Strong consistency of nonparametric Bayes density estimation on compact metric spaces

Abhishek Bhattacharya and David Dunson[†]
Department of Statistical Science, Duke University
[†] *dunson@stat.duke.edu*

Abstract. This article considers a broad class of kernel mixture density models on compact metric spaces and manifolds. Following a Bayesian approach with a nonparametric prior on the location mixing distribution and bandwidth, sufficient conditions are obtained on the kernel, prior and the underlying space for strong posterior consistency at any positive continuous density. The prior is also allowed to depend on the sample size n and sufficient conditions are obtained for weak and strong consistency. These conditions are verified on the hypersphere using a von Mises-Fisher kernel and on the planar shape space using complex Watson kernels.

1. Introduction

Density estimation on compact metric spaces, such as manifolds, is a fundamental problem in nonparametric inference on non-Euclidean spaces. Some applications include directional data analysis, spatial modeling, shape analysis and dimensionality reduction problems in which the data lie on an unknown lower dimensional space. However, the literature on statistical theory and methods of density estimation in non-Euclidean spaces is still under-developed. Our focus is on Bayesian nonparametric approaches.

For nonparametric Bayes density estimation on the real line $\Re$, there is a rich literature, with Dirichlet process mixtures of Gaussian kernels providing a commonly-used approach ([**6**]) that leads to dense support ([**15**]) and weak and strong posterior consistency ([**9**]). From the celebrated theorem of [**16**], weak posterior consistency results when the true density f_0 is in the Kullback-Leibler (KL) support of the prior, meaning that all KL neighborhoods around f_0 are assigned positive probability. In general, it is quite difficult to show KL support for new priors for a density, though [**19**] provide useful conditions for a class of kernel mixture priors, with [**3**] extending these conditions to general compact metric spaces. It is widely accepted that weak consistency is an insufficient property when the focus is on density estimation. For example, if f_0 is a density with respect to Lebesgue measure, weak consistency d

Key words and phrases. Nonparametric Bayes; Density Estimation; Posterior consistency; Sample dependent prior; Riemannian manifold; Hypersphere; Planar shape space.

oes not even ensure that the posterior assigns positive probability to the set of densities with respect to Lebesgue measure. Hence, it is important to provide stronger results.

Until very recently, essentially all the literature on theory of nonparametric Bayes density estimation focused on one-dimensional Euclidean spaces. An important development in multivariate Euclidean spaces is the article of [**20**] who provide sufficient conditions for strong consistency in nonparametric Bayes density estimation from Dirichlet process mixtures of multivariate Gaussian kernels. The theory developed in their paper is specialized and cannot be easily generalized to arbitrary kernel mixtures on more general spaces.

We are particularly interested in density estimation in the special case in which the compact metric space M corresponds to a Riemannian manifold, such as a unit hypersphere or landmark-based planar shape space. In order to extend kernel mixture models used in Euclidean spaces to manifolds M, the kernel needs to be carefully chosen. One approach is to introduce an invertible coordinate map between a subset of M and a Euclidean space ([**11**]). Under such an approach, the density prior on M can be induced through a kernel mixture model in a Euclidean space. However, several major problems arise in using such an approach. Firstly, it is not possible to cover the entire manifold with a single smooth coordinate chart except for very simple manifolds, so unless the data are very concentrated one may obtain poor performance. Different local charts can be patched together to form an atlas, but this may introduce artifactual discontinuities in the resulting den sity. Because the coordinate map is not isometric, the geometry of the manifold can be heavily distorted. As good choices of coordinate frames necessarily depend on the observations, additional uncertainty is automatically induced. Due to these and other shortcomings of coordinate based methods, we focus on modeling approaches that are coordinate free in the sense that we build density models with respect to the invariant volume form on the manifold.

In [**3**], a density model is presented on a general compact metric space with respect to any fixed base measure using a random mixture of probability kernels. Under mild conditions on the kernel and the mixing prior, it is shown that the prior probability of any uniform neighborhood of any continuous density f_0 is positive and if f_0 is positive everywhere, it lies in the KL support of the prior. Density estimation on the planar shape space is presented as a special case. In [**2**], such a density model is used to carry out classification with features on some non-Euclidean manifold and nonparametric Bayes hypothesis testing with observations on the manifold. Consistency results are proved and for illustration, the methods are applied to hyperspheres.

Focusing on kernel mixture priors for densities on a compact metric space M, in this article, we provide sufficient conditions on the kernel, prior and the underlying space to ensure strong consistency. Theorem 2.4 and Corollary 2.5 provide sufficient conditions to ensure that all total variation neighborhoods around f_0 will be assigned probability converging to one as the sample size increases. The theoretical development relies on the method of sieves and exponentially consistent tests reviewed in [**8**]. However, applying this framework outside Euclidean spaces is not standard and requires careful use of differential geometry. To illustrate the theory, we focus on density estimation on the unit hypersphere using von Mises-Fisher kernels and on the planar shape space using complex Watson kernels. In both these cases, it is shown that the kernels satisfy the sufficient conditions. The results also apply to Gaussian mixture densities on $\Re^d$ whenever the true density has compact support.

When the manifold is high-dimensional, priors satisfying conditions for strong consistency tend to put too little probability near bandwidths close to 0, which is undesirable for applications. A gamma prior on the inverse-bandwidth, for example, cannot be shown to satisfy the conditions. Hence, we extend the consistency results to cover cases with priors depending on the sample size n. Theorem 2.6 extends the Schwartz theorem to prove weak consistency, while Theorem 2.9 proves strong consistency using such priors. A gamma prior with scale decreasing with n at an appropriate rate satisfies the conditions for both weak and strong posterior consistency at an exponential rate.

2. Consistency theorems on compact metric spaces

2.1. Weak posterior consistency. Let (M, ρ) be a compact metric space, ρ being the distance metric, and let X be a random variable on M (from some measurable space $(\Omega, \mathcal{B}, Q)$). We assume that the distribution of X has a density with respect to some fixed finite base measure λ on M. The natural choice for such a λ when M is a Riemannian manifold is the invariant volume form. We are interested in modelling this unknown density via a flexible model. Let $K(m; \mu, \kappa)$ be a probability kernel on M with location $\mu \in M$ and inverse-scale $\kappa \in [0, \infty)$, with $\int_M K(m; \mu, \kappa)\lambda(dm) = 1$. Then a location mixture density model for X is defined as

$$f(m; P, \kappa) = \int_M K(m; \mu, \kappa) P(d\mu) \tag{2.1}$$

with parameters P in the space $\mathcal{M}(M)$ of all probability distributions on M and $\kappa \geq 0$. Kernel mixture models are used routinely in Bayesian density estimation in Euclidean spaces, with [**14**] applying such an approach to bivariate angular data and [**2, 3**] considering kernel mixtures on general metric spaces.

A prior Π_1 on (P, κ) induces a prior Π on the space of densities $\mathcal{D}(M)$ on M through the model (2.1). Given a random realization $X_1, \ldots, X_n$ of X, we can compute the posterior of f. The Schwartz theorem([**16**]) provides a useful tool in proving that the posterior assigns probability converging to one in arbitrarily small neighborhoods of the true density f_0 as the sample size $n \to \infty$. Let F_0 denote the probability distribution corresponding to f_0, let $\mathrm{KL}(f_0; f) = \int_M f_0(m) \log\{f_0(m)/f(m)\}\lambda(dm)$ denote the KL divergence of another density f from f_0, and let $K_\epsilon(f_0)$ denote the KL neighborhood $\{f \in \mathcal{D}(M) : \mathrm{KL}(f_0; f) < \epsilon\}$. f_0 is said to be in the KL support of Π if $\Pi\{K_\epsilon(f_0)\} > 0$ for all $\epsilon > 0$.

