_{hj} < \pi.$$

Then $\mathbf{z}^*\mathbf{A}\mathbf{z} = \mathbf{u}^T\mathbf{B}\mathbf{u}$ so that a complex Bingham distribution for $\mathbf{z}$ is equivalent to a real Bingham distribution for $\mathbf{u}$. Equivalently, we can write

$$\mathbf{u}^T\mathbf{B}\mathbf{u} = \Sigma\alpha_{ij}(x_ix_j + y_iy_j) + \Sigma\beta_{ij}(x_jy_i - x_iy_j)$$

which after diagonalization reduces to

$$\mathbf{u}^T\mathbf{B}\mathbf{u} = \sum_{i=1}^{k} \lambda_i^2(x_i^2 + y_i^2). \tag{4.15}$$

Thus the complex Bingham distribution is the same as the real distribution on $S^{2k-1}$ with canonical parameter matrix diag $(\lambda_1, \lambda_1; \lambda_2, \lambda_2; ....; \lambda_k, \lambda_k)$. Also, because the surface measure on $S^{2k-1}$ is invariant under the orthogonal group $O(2k)$, it is also invariant under the unitary group $U(k)$.

**(c)** Relationship with the Fisher Distribution In Section 4.2, we introduced the Fisher distribution for triangles. Following Kent (1994), we now establish links between the circular complex Bingham distribution and the real Fisher distribution. First, consider for $k = 2$ which is Hermitian, so that there are four independent variables. Since we can replace $\mathbf{A}$ by Hermitian $\mathbf{A} + \alpha\mathbf{I}$ where $\alpha$ is real, thus there are only three independent real variables so that we may reduce the rank of $\mathbf{A}$ to 1. Hence we can take

$$\mathbf{A} = \lambda \begin{bmatrix} u \\ v \end{bmatrix} [u, v^*], u^2 + |v|^2 = 1. \tag{4.16}$$

Further, we can choose $\lambda > 0$. (An equivalent choice of $\mathbf{A}$ is to replace $\lambda$ by $-\lambda$ and $(u, v)$ by $(-v, u), \lambda > 0, u \geq 0$.) Further, rotating the eigenvector $(u, v)$ does not alter $\mathbf{A}$ so that we can select $u > 0$. Hence we can write

$$u = \cos\zeta, v = \sin\zeta(\cos\eta + i\sin\eta), 0 < \zeta < \frac{\pi}{2}, 0 < \eta < 2\pi. \tag{4.17}$$

Similarly, we can transform $(z_1, z_2)$ to $(\psi, \theta, \phi)$ by

$$z_1 = e^{i\psi}\cos\theta, z_2 = e^{i\psi}\sin\theta(\cos\phi + i\sin\phi), 0 < \theta < \pi/2, 0 < \phi < 2\pi; 0 < \psi < 2\pi \tag{4.18}$$

satisfying $|z_1|^2 + |z_2|^2 = 1$. To obtain the Jacobian of this transformation, consider the transformation of $\mathbb{R}^4$ by

$$z_1 = r\cos\theta e^{i\psi}\ ,\ z_2 = r\sin\theta e^{i(\psi+\phi)}, \tag{4.19}$$

where $r > 0, 0 < \theta < \pi/2, 0 < \psi, \phi < 2\pi$ and $|z_1|^2 + |z_2|^2 = r^2$. To find the Jacobian of this transformation, a convenient way is to define $\zeta = (\psi + \phi) \bmod 2\pi, 0 < \zeta < 2\pi$ so that in real coordinates

$$x_1 = r\cos\theta\cos\psi, x_2 = r\cos\theta\sin\psi, x_3 = r\sin\theta\cos\zeta, x_4 = r\sin\theta\sin\zeta.$$

After some algebra, it is found that the Jacobian of this transformation is $|J| = r^2\sin\theta\cos\theta$. Thus the volume element is $dx_1dx_2dx_3dx_4 = \frac{1}{2}r^3\sin 2\theta d\theta d\theta d\phi d\psi$. Hence the uniform probability element of $(\theta, \phi, \psi)$ is $c\sin 2\theta d\theta d\phi d\psi$ . It is seen that $c = (2\pi^2)^{-1}$. Hence the uniform probability element on $S^3$ in $(\theta, \phi, \psi)$ representation is

$$(4\pi^2)^{-1}\sin 2\theta d\theta d\phi d\psi. \tag{4.20}$$

We have from (4.3.15) - (4.3.17)

$$\mathbf{z}^*\mathbf{A}\mathbf{z} = \lambda|(u, v^*)\begin{pmatrix} z_1 \\ z_2 \end{pmatrix}|^2$$

$$= \lambda\{\cos^2\zeta\cos^2\theta + \sin^2\zeta\sin^2\theta + 2\sin\zeta\cos\zeta\sin\theta\cos\theta\cos(\phi-\eta)\}. \quad (4.21)$$

Using (4.17) in (4.2), the spherical shape vector becomes

$$\mathbf{t} = (\cos 2\theta, \sin 2\theta\cos\phi, \sin 2\theta\sin\phi)^T. \quad (4.22)$$

Define similarly from $(u, v)$ a new vector

$$\boldsymbol{\mu} = (\cos 2\zeta, \sin 2\zeta\cos\eta, \sin 2\zeta\sin\eta)^T. \quad (4.23)$$

Using (4.22) and (4.23) together with (4.21) it can be shown that

$$\mathbf{z}^*\mathbf{A}\mathbf{z} = \frac{\lambda}{2}(1 + \boldsymbol{\mu}^T\mathbf{t}). \quad (4.24)$$

