Sufficient conditions for Bayesian consistency

Yang Xing and Bo Ranneby

Research Report
Centre of Biostochastics

Swedish University of Agricultural Sciences

Report 2008:6
ISSN 1651-8543

Sufficient conditions for Bayesian consistency

Yang Xing[1] and Bo Ranneby

Centre of Biostochastics
Swedish University of Agricultural Sciences
SE-901 83 Umeå, Sweden

Abstract

We introduce the Hausdorff α-entropy to study the strong Hellinger consistency of posterior distributions. We obtain general Bayesian consistency theorems which extend the well-known results of Barron, Schervish and Wasserman (1999), Ghosal, Ghosh and Ramamoorthi (1999) and Walker (2004). As an application we strengthen previous results on Bayesian consistency of the (normal) mixture models.

Keywords: Hellinger consistency, posterior distribution, sieve, infinite-dimensional model.

AMS classification: 62G07, 62G20, 62F15.

[1]E-mail address to the correspondence author: yang.xing@sekon.slu.se

1 Introduction

Let $\mathbb{X}$ be a Polish space endowed with a σ-algebra $\mathcal{X}$. We consider a family of probability measures dominated by a σ-finite measure μ in $\mathbb{X}$. Let $X_1, X_2, \ldots, X_n$ stand for an independent identically distributed (i.i.d.) sample of n random variables, taking values in $\mathbb{X}$ and having a common probability density function f_0 with respect to the dominating measure μ. For two probability density functions f and g we denote the Hellinger distance

$$H(f,g) = \left(\int_{\mathbb{X}} \left(\sqrt{f(x)} - \sqrt{g(x)} \ \right)^2 \mu(dx) \right)^{\frac{1}{2}}$$

and the Kullback-Leibler divergence

$$K(f,g) = \int_{\mathbb{X}} f(x) \log \frac{f(x)}{g(x)} \ \mu(dx).$$

Assume that the space $\mathbb{F}$ of probability density functions is separable with respect to the Hellinger metric and that $\mathcal{F}$ is the Borel σ-algebra of $\mathbb{F}$. Denote

$$A_\varepsilon = \left\{ f : \ H(f_0, f) \geq \varepsilon \right\},$$

$$N_\delta = \left\{ f : K(f_0, f) < \delta \right\}.$$

Let Π be a prior distribution on $\mathbb{F}$. It is known that the posterior distribution Π_n of the Π given $X_1, X_2, \ldots, X_n$ coincides with

$$\Pi_n(A) = \frac{\int_A \prod_{i=1}^n f(X_i)\, \Pi(df)}{\int_{\mathbb{F}} \prod_{i=1}^n f(X_i)\, \Pi(df)} \qquad \text{for all measurable subsets} \quad A \subset \mathbb{F}.$$

A more useful expression of the posterior distribution is the following

$$\Pi_n(A) = \frac{\int_A R_n(f)\, \Pi(df)}{\int_{\mathbb{F}} R_n(f)\, \Pi(df)},$$

where $R_n(f) = \prod_{i=1}^n \left\{ f(X_i)/f_0(X_i) \right\}$ stands for the likelihood ratio.

A key point to the area of Bayesian nonparametric inference is to establish consistency of posterior distributions with respect to some metric, typically the Hellinger metric. Early works on consistency of posterior distributions were concerned with weak consistency. Freedman (1963) and Diaconis and

Freedman (1986) had demonstrated that a prior distribution having positive mass on all weak neighborhoods of the true density function f_0 is not necessarily weakly consistent. A sufficient condition for weak consistency was suggested by Schwartz (1965). Recall that f_0 is said to be in the Kullback-Leibler support of the prior distribution Π if $\Pi(N_\delta) > 0$ for all $\delta > 0$. Schwartz (1965) proved that, if f_0 is in the Kullback-Leibler support of Π, then the sequence of posterior distributions accumulates in all weak neighborhoods of f_0. Schwartz's theorem provides a powerful tool in establishing posterior consistency, see, for example, Barron (1999). However, it seems not to be useful for establishing strong consistency. In many applications like density estimation it is natural to ask for strong consistency of Bayesian procedures. Recent attention has switched to studying the strong consistency. It is known that the condition of f_0 being in the Kullback-Leibler support is not enough to guarantee F_0^∞-almost sure consistency of posterior distributions with respect to the Hellinger distance, where F_0^∞ stands for the infinite product distribution of the probability distribution F_0 associated with f_0. Some additional restrictions must be needed to obtain that, for any given $\varepsilon > 0$, $\Pi_n(A_\varepsilon)$ tends to zero F_0^∞-almost surely as $n \to \infty$. Barron et al. (1999), Ghosal et al. (1999) and Walker (2004) have made important contributions in this direction. The results of Barron et al. (1999) and Ghosal et al. (1999) rely upon construction of suitable sieves and computation of metric entropies, which measures the size of the density space $\mathbb{F}$. The sieve approach was discussed in great detail in the monograph by Ghosh and Ramamoorthi (2003), see also the nice review of Wasserman (1998). Walker's approach relies upon summability of prior probability of suitable coverings. In this paper, in order to deal with Bayesian consistency we introduce the Hausdorff α-entropy which is less than the metric entropies provided by Barron et al. (1999) and Ghosal et al. (1999). The Hausdorff α-entropy includes some information on the prior distribution. One of main advantages to use the Hausdorff α-entropy is that in many important cases the Hausdorff α-entropy of the whole density space is finite, whereas the corresponding metric entropies usually take infinite value. We present a more general sufficient condition for strong Hellinger consistency. This extends results given in Barron et al. (1999) and Ghosal et al. (1999). Furthermore, our result also implies Walker's theorem (2004).

The following elementary equality plays an important role in our estimation of the numerator of $\Pi_n(A)$

$$\int_A R_n(f)\,\Pi(df) = \Pi(A)\prod_{k=0}^{n-1}\frac{\int_A R_{k+1}(f)\,\Pi(df)}{\int_A R_k(f)\,\Pi(df)},$$

where we assume that $R_0(f) = 1$ and all denominators on the right hand side do not equal zero. By Lemma 1 of Barron et al. (1999) we know that, if f_0 is in the Kullback-Leibler support of Π, the last product is almost surely well defined. Following Walker (2004) we shall use the function

$$f_{k\,A}(x) = \frac{\int_A f(x) \prod_{i=1}^{k} f(X_i)\,\Pi(df)}{\int_A \prod_{i=1}^{k} f(X_i)\,\Pi(df)} = \frac{\int_A f(x)\,R_k(f)\,\Pi(df)}{\int_A R_k(f)\,\Pi(df)}$$

for each measurable set A of $\mathbb{F}$ with the non-zero denominator. The function $f_{k\,A}$ can be considered as the predictive density of f with a normalized posterior distribution, restricted on the set A. Now we can write

$$\int_A R_n(f)\,\Pi(df) = \Pi(A) \prod_{k=0}^{n-1} \frac{f_{k\,A}(X_{k+1})}{f_0(X_{k+1})}.$$

Our purpose is to apply the Hausdorff α-entropy to deal with the estimation of the last product. We develop Walker's approach (2004). For the denominator of $\Pi_n(A_\varepsilon)$ we apply the known result that the denominator is bounded below by $e^{-n\,c}$ for any given constant $c > 0$ if f_0 is in the Kullback-Leibler support of Π, see Lemma 4 of Barron et al. (1999).

