Proceedings 59th ISI World Statistics Congress, 25-30 August 2013, Hong Kong (Session IPS070)

Nonparametric Density Estimation on Manifolds with Applications to Shape Analysis

Abhishek Bhattacharya
Indian Statistical Institute

August 29, 2013

Proceedings 59th ISI World Statistics Congress, 25-30 August 2013, Hong Kong (Session IPS070)

Contents

Proceedings 59th ISI World Statistics Congress, 25-30 August 2013, Hong Kong (Session IPS070)

Motivation and Goal

- Statistical analysis of landmark based shapes has diverse applications in morphometrics, medical diagnostics, machine vision e.t.c.
- These shape spaces are non-Euclidean quotient manifolds.
- To conduct nonparametric (np) inference, one may define notions of center and spread on the manifold and work with their estimates.
- However it is useful to consider full likelihood-based methods which allow nonparametric density estimation.

- This talk presents a broad class of kernel mixture density models on a general metric space and on the **planar shape space** Σ_2^k in particular.
- Using a Bayesian approach with a np prior on the density, the density estimate is shown to be consistent.
- Gibbs sampling methods are used for posterior computations.

- The methods are applied to problems in density estimation and classification with shape-based predictors.
- Simulation studies show improved estimation performance relative to existing approaches.
- This talk outlines the work in *Bhattacharya & Dunson 2010, 2012* and Chapter 13, *Bhattacharya & Bhattacharya 2012*.

- Consider a set of k points/landmarks, $k > 2$, on a 2D image, not all points identical. Such a set referred as a k-ad.
- The similarity shape of this k-ad is what remains after removing the Euclidean rigid body motions of translation, rotation and scaling.
- The k-ad denoted by a complex k-vector z in $\mathcal{C}^k$. To remove effect of translation, let $z_c = z - \overline{z}$, with $\overline{z} = (\sum_{j=1}^{k} z_j)/k$. The centered k-ad z_c lies in a $k-1$ dimensional complex subspace, and hence we can use $k-1$ complex coordinates.
- Efect of scaling then removed by normalizing the coordinates of z_c to obtain a point w on the complex unit sphere $\mathcal{CS}^{k-2}$ in $\mathcal{C}^{k-1}$.

- Since w contains shape info. of z along with rotation, it is called the **preshape** of z.
- Then the similarity shape of z is the orbit of w under all rotations in 2D i.e.

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

- This shape representation first proposed by [6, *Kendall 1984*].
- This represents a shape as the set of all intersection points of a unique complex line passing through the origin with $\mathcal{CS}^{k-2}$ and Σ_2^k is then the set of all such shapes.
- Hence Σ_2^k can be identified with the space of all complex lines passing through the origin in $\mathcal{C}^{k-1}$ which is a compact Riemannian manifold of dimension $2k-4$.

- Σ_2^k can be embedded into the space of all order $k-1$ complex Hermitian matrices via the embedding $J([w]) = ww^*$.
- This embedding induces a distance on Σ_2^k called **extrinsic distance** which generates the manifold topology and is given by

$$d_E([u],[v]) = \|J([u]) - J([v])\| = \sqrt{2(1-|u^*v|^2)} \quad ([u],[v] \in \Sigma_2^k).$$

- Let Q be a probability on Σ_2^k. The extrinsic/Fréchet mean of Q is defined as the minimizer of the loss function $F(p) = \int_{\Sigma_2^k} d_E^2(m,p)Q(dm)$, $p \in \Sigma_2^k$, provided F has a unique minimizer.
- The minimum value of F is called the extrinsic/Fréchet variation/spread of Q.
- Let $\tilde{\mu} = \int_{\Sigma_2^k} J(m)Q(dm)$, λ be its largest eigenvalue and U be a corresponding unit norm eigenvector.
- Then the ex. variation equals $2(1-\lambda)$ and the ex. mean given by $[U]$ provided λ has multiplicity 1.
- Given a random sample from Q, one can define the sample ex. mean and variation analogously.
- For more details, see [1, Bhattacharya & Bhattacharya 2012] Chapter 8 & the references cited therein.

- Let $K(m; \mu, \kappa)$ be a density model (kernel) with a known parametric form on a seperable m.s. (M, ρ) w.r.t. some fixed base measure λ.
- When M is a Riemannian manifold such as Σ_2^k, λ is chosen to be its invariant volume form.
- The kernel K has variable $m \in M$, parameters $\mu \in M$ and $\kappa \in N$, N being a polish space.
- Usually this parameter μ turns out to be the Fréchet mean of K. Hence μ is called the kernel location.
- Parameter κ comprises of all other parameters on appropriate spaces determining kernel shape, spread and so on.
- In the illustrated examples N is $(0, \infty)$ and κ a decreasing function of the Fréchet variation. Such kernels are called **location-scale kernels**.