Further, using the Jacobian (4.19) and (4.23), the probability element of $(\theta, \phi, \psi)$ becomes

$$\tfrac{1}{2}\{c(\lambda)\}^{-1}e^{\mathbf{z}^*\mathbf{A}\mathbf{z}}d\mathbf{z}$$

$$= \frac{1}{2}\{c(\lambda)\}^{-1}\sin 2\theta e^{\frac{1}{2}\lambda(1+\boldsymbol{\mu}^T\mathbf{t})}d\psi d\theta d\phi,\ 0 < 2\theta < \pi, 0 < \phi < 2\pi, -\pi/2 < \psi < \pi/2, \quad (4.25)$$

where

$$c(\lambda) = 2\pi^2(e^\lambda - 1) > /\lambda. \quad (4.26)$$

Hence $\psi$ is independent of $(\theta, \phi)$ and uniformly distributed on $(-\pi/2, \pi/2)$, and $(2\theta, \phi)$ has a Fisher distribution with the mean direction $\boldsymbol{\mu}$ given by (13.4.22) and $\kappa = \lambda/2$. Finally, if we are given $\kappa$ and the mean vector $(\cos 2\zeta, \sin 2\zeta\cos\eta, \sin 2\zeta\sin\eta)$ we can obtain $\lambda$ from

$$\lambda = 2\kappa \quad (4.27)$$

and the Procrustes mean direction from

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \cos\zeta \\ \sin\zeta\cos\eta + i\sin\zeta\sin\eta \end{pmatrix}. \quad (4.28)$$

In practice, we can obtain $\hat{\kappa}$ and $\hat{\boldsymbol{\mu}}^T = (\cos\hat{\psi}, \sin\hat{\psi}\cos\hat{\eta}, \sin\hat{\psi}\sin\hat{\eta})$ from the standard method, and therefore $\hat{\lambda} = 2\hat{\kappa}$ and the Procrustes sample mean is given by (4.22) on replacing $\zeta$ by $\hat{\psi}/2$ and $\eta$ by $\hat{\eta}$. Also note that for $\boldsymbol{\mu} = \mathbf{0}$, from (4.23) $\mathbf{z}^*\mathbf{A}\mathbf{z} = \lambda\cos^2\theta = \frac{\lambda}{2}(1 + \cos 2\theta)$.

## (d) Some Further Properties

1. We have from the normalizing constant section (a) that $\theta_1, ..., \theta_k$ are independently uniform variables on $[0, 2\pi)$ and are independent of $s^T = (s_1, ..., s_{k-1})$.

2. The vector s follows a 'truncated exponential distribution' with pdf

$$f(s) = 2\pi^k C(\mathbf{\Lambda})^{-1} \exp(\sum_{j=1}^{k-1} \lambda_j s_j), \quad \mathbf{s} \in S_{k-1} \tag{4.29}$$

Thus the moment generating function of s

$$E[\exp(\Sigma \delta_j s_j)] = \phi(\boldsymbol{\delta}; \mathbf{\Lambda}) \text{ say}$$

is given by

$$\phi(\boldsymbol{\delta}; \mathbf{\Delta}) = C(\mathbf{\Lambda} + \mathbf{\Delta})/C(\mathbf{\Lambda}) \tag{4.30}$$

where $\boldsymbol{\delta} = (\delta_1, ..., \delta_{k-1})^T \in \mathbb{C}^{k-1}$ and $\mathbf{\Delta} =$ diag $(\boldsymbol{\delta}^T, \mathbf{0})$.

Thus

$$E(s_j) = \{C(\mathbf{\Lambda})\}^{-1} \partial C(\mathbf{\Lambda})/\partial \lambda_j, j = 1, ..., k-1. \tag{4.31}$$

3. To model high concentration, replace $\mathbf{A}$ by $\kappa \mathbf{A}$ so that from (4.29)

$$f(\mathbf{s}) = 2\pi^k C(k\mathbf{\Lambda}) \exp(\kappa \Sigma \lambda_j s_j), \quad \mathbf{s} \in \mathcal{S}_{k-1}. \tag{4.32}$$

As $\kappa \to \infty, u_i = \kappa s_i, i = 1, ..., k-1$ tend to independent exponential

$$\lambda_i^{-1} \exp(\lambda_i u_i) \quad , \quad \lambda_i < 0, u_i > 0.$$

Thus for large $\kappa$,

$$C(\mathbf{\Lambda}) = 2\pi^k \Pi_{j=1}^{k-1} (-\lambda_j)^{-1}. \tag{4.33}$$

As $\lambda_i \to 0$ for all $i$, we have the uniform distribution on $\mathbb{C}P^{k-1}$ or $S^{2k-1}$.

4. Another way to look at the high concentration about the axis $\boldsymbol{\gamma}_k$, is to align $\mathbf{z}$ with $\boldsymbol{\gamma}_k$ and then to project $\mathbf{z}$ onto the tangent plane at $\boldsymbol{\gamma}_k$. Given $\mathbf{z} \in \mathbb{C}S^{k-1}$, we construct $\mathbf{v} \in \mathbb{C}^k$ given by

$$\mathbf{v} = \exp(-i\theta_k)(\mathbf{I} - \boldsymbol{\gamma}_k \boldsymbol{\gamma}_k^*)\mathbf{z}, \theta_k = \arg(\boldsymbol{\gamma}_k^* \mathbf{z}). \tag{4.34}$$

