NONPARAMETRIC BAYESIAN INFERENCE ON PLANAR SHAPES

Author: Abhishek Bhattacharya
Coauthor: David Dunson
Department of Statistical Science, Duke University

7th Workshop on Bayesian Nonparametrics
Collegio Carlo Alberto, Moncalieri, Italy
June 24 2009

INTRODUCTION

The goal of this poster is to present recently developed Nonparametric Bayes methods for analysis of landmark based 2D shapes. Among significant applications are those in the discrimination, clustering and classification of biological shapes and medical images, and in machine vision.

Each observation under study consists of a set of k points, or landmarks on a 2D image. Such an observation is called a k-ad or a configuration of k points. In order to compare samples of such data from different groups, collected with possibly different magnifications, different camera angles and orientations, one needs to remove the effects of scale, location and rotation from each k-ad, yielding its shape.

The space of such shapes is called the Planar Shape Space which is a special type of geometric space called Riemannian manifold.

A lot of inference has been done on shape spaces using parametric models, see Dryden et al.[3] and the references cited therein, for more details.

Recently there has also been a surge in frequentist nonparametric methods of inference on manifolds, which avoid a complete likelihood specification, and rely on asymptotic and bootstrap distribution of the statistics. Refer, for example to Bhattacharya & Bhattacharya[1] and the references cited therein. Such methods are based on notions of center and spread, which are appropriately defined on non-Euclidean manifolds.

However in many applications, aspects of the distribution other than center and spread may also be important. In addition, Bayesian likelihood-based methods provide a full probabilistic characterization of uncertainty, which is valid even in small samples.

In the paper Bhattacharya & Dunson[2], we propose a kernel mixture model of the form

$$f(y; P) = \int K(y; \theta) P(d\theta), \tag{1.1}$$

where K is a kernel and P is a random mixture distribution on the manifold. A common choice of prior for P is the Dirichlet process (DP).

The kernel must be carefully chosen, so that the induced prior will have large support, meaning that the prior assigns positive probability in arbitrarily small neighborhoods around any density f_0.

Prior positivity of Kullback-Leibler (KL) neighborhoods around the true density f_0 implies that the posterior probability in any weak neighborhood of f_0 converges to one as $n \to \infty$. Such a prior is said to satisfy the KL condition.

We prove this condition for a kernel defined on arbitrary manifolds, and also for a so called Complex-Watson (CW) kernel on the Planar Shape Space.

This poster highlights the major results and findings of [2].

The poster ends with an application in morphometrics where we estimate the densities for shapes of male and female gorilla skulls and use that to build a classifier to determine the gender of a gorilla from the shape of its skull.

THE PLANAR SHAPE SPACE Σ_2^k

Consider a k-ad

$$\mathbf{z} = \{(x_j, y_j), 1 \leq j \leq k\}$$

on a 2D image. The planar shape space Σ_2^k comprises the equivalence classes of all such k-ads under translation, rotation and scaling.

For mathematical convenience, we view $\mathbf{z}$ as an element of $\mathbb{C}^k$, namely, $\mathbf{z} = (x_j + iy_j)_{1 \leq j \leq k}$. Then the effect of translation is removed by bringing the center $\bar{\mathbf{z}}$ to the origin. The centered k-ad $\mathbf{z}_c$ lies in the (k-1)-dim. subspace H_{k-1} of $\mathbb{C}^k$ of vectors orthogonal to $\mathbf{1}_k$. One considers coordinates of $\mathbf{z}_c$ into H_{k-1}, say $\tilde{\mathbf{z}}_c$ $(\in \mathbb{C}^{k-1})$.

To remove the effect of scaling, we divide $\tilde{\mathbf{z}}_c$ by its total norm to get what is called the preshape $\mathbf{w}$ of the original k-ad $\mathbf{z}$. The set of preshapes is then naturally identified with the complex unit sphere $\mathbb{C}S^{k-2}$ in $\mathbb{C}^{k-1}$.

Finally, the shape of **z** or **w** is the orbit of **w** under 2D rotations, namely,

$$[\mathbf{z}] = [\mathbf{w}] = \{e^{i\theta}\mathbf{w} : \ \theta \in (-\pi, \pi]\}.$$

The set of all such shapes or orbits is Σ_2^k.

It can be identified with the complex projective space $\mathbb{C}P^{k-2}$ - the space of all (complex) lines passing through the origin in $\mathbb{C}^{k-1}$, which is a Riemannian manifold of dimension $2k - 4$.

Two commonly used notions of center of a probability distribution Q on Σ_2^k are the intrinsic and extrinsic mean shapes, based on the appropriate distance chosen on Σ_2^k. The spread of Q is measured by its intrinsic or extrinsic variation. For details, see [1].

A commonly used parametric model, proposed by Kent[4], is the Complex Bingham distribution.