Proceedings 59th ISI World Statistics Congress, 25-30 August 2013, Hong Kong (Session IPS070)

- Using this kernel K and a probability P on M, a **location mixture density** model $f(.; P, \kappa)$ can be defined on M as

$$f(.; P, \kappa) = \int_M K(.; \mu, \kappa) P(d\mu)$$

  with parameters P and κ.
- For np Bayes inference we set priors on P and κ i.e. a prior Π_1 on $\mathcal{M}(M) \times N$, s.t. the induced prior Π on f has full support and the posterior is consistent.
- To talk about support of a prior on probabilities, we need to introduce a topology on the space $\mathcal{M}(M)$ of probabilities on M.
- We use three, namely **weak**, **strong**, and **Kullback-Leibler** (KL) neighborhoods.

Proceedings 59th ISI World Statistics Congress, 25-30 August 2013, Hong Kong (Session IPS070)

- A sequence of probabilities $\{P_n\}$ converges weakly to P if $\int_M \phi dP_n \longrightarrow \int_M \phi dP$ for any continuous $\phi : M \to [-1, 1]$.
- The strong/total variation/L^1 distance between two probabilities P & Q is $\sup_\phi |\int \phi dP - \int \phi dQ|$.
- The KL divergence between P & Q with densities p & q w.r.t. some base measure λ is $\int p \log \frac{p}{q} d\lambda$, the definition being invariant to choice of λ.
- KL convergence implies L^1 convergence which implies weak convergence.

- Let f_t be a continuous density on M with prob. P_t.
- Since most shape spaces (including Σ_2^k) are compact non-Euclidean, from now on we assume M is compact.
- For similar results on Euclidean spaces, refer to [8, *Wu & Ghosal 2008*] and other other related works.
- To show that f_t is in the support of Π, assume

A1 Kernel K is continuous in its arguments.

A2 For any cont. $\phi : M \longrightarrow \Re$, for any $\epsilon > 0$, $\exists$ a compact subset N_ϵ of N (domain of density parameter κ) with nonempty interior N_ϵ^0, s.t.

$$\sup_{m \in M, \kappa \in N_\epsilon} |\phi(m) - \int_M K(m; \mu, \kappa)\phi(\mu)\lambda(d\mu)| < \epsilon.$$

A3 $\forall\ \epsilon > 0$, the set $\{P_t\} \times N_\epsilon^0$ intersects with the weak support of Π_1. This holds for e.g. if Π_1 is a full (weak) support product prior on $\mathcal{M}(M) \times N$.

- If the chosen kernel $K(m; \mu, \kappa)$ is symmetric in m & μ, **A2** implies that as a prob. on M, $K(.; \mu, \kappa)$ can be made arbitrarily close in weak sense to the point mass δ_μ at μ, uniformly in μ, for appropriate κ choice. This further justifies the name **location** for μ. Most standard parametric models on Σ_2^k and other manifolds satisfy this condition.

- A common choice for Π_1 satisfying **A3** is a **Dirichlet process** (DP) prior on P with a full support base probability and an independent full support prior on κ.

- The DP($w_0 P_0$) prior has a base measure $w_0 P_0$, P_0 the base probability and $w_0 > 0$ called the precision parameter.
- The DP prior was introduced by [5, *Ferguson 1973*] and its properties investigated by many others.

Theorem (1.Support of Mixture Density Model)

*Under Assumpts.**A1**-**A3**,* $\forall\ \epsilon > 0$,

$$\Pi\{f : \sup_{m \in M} |f_t(m) - f(m)| < \epsilon\} > 0$$

which means that Π *assigns* $+$*ve probability to arbitrarily small uniform nbhoods of any cont. density* f_t.

- For proof see [2, *Bhattacharya & Dunson 2010*].
- Thm 1 implies that f_t is in the KL supp. of Π (and hence in its strong & weak supp.). Then Π is said to satisfy the **KL condn.** at f_t.
- To achieve more flexibility, in the mixture model mix across κ as well, i.e., replace $P(d\mu)$ by $P(d\mu d\kappa)$. Then KL condn. and posterior consistency is verified in [2].

- Let $X_1, \ldots, X_n$ be iid realisations on M from some cont. density. Let f_t denote the true 'density'.
- The posterior prob. of any set U can be expressed as

$$\Pi_n(U) \doteq \Pi(U|X_1, \ldots, X_n) = \frac{\int_U \prod f(X_i)\Pi(df)}{\int \prod f(X_i)\Pi(df)}.$$

- Then using the [9, *Schwartz 1965*] theorem, Thm 1 implies WPC at an exponential rate, i.e., for any weakly open nbhood U of f_t, f.s. $\alpha > 0$,

$$\exp(n\alpha)\Pi_n(U) \longrightarrow 0 \text{ a.s. } f_t.$$

SPC means that the posterior prob. of any total variation nbhood of f_t converges to 1 a.s. f_t.