It can be checked that $\theta_k$ and $\mathbf{v}$ are independently distributed with $\theta_k$ uniform on $[0, 2\pi)$. Writing $\mathbf{A} = \sum_{j=1}^{k-1} \lambda_i \boldsymbol{\gamma}_i \boldsymbol{\gamma}_i^*$, we have

$$k\mathbf{z}^* \mathbf{A} \mathbf{z} = k\mathbf{v}^* \mathbf{A} \mathbf{v}$$

and therefore $\mathbf{v}$ has a complex normal distribution in $k-1$ complex dimensions with

$$\mathbf{\Sigma} = (-2\mathbf{A})^+ = -\frac{1}{2} \sum_{j=1}^{k-1} (-\lambda_j)^{-1} \boldsymbol{\gamma}_j \boldsymbol{\gamma}_j^* \tag{4.35}$$

where the last expression is the Moore-Penrose generalized inverse of $-2\mathbf{A}$. Here $\mathbf{\Sigma}$ is singular because $\mathbf{v}$ lies in the tangent plane $\boldsymbol{\gamma}_k^* \mathbf{v} = 0$. Thus $\mathbf{z}$ is $CN(\boldsymbol{\gamma}_k, \mathbf{\Sigma})$ so that $\mathbf{\Sigma}$ is determined by the rest of the eigenvectors apart from $\boldsymbol{\gamma}_k$.

The asymptotic distribution shows that $\mathbf{v}_i$ is distributed as $N_2(\mathbf{0}, \mathbf{I}\sigma_i^2)$. Hence the complex Bingham distribution imposes an isotropic distribution on the marginals. However, it does allow correlation between landmarks. Among various distributions considered so far, this is the most tractable distribution for shape.

(e) Inference

Let $z_1, ..., z_n$ be a random sample from (4.5) where, $n \geq k$. Set

$$\mathbf{S} = \sum_{j=1}^{n} z_j z_j^*$$

to be the $k \times k$ complex sum of squares and products matrix. Suppose that the eigenvalues of $\mathbf{S}$ are positive and distinct, $0 < l_1 < ... < l_k$ and let $\mathbf{g}_1, ..., \mathbf{g}_k$ denote the corresponding eigenvectors. Note that $\Sigma l_j = n$. Let $\boldsymbol{\gamma}_1, ..., \boldsymbol{\gamma}_k$ denote the standardized eigenvectors of $\mathbf{A}$ so that $\boldsymbol{\gamma}_j^* \boldsymbol{\gamma}_j = 1$ and $\mathbf{A}\boldsymbol{\gamma}_j = \lambda_j \boldsymbol{\gamma}_j$.

We can obtain the m.l.e. of $\boldsymbol{\Lambda} = \text{ diag}(\lambda_1, ..., \lambda_k)$ and $\mathbf{g}_1, ..., \mathbf{g}_k$ as in the real Bingham distribution. It can be seen that when the eigenvalues $\lambda_1, ..., \lambda_k$ are distinct then

$$\hat{\boldsymbol{\gamma}}_j = \mathbf{g}_j, j = 1, ..., k \tag{4.36}$$

since the loglikelihood is,

$$L = \Sigma l_j \lambda_j - n \log C(\boldsymbol{\Lambda}),$$

the m.l.e.'s of the eigenvalues are found by solving

$$\frac{\partial \log C(\boldsymbol{\Lambda})}{\partial \lambda_j} = \frac{1}{n} l_j, \; j = 1, ..., k-1.$$

and the dominant eigenvector $\hat{\boldsymbol{\gamma}}_k$ can be regarded as the "average axis" of the data. This estimate is the same as from Procrustes shape analysis (Kent, 1994). ■

Note that under high concentration we have from (4.32)

$$\log C(\Lambda) \simeq \text{ const } - \sum_{j=1}^{k-1} \log(-\lambda_j)$$

so that

$$\hat{\lambda}_j \cong -n/l_j \quad , \quad j = 1, ..., k-1$$

Note that analogously the p.d.f. of the complex Fisher distribution can be defined as

$$\text{const} \exp\{\kappa \text{ Re}(\boldsymbol{\mu}^* \mathbf{z})\}$$

on the complex sphere $\mathbb{C}S^{k-1}$. However $\kappa \boldsymbol{\mu}^* \mathbf{z} = \kappa(\boldsymbol{\mu}^T \mathbf{x} + \boldsymbol{\nu}^T \mathbf{y})$. Thus the complex von Mises-Fisher distribution is the real von Mises-Fisher distribution on $S^{2k-1}$. Thus it produces no new distribution. (The situation is different for the matrix von Mises-Fisher distribution. See Bingham et al (1992).

# 5 The Complex Bingham Distribution and a Test of Uniformity

## 5.1 Test of Uniformity

We now give a test of uniformity against the alternative of a complex Bingham distribution. Thus if shapes are random, we would be able to assess the hypothesis.

Let $\mathbf{z}_1, ..., \mathbf{z}_n$ be a random sample from 'complex uniform' $\mathbf{z}$ on $\mathbb{C}^k$, i.e. $\mathbf{z}\mathbf{z}^* = 1$. Consider the matrix of the sum of squares and products

$$\mathbf{Z} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{z}_i \mathbf{z}_i^*. \tag{5.1}$$

Let $\mathbf{z}_i = \mathbf{x}_i + i\mathbf{y}_i$ then

$$\mathbf{Z} = \mathbf{T} + i\mathbf{U} \ ,$$

where

$$\mathbf{T} = \frac{1}{n}\Sigma(\mathbf{x}_i\mathbf{x}_i^T + \mathbf{y}_i\mathbf{y}_i^T), \mathbf{U} = \frac{1}{n}\Sigma(\mathbf{y}_i\mathbf{x}_i^T - \mathbf{x}_i\mathbf{y}_i^T). \tag{5.2}$$