Proposition 2.1 (Schwartz). *If (1) f_0 is in the KL support of Π, and (2) $U \subset \mathcal{D}(M)$ is such that there exists a uniformly exponentially consistent sequence of test functions for testing H_0: $f = f_0$ versus H_1: $f \in U^c$, then $\Pi(U|X_1, \ldots, X_n) \to 1$ as $n \to \infty$ a.s. F_0^∞.*

The posterior probability of U^c can be expressed as

$$\Pi(U^c|X_1, \ldots, X_n) = \frac{\int_{U^c} \prod_{i=1}^n \frac{f(X_i)}{f_0(X_i)} \Pi(df)}{\int \prod_1^n \frac{f(X_i)}{f_0(X_i)} \Pi(df)} \tag{2.2}$$

Condition (1), known as the KL condition, ensures that for any $\beta > 0$,

$$\liminf_{n\to\infty} \exp(n\beta) \int \prod_{i=1}^{n} \frac{f(X_i)}{f_0(X_i)} \Pi(df) = \infty \text{ a.s.} \tag{2.3}$$

while condition (2) implies that

$$\lim_{n\to\infty} \exp(n\beta_0) \int_{U^c} \prod_{i=1}^{n} \frac{f(X_i)}{f_0(X_i)} \Pi(df) = 0 \text{ a.s.}$$

for some $\beta_0 > 0$ and therefore

$$\lim_{n\to\infty} \exp(n\beta_0/2) \Pi(U^c|X_1, \ldots, X_n) = 0 \text{ a.s.}$$

Hence Proposition 2.1 provides conditions for posterior consistency at an exponential rate. Proposition 2.2, proved in [**3**], derives sufficient conditions on the kernel and the prior so that f_0 is in the KL support of Π. They are

A1 The kernel K is continuous on $M \times M \times (\kappa_0, \infty)$ for some $\kappa_0 \geq 0$.

A2

$$\lim_{\kappa\to\infty} \sup_{m\in M} \left| f_0(m) - \int_M K(m; \mu, \kappa) f_0(\mu) \lambda(d\mu) \right| = 0.$$

A3 For any $P \in \mathcal{M}(M)$ and $\kappa > 0$, there exists $\tilde{\kappa} \geq \kappa$ such that $(P, \tilde{\kappa}) \in \text{supp}(\Pi_1)$ with $\text{supp}(\Pi_1)$ denoting the weak support of Π_1.

A4 f_0 is strictly positive and continuous everywhere.

Proposition 2.2. *Under assumptions* ***A1-A4****, for any $\epsilon > 0$,*

$$\Pi\Big\{ f : \sup_{m\in M} |f(m) - f_0(m)| < \epsilon \Big\} > 0,$$

which implies that f_0 is in the KL support of Π.

When U is a weakly open neighborhood of f_0, condition (2) in Proposition 2.1 is always satisfied. Hence under assumptions **A1-A4**, from Proposition 2.2, weak posterior consistency at an exponential rate follows. We will provide examples of kernels on some compact manifolds which satisfy **A1** and **A2**. **A3** imposes a mild support condition on the prior on the mixing distribution and bandwidth which is easily satisfied by several priors. A common choice is $\Pi_{11} \otimes \pi_1$ with Π_{11} being a Dirichlet process $\text{DP}(w_0 P_0)$ with $\text{supp}(P_0) = M$ and π_1 being a density on $\Re^+$ giving non-zero probability near infinty.

2.2. Strong consistency. When U is a total variation neighborhood of f_0, [**13**] and [**1**] show that condition (2) of Proposition 2.1 will not be satisfied in most cases. In [**1**] (also see [**9**]), a sieve method is considered to obtain sufficient conditions for the numerator in (2.2) to decay at an exponential rate and hence get strong posterior consistency at an exponential rate. This is stated in Proposition 2.3. In its statement, for $\mathcal{F} \subseteq \mathcal{D}(M)$ and $\epsilon > 0$, the L_1-metric entropy $N(\epsilon, \mathcal{F})$ is defined as the logarithm of the minimum number of ϵ-sized (or smaller) L_1 subsets needed to cover $\mathcal{F}$.

Proposition 2.3. *If there exists a $\mathcal{D}_n \subseteq \mathcal{D}(M)$ such that (1) for n sufficiently large, $\Pi(\mathcal{D}_n^c) < exp(-n\beta)$ for some $\beta > 0$, and (2) $N(\epsilon, \mathcal{D}_n)/n \to 0$ as $n \to \infty$ for any $\epsilon > 0$, then for any total variation neighborhood U of f_0, there exists a $\beta_0 > 0$ such that* $\limsup_{n\to\infty} \exp(n\beta_0) \int_{U^c} \prod_1^n \frac{f(X_i)}{f_0(X_i)} \Pi(df) = 0$ *a.s. F_0^∞. Hence if f_0 is in* *the KL support of* Π*, the posterior probability of any total variation neighborhood of* f_0 *converges to 1 almost surely.*

Theorem 2.4, which is the main theorem of this paper, describes a $\mathcal{D}_n$ which satisfies condition (2). We impose the following additional restrictions on the kernel K and the space M.

A5 There exist positive constants $\mathcal{K}_1, a_1, A_1$ such that for all $\mathcal{K} \geq \mathcal{K}_1$, $\mu, \nu \in M$,

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

A6 There exists positive constants a_2, A_2 such that for all $\kappa_1, \kappa_2 \in [0, \mathcal{K}]$, $\mathcal{K} \geq \mathcal{K}_1$,

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

A7 There exist positive constants a_3, A_3, A_4 such that given any $\epsilon > 0$, M can be covered by $A_3 \epsilon^{-a_3} + A_4$ or fewer subsets of diameter at most ϵ.

Theorem 2.4. *For a positive sequence* $\{\kappa_n\}$ *diverging to* ∞*, define*

$$\mathcal{D}_n = \left\{ f(P,\kappa) : P \in \mathcal{M}(M),\ \kappa \in [0, \kappa_n] \right\}.$$

Under assumptions ***A5-A7****, given any* $\epsilon > 0$*, for* n *sufficiently large,* $N(\epsilon, \mathcal{D}_n) \leq C(\epsilon)\kappa_n^{a_1 a_3}$ *for some* $C(\epsilon) > 0$*. Hence* $N(\epsilon, \mathcal{D}_n)$ *is* $o(n)$*, that is,* $\lim_{n\to\infty} N(\epsilon, \mathcal{D}_n)/n = 0$*, whenever* $\kappa_n = o\left(n^{(a_1 a_3)^{-1}}\right)$*.*

As a corollary, we derive conditions on the prior Π_1 on (P, κ) under which strong posterior consistency at an exponential rate follows.

Corollary 2.5. *Under assumptions* ***A1-A7*** *and*
A8 $\Pi_1(\mathcal{M}(M) \times (n^a, \infty)) < \exp(-n\beta)$ *for some* $a < (a_1 a_3)^{-1}$ *and* $\beta > 0$*,*
the posterior probability of any total variation neighborhood of f_0 *converges to 1 a.s.* F_0^∞*.*

When we choose $\Pi_1 = \Pi_{11} \otimes \pi_1$ with a Dirichlet process Π_{11} as in Section 2.1, a choice for π_1 for which assumptions **A3** and **A8** are satisfied is a Weibull density with shape parameter exceeding $a_1 a_3$.