To show that assume $\exists$ a cont. $\phi : N \to [0, \infty)$ for which the following regularity condns. hold.

A4 $\exists$ +ve constants $\mathcal{K}_1$, a_1, A_1 s.t. $\forall \mathcal{K} \geq \mathcal{K}_1$, $\mu, \nu \in M$,

$$\sup_{m \in M, \kappa \in \phi^{-1}[0,\mathcal{K}]} |K(m; \mu, \kappa) - K(m; \nu, \kappa)| \leq A_1 \mathcal{K}^{a_1} \rho(\mu, \nu).$$

A5 $\exists$ +ve constants a_2, A_2 s.t. $\forall \kappa_1, \kappa_2 \in \phi^{-1}[0, \mathcal{K}]$, $\mathcal{K} \geq \mathcal{K}_1$,

$$\sup_{m, \mu \in M} |K(m; \mu, \kappa_1) - K(m; \mu, \kappa_2)| \leq A_2 \mathcal{K}^{a_2} \rho_2(\kappa_1, \kappa_2),$$

ρ_2 metrizing the topology of N.

A6 $\phi^{-1}[0, \mathcal{K}]$ is compact $\forall\ \mathcal{K} \geq \mathcal{K}_1$, and, given $\epsilon > 0$, the min. number of ϵ radius balls covering it (the **ϵ-covering number**) is bounded by $(\mathcal{K}\epsilon^{-1})^{b_2}$ f.s. $b_2 > 0$ (independent of $\mathcal{K}$ and ϵ choice).

A7 $\exists\ a_3, A_3 > 0$ s.t. the ϵ-covering number of M is bounded by $A_3\epsilon^{-a_3}\ \forall \epsilon > 0$.

A8 $\Pi_1(\mathcal{M} \times \phi^{-1}(n^a, \infty)) < \exp(-n\beta)$ f.s. $a < (a_1 a_3)^{-1}$ and $\beta > 0$.

Theorem (2.Strong Posterior Consistency (SPC))

*Under Assumpts.**A1**-**A8**, SPC at an exponential rate follows.*

For proof see [3, *Bhattacharya & Dunson 2012*].

Proceedings 59th ISI World Statistics Congress, 25-30 August 2013, Hong Kong (Session IPS070)

- When using a location-scale kernel, i.e. $N = (0, \infty)$, with a full support product prior $\Pi_1 = \Pi_{11} \otimes \pi_1$ on (P, κ), one may set ϕ to be the identity map.
- Then a π_1 satisfying **A8** is the Weibull density $\text{Weib}(\kappa; a, b) \propto \kappa^{a-1} \exp(-b\kappa^a)$, whenever shape hyper-parameter $a > a_1 a_3$. Π_{11} can be any full supp. prior such as full supp. Dirichlet Process.
- $M = \Sigma_2^k$ satisfies **A7** with $a_3 = 2k - 3$. See Thm 4.
- Then SPC follows on Σ_2^k with appropriate kernel choice.
- A Gamma prior on κ doesn't satisfy **A8** (unless $a_1 a_3 < 1$ which is unlikely).

Proceedings 59th ISI World Statistics Congress, 25-30 August 2013, Hong Kong (Session IPS070)

- On a high dimensional manifold, such as shapes with large number of lmks., the constraints on the shape hyper-parameter *a* of Weibull prior become overly restrictive.

- SPC assumptions require *a* to be very large, implying a prior on bandwidth $1/\kappa$ placing very small probability in neighborhoods of zero, which is undesirable.

- [3, *Bhattacharya & Dunson 2012*] propose an alternative by allowing prior Π_1 to depend on sample size *n*. Then for e.g. a DP prior on *P* and independent Gamma prior on κ with scale hyper-parameter of order $\log(n)/n$ satisfys requirements for SPC.

A common kernel choice K on Σ_2^k is the Complex Watson (CW) density introduced in [7, *Mardia & Dryden 1999*] and given by

$$K(m;\mu,\kappa) \equiv \mathrm{CW}(m;\mu,\kappa) = c^{-1}(\kappa)\exp(\kappa|w^*\nu|^2),\ (m=[w],\mu=[\nu]),$$

$$m,\mu \in \Sigma_2^k,\ \kappa \in (0,\infty),\ c(\kappa) = (\pi\kappa^{-1})^{k-2}\{\exp(\kappa) - \sum_{r=0}^{k-3}\kappa^r/r!\}.$$

It has extrinsic mean μ and variation a decreasing function of κ.

