6

Shape Distributions and Inference

Of fundamental interest are probability distributions in shape spaces, which provide models for shape analysis. There are several issues to consider and there are various difficulties to overcome. Since the shape space is non-Euclidean special care is required.

6.1 Uniform Distribution

Result 6.1 *Using Kendall coordinates* $U = (U_3^K, ..., U_k^K, V_3^K, ..., V_k^K)^{\mathrm{T}}$ *we have the 'uniform' density on the shape space given by*

$$f_\infty(u) = \frac{(k-2)!\pi}{\{\pi(1+u^{\mathrm{T}}u)\}^{k-1}}. \qquad (6.1)$$

If the original landmarks are independent identically distributed with a rotationally symmetric distribution, then the resulting shape distribution is uniform in the shape space.

6.1.1 The offset normal distribution in two dimensions

An i.i.d. model (with the same mean for each point) is often not appropriate. For example, in many applications one would be interested in a model with different means at each landmark. The first extension that we consider is when the landmarks $(X_j, Y_j)^{\mathrm{T}}, j = 1, \ldots, k$, are independent isotropic bivariate normal, with different means but the same variance at each landmark. The isotropic normal model is

$$X = (X_1, \ldots, X_k, Y_1, \ldots, Y_k)^{\mathrm{T}} \sim N_{2k}(\mu, \sigma^2 I_{2k}), \quad (6.2)$$

where $\mu = (\mu_1, \ldots, \mu_k, \nu_1, \ldots, \nu_k)^{\mathrm{T}}$. This model is particularly appropriate when the major source of variability is measurement error at each landmark. For example, consider Figure 65 where the points of a triangle are independently perturbed by isotropic normal errors. We wish to find the resulting perturbed shape distribution, using say Bookstein or Kendall coordinates after translating, rotating and rescaling two points to a fixed baseline.

We shall also use the variance parameter

$$\tau^2 = \sigma^2/\|\mu_2 - \mu_1\|^2,$$

which is particularly useful when working with Bookstein or Kendall co-ordinates.

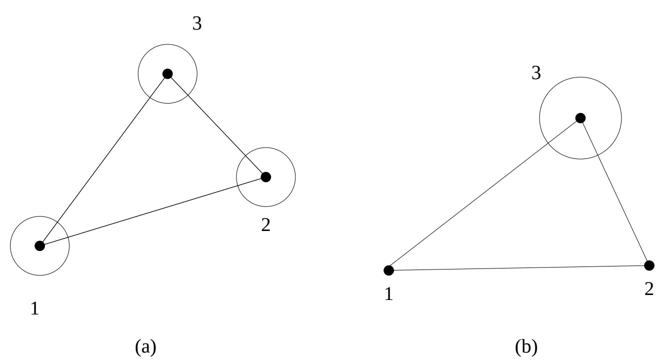

Figure 65 The isotropic model is appropriate for independent isotropically perturbed landmarks (a). The resulting shape distribution can be thought of as the distribution of the landmarks after translating, rotating and rescaling so that the baseline is sent to a fixed position (b).

Result 6.2 *The* **offset normal** *shape density under the* *isotropic normal model of Equation (6.2), with respect to the uniform measure in the shape space, is*

$$ {}_1F_1\{2-k; 1; -\kappa(1+\cos 2\rho(X,\mu))\} \exp\{-\kappa(1-\cos 2\rho(X,\mu))\}, \tag{6.3}$$

where

$$\kappa = S^2(\mu)/(4\sigma^2)$$

is a concentration parameter, $S(\mu)$ is the population centroid size, ρ is the Procrustes distance between the observed shape $[X]$ and the population shape $[\mu]$, and the confluent hypergeometric function ${}_1F_1(\cdot)$ is given by

$$ {}_1F_1(a;b;x) = 1+\frac{a}{b}\frac{x}{1!}+\frac{a(a+1)}{b(b+1)}\frac{x^2}{2!}+\frac{a(a+1)(a+2)}{b(b+1)(b+2)}\frac{x^3}{3!}+\ldots \tag{6.4}$$

6.2 Likelihood-based Inference

Inference can be carried out using maximum likelihood estimators or Bayesian inference, provided we can specify some suitable distributions in the shape space.

Let $u_1^K, \ldots, u_n^K$ be independent observations from a distribution with density $f(u; \theta, \tau)$ with mean shape θ and variation parameter τ, then the likelihood is given by

$$L(u_1^K, \ldots, u_n^K; \theta, \tau) = \prod_{i=1}^{n} f(u_i^K; \theta, \tau).$$

The maximum likelihood estimator of shape is then given by

$$\hat{\theta} = \mathrm{argsup}_\theta \, L(u_1^K, \ldots, u_n^K; \theta, \tau).$$

If we have some prior knowledge about θ, τ which we can express with a prior density $\pi(\theta, \tau)$, then Bayesian inference can proceed by using the posterior density

$$\pi(\theta, \tau | u_1, \ldots, u_n) \propto \pi(\theta, \tau) L(u_1^K, \ldots, u_n^K; \theta, \tau).$$

6.2.1 A Rotationally Symmetric Shape Family

A rotationally symmetric family of shape distributions can be obtained with densities as functions of shape distance to a mean configuration. The class of densities is given by

$$c_\phi(\kappa)^{-1} \exp(-\kappa \phi(d_F^2)),$$

where $\phi(\cdot) \geq 0$ is a suitable penalty function, and it is assumed that $\phi(d_F)$ is an increasing function of d_F and $\phi(0) = 0$.

A particular sub-class is given by the densities with

$$\phi(d_F^2) = (1 - (1 - d_F^2)^h)/h, \tag{6.5}$$

which give the same MLE as a class of shape estimators proposed by Kent (1992). The estimators become more resistant to outliers as h increases.