Remark 2.1. A gamma prior on κ satisfies **A3** but not **A8** (unless $a_1 a_3 < 1$). However that does not prove that it is not eligible for strong consistency because Corollary 2.5 provides only sufficient conditions.

When the underlying space is non-compact (but separable) such as $\Re^d$, Corollary 2.5 applies to any true density f_0 with compact support, say M. Then the kernel can be chosen to have non-compact support, such as Gaussian, but the prior on the location mixing distribution needs to have support in $\mathcal{M}(M)$. We may even weaken assumption **A5** to

$$\textbf{A5'} \quad \sup_{\kappa \in [0,\mathcal{K}]} \left\| K(\mu,\kappa) - K(\nu,\kappa) \right\| \leq A_1 \mathcal{K}^{a_1} \rho(\mu,\nu)$$

where $\|f - g\|$ denotes the L_1 distance. The proof of Theorem 2.4 can be easily modified to show consistency under this assumption and is left to the reader. In such a case, we are modelling a compactly supported density with a mixture density possibly having full support but with locations drawn from a compact domain. Using a location mixture of Gaussian kernels on $\Re^d$, a_1 and a_3 from Assumptions **A5'** and **A7** can be shown to be $d/2$ and d respectively. Hence we can take π_1 to be Weibull with shape parameter exceeding $d^2/2$ which can be the gamma prior in one dimension.

Remark 2.2. Unlike in [**9**] and [**20**], Corollary 2.5 imposes no support restriction on the scale parameter. It will be extended to cover densities with non-compact support, in particular $\Re^d$ in later works. Since most of the non-Euclidean manifolds arising in applications are compact, that is not a high priority.

2.3. Consistency with sample size-dependent priors. When the dimension of the manifold is large, as is the case in shape analysis with a large number of landmarks, the constraints on the shape parameter in the proposed Weibull prior on the inverse bandwidth become overly-restrictive. In particular, for strong posterior consistency, the shape parameter needs to be very large in high-dimensional cases, implying a prior on the bandwidth that places very small probability in neighborhoods close to zero, which is undesirable in many applications. By instead allowing the prior to depend on sample size n, we can potentially obtain priors that may have better small sample operating characteristics, while still leading to strong consistency. However, for n-dependent priors, the KL condition is no longer sufficient to ensure that (2.3) holds and hence the Schwartz theorem breaks down. In this section, we will modify the conditions and derive weak and strong consistency results for n-dependent priors.

As recommended in earlier sections, we let P and κ be independent under Π_1. Then, assuming $P \sim \Pi_{11}$ is a constant prior, we focus on the case in which the inverse-bandwidth has a sample size-dependent prior distribution on $\Re^+$, $\kappa \sim \pi_n$. Denote the resulting sequence of induced priors on $\mathcal{D}(M)$ as Π_n. Theorem 2.6 proves weak posterior consistency under the following assumptions on the prior.
A9 The prior Π_{11} has full support.
A10 For any $\beta > 0$, there exists a $\kappa_0 \geq 0$, such that for all $\kappa \geq \kappa_0$,

$$\liminf_{n\to\infty} \exp(n\beta)\pi_n(\kappa) = \infty.$$

Theorem 2.6. *Under assumptions **A1** and **A2** on the kernel, **A9** and **A10** on the prior and **A4** on the true density f_0, the posterior probability of any weak neighborhood of f_0 converges to one a.s. F_0^∞.*

The proof is immediate from the following two lemmas.

Lemma 2.7. *Under assumptions **A1**-**A2**, **A4** and **A9**-**A10**, for any $\beta > 0$,*

$$\liminf_{n\to\infty} \exp(n\beta) \int \prod_1^n \frac{f(X_i)}{f_0(X_i)} \Pi_n(df) = \infty$$

a.s. F_0^∞.

Lemma 2.8. *If there exists a uniformly exponentially consistent sequence of test functions for testing H_0: $f = f_0$ versus H_1: $f \in U^c$, and $\Pi_n(U^c) > 0$ for all n, then for some $\beta_0 > 0$,*

$$\lim_{n\to\infty} \exp(n\beta_0) \int_{U^c} \prod_1^n \frac{f(X_i)}{f_0(X_i)} \Pi_n(df) = 0$$

a.s. F_0^∞.

The proof of Lemma 2.8 is related to that of Lemma 4.4.2.[**10**] which is stated for a constant prior Π but with the set U^c depending on n, they call this V_n. There it is assumed that $\liminf_{n\to\infty}\Pi(V_n) > 0$ but that is not necessary as long as $\Pi(V_n) > 0$ for all $n > C$ with C a sufficiently large constant.

A gamma prior $\pi_n(\kappa) \propto \exp(-\beta_n\kappa)\kappa^{\alpha-1}$, $\alpha, \beta_n > 0$, denoted by $\text{Gam}(\alpha, \beta_n)$ satisfies assumption **A10** as long as β_n is $o(n)$.

For strong consistency, we impose the following additional condition on π_n. Let a_1 and a_3 be as in assumptions **A5** and **A7**.
A11 For some $\beta_0 > 0$ and $a < (a_1a_3)^{-1}$,

$$\lim_{n\to\infty} \exp(n\beta_0)\pi_n\{(n^a, \infty)\} = 0.$$

Theorem 2.9. *Under assumptions **A1-A2**, **A4-A7** and **A9-A11**, the posterior probability of any total variation neighborhood of f_0 converges to 1 a.s F_0^∞.*

The proof is very similar to that of Corollary 2.5 and hence is omitted.

A $\text{Gam}(\alpha, \beta_n)$ prior satisfies **A11** when $n^{1-(a_1a_3)^{-1}}$ is $o(\beta_n)$. Hence, for example, we have weak and strong posterior consistency with $\beta_n = b_1 n/\{\log(n)\}^{b_2}$ for any $b_1, b_2 > 0$.

In the subsequent sections, we consider density estimation on two specific compact manifolds, namely the hypersphere and the planar shape space. We construct mixture models using suitable kernels which satisfy the requirements for weak and strong consistency.

3. Application to unit hypersphere

Let M be the unit sphere S^d embedded in $\Re^{d+1}$. It is a compact Riemannian manifold of dimension d and a compact metric space under the chord distance $\rho(u, v) = \|u - v\|_2$, $\|.\|_2$ denoting the L^2-norm. To define a probability density model as in (2.1) with respect to the volume form V, we need a suitable kernel which satisfies the assumptions in Section 2. One of the most commonly used probability densities on this space is the von Mises-Fisher (vMF) density which is given by

$$\text{vMF}(m; \mu, \kappa) = c^{-1}(\kappa)\exp(\kappa m^T\mu), \tag{3.1}$$

with c being the normalizing constant which can be derived to be

$$\frac{2\pi^{d/2}}{\Gamma(\frac{d}{2})}\int_{-1}^{1}\exp(\kappa t)(1-t^2)^{d/2-1}dt. \tag{3.2}$$

The vMF density on S^1 was first derived in [**17**] and the density in case of S^2 was given by [**7**]. [**18**] generalized this distribution to S^d and examined many of its properties. It can be shown that the parameter μ is the extrinsic mean (as defined in [**4**]), and hence can be interpreted as the distribution location. The parameter κ is a measure of concentration, with $\kappa = 0$ corresponding to the uniform distribution having constant density equal to $1/\int_{S^d} V(dm)$. As κ diverges to ∞, the vMF distribution converges to a point mass at μ in an L^1 sense uniformly. This is proved in Theorem 3.1.