Theorem (3.WPC on Σ_2^k)

*For the CW kernel, assumptions **A1** & **A2** of Thm. 1 are satisfied.*

For pf. see [2, *Bhattacharya & Dunson 2010*].

Proceedings 59th ISI World Statistics Congress, 25-30 August 2013, Hong Kong (Session IPS070)

Hence with full support priors on mixture density parameters (P, κ) the resulting density prior includes all cont. densities in its L^∞ (hence KL, strong, & weak)supp.

As a result WPC follows from Thm. 1 if the data generating density f_t is assumed to be cont.

A prior that satisfys the assumt. & leads to conditional conjugacy hence simplifying posterior computations can be $P \sim DP(w_0 P_0)$, with CW base P_0 & precision $w_0 > 0$, and independently $\kappa \sim$Gamma.

Proceedings 59th ISI World Statistics Congress, 25-30 August 2013, Hong Kong (Session IPS070)

- The posterior computations given a random sample can proceed via Gibbs sampling algo.
- A point estimate for the density can be the mean of the posterior.
- For detail see [2, *Bhattacharya & Dunson 2010*].

Theorem (4.SPC on Σ_2^k)

Take ϕ in Thm. 2 to be identity map. Then the CW kernel satisfys assumpt. ***A4*** *with $a_1 = k - 1$ &* ***A5*** *with $a_2 = 3k - 8$. The compact m.s. (Σ_2^k, d_E) satisfys* ***A7*** *with $a_3 = 2k - 3$. Hence SPC holds with $P \sim DP(w_0 P_0)$, $\kappa \sim Weib(a, b)$ whenever $a > (2k - 3)(k - 1)$. Alternatively use a Gamma prior on κ with scale of order $\log(n)/n$. The alternative preseves conjugacy & hence more convinient to use.*

For pf. see [3, *Bhattacharya & Dunson 2012*].

Proceedings 59th ISI World Statistics Congress, 25-30 August 2013, Hong Kong (Session IPS070)

Draw $X_i \sim 0.5\text{CW}(\mu_1, \kappa) + 0.5\text{CW}(\mu_2, \kappa)$ iid $i = 1, \ldots, 200$, on Σ_2^4, with $\kappa = 1000$, $\mu_i = [\nu_i]$, $\nu_1 = (1, 0, 0)'$, $\nu_2 = \{r, (1 - r^2)^{1/2}, 0\}^T$ & $r = 0.9975$ so that the extrinsic distance between μ_1 and μ_2 is 0.1.

We compare the Bayesian np density estimate based on DP mixture of CW kernels to a MLE under a parametric CW model and to the frequentist kernel density estimate (KDE) using a CW kernel.

Generate 20 simulated data sets, with the performance evaluated based on the L^1 distance and KL divergence estimated by averaging over the data points.

The Bayesian np approach implemented with the MCMC algo. run for 100,000 iterations with the first 15,000 discarded as a burn-in.

With $P \sim \text{DP}(w_0\text{CW}(\mu_0, \kappa_0))$ & $\kappa \sim \text{Gamm}(a, b)$, the hyperparameters chosen by setting $w_0 = 1$, $\mu_0 =$ sample extrinsic mean, $\kappa_0 = 10$, and $a = b = 0.1$.

By using the data to estimate the location of the base distribution, while choosing a moderate precision, we ensure that the prior introduces clusters close to the support of the data.

The DP precision parameter $w_0 = 1$ is a commonly-used default favoring a sparse representation with few clusters.

Proceedings 59th ISI World Statistics Congress, 25-30 August 2013, Hong Kong (Session IPS070)

Summaries of the results across the 20 simulated data sets presented in the table below.

The proposed np Bayes estimator has consistently better performance across the data sets and for each choice of criterion.

For the frequentist KDE, results presented for a bandwidth of $\sigma = 0.001$. The performance is similar or worse for other choices of bandwidth, including setting σ equal to the MLE under the parametric CW model and the posterior mean of $\sigma = \kappa^{-1}$ from the Bayes analysis.

Proceedings 59th ISI World Statistics Congress, 25-30 August 2013, Hong Kong (Session IPS070)

Table : Summaries of estimated distances from the true density across the 20 simulations.

	Bayes		MLE		KDE	
	L^1	KL	L^1	KL	L^1	KL
min	0.27	0.02	0.60	0.33	0.56	0.14
25th	0.35	0.07	0.68	0.36	0.74	0.24
50th	0.42	0.08	0.73	0.43	0.87	0.26
75th	0.48	0.16	0.83	0.46	1.20	0.27
max	0.91	0.39	0.94	0.52	2.72	0.32
mean	0.44	0.13	0.75	0.41	1.03	0.25

Proceedings 59th ISI World Statistics Congress, 25-30 August 2013, Hong Kong (Session IPS070)

The method is applied to data on shapes of 29 male and 30 female gorilla skulls, with 8 lmks. chosen on the midline plane of 2D images of each skull [4, *Dryden & Mardia 1998*].