The paper is organized as follows. In Section 2 we first introduce the Hausdorff α-entropy and discuss properties on it. Then general Bayesian consistency theorems are presented. In Section 3 we apply our results to several examples. Our theorems lead to some improvements of known results in these examples. Some other closing remarks are included in Section 4. The final section is a technical appendix.

2 Consistency of posteriors

Barron et al.(1999) provide an elegant general result on strong Hellinger consistency that uses the upper bracketing L_μ-entropy with the following definition. Let L_μ be the space of all nonnegative integrable functions with respect to a measure μ and $||f|| = \int \big|f(x)\big|\,\mu(dx)$ be the standard norm in L_μ. For $\mathcal{G} \subset \mathbb{F}$ and $\delta > 0$, the upper bracketing L_μ-entropy $J_1(\delta, \mathcal{G})$ is defined as the logarithm of the minimum of all numbers N such that there exist $f_1, f_2, \ldots, f_N$ in L_μ with the properties: (a) $\int f_j(x)\,\mu(dx) \le 1 + \delta$ for all j; (b) For each $f \in \mathcal{G}$ there exists some f_j with $f \le f_j$. Motivated by this definition, Ghosal et al.(1999) introduce the L_μ-metric entropy $J_2(\delta, \mathcal{G})$ which is the logarithm of the minimum of all numbers N such that there exist $f_1, f_2, \ldots, f_N$ in L_μ satisfying $\mathcal{G} \subset \bigcup_{i=1}^{N} \{f \in L_\mu : ||f - f_i|| < \delta\}$, see also Definition 4.4.5 in Ghosh et al. (2003). They obtained the following result.

Theorem A. (Ghosal et al., 1999) *Suppose that the true density function* f_0 *is in the Kullback-Leibler support of* Π *and suppose that for any* $\varepsilon > 0$ *there exist* $0 < \delta < \varepsilon,\ c_1,\, c_2 > 0,\ 0 < \beta < \frac{\varepsilon^2}{2}$*, and* $\mathcal{G}_n \subset \mathbb{F}$ *such that for all large* n*,*

(i) $\Pi\big(\mathcal{G}_n^c\big) < c_1\, e^{-n\, c_2}$;

(ii) $J_2(\delta, \mathcal{G}_n) < n\, \beta$.

Then for any $\varepsilon > 0$*,* $\Pi_n(A_\varepsilon)$ *tends to zero almost surely as* $n \to \infty$*.*

Since the inequality $J_2(2\,\delta, \mathcal{G}_n) \leq J_1(\delta, \mathcal{G}_n)$ holds for any $\delta > 0$, Theorem A is essentially stronger than the convergence result of Barron et al.(1999). Later, Walker (2004) used a different condition to give a strong Hellinger consistency result for posterior distributions.

Theorem B. (Walker, 2004) *Suppose that the true density* f_0 *is in the Kullback-Leibler support of* Π *and suppose that for any* $\varepsilon > 0$ *there exist a covering* $\{A_1,\, A_2, \ldots, A_j \ldots\}$ *of* A_ε *and* $0 < \delta < \varepsilon$ *such that* $\sum_{j=1}^{\infty} \sqrt{\Pi(A_j)} < \infty$ *and each* $A_j \subset \{f : H(f_j, f) < \delta\}$ *for some density function* f_j *satisfying* $H(f_j, f_0) > \varepsilon$*. Then for any* $\varepsilon > 0$*,* $\Pi_n(A_\varepsilon)$ *tends to zero almost surely as* $n \to \infty$*.*

Walker, Lijoi and Prunster (2005) state that the square root of Theorem B can be replaced by any $0 < \alpha < 1$. Theorem A and Theorem B both have been shown to be extremely useful in the theory of Bayesian consistency. In this section we introduce the Hausdorff α-entropy in studying Hellinger consistency of posterior distributions. Using the Hausdorff α-entropy as a tool we prove a Bayesian consistency theorem which essentially implies both Theorem A and Theorem B (up to a constant multiple). Our result relaxes the entropy condition of Theorem A and finiteness of the series with the square roots of Theorem B. For convenience of computation, it is worth pointing out that our result also implies an analogue of Theorem B, in which we take away the restriction that the centers of Hellinger balls locate in the set $\{f : H(f, f_0) > \varepsilon\}$ (of course, we need to shrink a little the common radius of Hellinger balls). Denote $\log 0 = -\infty$. Now we define

Definition. *Let $\alpha \geq 0$ and $\mathcal{G} \subset \mathbb{F}$. For $\delta > 0$ we define the Hausdorff α-entropy $J(\delta, \mathcal{G}, \alpha)$ with respect to Π as*

$$J(\delta, \mathcal{G}, \alpha) = \log \inf \sum_{j=1}^{N} \Pi(A_j)^\alpha,$$

where the infimum is taken over all coverings $\{A_1, A_2, \ldots, A_N\}$ of $\mathcal{G}$, where N may be ∞, such that each A_j is contained in $\{f : H(f_j, f) < \delta\}$ for some $f_j \in L_\mu$.

Note that $f_1, f_2, \ldots, f_N$ in the definition are not necessarily density functions, however, it is no problem to define the Hellinger distance of functions in L_μ. The definition of Hausdorff α-entropy $J(\delta, \mathcal{G}, \alpha)$ is motivated by that of the standard Hausdorff α-measure. Clearly, $J(\delta, \mathcal{G}, \alpha) \leq J(\delta, \mathcal{G}, 0) = J_3(\delta, \mathcal{G})$ for all $\alpha \geq 0$, where $J_3(\delta, \mathcal{G})$ stands for the logarithm of the minimum of all numbers $N = N(\delta, \mathcal{G})$ such that there exist functions $f_1, f_2, \ldots, f_N$ in L_μ satisfying $\mathcal{G} \subset \bigcup_{i=1}^{N} \big\{f : H(f_i, f) < \delta\big\}$. Moreover, we have

Lemma 1. *The following statements are true.*

(i) *The inequality*

$$\alpha \log \Pi(\mathcal{G}) \leq J(\delta, \mathcal{G}, \alpha) \leq \alpha \log \Pi(\mathcal{G}) + (1 - \alpha) J_3(\delta, \mathcal{G})$$

holds for all $0 \leq \alpha \leq 1$ and $\mathcal{G} \subset \mathbb{F}$.

(ii) *If $\mathcal{G} \subset \cup_{k=1}^{m} \mathcal{G}_k$ with $1 \leq m \leq \infty$, then*

$$e^{J(\delta, \mathcal{G}, \alpha)} \leq \sum_{k=1}^{m} e^{J(\delta, \mathcal{G}_k, \alpha)}.$$

(iii) *If $0 \leq \alpha_1 \leq \alpha_2 \leq 1$ then*

$$J_3(\delta, \mathcal{G}) = J(\delta, \mathcal{G}, 0) \geq J(\delta, \mathcal{G}, \alpha_1) \geq J(\delta, \mathcal{G}, \alpha_2) \geq J(\delta, \mathcal{G}, 1) = \log \Pi(\mathcal{G}).$$

Since $\log \Pi(\mathcal{G}) \leq 0$, assertion (i) of Lemma 1 implies that if α is close to one then the Hausdorff α-entropy $J(\delta, \mathcal{G}, \alpha)$ is much smaller than $J_3(\delta, \mathcal{G})$. Assertion (ii) states in fact that $J(\delta, \mathcal{G}, \alpha)$ is an increasing subadditive function of $\mathcal{G}$. Now we present the main results of this paper.