Theorem 3.1. *The vMF kernel satisfies assumption* **A1** *with $\kappa_0 = 0$ and assumption* **A2** *for any continuous f_0.*

Hence from Proposition 2.2, weak posterior consistency follows using the location mixture density model (2.1) with a Dirichlet Process prior on P and an independent gamma prior on κ. In the $d=2$ special case, [**14**] proposed a closely related model but did not consider theoretical properties. Theorem 3.2 verifies the assumptions for strong consistency.

Theorem 3.2. *The vMF kernel on S^d satisfies assumption* **A5** *with $a_1 = d/2+1$ and* **A6** *with $a_2 = d/2$. The compact metric-space (S^d, ρ) satisfies assumption* **A7** *with $a_3 = d$.*

As a result a Weibull prior on κ with shape parameter exceeding $(d + d^2/2)^{-1}$ satisfies the condition of Corollary 2.5 and strong posterior consistency follows. The proofs of Theorems 3.1 and 3.2 use the following lemma which establishes certain properties of the normalizing constant.

Lemma 3.3. *Define $\tilde{c}(\kappa) = \exp(-\kappa)c(\kappa)$, $\kappa \geq 0$. Then $\tilde{c}$ is decreasing and for $\kappa \geq 1$,*

$$\tilde{c}(\kappa) \geq C\kappa^{-d/2}$$

for some appropriate positive constant C.

When d is large, as is often the case for spherical data, a more appropriate prior on κ for which weak and strong consistencies hold can be $\mathrm{Gam}(\alpha, \beta_n)$ as mentioned at the end of §2.3.

4. Planar Shape Space

4.1. Background. Let M be the planar shape space Σ_2^k which is defined as follows. Consider a set of k landmark locations, $k > 2$, on a 2D image, not all points being the same. We refer to such a set 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. We use the following shape representation first proposed by [**12**]. Denote the k-ad by a complex k-vector z in $\mathcal{C}^k$. To remove the effect of translation from z let $z_c = z - \overline{z}$, with $\overline{z} = (\sum_{j=1}^k z_j)/k$ being the centroid. The centered k-ad z_c lies in a $k-1$ dimensional complex subspace, and hence we can use $k-1$ complex coordinates. The effect of scaling is 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 the shape information of z along with rotation, it is called the preshape of z. The similarity shape of z is the orbit of w under all rotations in 2D which is

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

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 the planar shape space 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 the complex projective space and is a compact Riemannian manifold of dimension $2k-4$. The Σ_2^k can be embedded into the space of all order $k-1$ complex Hermitian matrices via the embedding $J([w]) = ww^*$, $*$ denoting the complex conjugate transpose. This embedding induces a distance on Σ_2^k called the extrinsic distance which generates the manifold topology and is given by

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

For more details, see [**3**] and the references cited therein.

4.2. Density model. We define a location-mixture density on Σ_2^k as in (2.1) with respect to the Riemannian volume form V and the kernel being a complex Watson density. This complex Watson density was used in [**5**] for parametric density modelling and is given by

$$\mathrm{CW}(m;\mu,\kappa) = c^{-1}(\kappa)\exp\{\kappa(|z^*\nu|^2 - 1)\} \quad (m = [z], \mu = [\nu]) \tag{4.1}$$

$$\text{with } c(\kappa) = \pi^{k-2}\kappa^{2-k}\left(1 - \exp(-\kappa)\sum_{r=0}^{k-3}\frac{\kappa^r}{r!}\right). \tag{4.2}$$

It is shown in [**3**] that the complex Watson kernel satisfies assumptions **A1** and **A2** in §2. Using a Dirichlet Process prior on the location mixing distribution and an independent gamma prior on the inverse-scale parameter, Proposition 2.2 implies that the density model (2.1) has full support in the space of all positive continuous densities on Σ_2^k in uniform and KL sense and hence the posterior is weakly consistent.

Theorem 4.1 verifies that the complex Watson kernel also satisfies the regularity conditions in **A5** and **A6**.

Theorem 4.1. *The complex Watson kernel CW$(m;\mu,\kappa)$ on the compact metric space Σ_2^k endowed with the extrinsic distance d_E satisfies assumption* **A5** *with $a_1 = k-1$ and* **A6** *with $a_2 = 3k-8$.*

The proof uses Lemma 4.2 which verifies certain properties of the normalizing constant.

Lemma 4.2. *Let $c(\kappa)$ be the normalizing constant for CW(μ,κ) as defined in* (4.2). *Then c is decreasing on $[0,\infty)$ with*

$$\lim_{\kappa\to 0} c(\kappa) = \frac{\pi^{k-2}}{(k-2)!} \text{ and } \lim_{\kappa\to\infty} c(\kappa) = 0.$$

If we define

$$\tilde{c}(\kappa) = 1 - \exp(-\kappa)\sum_{r=0}^{k-3}\frac{\kappa^r}{r!},$$

it follows that $\tilde{c}$ is increasing with

$$\lim_{\kappa\to 0}\tilde{c}(\kappa) = 0, \ \lim_{\kappa\to\infty}\tilde{c}(\kappa) = 1 \text{ and}$$
$$\tilde{c}(\kappa) \geq (k-2)!^{-1}\exp(-\kappa)\kappa^{k-2}.$$

PROOF. Follows from direct computations. □

Theorem 4.3 verifies that assumption **A7** holds on Σ_2^k.

Theorem 4.3. *The compact metric space (Σ_2^k, d_E) satisfies assumption* ***A7*** *with $a_3 = 2k-3$.*

As a result, Corollary 2.5 implies that strong posterior consistency holds with $\Pi_1 = (DP)(\omega_0 P_0)\otimes\pi_1$, for Weibull π_1 with shape parameter exceeding $(2k-3)(k-1)$. Alternatively one may use a gamma prior on κ with inverse-scale increasing with n at a suitable rate and we have consistency using Theorems 2.6 and 2.9.

5. Summary

We consider kernel mixture density models on general compact metric spaces and obtain sufficient conditions on the kernel, priors and the space for the density estimate to be strongly consistent. Thereby we extend the existing literature on strong posterior consistency on $\Re$ using Gaussian kernels to more general non-Euclidean manifolds. The conditions are verified for specific kernels on two important manifolds, namely the hypersphere and the planar shape space. We also allow the prior to depend on the sample size and obtain sufficient conditions for weak and strong consistency. The assumption that the true density is positive everywhere can be relaxed if the locations for the mixture density model are drawn from the support of the truth.

6. Appendix

6.1. Proof of Theorem 2.4. In this proof and the subsequent ones, we shall use a general symbol C for any constant not depending on n (but possibly on ϵ).