**Result** Let $\bar{l}_1, ..., \bar{l}_n$ be the eigenvalues of $\mathbf{z}$. For large $n$, we have

$$K = nk(k+1)\sum_{i=1}^{k}(\bar{l}_i - \frac{1}{k})^2 \simeq \chi^2_{k^2-1}, \tag{5.3}$$

where

$$\text{tr } \mathbf{Z} = \sum_{i=1}^{k} \bar{l}_i = 1. \tag{5.4}$$

**Proof** Consider the vector

$$\mathbf{v}^T = (t_{11}, t_{22}, ..., t_{kk}, t_{12}, t_{13}, ..., t_{1k-1}, ..., t_{k-1,k}; u_{12}, u_{13}, ..., u_{k-1,k}).$$

Let $\mathbf{z} = \mathbf{x} + i\mathbf{y}$ so that $x_1, ..., x_k, \ \ y_1, ..., y_k$ are uniform on $S^{2k-1}$. For the uniform distribution, we have $E(x_i^2) = 1/2k, E(x_i x_j) = 0$, so that

$$E(t_{ii}) = E(x_i^2 + y_i^2) = \frac{1}{k}, E(t_{ij}) = E(x_i x_j + y_i y_j) = 0, i \neq j.$$

Similarly, $E(u_{ij}) = 0$.

Hence

$$E(\mathbf{v}^T) = (\mathbf{1}_k^T, \mathbf{0}_p^T, \mathbf{0}_p^T) \tag{5.5}$$

where $\mathbf{1}_k$ is a $k \times 1$ vector of unity, and $\mathbf{0}_p$ is a vector of zeros where $p = k(k-1)/2$. Now $\text{var}(t_{ii}) = \frac{1}{n}\text{ var}(x_i^2 + y_i^2) = \frac{1}{n}E(x_i^2 + y_i^2)^2 - \{E(x_i^2 + y_i^2)\}^2$. Now from the results on the uniform distribution on $S^{2k-1}$ (Ch.9), we have

$$E(x_i^4) = 3\{4k(k+1)\}^{-1}, E(x_i^2 y_i^2) = \{4k(k+1)\}^{-1}.$$

Hence

$$\text{var}(t_{ii}) = a(k-1)/k, a = \{nk(k+1)\}^{-1}, \tag{5.6}$$

$$\text{cov}(t_{ii}, t_{jj}) = \frac{1}{n}\{E(x_i^2 + y_i^2)(x_j^2 + y_j^2) - \{E(x_i^2 + y_i^2)\}^2\} = -a/k. \tag{5.7}$$

Similarly

$$\text{var}(t_{ij}) = \text{var}(u_{ij}) = a/2, \tag{5.8}$$

$$\text{cov}(t_{ij}, t_{i'j'}) = \text{cov}(u_{ij}, u_{i'j'}) = 0, i \neq i', j \neq j', \tag{5.9}$$

and

$$\text{cov}(t_{ij}, u_{i'j'}) = 0 \text{ for all } i, i', j, j'. \tag{5.10}$$

Hence

$$\mathbf{A} = \text{ cov}(\mathbf{v}) = a(\mathbf{I}_k - \mathbf{1}_k\mathbf{1}_k' k^{-1}, \frac{1}{2}\mathbf{I}_p, \frac{1}{2}\mathbf{I}_k). \tag{5.11}$$

Now,

$$\mathbf{A}^- = a^{-1}(\mathbf{I}_k, 2\mathbf{I}_p, 2\mathbf{I}_p) \tag{5.12}$$

and rank of $\mathbf{A}$ is $k - 1 + \{k(k-1)/2\} + \{k(k-1)/2\} = k^2 - 1$.

Hence for large $n$, by the central limit theorem

$$Q = \{\mathbf{v} - E(\mathbf{v})\}^T\mathbf{A}^-\{\mathbf{v} - E(\mathbf{v})\} \simeq \chi^2_{k^2-1} \tag{5.13}$$

or from (5.5) and (5.12), we have

$$Q = a^{-1}\left\{\sum_{i=1}^{k}(t_{ii} - \frac{1}{k})^2 + 2\sum_{i<j} t_{ij}^2 + 2\sum_{i<j} u_{ij}^2\right\},$$

which can be rewritten as

$$Q = a^{-1}\{\text{tr}(\mathbf{T} - \frac{1}{k}\mathbf{I})^2 + \text{ tr } \mathbf{U}^2\} \tag{5.14}$$

or equivalently

$$Q = a^{-1} \text{ tr}(\mathbf{Z} - \frac{1}{k}\mathbf{I})(\mathbf{Z} - \frac{1}{k}\mathbf{I})^*. \tag{5.15}$$

Hence the result follows from (5.13) and (5.15).

Note that $E[\text{tr}(\mathbf{Z} - \frac{1}{k}\mathbf{I})(Z - \frac{1}{k}\mathbf{I})^*] = (k-1)/(nk)$ here is the same as for the real Bingham Case in $\mathbb{C}^k$. Also, using the canonical representation of the complex Bingham distribution, it can be seen that (5.3) is a score statistic.