Theorem 1. *Let $\varepsilon > 0$. Suppose that the true density function f_0 is in the Kullback-Leibler support of Π and suppose that there exist $0 \le \alpha < 1$, $0 < \delta < \varepsilon(1-\alpha)/7$, $c_1, c_2 > 0$, $0 < \beta < \varepsilon^2/4$, and $\mathcal{G}_n \subset \mathbb{F}$ such that for all large n,*

(i) $\Pi\big(A_\varepsilon \setminus \mathcal{G}_n\big) < c_1\, e^{-n\, c_2}$;

(ii) $J(\delta, \mathcal{G}_n, \alpha) < n\, \beta$.

Then $\Pi_n(A_\varepsilon)$ tends to zero almost surely as $n \to \infty$.

Theorem 1 fails for $\alpha \ge 1$ as shown in the following: assume that $\alpha \ge 1$ and that we do not have Bayesian consistency for some prior Π. Since $\mathbb{F}$ is separable with respect to the Hellinger distance, there exist at most countable subsets $E_1, E_2, \ldots,$ which form a covering of $\mathbb{F}$ and have Hellinger diameters less than any given positive constant 2δ. Denote $A_1 = E_1$ and $A_j = E_j \setminus (A_1 \cup \cdots \cup A_{j-1})$ for $j = 2, 3, \ldots.$ Then all sets A_j are disjoint, $\cup_j A_j = \mathbb{F}$ and the Hellinger diameter of each A_j does not exceed 2δ. Hence $e^{J(\delta, \mathbb{F}, \alpha)} \le \sum_j \Pi(A_j)^\alpha \le \sum_j \Pi(A_j) = \Pi(\mathbb{F}) = 1$, which yields that conditions (i) and (ii) of Theorem 1 are fulfilled for the $\alpha \ge 1$.

Let $N(\delta, \mathcal{G}_{nj})$ be the minimal number of Hellinger balls of radius δ needed to cover $\mathcal{G}$, that is, $N(\delta, \mathcal{G}_{nj}) = e^{J_3(\delta, \mathcal{G})}$. An application of Theorem 1 and Lemma 1 yields the following extension of Theorem A and Theorem B.

Theorem 2. *Let $\varepsilon > 0$. Suppose that the true density function f_0 is in the Kullback-Leibler support of Π and suppose that there exist $0 \le \alpha < 1$, $0 < \delta < \varepsilon(1-\alpha)/7$, $c_1, c_2 > 0$, $0 < \beta < \varepsilon^2/4$, and a sequence $\{\mathcal{G}_n\}_{n=1}^\infty$ of subsets on $\mathbb{F}$ such that each $\mathcal{G}_n$ is contained in $\cup_{j=1}^\infty \mathcal{G}_{nj}$. If*

(i) $\Pi\big(A_\varepsilon \setminus \mathcal{G}_n\big) < c_1\, e^{-n\, c_2}$;

(ii) $\sum\limits_{j=1}^{\infty} N(\delta, \mathcal{G}_{nj})^{1-\alpha}\, \Pi(\mathcal{G}_{nj})^\alpha < e^{n\, \beta}$,

then $\Pi_n(A_\varepsilon)$ tends to zero almost surely as $n \to \infty$.

Proof. From conditions (i) and (ii) of Lemma 1 it turns out that

$$e^{J(\delta, \mathcal{G}_n, \alpha)} \le \sum_{j=1}^{\infty} e^{J(\delta, \mathcal{G}_{nj}, \alpha)} \le \sum_{j=1}^{\infty} N(\varepsilon_n, \mathcal{G}_{nj})^{1-\alpha}\, \Pi(\mathcal{G}_{nj})^\alpha.$$

Then Theorem 2 follows directly from Theorem 1. □

As a direct application we have

Corollary 1. *Suppose that f_0 is in the Kullback-Leibler support of* Π *and suppose that for any $\varepsilon > 0$ there exist $0 < \alpha < 1$ and a covering $\{A_1, A_2, \ldots, A_j \ldots\}$ of A_ε such that*

(i) $\sum_{j=1}^{\infty} \Pi(A_j)^\alpha < \infty$;

(ii) *each $A_j \subset L_\mu$ is included in some Hellinger ball with radius $\frac{\varepsilon(1-\alpha)}{8}$.*

Then for any $\varepsilon > 0$, $\Pi_n(A_\varepsilon)$ tends to zero almost surely as $n \to \infty$.

Proof. Given $\varepsilon > 0$, take $\mathcal{G}_n = \mathbb{F} \cap A_\varepsilon$ and $\mathcal{G}_{nj} = A_j \cap A_\varepsilon$. Then it is clear to check conditions (i) and (ii) of Theorem 2 for $\delta = \frac{\varepsilon(1-\alpha)}{8}$, which concludes the proof. $\square$

As another consequence of Theorem 2 (for $\alpha = 0$) we obtain the strong Hellinger consistency by means of the entropy $J_3(\delta, \mathcal{G})$.

Corollary 2. *Suppose that the true density function f_0 is in the Kullback-Leibler support of Π and suppose that for any $\varepsilon > 0$ there exist $0 < \delta < \varepsilon/7$, $c_1, c_2 > 0$, $0 < \beta < \varepsilon^2/4$, and $\mathcal{G}_n \subset \mathbb{F}$ such that for all large n,*

(i) $\Pi(\mathcal{G}_n^c) < c_1 e^{-n c_2}$;

(ii) $J_3(\delta, \mathcal{G}_n) < n \beta$.

Then for any $\varepsilon > 0$, $\Pi_n(A_\varepsilon)$ tends to zero almost surely as $n \to \infty$.

Remark. By the inequality $H(f,g)^2 \leq ||f - g||$ for all $0 \leq f$, $g \in L_\mu$ we have $J_3(\sqrt{\delta}, \mathcal{G}_n) \leq J_2(\delta, \mathcal{G}_n)$. On the other hand, the inverse inequality $||f - g|| \leq 2\,H(f,g)$ holds for all f, g in $\mathbb{F}$, which together with the triangle inequality yields that $J_2(4\,\delta, \mathcal{G}_n) \leq J_3(\delta, \mathcal{G}_n)$. Therefore, Corollary 2 is in fact an analogue of Theorem A. Walker (2003) has given a nice proof of Corollary 2.

3 Illustrations

In this section we present several examples illustrating our theorems. In particular, we consider two types of mixture models and the infinite-dimensional exponential family. Our theorems lead to some improvements of known results on these examples.