Proof. Given $\delta_1 > 0$ ($\equiv \delta_1(\epsilon, n)$), cover M by N_1 ($\equiv N_1(\delta_1)$) many disjoint subsets of diameter at most δ_1: $M = \cup_{i=1}^{N_1} E_i$. Assumption **A7** implies that for δ_1 sufficiently small, $N_1 \leq C\delta_1^{-a_3}$. Pick $\mu_i \in E_i$, $i = 1, \ldots, N_1$, and define for a probability P,

$$P_n = \sum_{i=1}^{N_1} P(E_i)\delta_{\mu_i},\ P_n(\mathbf{E}) = (P(E_1), \ldots, P(E_{N_1}))^T. \tag{6.1}$$

Denoting the L_1-norm as $\|.\|$, for any $\kappa \leq \kappa_n$,

$$\begin{aligned} \|f(P,\kappa) - f(P_n,\kappa)\| &\leq \sum_{i=1}^{N_1} \int_{E_i} \|K(\mu,\kappa) - K(\mu_i,\kappa)\| P(d\mu) \\ &\leq C \sum_i \int_{E_i} \sup_{m \in M} |K(m;\mu,\kappa) - K(m;\mu_i,\kappa)| P(d\mu) \quad (6.2)\\ &\leq C\kappa_n^{a_1}\delta_1. \quad (6.3) \end{aligned}$$

The inequality in (6.3) follows from (6.2) using Assumption **A5**.

For $\kappa, \tilde{\kappa} \leq \kappa_n$, $P \in \mathcal{M}(M)$,

$$\begin{aligned} \|f(P,\kappa) - f(P,\tilde{\kappa})\| &\leq C \sup_{m,\mu \in M} |K(m;\mu,\kappa) - K(m;\mu,\tilde{\kappa})| \\ &\leq C\kappa_n^{a_2}|\kappa - \tilde{\kappa}|, \quad (6.4) \end{aligned}$$

the inequality in (6.4) following from Assumption **A6**. Given $\delta_2 > 0$ ($\equiv \delta_2(\epsilon, n)$), cover $[0, \kappa_n]$ by finitely many subsets of length at most δ_2, the number of such subsets required being at most $\kappa_n\delta_2^{-1}$. Call the collection of these subsets $W(\delta_2, n)$.

Letting $S_d = \{x \in [0,1]^d : \sum x_i \leq 1\}$, S_d is compact under the L^1-metric ($\|x\|_{L^1} = \sum |x_i|$, $x \in \Re^d$), and hence given any $\delta_3 > 0$ ($\equiv \delta_3(\epsilon)$), can be covered by finitely many subsets of the cube $[0,1]^d$ each of diameter at most δ_3. In particular cover S_d with cubes of side length δ_3/d lying partially or totally in S_d. Then an upper bound on the number $N_2 \equiv N_2(\delta_3, d)$ of such cubes can be shown to be $\frac{\lambda(S_d(1+\delta_3))}{(\delta_3/d)^d}$, λ denoting the Lebesgue measure on $\Re^d$ and $S_d(r) = \{x \in [0,\infty)^d :$

$\sum x_i \leq r\}$. Since $\lambda(S_d(r)) = r^d/d!$, hence

$$N_2(\delta_3, d) \leq \frac{d^d}{d!}\left(\frac{1+\delta_3}{\delta_3}\right)^d.$$

Let $\mathcal{W}(\delta_3, d)$ denote the partition of S_d as constructed above.

Let $d_n = N_1(\delta_1)$. For $1 \leq i \leq N_2(\delta_3, d_n)$, $1 \leq j \leq \kappa_n \delta_2^{-1}$, define

$$\mathcal{D}_{ij} = \{f(P,\kappa) : P_n(\mathbf{E}) \in \mathcal{W}_i,\ \kappa \in W_j\},$$

with $\mathcal{W}_i$ and W_j being elements of $\mathcal{W}(\delta_3, d_n)$ and $W(\delta_2, n)$ respectively. We claim that this subset of $\mathcal{D}_n$ has L^1 diameter of at most ϵ. For $f(P,\kappa)$, $f(\tilde{P},\tilde{\kappa})$ in this set, $\|f(P,\kappa) - f(\tilde{P},\tilde{\kappa})\| \leq$

$$\begin{aligned} &\|f(P,\kappa) - f(P_n,\kappa)\| + \|f(P_n,\kappa) - f(\tilde{P}_n,\kappa)\| + \\ &+\|f(\tilde{P}_n,\kappa) - f(\tilde{P},\kappa)\| + \|f(\tilde{P},\kappa) - f(\tilde{P},\tilde{\kappa})\|. \end{aligned} \tag{6.5}$$

¿From inequality (6.3), it follows that the first and third terms in (6.5) are at most $C\kappa_n^{a_1}\delta_1$. The second term can be bounded by

$$\sum_{i=1}^{d_n} |P(E_i) - \tilde{P}(E_i)| < \delta_3$$

and from the inequality in (6.4), the fourth term is bounded by $C\kappa_n^{a_2}\delta_2$. Hence the claim holds if we choose $\delta_1 = C\kappa_n^{-a_1}$, $\delta_2 = C\kappa_n^{-a_2}$, and $\delta_3 = C$. The number of such subsets covering $\mathcal{D}_n$ is at most $N_2(\delta_3, d_n)\kappa_n\delta_2^{-1}$. ¿From Assumption **A7**, it follows that for n sufficiently large,

$$d_n = N_1(\delta_1) \leq C\kappa_n^{a_1 a_3}.$$

Using the Stirling's formula, we can bound $\log(N_2(\delta_3, d_n))$ by Cd_n. Also $\kappa_n\delta_2^{-1}$ is bounded by $C\kappa_n^{a_2+1}$, so that $N(\epsilon, \mathcal{D}_n) \leq$

$$C + C\log(\kappa_n) + Cd_n \leq C\kappa_n^{a_1 a_3}$$

for n sufficiently large. This completes the proof. □

6.2. Proof of Lemma 2.7.

PROOF. Under assumptions **A1** and **A2**, from the proof of Proposition 2.2, it follows that given $\epsilon > 0$, for any $\kappa_0 \geq 0$, there exist $\kappa_2 > \kappa_1 > \kappa_0$ and a weakly open neighborhood $\mathcal{W}$ of F_0 (all depending on ϵ), such that $K_\epsilon(f_0)$ contains $\{f(P,\kappa) : P \in \mathcal{W}, \kappa \in (\kappa_1, \kappa_2)\}$. Hence

$$\begin{aligned} \int \prod_1^n \frac{f(X_i)}{f_0(X_i)}\Pi_n(df) &\geq \int_{K_\epsilon(f_0)} \prod_1^n \frac{f(X_i)}{f_0(X_i)}\Pi_n(df) \\ &\geq \int_{\mathcal{W}\times(\kappa_1,\kappa_2)} \prod_1^n \frac{f(X_i;P,\kappa)}{f_0(X_i)}\pi_n(\kappa)\Pi_{11}(dP)d\kappa. \end{aligned}$$

By the law of large numbers, for any $f \in K_\epsilon(f_0)$,

$$\frac{1}{n}\sum_i \log\{(f_0/f)(X_i)\} \to \mathrm{KL}(f_0; f) < \epsilon$$

a.s. F_0^∞ as $n \to \infty$. Therefore for any $P \in \mathcal{W}$ and $\kappa \in (\kappa_1, \kappa_2)$,

$$\liminf_n \exp(2n\epsilon) \prod_1^n \frac{f(X_i; P, \kappa)}{f_0(X_i)} =$$

$$\liminf_n \exp[n[2\epsilon - (1/n) \sum_i \log\{f_0(X_i)/f(X_i; P, \kappa)\}]] = \infty \text{ a.s. } F_0^\infty.$$