A special case is the Complex Watson distribution which has the following density:

$$f([z];[\mu],\sigma) = c^{-1}(\sigma)\exp(|z^*\mu|^2/\sigma),\ z,\mu \in \mathbb{C}S^{k-2},\ \sigma \in \mathbb{R}^+$$

where

$$c(\sigma) = (\pi\sigma)^{(k-2)}[e^{1/\sigma} - \sum_{r=0}^{k-3}\frac{\sigma^{-r}}{r!}].$$

It can be shown that the extrinsic mean of the above distribution is $[\mu]$ and for σ small, the extrinsic variation is approximately a constant multiple of σ.

DENSITY ESTIMATION ON Σ_2^k

Consider a mixture model density f of the form (1.1) with a Complex-Watson (CW) kernel

$$K([z];[\mu],\sigma) = c^{-1}(\sigma)\exp\left(\frac{|z^*\mu|^2}{\sigma}\right). \tag{3.1}$$

Then

$$f([z];P) = \int_{\Sigma_2^k\times\mathbb{R}^+} c^{-1}(\sigma)\exp\left(\frac{|z^*\mu|^2}{\sigma}\right)P(\mathrm{d}[\mu]\mathrm{d}\sigma). \tag{3.2}$$

For a given distribution P on $\Sigma_2^k \times \mathbb{R}^+$, f defines a valid probabilty density on Σ_2^k (see [2]). For a given prior Π on P, the DP prior for example, we generate a family of distributions absolutely continuous wrt the volume measure. For this to be a rich class of densities, we need to verify that the KL condition holds.

This is proved in the theorem below. Then from the Schwarz theorem, posterior consistency follows.

Theorem (Proof: See [2])

Let f_0 be a continuous density on Σ_2^k which is positive everywhere and let F_0 be the corresponding probability distribution. Let Π be a prior on the space of probabilities P on $\Sigma_2^k \times \mathbb{R}^+$ whose support contains $F_0 \otimes \delta_0$. Then Π satisfies the KL condition at f_0. In other words, given any $\epsilon > 0$,

$$\Pi\left(P:\ \int_{\Sigma_2^k} f_0(m) \log\left(\frac{f_0(m)}{f(m;P)}\right) V(dm) < \epsilon\right) > 0.$$

Given an iid sample on Σ_2^k from $f(m;P)$, to estimate f, we set a prior for P such as the DP prior. Then we compute the Bayes estimate for f using some standard algorithm such as the Polya Urn or Block-Gibbs sampling. For prior selection and computational details, see [2].

DENSITY ESTIMATION ON GENERAL MANIFOLD

In [2] we also present a mixture density model on a general compact manifold M using the kernel

$$K(m; \mu, \sigma) = \mathcal{X}(\frac{d_g(m, \mu)}{\sigma})\sigma^{-d} G_\mu^{-1}(m)$$

where $m, \mu \in M$, $\sigma \in \mathbb{R}^+$, $\mathcal{X} : \mathbb{R}^+ \to \mathbb{R}^+$, d_g denotes the geodesic distance on M, d is the dim. of M and $G_\mu(m)$ is the density of the volume-form of M pulled to the tangent space at μ using normal coordinates evaluated at the coordinates of m.

This kernel is used in Pelletier[5] to perform frequentist Kernel Density Estimation.

We prove the KL condition for this model as well.

Since Σ_2^k is a compact manifold, this density model can also be used. However due to computational complexities and support restrictions imposed on $\mathcal{X}$ and σ (see [2]), we prefer using the CW kernel.

MORPHOMETRICS: CLASSIFICATION OF GORILLA SKULLS

To test the difference in shapes of skulls of male and female gorillas, eight landmarks are chosen on the midline plane of the skull of 30 female and 29 male gorillas. The data can be found in [3].

Figure 1 shows the plot of the preshapes of the k-ads along with the mean shapes for the 2 groups. The sample pre-shapes have been rotated appropriately to align them with the means' preshapes.

Figure 2 shows the plot of the estimated shape density for the two groups along with a 95% credible region. The plots show the densities conditioned to the geodesic joining the mean shapes for the two groups.

FIG 1: 8 LANDMARKS FROM SKULLS OF 30 FEMALE & 29 MALE GORILLAS ALONG WITH THE EXTRINSIC MEANS (red)