3.1 Normal mixtures for Bayesian density estimation

The normal mixture model is given by

$$f_{\sigma,P}(x) = \phi_\sigma * P = \int \phi_\sigma(x-z)\, P(dz),$$

where ϕ_σ denotes the normal density with mean 0 and variance σ^2, and P is a random probability measure on $\mathbb{R}$ with law Λ selecting discrete distributions almost surely. These models consist of a prior distribution μ for σ and the independent prior distribution Λ, which induces a prior $\Pi = \mu \times \Lambda$ through the mapping $(\sigma, P) \longmapsto f_{\sigma,P}$. Normal mixture models include many important models such as the mixture of Dirichlet process (Ferguson (1973) and Lo (1984)) in which P is the Dirichlet process with parameter measure α, a finite nonnull measure. See Ghosal et al. (1999) and Lijor et al. (2005) for a detailed description of normal mixture models. If the aim is density estimation, it is natural to study Bayesian strong consistency for such models. Applying Theorem 1, we shall prove

Theorem 3. *Suppose that the prior distribution μ has support in $[0, M]$ and suppose that the true density function f_0 is in the Kullback-Leibler support of Π. Let $\beta > 0$. If for any $\delta > 0$ there exist $c_1 > 0$, $c_2 > 0$ and two sequences $a_n \nearrow \infty$, $\sigma_n \searrow 0$ such that for all large n,*

(i) $\Lambda\{P : \, P[-a_n, a_n] < 1 - \delta\} \le e^{-c_1 n}$;

(ii) $\mu\{\sigma < \sigma_n\} \le e^{-c_2 n}$;

(iii) $a_n/\sigma_n \le \beta\, n$,

then for any $\varepsilon > 0$, $\Pi_n(A_\varepsilon)$ tends to zero almost surely as $n \to \infty$.

Theorem 3 strengthens slightly Theorem 7 of Ghosal et al. (1999), where they have the same conditions (i)-(ii) as ours except the last condition (iii), in which they need an arbitrarily small coefficient β (this is essentially equivalent to $a_n/n\sigma_n = \mathrm{o}(1)$ as $n \to \infty$), whereas our condition (iii) is that $a_n/n\sigma_n = \mathrm{O}(1)$ as $n \to \infty$.

3.2 Mixtures of priors

Another type of mixture of priors is defined by

$$\Pi(\cdot) = \sum_{j=1}^{\infty} \rho_j \, \Pi_{B_j}(\cdot),$$

where ρ_j are positive constants with $\sum_{j=1}^{\infty} \rho_j = 1$, and $\Pi_{B_j}(\cdot)$ stands for a probability measure supported on $B_j \subset \mathbb{F}$. Petrone and Wasserman (2002) studied these type priors by terms of Bernstein polynomials. See also Walker (2004) for a convergency result of such priors. Now we apply Theorem 2 to get a sufficient condition of the Bayesian Hellinger consistency. Take $\mathcal{G}_n = \cup_{j=1}^{\infty} B_j$. Condition (i) of Theorem 2 is trivially fulfilled, since the prior distribution Π is supported on $\mathcal{G}_n$. To see (ii), choosing $\mathcal{G}_{nj} = B_j$, it is enough to assume that

$$\sum_{j=1}^{\infty} N(\delta, B_j)^{1-\alpha} \, \Pi(B_j)^{\alpha} = \sum_{j=1}^{\infty} N(\delta, B_j)^{1-\alpha} \, \rho_j^{\alpha} < \infty$$

for any $\delta > 0$. So this condition implies that the posterior distribution is Hellinger consistent at the true density function f_0 if f_0 is in the Kullback-Leibler support of the prior Π.

For example, in the case that $N(\delta, B_j) = (c/\delta)^j$ for some fixed constant $c > 0$ (just like the case of Bernstein polynomials), we need to assume that $\sum_{j=1}^{\infty} (c/\delta)^{(1-\alpha)\,j} \, \rho_j^{\alpha} < \infty$ for each $\delta > 0$. This holds if $\rho_j \le c_1 e^{-c_2 j}$ for all large j, where c_1 and c_2 are two fixed positive constants. The last inequality strengthens the result provided by Walker (2004), who assume that $\rho_j \le c_1 e^{-cj}$ for all $c > 0$ and for all large j, where c_1 is a fixed positive constant.

3.3 Infinite-dimensional exponential families

Here we consider a sequence of independent random variables $\Theta = \{\theta_1, \theta_2, \ldots, \}$ with $\theta_j \sim N(0, \sigma_j^2)$. The infinite-dimensional exponential family of density functions $f_\Theta(x)$ on $[0,1]$ is given by

$$f_\Theta(x) = \exp\Big(\sum_{j=1}^{\infty} \theta_j \, \phi_j(x) - c(\Theta)\Big),$$

where $\{\phi_j(x)\}$ is an orthonormal basis of uniformly bounded functions with respect to the Lebesgue measure on $[0,1]$ and the constant $c(\Theta)$ is chosen such that the integral of $f_\Theta(x)$ on $[0,1]$ is equal to 1. Since any prior on the family $\Omega = \{\Theta\}$ induces naturally a prior on $\mathbb{F} = \{f_\Theta\}$, it is convenient to work directly with Ω. This family is originally studied by Leonard (1978) and Lenk (1988, 1991). Denote $a_j = \sup_{0\le x\le 1}|\phi_j(x)|$. To make $f_\Theta(x)$ to be a density function with probability 1, we assume that $\sum_{j=1}^\infty a_j\,\sigma_j < \infty$, which implies that $\sum_{j=1}^\infty \sigma_j \le \sum_{j=1}^\infty a_j\,\sigma_j < \infty$ since $a_j = \sup_{0\le x\le 1}|\phi_j(x)| \ge \left(\int_0^1 \phi_j(x)^2\right)^{1/2} = 1$. Under the additional condition $\sum_{j=1}^\infty b_j\,\sigma_j < \infty$ with $b_j = \sup_{0\le x\le 1}|\phi_j'(x)|$, Barron et al. (1999) obtained strong Hellinger consistency for the family Ω.

Here we construct a special covering of Ω. Given $0 < \beta < 1$, a positive integer s, and a sequence $\{\delta_j\}$ of positive numbers less than 1. By symmetry of the prior we can only consider the covering of the subfamily $\Omega^+ = \{\Theta : \theta_j \ge 0 \text{ for all } j\}$, which consists of all subsets of the following type

$$\prod_{j=1}^{s}\big\{\Theta:\ \theta_j \in A(n_j,l,\delta_j)\big\},$$

where $n_1, n_2, \ldots, n_s$ are arbitrary nonnegative integers;

$$A(n_j,l,\delta_j) = \Big\{\theta_j:\ (n_j+(l-1)\delta_j^{\frac{1-\beta}{\beta}})^\beta\delta_j^\beta \le \theta_j < (n_j+l\delta_j^{\frac{1-\beta}{\beta}})^\beta\delta_j^\beta\Big\}$$
$$\text{for}\quad l = 1,2,\ldots,N_j;$$
$$A(n_j,N_j+1,\delta_j) = \Big\{\theta_j:\ (n_j+N_j\delta_j^{\frac{1-\beta}{\beta}})^\beta\delta_j^\beta \le \theta_j < (n_j+1)^\beta\delta_j^\beta\Big\}$$