Also from Assumption **A10**, for κ_0 sufficiently large, $\liminf_n \exp(n\epsilon)\pi_n(\kappa) = \infty$ and hence

$$\liminf_n \exp(3n\epsilon) \prod_1^n \frac{f(X_i; P, \kappa)}{f_0(X_i)} \pi_n(\kappa) = \infty \text{ a.s. } F_0^\infty.$$

By Fubini-Tonelli theorem, there exists a $\Omega_0 \subset \Omega$ with probability 1 such that for any $\omega \in \Omega_0$,

$$\liminf_n \exp(3n\epsilon) \prod_1^n \frac{f(X_i(\omega); P, \kappa)}{f_0(X_i(\omega))} \pi_n(\kappa) = \infty$$

for all $(P, \kappa) \in \mathcal{W} \times (\kappa_1, \kappa_2)$ outside of a $\Pi_{11}(dP) \otimes d\kappa$ measure 0 subset. By Assumption **A9**, $\Pi_{11}(\mathcal{W}) > 0$. Therefore using the Fatou's lemma, we conclude that

$$\liminf_n \exp(3n\epsilon) \int \prod_1^n \frac{f(X_i)}{f_0(X_i)} \Pi_n(df) \geq$$

$$\int_{\mathcal{W} \times (\kappa_1, \kappa_2)} \liminf_n \{\exp(3n\epsilon) \prod_1^n \frac{f(X_i; P, \kappa)}{f_0(X_i)} \pi_n(\kappa)\} \Pi_{11}(dP) d\kappa = \infty \text{ a.s. } F_0^\infty.$$

Since ϵ was arbitrary, the proof is completed. □

6.3. Proof of Lemma 3.3.

Proof. Express $\tilde{c}(\kappa)$ as

$$C \int_{-1}^1 \exp\{-\kappa(1-t)\}(1-t^2)^{d/2-1} dt$$

and it is clear that it is decreasing. This expression suggests that

$$\begin{aligned}
\tilde{c}(\kappa) &\geq C \int_0^1 \exp\{-\kappa(1-t)\}(1-t^2)^{d/2-1} dt \\
&\geq C \int_0^1 \exp\{-\kappa(1-t^2)\}(1-t^2)^{d/2-1} dt \\
&= C \int_0^1 \exp(-\kappa u) u^{d/2-1} (1-u)^{-1/2} du \\
&\geq C \int_0^1 \exp(-\kappa u) u^{d/2-1} du \\
&= C \kappa^{-d/2} \int_0^\kappa \exp(-v) v^{d/2-1} dv \\
&\geq C \Big\{ \int_0^1 \exp(-v) v^{d/2-1} dv \Big\} \kappa^{-d/2}
\end{aligned}$$

if $\kappa \geq 1$. This completes the proof. □

6.4. Proof of Theorem 3.1.

PROOF. Denote by M the unit sphere S^d and by ρ the chord distance on it. Express the vMF kernel as

$$K(m;\mu,\kappa) = c^{-1}(\kappa)\exp\left[\kappa\big\{1-\rho^2(m,\mu)/2\big\}\right] \quad (m,\mu\in M;\kappa\in[0,\infty)).$$

Since ρ is continuous on the product space $M\times M$ and c is continuous and non-vanishing on $[0,\infty)$, K is continuous on $M\times M\times[0,\infty)$ and assumption **A1** follows.

For a given continuous function ϕ on M, $m\in M$, $\kappa\geq 0$, define

$$I(m,\kappa) = \phi(m) - \int_M K(m;\mu,\kappa)\phi(\mu)V(d\mu) = \int_M K(m;\mu,\kappa)\{\phi(m)-\phi(\mu)\}V(d\mu).$$

Then showing assumption **A2** for $f_0=\phi$ is equivalent to showing

$$\lim_{\kappa\to\infty}\big(\sup_{m\in M}|I(m,\kappa)|\big) = 0.$$

To simplify $I(m,\kappa)$, make a change of coordinates $\mu\mapsto\tilde\mu = U(m)^T\mu$, $\tilde\mu\mapsto\theta\in\Theta_d\equiv(0,\pi)^{d-1}\times(0,2\pi)$ where $U(m)$ is an orthogonal matrix with first column equal to m and $\theta=(\theta_1,\ldots,\theta_d)^T$ are the spherical coordinates of $\tilde\mu\equiv\tilde\mu(\theta)$ which are given by

$$\tilde\mu_j = \cos\theta_j\prod_{h<j}\sin\theta_h,\ j=1,\ldots,d,\quad \tilde\mu_{d+1} = \prod_{j=1}^d\sin\theta_j.$$

Using these coordinates, the volume form can be written as

$$V(d\mu) = V(d\tilde\mu) = \sin^{d-1}(\theta_1)\sin^{d-2}(\theta_2)\ldots\sin(\theta_{d-1})d\theta_1\ldots d\theta_d$$

and hence $I(m,\kappa)$ equals

$$\int_{\Theta_d} c^{-1}(\kappa)\exp\big\{\kappa\cos(\theta_1)\big\}\big\{\phi(m)-\phi\left(U(m)\tilde\mu\right)\big\}\sin^{d-1}(\theta_1)\ldots\sin(\theta_{d-1})d\theta_1\ldots d\theta_d$$

$$= c^{-1}(\kappa)\int_{\Theta_{d-1}\times(-1,1)}\exp(\kappa t)\big\{\phi(m)-\phi\left(U(m)\tilde\mu\right)\big\}(1-t^2)^{d/2-1} \sin^{d-2}(\theta_2)\ldots\sin(\theta_{d-1})d\theta_2\ldots d\theta_d dt \tag{6.6}$$

where $t=\cos(\theta_1)$, $\tilde\mu=\tilde\mu\big(\theta(t)\big)$ and $\theta(t)=(\arccos(t),\theta_2,\ldots,\theta_d)^T$. In the integrand in (6.6), the distance between m and $U(m)\tilde\mu$ is $\sqrt{2(1-t)}$. Substitute $t=1-\kappa^{-1}s$ in the integral with $s\in(0,2\kappa)$. Define

$$\Phi(s,\kappa) = \sup\big\{|\phi(m)-\phi(\tilde m)|:\ m,\tilde m\in M,\ \rho(m,\tilde m)\leq\sqrt{2\kappa^{-1}s}\big\}.$$

Then

$$\big|\phi(m)-\phi\left(U(m)\tilde\mu\right)\big| \leq \Phi(s,\kappa).$$

Since ϕ is uniformly continuous on (M,ρ), therefore Φ is bounded on $(\Re^+)^2$ and $\lim_{\kappa\to\infty}\Phi(s,\kappa)=0$. Hence from (6.6), we deduce that $\sup_{m\in M}|I(m,\kappa)|\leq$

$$c^{-1}(\kappa)\kappa^{-1}\int\limits_{\Theta_{d-1}\times(0,2\kappa)}\exp(\kappa-s)\Phi(s,\kappa)(\kappa^{-1}s(2-\kappa^{-1}s))^{d/2-1}$$

$$\sin^{d-2}(\theta_2)\dots\sin(\theta_{d-1})d\theta_2\dots d\theta_d ds\leq$$

$$C\kappa^{-d/2}\tilde{c}^{-1}(\kappa)\int\limits_0^\infty\Phi(s,\kappa)e^{-s}s^{d/2-1}ds. \tag{6.7}$$

¿From Lemma 3.3, it follows that

$$\limsup_{\kappa\to\infty}\kappa^{-d/2}\tilde{c}^{-1}(\kappa)<\infty.$$

This in turn, using the Lebesgue Dominated Convergence Theorem implies that the expression in (6.7) converges to 0 as $\kappa\to\infty$. This verifies assumption **A2** and completes the proof. □

6.5. Proof of Theorem 3.2. In the proof, $B_d(r)$ denotes the ball of radius r around 0 in $\Re^d$:

$$B_d(r)=\{x\in\Re^d\colon \|x\|_2\leq r\}$$

and B_d refers to $B_d(1)$.