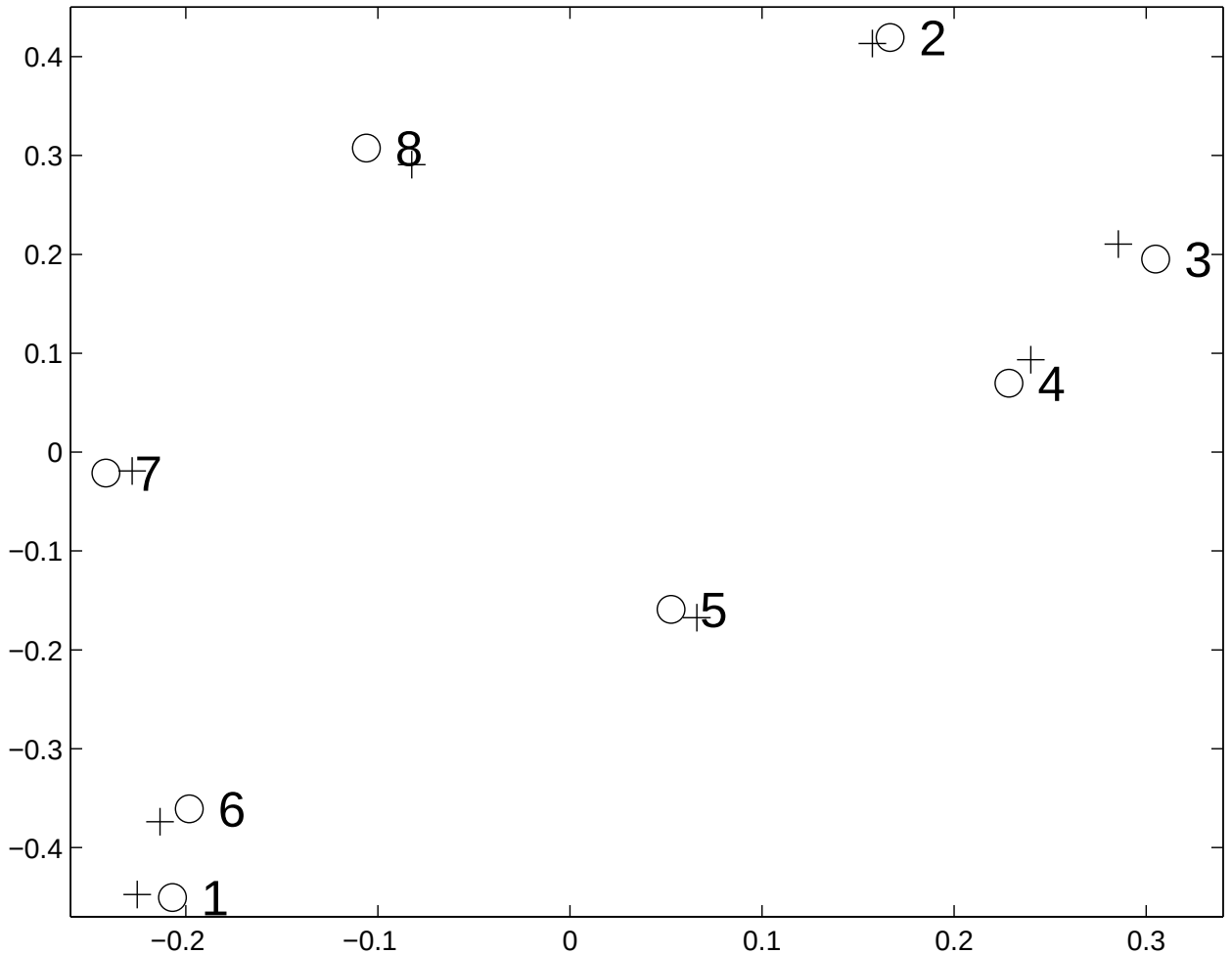
Goal is to study how the shapes of the skulls vary between males and females, and build a classifier to predict gender.

The shape samples lie on Σ_2^k, $k = 8$.

25 individuals of each gender are randomly picked as training sample, & the 9 remaining used as test data.

Figure below shows preshapes of the sample extrinsic means for the female and male training groups. The preshape of the male mean $\hat{\mu}_2$ has been rotated appropriately so as to bring it closest to the preshape of the female mean $\hat{\mu}_1$.