with $N_j = [\delta_j^{1-1/\beta}]$ for $n_j \ge 1$ and $N_j = 0$ for $n_j = 0$. Clearly, for any fixed $s \ge 1$ the union of all these products builds a covering of Ω^+. However, in order to keep uniformly small Hellinger diameters of the covering sets, we are most interesting in the case of $s = \infty$. Unfortunately, such a covering with $s = \infty$ consists of uncountably many sets in which theorem 1 fails to be applied. Hence we have to take an (large) integer s to get a suitable countable covering. Now we check condition (ii) of Theorem 1 for such a covering. Let $2\,\delta$ be the largest Hellinger diameter of all sets in the covering. Assume that δ is a finite number. By the definition of the Hausdorff 1/2-entropy we have

$$\begin{aligned}
e^{J(\delta,\Omega,1/2)} &\le \sum_{n_1=0}^\infty\cdots\sum_{n_s=0}^\infty\prod_{j=1}^s\sum_{l=1}^{N_j+1}\sqrt{Pr\{\theta_j\in A(n_j,l,\delta_j)\}}\\
&\le \prod_{j=1}^s\sum_{n=0}^\infty\sum_{l=1}^{N_j+1}\sqrt{Pr\{\theta_j\in A(n,l,\delta_j)\}}\\
&= \prod_{j=1}^s\Big(\ 1+\sum_{n=1}^\infty\sum_{l=1}^{N_j+1}\sqrt{Pr\{\theta_j\in A(n,l,\delta_j)\}}\ \Big).
\end{aligned}$$

From the inequality $|a^\beta - b^\beta| \le |a-b|^\beta$ it turns out that $|\theta_{1j} - \theta_{2j}| \le \delta_j$ for all θ_{1j}, θ_{2j} in $A(n_j, l, \delta_j)$ with $j = 1, 2, \ldots, s$, which yields

$$\begin{aligned}
&\sum_{n=1}^{\infty} \sum_{l=1}^{N_j+1} \sqrt{Pr\{\theta_j \in A(n, l, \delta_j)\}} \\
&\le \; \Big(\frac{1}{2\pi}\Big)^{1/4} \Big(\frac{\delta_j}{\sigma_j}\Big)^{1/2} \sum_{n=1}^{\infty} \sum_{l=1}^{N_j+1} \exp\Big(-\frac{\big(n + (l-1)\delta_j^{\frac{1-\beta}{\beta}}\big)^{2\beta} \delta_j^{2\beta}}{4\sigma_j^2}\Big) \\
&\le \; \Big(\frac{1}{2\pi}\Big)^{1/4} \Big(\frac{\delta_j}{\sigma_j}\Big)^{1/2} (N_j + 1) \sum_{n=1}^{\infty} \exp\Big(-\frac{n^{2\beta}\delta_j^{2\beta}}{4\sigma_j^2}\Big) \\
&\le \; \Big(\frac{1}{2\pi}\Big)^{1/4} \Big(\frac{\delta_j}{\sigma_j}\Big)^{1/2} 2\, \delta_j^{1-1/\beta} \frac{m! 4^m \sigma_j^{2m}}{\delta_j^{2\beta m}} \sum_{n=1}^{\infty} \frac{1}{n^{2\beta m}} \\
&= \; 2\, m!\, 4^m \Big(\frac{1}{2\pi}\Big)^{1/4} \frac{\sigma_j^{2m-\frac{1}{2}}}{\delta_j^{2\beta m + \frac{1}{\beta} - \frac{3}{2}}} \sum_{n=1}^{\infty} \frac{1}{n^{2\beta m}},
\end{aligned}$$

where the last inequality follows from $e^x \ge x^m/m!$ for each m and $x \ge 0$. Then for $m > (2\beta)^{-1}$ we have that $d = 2\, m! 4^m \, (2\pi)^{-1/4} \sum_{n=1}^{\infty} n^{-2\beta m} < \infty$ and hence, for any s,

$$e^{J(\delta, \Omega, 1/2)} \le \prod_{j=1}^{\infty} \Big(1 + \frac{d\, \sigma_j^{2m-\frac{1}{2}}}{\delta_j^{2\beta m + \frac{1}{\beta} - \frac{3}{2}}}\Big) \le \exp\Big(d \sum_{j=1}^{\infty} \frac{\sigma_j^{2m-\frac{1}{2}}}{\delta_j^{2\beta m + \frac{1}{\beta} - \frac{3}{2}}}\Big)$$

which is finite if

$$\sum_{j=1}^{\infty} \frac{\sigma_j^{2m-\frac{1}{2}}}{\delta_j^{2\beta m + \frac{1}{\beta} - \frac{3}{2}}} < \infty.$$

Given $d_0 > 0$, let $\delta_j = d_0\, \sigma_j \big/ \sum_{j=1}^{\infty} \sigma_j$. Then the above condition is equivalent to

$$\sum_{j=1}^{\infty} \sigma_j^{2(1-\beta)m - \frac{1}{2} - \frac{1}{\beta} + \frac{3}{2}} < \infty,$$

which holds if we choose a positive $\beta < 1$ and a sufficiently large m such that the exponent in the last sum is bigger than 1.

Now we prove the Bayesian consistency for the family

$$\Omega_C = \Big\{\Theta : \sum_{j=1}^{\infty} c_j\, |\theta_j| \le c_0\Big\},$$

where the positive sequence $C = \{c_0, c_1, \dots\}$ satisfies $\lim_{j\to\infty} c_j = \infty$. Since the corresponding density functions $f_\Theta(x)$ are bounded in $[0, 1]$, it follows from Barron et al.(1999) that the true density function is always in the Kullback-Leibler support of the prior Π. Using the argument above we only need to show that the largest Hellinger diameter $2\,\delta$ of the covering sets can become arbitrarily small if d_0 is small. Let $\Theta_1 = \{\theta_{1j}\}$ and $\Theta_2 = \{\theta_{2j}\}$ belong in some covering set. Then $|\theta_{1j} - \theta_{2j}| \leq \delta_j^\beta = d_0^\beta \left(\sigma_j / \sum_{j=1}^\infty \sigma_j\right)^\beta \leq d_0^\beta$ for $j = 1, 2, \dots, s$. Hence we have

$$
\begin{aligned}
&\sup_{0\leq x\leq 1} \Big| \sum_{j=1}^\infty \theta_{1j}\phi_j(x) - \sum_{j=1}^\infty \theta_{2j}\phi_j(x) \Big| \\
\leq\ & \Big(s\, d_0^\beta + \sum_{j=s+1}^\infty \big(|\theta_{1j}| + |\theta_{2j}|\big)\Big) \max_j a_j \\
\leq\ & \Big(s\, d_0^\beta + \frac{1}{\max\limits_{j\geq s+1} c_j} \sum_{j=s+1}^\infty \big(c_j\,|\theta_{1j}| + c_j\,|\theta_{2j}|\big)\Big) \max_j a_j \\
\leq\ & \Big(s\, d_0^\beta + \frac{2\,c_0}{\max\limits_{j\geq s+1} c_j}\Big) \max_j a_j,
\end{aligned}
$$

which can be arbitrarily small if we first take a large s and then let d_0 be small enough. Therefore, together with

$$
H(f_{\Theta_1}, f_{\Theta_2})^2 = 2 - 2\int \sqrt{f_{\Theta_1}\, f_{\Theta_2}} = 2 - 2E_{\Theta_2}\big(\frac{f_{\Theta_1}}{f_{\Theta_2}}\big)^{\frac{1}{2}},
$$

we have obtain that the Hellinger diameters of the above covering sets can be made uniformly small. Thus, the strong consistency of posterior distributions follows from Theorem 1.