Proof. It is clear from (3.1) and (3.2) that the vMF kernel K is continuously differentiable on $\Re^{d+1}\times\Re^{d+1}\times[0,\infty)$. Hence

$$\sup_{m\in S^d,\kappa\in[0,\mathcal{K}]}\big|K(m;\mu,\kappa)-K(m;\nu,\kappa)\big|\leq\sup_{m\in S^d,x\in B_{d+1},\kappa\in[0,\mathcal{K}]}\Big\|\frac{\partial}{\partial x}K(m;x,\kappa)\Big\|_2\|\mu-\nu\|_2.$$

Since

$$\frac{\partial}{\partial x}K(m;x,\kappa)=\kappa\tilde{c}^{-1}(\kappa)\exp\{-\kappa(1-m^Tx)\}m,$$

its norm is bounded by $\kappa\tilde{c}^{-1}(\kappa)$. Lemma 3.3 implies that this in turn is bounded by

$$\mathcal{K}\tilde{c}^{-1}(\mathcal{K})\leq C\mathcal{K}^{d/2+1}$$

for $\kappa\leq\mathcal{K}$ and $\mathcal{K}\geq 1$. This proves assumption **A5** with $a_1=d/2+1$.

To verify **A6**, given $\kappa_1,\kappa_2\leq\mathcal{K}$, use the inequality,

$$\sup_{m,\mu\in S^d}\big|K(m;\mu,\kappa_1)-K(m;\mu,\kappa_2)\big|\leq\sup_{m,\mu\in S^d,\kappa\leq\mathcal{K}}\Big|\frac{\partial}{\partial\kappa}K(m;\mu,\kappa)\Big||\kappa_1-\kappa_2|.$$

By direct computations, one can show that

$$\frac{\partial}{\partial\kappa}K(m;\mu,\kappa)=-\frac{\partial}{\partial\kappa}\tilde{c}(\kappa)\tilde{c}^{-2}(\kappa)\exp\{-\kappa(1-m^T\mu)\}$$
$$-\tilde{c}^{-1}(\kappa)\exp\{-\kappa(1-m^T\mu)\}(1-m^T\mu),$$

$$\frac{\partial}{\partial\kappa}\tilde{c}(\kappa)=-C\int_{-1}^1\exp\{-\kappa(1-t)\}(1-t)(1-t^2)^{d/2-1}dt,$$

$$\Big|\frac{\partial}{\partial\kappa}\tilde{c}(\kappa)\Big|\leq C\tilde{c}(\kappa).$$

Therefore, using Lemma 3.3,

$$\Big|\frac{\partial}{\partial\kappa}K(m;\mu,\kappa)\Big| \le C\tilde{c}^{-1}(\kappa) \le C\tilde{c}^{-1}(\mathcal{K}) \le C\mathcal{K}^{d/2}$$

for any $\kappa \le \mathcal{K}$ and $\mathcal{K} \ge 1$. Hence **A6** is verified with $a_2 = d/2$.

Finally to verify **A7**, note that $S^d \subset B_{d+1} \subset [-1,1]^{d+1}$ which can be covered by finitely many cubes of side length $\epsilon/(d+1)$. Each such cube has L_2 diameter ϵ. Hence their intersections with S^d provides a finite ϵ-cover for this manifold. If $\epsilon < 1$, such a cube intersects with S^d only if it lies entirely in $B_{d+1}(1+\epsilon) \cap B_{d+1}(1-\epsilon)^c$. The number of such cubes, and hence the ϵ-cover size can be bounded by

$$C\epsilon^{-(d+1)}\{(1+\epsilon)^{d+1} - (1-\epsilon)^{d+1}\} \le C\epsilon^{-d}$$

for some $C > 0$ not depending on ϵ. This verifies **A7** for appropriate positive constants A_3, A_4 and $a_3 = d$ and completes the proof. □

6.6. Proof of Theorem 4.1.

Proof. Express the complex Watson kernel as

$$K(m;\mu,\kappa) = c^{-1}(\kappa)\exp\left(\frac{-\kappa}{2}d_E^2(m,\mu)\right).$$

Given $\kappa \ge 0$, define

$$\phi(t) = \exp\left(\frac{-\kappa}{2}t^2\right),\ t \in [0,\sqrt{2}].$$

Then $|\phi^{'}(t)| \le \sqrt{2}\kappa$, so that

$$|\phi(t) - \phi(s)| \le \sqrt{2}\kappa|s-t|,\ s,t \in [0,\sqrt{2}]$$

which implies that

$$\begin{aligned} |K(m;\mu,\kappa) - K(m;\nu,\kappa)| &\le c^{-1}(\kappa)\sqrt{2}\kappa|d_E(m,\mu) - d_E(m,\nu)| \\ &\le \sqrt{2}\kappa c^{-1}(\kappa)d_E(\mu,\nu). \end{aligned} \tag{6.8}$$

For $\kappa \le \mathcal{K}$, from Lemma 4.2, it follows that

$$\begin{aligned} \kappa c^{-1}(\kappa) \le \mathcal{K}c^{-1}(\mathcal{K}) &= \pi^{2-k}\mathcal{K}^{k-1}\tilde{c}^{-1}(\mathcal{K}) \\ &\le \pi^{2-k}\mathcal{K}^{k-1}\tilde{c}^{-1}(1) \end{aligned}$$

provided $\mathcal{K} \ge 1$. Hence for any $\mathcal{K} \ge 1$,

$$\sup_{\kappa\in[0,\mathcal{K}]} \kappa c^{-1}(\kappa) \le C\mathcal{K}^{k-1}$$

and from inequality (6.8), $a_1 = k-1$ follows.

By direct computation, one can show that

$$\begin{aligned} \frac{\partial}{\partial\kappa}K(m;\mu,\kappa) &= \pi^{k-2}\exp\Big\{-\frac{1}{2}\kappa d_E^2(m,\mu) - \kappa\Big\}\times \\ &c^{-2}(\kappa)\kappa^{2-k}\Big[\sum_{r=k-1}^{\infty}\frac{\kappa^{r-1}}{r!}\Big\{k-2-\frac{r}{2}d_E^2(m,\mu)\Big\}\Big]. \end{aligned} \tag{6.9}$$

Denote by S the sum in the second line of (6.9) and by T_r its rth term, $r \geq k-1$. Since $d_E^2(m,\mu) \leq 2$, it can be shown that

$$|T_r| \leq \begin{cases} k-2 & \text{if } k-1 \leq r \leq 2k-4, \\ r-k+2 & \text{if } 2k-3 \leq r, \end{cases}$$

so that

$$\begin{aligned} |S| &\leq (k-2)\sum_{r=k-1}^{2k-4} \frac{\kappa^{r-1}}{r!} + \sum_{r=2k-3}^{\infty} \frac{\kappa^{r-1}}{r!}(r-k+2) \\ &= (k-2)\kappa^{k-2}\sum_{r=0}^{k-3} \frac{\kappa^r}{(r+k-1)!} + \kappa^{2k-4}\sum_{r=0}^{\infty} \frac{\kappa^r}{(r+2k-3)!}(r+k-1) \\ &\leq C\kappa^{k-2}e^{\kappa} + \kappa^{2k-4}e^{\kappa}. \end{aligned}$$