Proceedings 59th ISI World Statistics Congress, 25-30 August 2013, Hong Kong (Session IPS070)



Landmarks from preshapes of $\hat{\mu}_1$ (female, $\circ$) & $\hat{\mu}_2$ (males, $+$)

Proceedings 59th ISI World Statistics Congress, 25-30 August 2013, Hong Kong (Session IPS070)

Applying np discriminant analysis, we assume the prob. of being female is 0.5 and use separate DP CW mixture models for the shape density in the male and female groups.

Letting f_1 and f_2 denote the female and male shape densities, the conditional prob. of being female given shape data z is $p(z) = 1/\{1 + f_2(z)/f_1(z)\}$.

To estimate the posterior prob., we average $p(z)$ across MCMC iterations to obtain $\hat{p}(z)$.

The analysis implemented as in the simulation example, but with hyperparameters $\kappa_0 = 1000$, $a = 1.01$ and $b = 0.001$ elicited based on our prior expectation for the gorilla example.

Proceedings 59th ISI World Statistics Congress, 25-30 August 2013, Hong Kong (Session IPS070)

Table below presents the estimated posterior probs. of being female for each of the gorillas in the test sample along with a 95% credible interval (CI) for $p(z)$.

For most gorillas, there is a high posterior probability of assigning the correct gender. There is misclassification only in the 3rd female and 3rd male.

For the 3rd female, the credible interval includes 0.5, suggesting that there is insufficient information to be confident in the classification.

However for the 3rd male, the CI suggests a high degree of confidence that this individual is female. Perhaps this individual is an outlier and there is something unusual about the shape of his skull, with such characteristics not represented in the training data, or alternatively he was labeled incorrectly.

Proceedings 59th ISI World Statistics Congress, 25-30 August 2013, Hong Kong (Session IPS070)

Table : Posterior prob. of being female for each gorilla in the test sample.

gender	$\hat{p}(z)$	95% CI	$d_E(z_i, \hat{\mu}_1)$	$d_E(z_i, \hat{\mu}_2)$
F	1.000	(1.000,1.000)	0.041	0.111
F	1.000	(0.999,1.000)	0.036	0.093
F	0.023	(0.021, 0.678)	0.056	0.052
F	0.998	(0.987, 1.000)	0.050	0.095
F	1.000	(1.000, 1.000)	0.076	0.135
M	0.000	(0.000, 0.000)	0.167	0.103
M	0.001	(0.000, 0.004)	0.087	0.042
M	0.992	(0.934, 1.000)	0.091	0.121
M	0.000	(0.000, 0.000)	0.152	0.094

$d_E(., \hat{\mu}_i)$ = extrinsic dist. from the mean shape in training group i, with $i = 1, 2$ for females & males respectively.

Proceedings 59th ISI World Statistics Congress, 25-30 August 2013, Hong Kong (Session IPS070)

In addition the extrinsic distance between each gorilla shape and the female and male sample extrinsic means are also displayed in the table.