## 5.2 The case of $\mathbb{C}P^1$

For $k = 2$, we note from (5.3) that

$$6n\{(l_1 - \frac{1}{2})^2 + (l_2 - \frac{1}{2})^2\} = 3n(l_1 - l_2)^2 \simeq \chi_3^2 \tag{5.16}$$

since $l_1 + l_2 = 1$. This result can be derived directly using the relationship of the complex Bingham to the Fisher distribution. We have by definition

$$\bar{R}^2 = \bar{t}_1^2 + \bar{t}_2^2 + \bar{t}_3^2$$

where

$$\bar{t}_i = \sum_{j=1}^{n} t_{ij}/n, i = 1, 2, 3; z_{1j} = x_{1j} + iy_{1j}, z_{2j} = x_{2j} + iy_{2j},$$

and

$$t_{ij} = |Z_{1j}|^2 - |Z_{2j}|^2, t_{2j} = 2 \text{ Re } (\bar{Z}_{1j} Z_{2j}), t_{3j} = 2 \text{ Im}(\bar{Z}_{1j} Z_{2j}).$$

If $n\mathbf{U}$ and $n\mathbf{V}$ are the sum of squares and products matrices for $(x_1, x_2)$ and $(y_1, y_2)$ respectively, and

$$w_{12} = \Sigma x_{1j} y_{2j}/n, \; w_{21} = \Sigma x_{2j} y_{1j}/n$$

then we have

$$\bar{R}^2 = (u_{11} + v_{11} - u_{22} - v_{22})^2 + 4(u_{12} - v_{12})^2 + 4(w_{12} + w_{21})^2. \tag{5.17}$$

From the standard results (see Ch.9), we have under uniformity

$$3n\bar{R}^2 \simeq \chi_3^3. \tag{5.18}$$

Now $l_1$ and $l_2$ are the eigenvalues of

$$\mathbf{Z} = \frac{1}{n} \sum_{i=1}^{n} (Z_{1i}, Z_{2i}) \begin{pmatrix} Z_{1i} \\ Z_{2i} \end{pmatrix} = \mathbf{S}_1 + i\mathbf{S}_2$$

where

$$\mathbf{S}_1 = \begin{pmatrix} u_{11} + v_{11} & u_{12} + v_{12} \\ u_{12} + v_{12} & u_{22} + v_{22} \end{pmatrix}, \mathbf{S}_2 = \begin{pmatrix} 0 & w_{21} - w_{12} \\ -w_{21} + w_{12} & 0 \end{pmatrix}.$$

We find that on writing the characteristic equation for $\mathbf{S}_1 + i\mathbf{S}_2$,

$$l_1 + l_2 = u_{11} + v_{11} + u_{22} + v_{22},$$

$$l_1 l_2 = (u_{11} + v_{11})(u_{22} + v_{22}) - \{(u_{12} + v_{12})^2 - (w_{21} - w_{12})^2\}^2,$$

Hence (5.17) can be written as

$$\bar{R}^2 = (l_1 - l_2)^2. \tag{5.19}$$

Thus, from (5.18) we have another derivation of (5.16).

Example 7 For the $T2$ bone data ($k = 5, n = 23$), we have given the eigenvalues in Example 4 so that

$$\bar{l}_1 = \bar{l}_2 = 0, \; \bar{l}_3 = 0.001, \; \bar{l}_4 = 0.003, \; \bar{l}_5 = 0.996 \; .$$

We have $K = 49.7$ with $\chi^2_{24}$. Now

$$\Pr(\chi^2_{24} > 45.6) = 0.005, \; \Pr(\chi^2_{24} > 51.2) = 0.001.$$

Hence there is evidence of non-uniformity. In view of the large values of $\bar{l}_5$, this result is obvious.

The Procrustes mean obtained before is now the maximum likelihood estimator of the population dominant eigenvector under the complex Bingham distribution. For large concentration, from (4.36),

$$\begin{aligned} \hat{\text{cov}}(\hat{\mathbf{v}}) &= \sum_{j=1}^{k-1} -\frac{1}{2l_j} \hat{\boldsymbol{\gamma}}_j \hat{\boldsymbol{\gamma}}_j^* = \frac{1}{n} \sum_{j=1}^{k-1} l_j \mathbf{g}_j \mathbf{g}_j^* \\ &= \frac{1}{n} (\mathbf{S} - l_k \mathbf{g}_k \mathbf{g}_k^*). \end{aligned}$$

Thus we can obtain cov $(\mathbf{H}\hat{\mathbf{v}})$ to get some idea of variances and intercorrelations. It is found from côv$(\mathbf{H}\hat{\mathbf{v}})$ that the variances $(\times 23)$ are

$$0.020, 0.021, 0.005, 0.036, 0.005, 0.007.$$

Thus the landmark 4 has most variance whereas the landmarks 1 and 2 are the next highest. There is almost no variability in the landmarks 3, 5 and 6 but note that the complex symmetry imposes isotropy at each landmark.

The estimated correlation matrix $\mathbf{H}\hat{\mathbf{v}}$ is found to be

$$\begin{bmatrix}
1 & & & & & \\
0.05-0.76i & 1 & & & & \\
-0.34-0.25i & 0.10-0.29i & 1 & & & \\
-0.60+0.68i & -0.60-0.62i & -0.09-0.38i & 1 & & \\
0.30+0.12i & -0.24+0.26i & -0.36-0.01i & -0.24-0.39i & 1 & \\
-0.35-0.08i & -0.34+0.16i & 0.50-0.08i & 0.08+0.04i & -0.08-0.21i & 1
\end{bmatrix}$$

This indicates that correlations between landmarks are high. Note that the covariance matrix is necessarily singular. From a shape analysis point of view, looking into real covariance matrix for the tangent space to the dominant eigenvector is worth exploring (see, Kent, 1994). The technique also allows correlations within landmarks to be explored - the model here forces zero correlations between $x_i$ and $y_i$ to be asymptotically zero.