Plug the above inequality in (6.9) to get

$$\begin{aligned} \left|\frac{\partial}{\partial\kappa}K(m;\mu,\kappa)\right| &\leq Cc^{-2}(\kappa)\kappa^{2-k}\exp\Big\{-\frac{1}{2}\kappa d_E^2(m,\mu)\Big\}(C\kappa^{k-2}+\kappa^{2k-4}) \\ &\leq Cc^{-2}(\kappa)(C+\kappa^{k-2}). \end{aligned} \tag{6.10}$$

For $\kappa \leq \mathcal{K}$ and $\mathcal{K} \geq 1$, using Lemma 4.2, we bound the expression in (6.10) by

$$\begin{aligned} Cc^{-2}(\mathcal{K})(C+\mathcal{K}^{k-2}) &= C\mathcal{K}^{2k-6}\tilde{c}^{-2}(\mathcal{K})(C+\mathcal{K}^{k-2}) \\ &\leq C\mathcal{K}^{2k-6}\tilde{c}^{-2}(1)(C+\mathcal{K}^{k-2}) \leq C\mathcal{K}^{3k-8} \end{aligned} \tag{6.11}$$

for $\mathcal{K}$ sufficiently large. Since K is a continuously differentiable in κ, from (6.11) it follows that there exists $\mathcal{K}_1 > 0$ such that for all $\mathcal{K} \geq \mathcal{K}_1$, $\kappa_1, \kappa_2 \leq \mathcal{K}$,

$$\begin{aligned} \sup_{m,\mu\in\Sigma_2^k} |K(m;\mu,\kappa_1) - K(m;\mu,\kappa_2)| &\leq \sup_{m,\mu\in\Sigma_2^k,\kappa\in[0,\mathcal{K}]} \left|\frac{\partial}{\partial\kappa}K(m;\mu,\kappa)\right| |\kappa_1-\kappa_2| \\ &\leq C\mathcal{K}^{3k-8}|\kappa_1-\kappa_2|. \end{aligned}$$

This proves Assumption **A6** with $a_2 = 3k-8$ and completes the proof. □

6.7. Proof of Theorem 4.3. In the proof, C_i, $i = 1, 2, \ldots$ denote positive constants possibly depending on k.

Proof. The preshape sphere $\mathcal{CS}^{k-2}$, as a real manifold, can be identified with the real unit sphere S^{2k-3}. Endow it with the chord distance induced by the L^2-norm

$$\|u\|_2 = \sqrt{\sum_{i=1}^{k-1}|u_i|^2} \quad (u = (u_1, \ldots, u_{k-1})^T).$$

Then from Theorem 3.2, it follows that given any $\delta > 0$, $\mathcal{CS}^{k-2}$ can be covered by finitely many subsets of diameter less than or equal to δ, the number of such subsets being bounded by $C_1\delta^{-(2k-3)} + C_2$. The extrinsic distance d_E on Σ_2^k can be bounded by the chord distance on $\mathcal{CS}^{k-2}$ as follows. For $u, v \in \mathcal{CS}^{k-2}$,

$$\begin{aligned} \|u-v\|_2^2 &= 2 - 2\mathrm{Re}(u^*v) \geq 2 - 2|u^*v| = 2(1-|u^*v|) \\ &\geq (1+|u^*v|)(1-|u^*v|) = \frac{1}{2}d_E^2([u],[v]). \end{aligned}$$

Hence $d_E([u],[v]) \leq \sqrt{2}\|u-v\|_2$, so that given any $\epsilon > 0$, the shape image of a δ-cover for $\mathcal{CS}^{k-2}$ with $\delta = \epsilon/\sqrt{2}$ provides an ϵ-cover for Σ_2^k. Hence the ϵ-covering size for Σ_2^k can be bounded by $C_1\epsilon^{-(2k-3)} + C_2$. This completes the proof. $\square$

References

[1] A. R. Barron. Uniformly powerful goodness of fit tests. *Ann. Statist.*, 17:107–24, 1989.

[2] A. Bhattacharya and D. Dunson. Nonparametric Bayes classification and testing on manifolds with applications on hypersphere. 2010. Discussion Paper, Department of Statistical Science, Duke University.

[3] A. Bhattacharya and D. Dunson. Nonparametric Bayesian density estimation on manifolds with applications to planar shapes. *Biometrika*, 2010. In Press.

[4] R. N. Bhattacharya and V. Patrangenaru. Large sample theory of intrinsic and extrinsic sample means on manifolds. *Ann. Statist.*, 31:1–29, 2003.

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

[6] M. D. Escobar and M. West. Bayesian density-estimation and inference using mixtures. *J. Am. Statist. Assoc.*, 90:577–588, 1995.

[7] R. A. Fisher. Dispersion on a sphere. *Proc. of the Royal Soc. of London Ser. A - Math. and Phy. Sci.*, 1130:295–305, 1953.

[8] S. Ghosal. *Bayesian Nonparametrics.* Cambridge University Press, 2010. In Press.

[9] S. Ghosal, J. K. Ghosh, and R. V. Ramamoorthi. Posterior consistency of dirichlet mixtures in density estimation. *Ann. Statist.*, 27:143–158, 1999.

[10] J.K. Ghosh and R.V. Ramamoorthi. *Bayesian Nonparametrics.* Springer, N.Y., 2003.

[11] M. Hirsch. *Differential Topology.* Springer Verlag, New York, 1976.

[12] D. G. Kendall. Shape manifolds, procrustean metrics, and complex projective spaces. *Bull. of the London Math. Soc.*, 16:81–121, 1984.

[13] L. LeCam. Convergence of estimates under dimensionality restrictions. *Ann. Statist.*, 1:38–53, 1973.

[14] K. P. Lennox, D. B. Dahl, M. Vannucci, and J. W. Tsai. Density estimation for protein conformation angles using a bivariate von Mises distribution and Bayesian nonparametrics. *J. Am. Statist. Assoc.*, 104:586–596, 2009.

[15] A. Y. Lo. On a class of Bayesian nonparametric estimates. 1. density estimates. *Ann. Statist.*, 12:351–357, 1984.

[16] L. Schwartz. On Bayes procedures. *Z. Wahrsch. Verw. Gebiete*, 4:10–26, 1965.

[17] R.V. von Mises. Uber die “Ganzzahligkeit” der Atomgewicht und verwandte Fragen. *Physik Z*, 19:490–500, 1918.

[18] G.S. Watson and E.J. Williams. Construction of significance tests on the circle and sphere. *Biometrika*, 43:344–52, 1953.

[19] Y. Wu and S. Ghosal. Kullback-Leibler property of kernel mixture priors in Bayesian density estimation. *Elec J. Statist.*, 2:298–331, 2008.

[20] Y. Wu and S. Ghosal. L_1 - consistency of dirichlet mixtures in multivariate bayesian density estimation on bayes procedures. 2010. To Appear.

Department of Statistical Science, Duke University, Durham, NC, USA