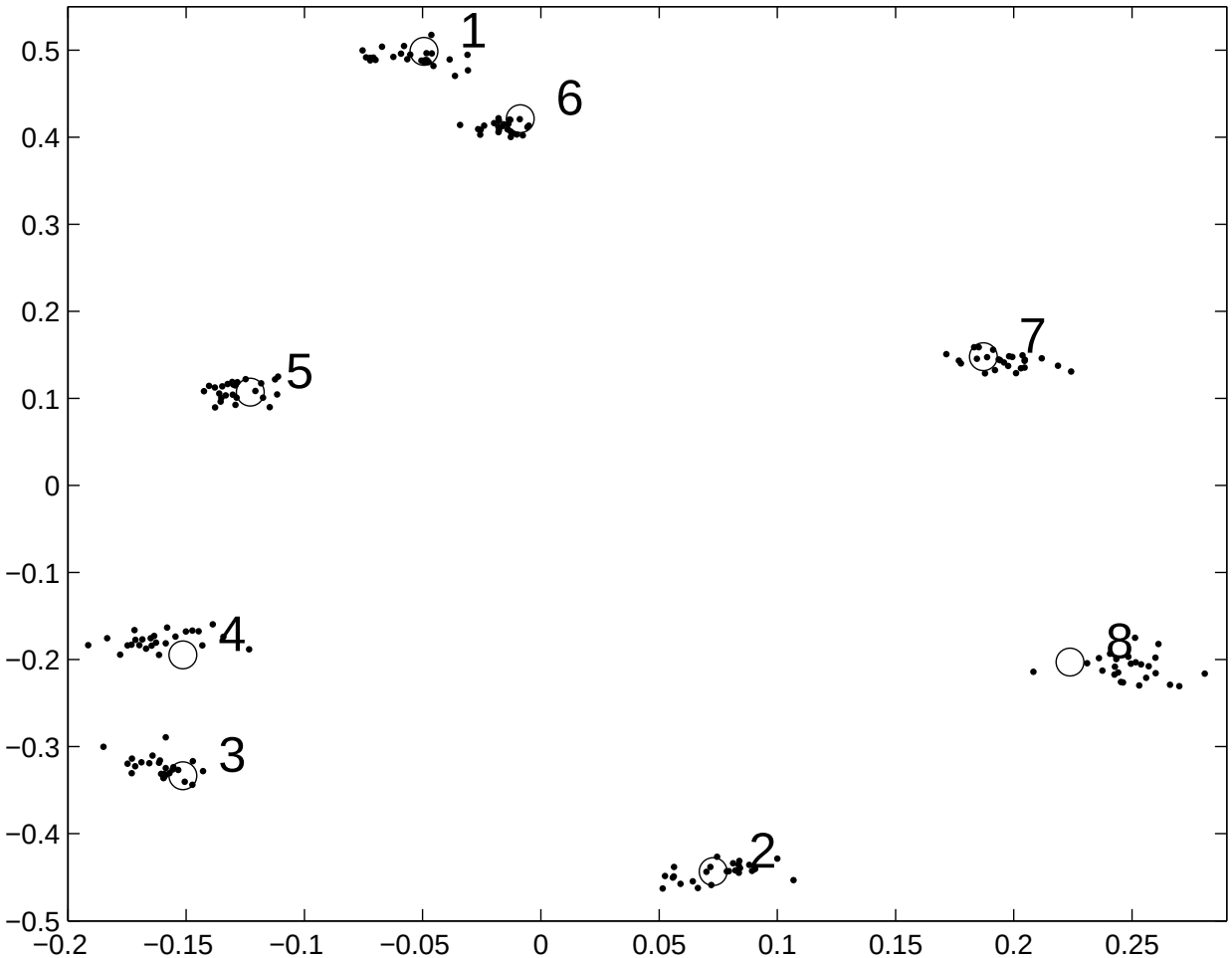
One could define a distance-based classifier which allocates a test subject to the group having mean shape closest to that subject's shape. The table suggests that such a classifier will yield consistent results with our np Bayes approach.

However, this distance-based classifier may be sub-optimal in not taking into account the variability within each group. In addition, the approach is deterministic and there is no measure of uncertainty in classification.

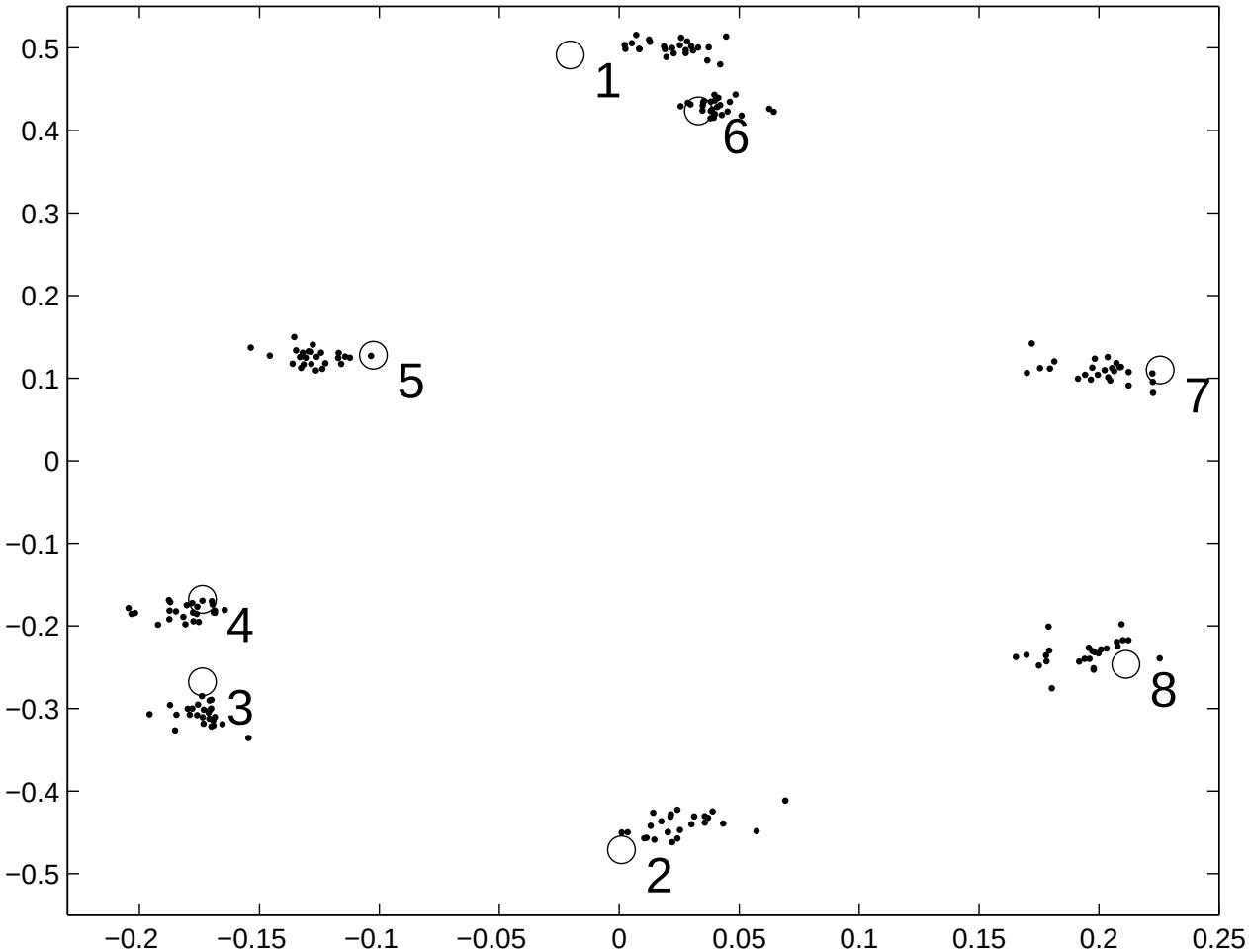
Proceedings 59th ISI World Statistics Congress, 25-30 August 2013, Hong Kong (Session IPS070)