## 6 The Complex Angular Gaussian Model

Another suitable model is an extension of the angular Gaussian distribution of Tyler (1987). The p.d.f. has the density (Kent, 1995)

$$f(\mathbf{z}) = |\mathbf{\Sigma}|^{-1}\{?\,(k)/(2\pi^k)\}(\mathbf{z}^*\mathbf{\Sigma}^{-1}\mathbf{z})^{-k}, \mathbf{z} \in \mathbb{C}S^{k-1}, \tag{6.1}$$

where the parameter $\mathbf{\Sigma}$ is a positive definite Hermitian matrix. Note that $f(\mathbf{z}) = f(e^{i\theta}\mathbf{z})$ for all $\theta$ so that $\mathbf{z} \in \mathbb{C}P^{k-1}$.

Also, it is invariant under rescaling $\mathbf{\Sigma} \to c\mathbf{\Sigma}, c > 0$. We can verify that the normalizing constant by the same method as for the complex Bingham distribution.

The mode of the density (6.1) is given by the dominant eigenvector of $\mathbf{\Sigma}$. Hence, it is natural to define an average shape from a set of landmark data by the dominant eigenvector of the MLE of $\mathbf{\Sigma}$. This MLE is given by the solution of the estimating equation,

$$\hat{\mathbf{\Sigma}} = \frac{k}{n}\sum_{j=1}^{n} \mathbf{z}_j\mathbf{z}_j^*/\mathbf{z}_j^*\hat{\mathbf{\Sigma}}^{-1}\mathbf{z}_j, \tag{6.2}$$

so that $\hat{\mathbf{\Sigma}}$ can be viewed as a weighted complex sum of squares and products matrix. This equation can be solved iteratively by the EM algorithm using

$$\mathbf{\Sigma}^{\nu+1} = \frac{1}{n}\sum_{j=1}^{n} v_j^{(\nu)}\mathbf{z}_j\mathbf{z}_j^* \quad , \quad \nu = 1, 2, ... \tag{6.3}$$

where

$$v_j^{(\nu)} = k/\mathbf{z}_j^* \{\boldsymbol{\Sigma}^{(\nu)}\}^{-1} \mathbf{z}_j. \tag{6.4}$$

Under mild regularity conditions the iterations in (5.23) converge to the MLE $\hat{\boldsymbol{\Sigma}}$, which is unique up to scaling. Note in equation (5.23) that outlying values of $z$ get downweighted when forming the complex sum of squares and products matrix. In order to link this result to conventional multivariate $M$-estimators, consider what happens in the case of highly concentrated data. Let the dominant eigenvector of $\boldsymbol{\Sigma}$ be given by $\boldsymbol{\nu} = (1, 0, ..., 0)^T$ for simplicity (any angular central Gaussian distribution can be rotated to this form by a suitable $k$-dimensional unitary transformation), and suppose $\boldsymbol{\Sigma}$ is of the form

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0 \\ 0 & \kappa^{-1}\mathbf{A} \end{bmatrix}, \boldsymbol{\Sigma}^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & \kappa\mathbf{B} \end{bmatrix} \tag{6.5}$$

where $\kappa > 0$ is large and $\mathbf{A}$ is a fixed $(k-1)$-dimensional positive definite matrix. In general, the projection of $\mathbf{z}$ onto the tangent plane of shape space at $\boldsymbol{\nu}$ is given by $\mathbf{z} \to (I - \boldsymbol{\nu}\boldsymbol{\nu}^*)e^{-i\theta}\mathbf{z} = \mathbf{w}$, say, where $\theta = \arg \boldsymbol{\nu}^*\mathbf{z}$. In this case $\mathbf{w} = e^{-i\theta}(z_2, ..., z_k)^T$ and $\theta = \arg z_1$. Under the angular central Gaussian model, $\theta$ is uniformly distributed on $(0, 2\pi)$ independently of $(z_2, ..., z_k)$, and the distribution of $(z_2, ..., z_k)$ is invariant under multiplication by $e^{i\theta}$. Further, if $\kappa$ is large, the distribution of $z$ becomes concentrated near $\nu$, and the distribution can be studied by projection onto the tangent plane. In particular $\sqrt{\kappa}\mathbf{w} = \mathbf{v}$, say, becomes distributed with density

$$\begin{aligned} f(v) &\propto \lim_{\kappa\to\infty}(\mathbf{z}^*\boldsymbol{\Sigma}^{-1}\mathbf{z})^{-k} \\ &= \lim_{\kappa\to\infty}(|z_1|^2 + \mathbf{v}^*\mathbf{B}\mathbf{v})^{-k} \\ &= \lim_{\kappa\to\infty}(1 - \kappa^{-1}\mathbf{v}^*\mathbf{v} + \mathbf{v}^*\mathbf{B}\mathbf{v})^{-k} \\ &= (1 + \mathbf{v}^*\mathbf{B}\mathbf{v})^{-k}\ ,\ \mathbf{v} \in \mathbb{C}^{k-1}. \end{aligned} \tag{6.6}$$

After switching to real coordinates in $\mathbb{R}^{2k-2}$, this density can be viewed as a multivariate $t$-density with $\nu = 2$ degrees of freedom and with scatter matrix given by re-expressing $\frac{1}{2}\mathbf{A}$ in real coordinates (Mardia et al, 1979, p.57). This result gives some insight into the extent to which the angular central Gaussian distribution can accommodate outliers or with outlying objects in shape analysis.