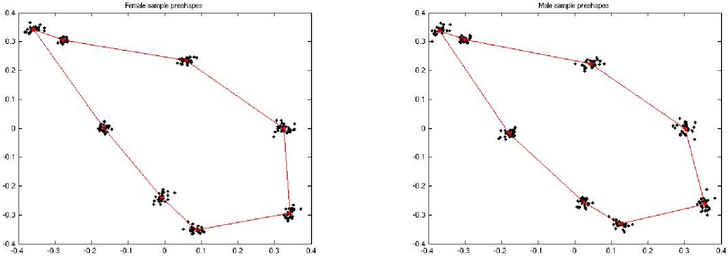
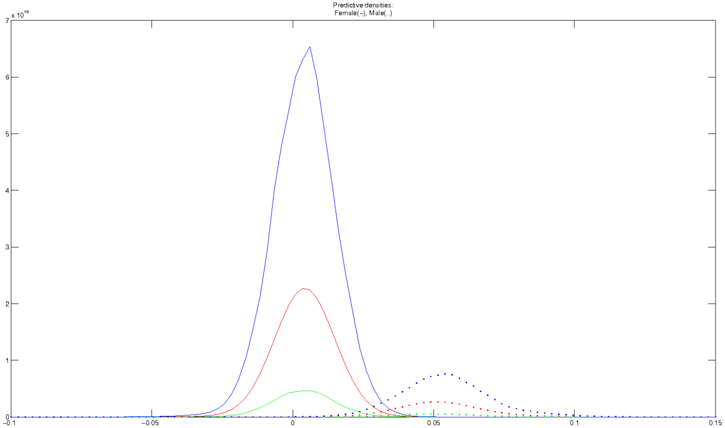


FIG 2: DENSITIES FOR GORILLA SKULL SHAPES: FEMALE (solid), MALE (dotted), POSTERIOR MEAN DENSITIES (red), 95% C.R. (blue/green)



To carry out a discriminant analysis, we randomly pick 25 shapes from each sample as training data sets and the remaining 9 are used as test data. Then we estimate the shape densities independently from the test data for each sex, and find the conditional probability of being female for each of the test sample shapes. Table 1 presents the probabilities along with a 95% Credible Interval for one such allocation.

The first 5 are females and the last 4 males. $\hat{p}$ denotes the estimated point prob. of being classified as female, **C.I.** is the Credible Interval for the prob., $d(., \mu_f)$ & $d(., \mu_m)$ denote extrinsic distances of the corresponding shape from female & male sample extrinsic mean respectively.

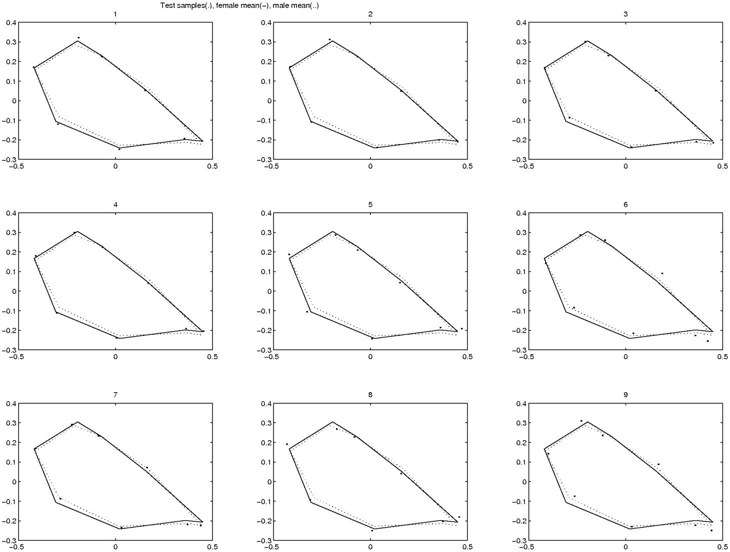
The table suggests that only shapes 3 and 8 are miss-classified. If we use a distance based classifier, we would reach the same conclusion.

TABLE 1

Shape	$\hat{p}$	**C.I.**	$d(., \mu_f)$	$d(., \mu_m)$
1	1	(1,1)	.041	.1109
2	.9999	(.9992,1)	.0362	.0934
3	.16	(.008,.602)	.056	.0517
4	.9958	(.968, 1)	.0495	.0952
5	1	(1, 1)	.0755	.135
6	.0001	(0, 0)	.1672	.1033
7	.0005	(0, .003)	.087	.0417
8	.983	(.8197, 1)	.0911	.1207
9	.0003	(0, 0)	.1523	.0935

Figure 3 plots the preshapes of these 9 samples along with the mean shapes from the male and female groups.

FIG. 3: Preshapes for test sample. Sample (.), Female mean (-), Male Mean (..)



References

[1] Bhattacharya, A. and Bhattacharya, R. (2008). 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.

[2] Bhattacharya, A. and Dunson, D. (2009). Nonparametric Bayesian Density Estimation on Manifolds with applications to Planar Shapes. *To Appear*.

[3] Dryden, I.L. and Mardia, K.V. (1998). *Statistical Shape Analysis*. Wiley N.Y.

[4] Kent, J.T. (1994). The complex Bingham distribution and shape analysis. *J. Roy. Statist. Soc. Ser. B* **56** no.2, 285-299.

[5] Pelletier, B. (2005). Kernel density estimation on Riemannian manifolds. *Stat. & Prob. Letters* **73** 297-304.

THANK YOU!

I am grateful to the BNP Conference committee for inviting me to present this poster and awarding a travel award to attend the conference.

I would like to thank my advisor, Prof. Dunson for providing me assistance while making the poster.

Above all, I thank Divine Mother Ma who is behind all this.