4 Discussion

A satisfactory covering (with $s = \infty$) in the example of section 3.3 consists of uncountable many sets. In fact, it is easy to see that these covering sets are not Hellinger open sets. It is worth to construct a suitable covering only consisting of Hellinger open subsets. Since $\mathbb{F}$ is separable with respect to the Hellinger distance, any (uncountable) covering must contain a countable subcovering for which Theorem 1 can be applied.

It is known that the Hellinger metric is essentially equivalent to the L^1-norm. So one can formulate Theorem 1 by using the Hausdorff α-entropy related to the L^1-norm instead of the Hellinger metric. An interesting problem is to get Bayesian consistency by means of the Hausdorff α-entropy related to the Kullback-Leibler divergence. Anyway, to make our result more useful, we should further understand the Hausdorff α-entropy.

We have not discussed rates of convergence in this paper. It is no problem to use the Hausdorff α-entropy as a tool to discuss rates of convergence of posterior distributions.

Appendix

Proof of Lemma 1. (i) The first inequality follows from

$$\begin{aligned} J(\delta, \mathcal{G}, \alpha) &= \log \inf \sum_{j=1}^{N} \Pi(A_j)^\alpha \geq \log \inf \Big(\sum_{j=1}^{N} \Pi(A_j) \Big)^\alpha \\ &\geq \log \inf \Big(\Pi\big(\bigcup_{j=1}^{N} A_j\big) \Big)^\alpha \geq \alpha \log \Pi(\mathcal{G}). \end{aligned}$$

To prove the second inequality, given $\varepsilon > 0$, take a partition $\{A_1, A_2, \ldots, A_N\}$ of $\mathcal{G}$ such that each A_j has the Hellinger diameter less than 2δ and $J_3(\delta, \mathcal{G}) + \varepsilon > \log N$. It then follows from Hölder's inequality that

$$\begin{aligned} J(\delta, \mathcal{G}, \alpha) &\leq \log \sum_{j=1}^{N} \Pi(A_j)^\alpha \leq \log \Big\{ \big(\sum_{j=1}^{N} \Pi(A_j)^{\alpha \cdot \frac{1}{\alpha}} \big)^\alpha \big(\sum_{j=1}^{N} 1 \big)^{1-\alpha} \Big\} \\ &= \log \Big\{ \big(\sum_{j=1}^{N} \Pi(A_j) \big)^\alpha N^{1-\alpha} \Big\} = \log \big\{ \Pi(\mathcal{G})^\alpha N^{1-\alpha} \big\} \\ &\leq \alpha \log \Pi(\mathcal{G}) + (1-\alpha) J_3(\delta, \mathcal{G}) + (1-\alpha)\, \varepsilon, \end{aligned}$$

which implies the second inequality.

(ii) For any $\delta_k > 0$ there exists $\cup_{j=1}^{N_k} A_{kj} \supset \mathcal{G}_k$ such that the Hellinger diameter of each A_{kj} is less than $2\,\delta$ and

$$\sum_{j=1}^{N_k} \Pi(A_{kj})^\alpha \leq (1 + \delta_k)\, e^{J(\delta, \mathcal{G}_k, \alpha)},$$

which yields that

$$e^{J(\delta, \mathcal{G}, \alpha)} \leq \sum_{k=1}^{m} \sum_{j=1}^{N_k} \Pi(A_{kj})^\alpha \leq \sum_{k=1}^{m} e^{J(\delta, \mathcal{G}_k, \alpha)} + \sum_{k=1}^{m} \delta_k\, e^{J(\delta, \mathcal{G}_k, \alpha)}.$$

By the arbitrariness of $\delta_k > 0$ we have obtained the required inequality.

(iii) The first equality is trivial and all the inequalities follows directly from the definition of Hausdorff α-entropy. To see the last quality, for any covering of $\mathcal{G}$ with the Hellinger diameters less than $2\delta > 0$ there exists a finer covering $A_1^\star$, $A_2^\star, \ldots$ of $\mathcal{G}$ containing at most countable many disjoint subsets of $\mathcal{G}$, since the space $\mathbb{F}$ is separable with respect to the Hellinger metric. This implies that

$$J(\delta, \mathcal{G}, 1) = \log \inf \sum_j \Pi(A_j^\star) = \log \Pi(\mathcal{G}).$$

The proof of Lemma 1 is complete. □

Proof of Theorem 1. Given $\varepsilon > 0$, we have

$$\Pi_n(A_\varepsilon) \leq \Pi_n\big(\mathcal{G}_n \cap A_\varepsilon\big) + \Pi_n\big(A_\varepsilon \setminus \mathcal{G}_n\big).$$

By Lemma 5 of Barron et al. (1999), assumption (i) implies that the second term $\Pi_n\big(A_\varepsilon \setminus \mathcal{G}_n\big)$ tends to zero almost surely as $n \to \infty$. From assumption (ii) it follows that there exist functions $f_1, f_2, \ldots, f_N$ in L_μ such that $\mathcal{G}_n \cap A_\varepsilon \subset \bigcup_{i=1}^N A_j$, where $A_j = \mathcal{G}_n \cap A_\varepsilon \cap \{f : H(f_j, f) < \delta\}$ and $\sum_{j=1}^N \Pi(A_j)^\alpha < e^{n\beta}$. Shrinking A_j if necessary, we assume that all sets A_j are disjoint. Assume also that $A_j \neq \emptyset$ for all j, otherwise we take away A_j in the covering. Taking $f_j^\star \in A_j$ and applying the triangle inequality, we get that $H(f_j, f_0) \geq H(f_j^\star, f_0) - H(f_j^\star, f_j) \geq \varepsilon - \delta$ for all j. Furthermore, by Jensen's inequality we have that

$$H^2(f_{k\,A_j}, f_j) \leq \frac{\int_{A_j} H^2(f, f_j)\, R_k(f)\, \Pi(df)}{\int_{A_j} R_k(f)\, \Pi(df)} \leq \delta^2,$$

which, together with the triangle inequality, yields that

$$H(f_{k\,A_j}, f_0) \geq H(f_j, f_0) - H(f_j, f_{k\,A_j}) \geq \varepsilon - 2\,\delta := \gamma > 0.$$

On the other hand, for any subset $A \subset \mathbb{F}$ the equality

$$\int_A R_n(f)\, \Pi(df) = \Pi(A) \prod_{k=0}^{n-1} \frac{f_{k\,A}(X_{k+1})}{f_0(X_{k+1})}$$

holds where $R_0(f) = 1$ and $R_k(f) = \prod_{i=1}^{k} \big\{f(X_i)/f_0(X_i)\big\}$ for $k \geq 1$. By (iii) of