# 7 Shape and Directional Statistics

Consider the QR decomposition (2.4) . Broadly speaking, in shape analysis we are interested only in $\mathbf{T}$ and not in $\mathbf{?} \in V_{k,m}$. However, in directional statistics we are interested only in $\mathbf{?} \in V_{k,m}$, namely, distributions on Stiefel manifold including rotation groups (Ch.12); the special interest is here on $V_{m,m} = O(m)$. Some techniques in directional statistics also play a role in shape analysis. The main approaches described here in this chapter based on the Fisher distribution and the complex Bingham distribution rely on directional statistics in an indirect way. However, from an analytical view-point, the directional statistics and shape analysis are complementary subjects as we now indicate (see also, Bingham et al, 1992). Kent (1992) has investigated how "average shape" and "directional averages" are connected (described above) together with some embeddings of shape space. In particular, techniques for calculating normalising constants for certain directional distributions are the same as those for computing shape densities.

For example, the matrix Fisher distribution (12.5) was introduced by Mardia and Khatri (1977). The normalizing constant is found by integration over $V_{k,m}$ as

$$\int_{V_{k,m}} \exp\{\mathrm{tr}(\mathbf{F}\mathbf{X}^T)\} d\mathbf{X} = {}_0F_1(m/2; \mathbf{F}\mathbf{F}^T/4),$$

where ${}_0F_1(.,.)$ is a matrix hypergeometric function. Alternatively consider the reflection size-and-shape density of Goodall and Mardia (1993), which contains the following integral in its proof:

$$\int_{V_{k,m}} \exp[\mathrm{tr}(\boldsymbol{\mu}^T \mathbf{T} \mathbf{?})] d\mathbf{?} \ \propto \ {}_0F_1(m/2; \mathbf{A}),$$

where $\mathbf{A} = (\boldsymbol{\mu}^T \mathbf{T}\mathbf{T}^T \boldsymbol{\mu})/4$ and $d\mathbf{?}$ denotes Haar measure. Both expressions depend on the hypergeometric function of the same type and order. Thus the **tool** for calculating the marginal reflection size-and-shape distribution was the same as that used for computing the integrating constant above. We summarize the situation as follows:

"In shape rotations are to be integrated out or to be optimized over, whilst they are the basis of directional statistics".

# 8 Discussion

Shape Analysis is a rapidly growing area. D.G. Kendall and his co-workers pioneered deep studies of the geometry of shape spaces using a differential-geometric framework. (See, D.G. Kendall (1977, 1983, 1984), W.S. Kendall (1988, 1990), Le (1991), Le & Kendall (1993), Stoyan et al (1987) and Stoyan and Stoyan (1994). Bookstein (1978) took a morphometric view point giving insight into shape statistics (see, Bookstein, 1991). Goodall (1991) proposed a Procrustes approach to shape analysis. Mardia and Dryden (1989a,b) and Kent (1994) have proposed a model-based approach which is related to directional statistics as described here. Mardia (1995) has given a recent review of shape analysis. This paper gives the viewpoint from directional statistics but of course, it is only a partial view of the field. A recent review has been given by Mardia (1975).

**Acknowledgements**

I am grateful to Peter Jupp and John Kent for their helpful comments.

# References

[1] Bingham, C., Change, T. and Richards, D. (1992). Approximating the matrix Fisher and Bingham distributions: applications to spherical regression and Procrustes analysis. *Jour. Multiv. Anal.*, **41**, 314-337.

[2] Bookstein, F.L. (1978). *The Measurement of Biological Shape and Shape Change*. Springer-verlag, Berlin.

[3] Bookstein, F.L. (1986). Size and shape spaces for landmark data in two dimensions (with discussion). *Statist. Sci.*, **1**, 181-242.

[4] Bookstein, F.L. (1991). *Morphometric Tools for Landmark Data: Geometry and Biology.* Cambridge, Cambridge University Press.

[5] Dryden, I.L., Faghihi, M. and Taylor, C.C. (1995). Investigating regularity in spatial point patterns using shape analysis. In *Current Issues in Statistical Shape Analysis.* Eds. Mardia, K.V. and Gill, C.A., University of Leeds Press, 40-48.

[6] Dryden, I.L. and Mardia, K.V. (1991). General shape distributions in a plane. *Adv. Appl. Prob.,* **23**, 259-276.

[7] Dryden, I.L. and Mardia, K.V. (1993). Multivariate shape analysis. *Sankhya*, **55**, Series A, part 3, 460-480.

[8] Goodall, C.R. (1991). Procrustes methods in the statistical analysis of shape (with discussion). *J. Roy. Statist. Soc.,* B, **53**, 285-339.

[9] Goodall, C.R. and Lange, N. (1989). Growth curve models for correlated triangular shapes. *Proc. 21st Interface*, 445-453.

[10] Goodall, C.R. and Mardia, K.V. (1993). Multivariate aspects of shape theory. *Ann. Statist.,* **21**, 848-866.

[11] Kendall, D.G. (1981). The Statistics of Shape. *Interpreting Multivariate Data* ed. V. Barnett, 75-80. Wiley: Chichester.

[12] Kendall, D.G. (1977). The diffusion of shape. *Adv. in Appl. Prob.,* **9**, 428-430.

[13] Kendall, D.G. (1983). The Shape of Poisson-Delaunay triangles. *Studies in Probability and related topics*. Demetrescu, H.C. and Iosifescu, M. (Eds.). Nagard, Sofia.

[14] Kendall, D.G. (1983). The Shape of Poisson-Delaunay triangles. In Papers in Honour of Octav Onicescv on his 90th birthday (M.C. Demetrescu and M. Iosifescu, eds.) 321-330, Negard, Rome.

[15] Kendall, D.G. (1984). Shape Manifolds, procrustean metrics and complex projective spaces. *Bull. Lond. Math. Soc., bf*, 81-121.

[16] Kendall, D.G. (1989). A survey of the statistical theory of shape (with discussion). Statist. Sc., **4**, 81-120.