Next 2 figures show the female and male training sample preshape clouds, along with the two misclassified test samples.

There seems to be a substantial deviation in the coordinates of these misclassified subjects from their respective gender training groups, especially for the male gorilla, even after having rotated each training preshape separately so as to bring each closest to the plotted test sample preshapes.



Lmks. from preshapes of training(.) & mis-classified female test sample(○)



Lmks. from preshapes of training(.) & mis-classified male test sample(○)

Proceedings 59th ISI World Statistics Congress, 25-30 August 2013, Hong Kong (Session IPS070)

Classification performance may be improved by also taking into account skull size. The proposed method can be easily extended to this case by using a DP mixture density with the kernel being the product of a CW kernel for the shape component and a log-Gaussian kernel for the size.

Such a model induces a prior with support on the space of densities on the **planar size & shape manifold** $\Sigma_2^k \times \Re^+$.

For more details see Chapter 8, [1, *Bhattacharya & Bhattacharya 2012*].

I would like to conclude with the following words from Srimad Bhagvad Gita, Chapter 15, verse 15.

sarvasya caham hrdi sannivisto, mattah smrtir jnanam apohanam ca; vedais ca sarvair aham eva vedyo, vedanta-krd veda-vid eva caham.

which means,
It is Lord who is seated in hearts of all, and from God comes remembrance, knowledge and forgetfulness. By all the sources of Knowledge it is God that is to be known; indeed God is the compiler of all Sciences, and God the knower of Knowledge.

THANKS TO GOD

References


Bhattacharya, A. & Bhattacharya, R. (2012). NONPARAMETRIC STATISTICS ON MANIFOLDS WITH APPLICATIONS TO SHAPE SPACES, IMS MONOGRAPH 2, Cambridge University Press.


BHATTACHARYA, A. & DUNSON, D. (2010). Nonparametric Bayesian Density Estimation on Manifolds with Applications to Planar Shapes. *Biometrika* 97, 4, 687-714.


BHATTACHARYA, A. & DUNSON, D. (2012). Strong consistency of nonparametric Bayes density estimation on compact metric spaces with applications to specific manifolds. *Ann Inst Stat Math* 64, 687-714.


Dryden, I. L. and Mardia, K. V. (1998). STATISTICAL SHAPE ANALYSIS, Wiley N.Y.

References

- FERGUSON, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. of Stats.* 1, 209-230.
- KENDALL, D. G. (1984). Shape Manifolds, Procrustean Metrics, and Complex Projective Spaces. *Bull. of the London Math. Soc.* 16, 81-121.
- MARDIA, K. V. & DRYDEN, I. L. (1999). The complex Watson distribution and shape analysis. *J. R. Statist. Soc. B* 61, Part 4, 913-926.
- WU, Y. & GHOSAL, S. (2008). Kullback-Leibler property of kernel mixture priors in Bayesian density estimation. *Elec J. Statist.* 2, 298-331

References

SCHWARTZ, L. (1965). On Bayes procedures. *Z. Wahrsch. Verw. Gebiete* 4, 10-26.