Lemma 1 it is no restriction to assume $0 < \alpha < 1$. Then we have

$$\begin{aligned}
\Pi_n\big(\mathcal{G}_n \cap A_\varepsilon\big) \;&\leq\; \big(\Pi_n(\mathcal{G}_n \cap A_\varepsilon)\big)^\alpha \leq \Big(\sum_{j=1}^N \Pi_n(A_j)\Big)^\alpha \\
&\leq\; \sum_{j=1}^N \Pi_n(A_j)^\alpha = \frac{\sum_{j=1}^N \Pi(A_j)^\alpha \prod_{k=0}^{n-1} \frac{f_{k\,A_j}(X_{k+1})^\alpha}{f_0(X_{k+1})^\alpha}}{\Big(\int_{\mathbb{F}} R_n(f)\,\Pi(df)\Big)^\alpha}.
\end{aligned}$$

We estimate the last numerator and denominator separately. For the numerator, given $b > 0$ we get

$$\begin{aligned}
&F_0^\infty\Big\{\sum_{j=1}^N \Pi(A_j)^\alpha \prod_{k=0}^{n-1} \frac{f_{k\,A_j}(X_{k+1})^\alpha}{f_0(X_{k+1})^\alpha} \geq e^{-n\,b\,\varepsilon^2}\Big\} \\
&\leq\; e^{n\,b\,\varepsilon^2}\, E \sum_{j=1}^N \Pi(A_j)^\alpha \prod_{k=0}^{n-1} \frac{f_{k\,A_j}(X_{k+1})^\alpha}{f_0(X_{k+1})^\alpha} \\
&=\; e^{n\,b\,\varepsilon^2} \sum_{j=1}^N \Pi(A_j)^\alpha\, E\Big(\prod_{k=0}^{n-1} \frac{f_{k\,A_j}(X_{k+1})^\alpha}{f_0(X_{k+1})^\alpha}\Big).
\end{aligned}$$

Let $\mathbb{F}_k = \sigma\{X_1, X_2, \ldots, X_k\}$. Then we have

$$\begin{aligned}
E\Big(\prod_{k=0}^{n-1} \frac{f_{k\,A_j}(X_{k+1})^\alpha}{f_0(X_{k+1})^\alpha}\Big) &= E\bigg(E\Big(\prod_{k=0}^{n-1} \frac{f_{k\,A_j}(X_{k+1})^\alpha}{f_0(X_{k+1})^\alpha} \;\Big|\; \mathbb{F}_{n-1}\Big)\bigg) \\
&= E\bigg(\prod_{k=0}^{n-2} \frac{f_{k\,A_j}(X_{k+1})^\alpha}{f_0(X_{k+1})^\alpha}\; E\Big(\frac{f_{n-1\,A_j}(X_n)^\alpha}{f_0(X_n)^\alpha} \;\Big|\; \mathbb{F}_{n-1}\Big)\bigg),
\end{aligned}$$

where by the conditional Hölder's inequality we get that with probability one,

$$\begin{aligned}
E\Big(\frac{f_{n-1\,A_j}(X_n)^\alpha}{f_0(X_n)^\alpha} \;\Big|\; \mathbb{F}_{n-1}\Big) &= E\Big(\frac{f_{n-1\,A_j}(X_n)^{\frac{\alpha}{2}}}{f_0(X_n)^{\frac{\alpha}{2}}}\,\frac{f_{n-1\,A_j}(X_n)^{\frac{\alpha}{2}}}{f_0(X_n)^{\frac{\alpha}{2}}} \;\Big|\; \mathbb{F}_{n-1}\Big) \\
&\leq E\Big(\frac{f_{n-1\,A_j}(X_n)^{\frac{\alpha}{2}\cdot\frac{2}{2-\alpha}}}{f_0(X_n)^{\frac{\alpha}{2}\cdot\frac{2}{2-\alpha}}} \;\Big|\; \mathbb{F}_{n-1}\Big)^{\frac{2-\alpha}{2}}\; E\Big(\frac{f_{n-1\,A_j}(X_n)^{\frac{\alpha}{2}\cdot\frac{2}{\alpha}}}{f_0(X_n)^{\frac{\alpha}{2}\cdot\frac{2}{\alpha}}} \;\Big|\; \mathbb{F}_{n-1}\Big)^{\frac{\alpha}{2}} \\
&= E\Big(\frac{f_{n-1\,A_j}(X_n)^{\frac{\alpha}{2-\alpha}}}{f_0(X_n)^{\frac{\alpha}{2-\alpha}}} \;\Big|\; \mathbb{F}_{n-1}\Big)^{\frac{2-\alpha}{2}}.
\end{aligned}$$

Take the smallest non-negative integer m satisfying $\frac{\alpha}{2^m(1-\alpha)+\alpha} \leq \frac{1}{2}$, i.e. $\frac{\alpha}{1-\alpha} \leq 2^m < \frac{2\alpha}{1-\alpha}$. Repeating the above procedure $m-1$ more times we obtain that with probability one,

$$E\Bigg(\frac{f_{n-1\,A_j}(X_n)^\alpha}{f_0(X_n)^\alpha} \;\Bigg|\; \mathbb{F}_{n-1} \Bigg) \leq E\Bigg(\frac{f_{n-1\,A_j}(X_n)^{\frac{\alpha}{2^m(1-\alpha)+\alpha}}}{f_0(X_n)^{\frac{\alpha}{2^m(1-\alpha)+\alpha}}} \;\Bigg|\; \mathbb{F}_{n-1} \Bigg)^{\frac{2^m(1-\alpha)+\alpha}{2^m}},$$

which by the conditional Hölder's inequality is less than

$$\begin{aligned} & E\Bigg(\frac{f_{n-1\,A_j}(X_n)^{\frac{1}{2}}}{f_0(X_n)^{\frac{1}{2}}} \;\Bigg|\; \mathbb{F}_{n-1} \Bigg)^{\frac{\alpha}{2^{m-1}}} \\ &= \Big(\int \sqrt{f_{n-1\,A_j}(X_n)\; f_0(X_n)}\; \mu(dX_n) \Big)^{\frac{\alpha}{2^{m-1}}} \\ &\leq (1-\frac{\gamma^2}{2})^{2^{1-m}\,\alpha} \leq e^{-2^{-m}\,\gamma^2\,\alpha}. \end{aligned}$$

Hence, with probability one, we have

$$E\Bigg(\prod_{k=0}^{n-1} \frac{f_{k\,A_j}(X_{k+1})^\alpha}{f_0(X_{k+1})^\alpha} \Bigg) \leq e^{-2^{-m}\,\gamma^2\,\alpha}\; E\Bigg(\prod_{k=0}^{n-2} \frac{f_{k\,A_j}(X_{k+1})^\alpha}{f_0(X_{k+1})^\alpha} \Bigg).$$

Repeat the same argument $n-1$ times and we get

$$E\Bigg(\prod_{k=0}^{n-1} \frac{f_{k\,A_j}(X_{k+1})^\alpha}{f_0(X_{k+1})^\alpha} \Bigg) \leq e^{-n\,2^{-m}\,\gamma^2\,\alpha}.$$