A density in this form is given, with respect to the uniform measure, is proportional to

$$\exp(\kappa \cos 2\rho), \tag{6.6}$$

where ρ is the Procrustes distance from the observed shape to the mean shape. The MLE of the mean shape is the same as the full Procrustes estimator of the mean shape. If $k = 3$, then the density reduces to

$$\frac{\kappa}{4\pi^2 \sinh(\kappa)} \exp(\kappa \cos 2\rho)$$

which is the Fisher distribution on the shape sphere.

An alternative distribution with $h = \frac{1}{2}$ is the distribution with density proportional to

$$\exp(4\kappa \cos \rho), \tag{6.7}$$

where the mean shape is the same as the partial Procrustes mean. We call this distribution the partial Procrustes shape distribution.

Both of the distributions of Equations (6.6) and (6.7) are asymptotically normal as $\kappa \to \infty$ and uniform if $\kappa = 0$.

6.2.2 *Likelihood ratio tests*

We can write down likelihood ratio tests for various problems, e.g. testing for differences in shape between two independent populations or for changes in variation parameter. Large sample standard likelihood ratio tests can be carried out in the usual way, using Wilks' theorem.

Consider, for example, testing whether $H_0 : \Theta \in \Omega_0$, versus $H_1 : \Theta \in \Omega_1$, where $\Omega_0 \subset \Omega_1 \subseteq \mathbb{R}^{2k-4}$, with $\dim(\Omega_0) = p < 2k-4$ and $\dim(\Omega_1) = q \le 2k-4$ Let

$$-2\,\log\Lambda = 2\sup_{H_1}\log L(\Theta,\tau) - 2\sup_{H_0}\log L(\Theta,\tau)$$

then according to Wilks' theorem $-2 \log \Lambda \approx \chi^2_{q-p}$ under H_0, for large samples (under certain regularity conditions).

If we had chosen different shape coordinates for the mean shape, then inference would be identical as there is a 1-1 linear correspondence between Kendall (or Bookstein) coordinates as seen in Equation (2.11). Also, if a different baseline was chosen, then there is a 1-1 correspondence between Kendall's coordinates and the alternative shape parameters. Therefore, any inference will not be dependent upon baseline or coordinate choice with this exact likelihood approach.

Example 6.1 Consider the schizophrenia data described in Section 1.2.3. The isotropic model could be considered reasonable for these data, because the Procrustes scatter plots are roughly circular and there are no strongly

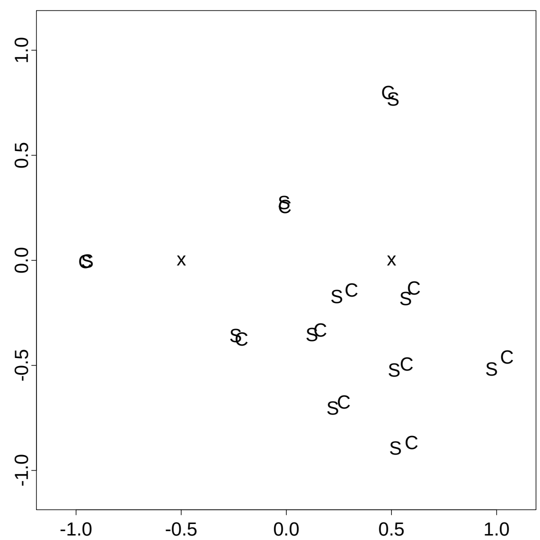

Figure 66 The exact isotropic MLE mean shapes for the schizophrenic patients (S) and control group (C), pictured in Bookstein coordinates with baseline 2, 1 (x).

dominant principal components (see later in Example 7.4). So, we consider the shape variables in group g to be random samples from the isotropic offset normal distribution, with $2k - 3 = 23$ parameters in each group Θ_g, τ_g, where $g = 1$ (control group) or $g = 2$ (schizophrenia group).

The isotropic offset normal MLEs can be obtained using R for the maximization. Plots of the mean shapes in Bookstein coordinates using baseline genu (landmark 2) and splenium (landmark 1) are given in Figure 66. The exact isotropic MLE of shape is very close to the full Procrustes mean shape – in particular the Procrustes distance ρ from the exact isotropic MLE to the full Procrustes mean shapes in both groups are both less than $\rho = 0.0002$. The MLE of the variation parameters are

$\hat{\tau}_1 = 0.049$ and $\hat{\tau}_2 = 0.051$.

Testing for equality in mean shape in each group we have $-2\log\Lambda$, distributed as χ^2_{22} under H_0, as 43.269. Since the p-value for the test is $P(\chi^2_{22} \geq 43.269) = 0.005$ we have strong evidence that the groups are different in mean shape, as is also seen in Example 7.4. Hence, there is evidence for a significant difference in mean shape between the controls and the schizophrenic patients. If we restrict the groups to have the same τ in H_0, then $-2\log\Lambda = 45.03$ and the p-value for the test is $P(\chi^2_{23} \geq 45.03) = 0.004$.

Of course, we may have some concern that our sample sizes may not be large enough for the theory to hold and alternative F and Monte Carlo permutation tests (with similar findings) are given in Example 7.4. □

6.3 Practical Inference

For isotropic models in two dimensions it is practical to use these likelihood-based or Bayesian methods but for more general covariance structures or models in more than two dimensions the procedure is very complicated.

Perhaps the most straightforward and preferred way to proceed with inference is to use Procrustes analysis to obtain an estimate of the mean shape. In order to assess the structure of variability it is best to proceed with a multivariate analysis approach in the real tangent space. We continue with this theme in the next chapter, where some classical hypothesis tests are explicitly stated in the tangent space.