[17] Kendall, W.S. (1988). Symbolic computation and the diffusion of shapes of triads. *Adv. in Appl. Probab.* **20**, 775-797.

[18] Kendall, W.S. (1990). The diffusion of Euclidean shape. In *Disorder in Physical Systems* (G. Grimmett and D. Welsh, eds.) Oxford Univ. Press.

[19] Kent, J.T. (1992). New directions in shape analysis. In *The Art of Statistical Science*, K.V. Mardia (ed.). Chichester, Wiley.

[20] Kent, J.T. (1994). The complex Bingham distribution and shape analysis. *Jour. Roy. Statist. Soc. B*, **56**, 285-299.

[21] Khatri, C.G. and Mardia, K.V. (1977). The von Mises-Fisher matrix distribution in orientation statistics. *Jour. Roy. Statist. Soc. B*, **39**, 95-106.

[22] Le, H. (1991). On geodesics in Euclidean shape spaces. *J. London Math. Soc.* 44, 360-372.

[23] Le, H. and Kendall, D.G. (1993). The Riemannian structure of Euclidean shape spaces: A novel environment for statistics. Ann. Statist. **21**, 1225-1271.

[24] Mardia, K.V. (1972). *Statistics of Directional Data.* London, Academic Press.

[25] Mardia, K.V. (1975). Statistics of directional data (with discussion). *Jour. Roy. Statist. Soc. B*, **37**, 349-393.

[26] Mardia, K.V. (1989). Shape analysis of triangles through directional techniques. *Jour. Roy. Statist. Soc. B*, **51**, 449-458.

[27] Mardia, K.V. (1995). Shape advances and future perspectives. In *Current Issues in Statistical Shape Analysis.* Eds. Mardia, K.V. and Gill, C.A. University of Leeds Press, 57-75.

[28] Mardia, K.V., Edwards, R. and Puri, M.L. (1977). Analysis of central place theory (with discussion). *Bull Int. Statist. Inst.,* **47**, 93-110, 139-146.

[29] Mardia, K.V. and Walder, A.N. (1988). On Kendall's spherical blackboard. Research Report Dept. of Statistics, University of Leeds.

[30] Mardia, K.V. and Dryden, I.L. (1989a). Statistical analysis of shape data. *Biometrika*, **76**, 271-281.

[31] Mardia, K.V. and Dryden, I.L. (1989b). Shape distributions for landmark data. *Adv. Appl. Prob.,* **21**, 742-755.

[32] Okabe, A, Boots, B. and Sugihara, K. (1992). Spatial Tessellations: Concepts and Applications of Vornoi Diagrams. John Wiley & Sons, Chichester.

[33] Stoyan, D., Kendall, W.S. and Mecke, J. (1987). *Stochastic Geometry and Its Applications.* Wiley, Chichester.

[34] Stoyan, D. and Stoyan, H. (1994). *Fractals, random shapes and point fields.* Chichester, Wiley. Subrahmonia, J., Keren, D. and Cooper, D.B. (1994) An algebraic and Bayesian technology for object recognition. *Statistics and Images*, Vol.2, ed. Mardia, K.V. Carfax publishing Co., Abingdon, 165-205.

[35] Tyler, D. (1987). Statistical analysis for the angular central Gaussian distribution on the sphere. *Biometrika*, **74**, 579-589.

# Titles of the Figures

**Fig.1** A microfossil Globorotolia truncatulinoides with three landmarks (Bookstein, 1986).

**Fig.2** A mouse vertebra bone (T2) with six landmarks.

**Fig.3** Procrustes mean shape (-) for the T2 bones with data superimposed ($n = 23$).

**Fig.4** Procrustes Mean Shape for the T2 bone data.

**Fig.5** Kendall's spherical shape space for triangles in two dimensions.

**Fig.6** Schmidt net of triangle shapes from a half-lune.

**Fig.7** Kendall's bell representation of triangle shapes.

**Fig.8** Delaunay triangles (63) of 44 Iowa Central Places.

**Fig.9** Schmidt net for the Iowa data $*, \odot, +$ mean directions for the data, simulated under Poisson assumption, simulated under uniform assumption respectively.

**Table 1** The fossil data in Bookstein coordinates $(u, v)$ with Kendall shape variables $(\theta, \phi)$.

| $u$ | $v$ | $\theta$ | $\phi$ |
|---|---|---|---|
| −0.16 | 0.66 | 20.2 | 134.0 |
| −0.13 | 0.65 | 18.9 | 141.1 |
| −0.08 | 0.77 | 9.3 | 132.1 |
| −0.09 | 0.71 | 13.1 | 142.8 |
| −0.16 | 0.75 | 14.3 | 119.5 |
| −0.13 | 0.68 | 16.9 | 137.7 |
| −0.10 | 0.72 | 13.1 | 138.2 |
| −0.08 | 0.73 | 11.3 | 143.0 |
| −0.09 | 0.63 | 19.1 | 152.2 |
| −0.08 | 0.62 | 20.2 | 156.2 |
| −0.08 | 0.65 | 17.8 | 153.3 |
| −0.07 | 0.59 | 22.1 | 160.6 |
| −0.10 | 0.61 | 21.1 | 153.1 |
| −0.08 | 0.63 | 18.9 | 155.6 |
| −0.08 | 0.54 | 26.7 | 161.7 |
| −0.07 | 0.66 | 16.9 | 155.2 |
| −0.05 | 0.57 | 24.2 | 165.8 |
| −0.07 | 0.52 | 28.5 | 163.9 |
| −0.10 | 0.47 | 33.7 | 160.1 |
| −0.07 | 0.44 | 37.1 | 166.3 |