Thus we have obtained that for all n,

$$\begin{aligned} & F_0^\infty\Big\{ \sum_{j=1}^N \Pi(A_j)^\alpha \prod_{k=0}^{n-1} \frac{f_{k\,A_j}(X_{k+1})^\alpha}{f_0(X_{k+1})^\alpha} \geq e^{-n\,b\,\varepsilon^2} \Big\} \\ &\leq e^{n\,(b\,\varepsilon^2 - 2^{-m}\,\gamma^2\,\alpha)} \sum_{j=1}^N \Pi(A_j)^\alpha \\ &\leq e^{n\,(\beta + b\,\varepsilon^2 - 2^{-m}\,\gamma^2\,\alpha)}, \end{aligned}$$

where m is an integer with $\frac{\alpha}{1-\alpha} \leq 2^m < \frac{2\alpha}{1-\alpha}$. By $\delta < \varepsilon\,(1-\alpha)/7$ and $\beta < \varepsilon^2/4$ we get $\beta < 2^{-m}\,\gamma^2\,\alpha$. Therefore, for any $0 < b < (2^{-m}\,\gamma^2\,\alpha - \beta)/\varepsilon^2$ we have that $\sum_{n=1}^{\infty} e^{n\,(\beta + b\,\varepsilon^2 - 2^{-m}\,\gamma^2\,\alpha)} < \infty$. Thus, the first Borel-Cantelli Lemma yields that

$$\sum_{j=1}^N \Pi(A_j)^\alpha \prod_{k=0}^{n-1} \frac{f_{k\,A_j}(X_{k+1})^\alpha}{f_0(X_{k+1})^\alpha} \leq e^{-n\,b\,\varepsilon^2}$$

holds almost surely for all n large enough.

The estimation of the denominator follows from Lemma 4 of Barron et al. (1999). They obtained that for any $\eta > 0$ the inequality

$$\Big(\int_{\mathbb{F}} R_n(f)\,\Pi(df)\Big)^{\alpha} \geq e^{-n\,\eta\,\alpha}$$

holds almost surely for all n large enough.

Finally, applying the above estimations for both numerator and denominator and choosing $\eta = \frac{b\,\varepsilon^2}{2\,\alpha}$, we obtain that

$$\Pi_n(\mathcal{G}_n \cap A_\varepsilon) \leq e^{-n\,b\,\varepsilon^2 + n\,\eta\,\alpha} \leq e^{-\frac{n\,b\,\varepsilon^2}{2}}$$

almost surely for all sufficiently large n and the proof of Theorem 1 is complete. □

Proof of Theorem 3. We need to construct sieves $\mathcal{G}_n$ satisfying conditions (i) and (ii) of Theorem 1. Given $\delta > 0$, following Ghosal et al. (1999) we choose

$$\mathcal{G}_n = \bigcup_{\sigma_n < \sigma < M} \big\{\phi_\sigma * P :\; P[-a_n, a_n] > 1 - \delta\big\}.$$

Then conditions (i) and (ii) implies that $\mathcal{G}_n$ fulfill condition (i) of Theorem 1. On the other hand, by the inequality $H(f,g)^2 \leq ||f - g||$ and Theorem 6 of Ghosal et al. (1999) we obtain that $J_3(\sqrt{\delta}, \mathcal{G}_n) \leq J_2(\delta, \mathcal{G}_n) \leq K\, a_n/\sigma_n \leq K\beta n$ for all n, where the last inequality follows from condition (ii) and K is some constant depending only on δ and M. It then turns out from Lemma 1 that

$$J(\sqrt{\delta}, \mathcal{G}_n, \alpha) \leq \alpha\, \log \Pi(\mathcal{G}_n) + (1-\alpha)K\beta n \leq n\,\big(\,\frac{\alpha}{n} + (1-\alpha)K\beta\,\big).$$

Taking α sufficiently close to one and then letting n be large enough one can make $\frac{\alpha}{n} + (1-\alpha)K\beta$ arbitrarily small, which implies condition (ii) of Theorem 1 and hence we have obtained strong consistency of the prior distributions Π_n. The proof of Theorem 3 is complete. □

References

[1] Barron, A. (1999). Information-theoretic characterization of Bayesian performance and the choice of priors in parametric and nonparametric problems. In *Bayesian Statistics 6*, Ed:Alcoceber, pages 27-52. Oxford University Press, New York.

[2] Barron, A., Schervish, M., and Wasserman, L. (1999). The consistency of posterior distributions in nonparametric problems. *Ann. Statist.* **27**, 536-561.

[3] Diaconis, P. and Freedman, D. (1986). On the consistency of Bayes estimates. *Ann. Statist.* **14**, 1-26.

[4] Ferguson, T.S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1**, 209-230.

[5] Freedman, D. (1963). On the asymptotic behavior of Bayes estimates in the discrete case. *Ann. Math. Statist.* **34**, 1386-1403.

[6] Ghosal, S., Ghosh, J.K., and Ramamoorthi, R.V. (1999). Posterior consistency of Dirichlet mixtures in density estimation. *Ann. Statist.* **27**, 143-158.

[7] Ghosh, J.K. and Ramamoorthi, R.V. (2003). *Bayesian Nonparametrics.* Springer-Verlag, New York.

[8] Lenk, P.J. (1988). The logistic normal distribution for Bayesian, nonparametric, predictive density. *J. Amer. Statist. Assoc.* **83**, 509-516.

[9] Lenk, P.J. (1991). Towards a practicable Bayesian nonparametric density estimator. *Biometrika* **78**, 531-543.

[10] Leonard, T. (1978). Density estimation, stochastic processes, and prior information. *J. Roy. Statist. Soc. Ser. B* **40**, 113-146.

[11] Lijoi, A., Prunster, I., and Walker, S. (2005). On consistency of nonparametric normal mixtures for Bayesian density estimation. *J. Amer. Statist. Assoc.* **100**, 1292-1296.

[12] Lo, A.Y. (1984). On a class of Bayesian nonparametric estimates I: density estimates. *Ann. Statist.* **12**, 351-357.

[13] Petrone, S. and Wasserman, L. (2002). Consistency of Bernstein polynomial posteriors. *J. Roy. Statist. Soc. Ser. B* **64**, 79-100.

[14] Schwartz, L. (1965). On Bayes procedures. *Z. Wahr. verw. Geb.* **4**, 10-26.

[15] Walker, S. (2003). On sufficient conditions for Bayesian consistency. *Biometrika* **90**, 482-488.

[16] Walker, S. (2004). New approaches to Bayesian consistency. *Ann. Statist.* **32**, 2028-2043.

[17] Walker, S., Lijoi, A., and Prunster, I. (2005). Data tracking and the understanding of Bayesian consistency. *Biometrika* **92**, 765-788.

[18] Wasserman, L. (1998). Asymptotic properties of nonparametric Bayesian procedures. In *Practical Nonparametric and Semiparametric Bayesian Statistics*, eds. D. Dey, P. Müller, and D. Sinha, New York: Springer-Verlag, 293-304.

[19] Wasserman, L. (2000). Asymptotic inference for mixture models using data-dependent priors. *J. Roy. Statist. Soc. Ser. B* **62**, 159-180.