Convergence rates of

nonparametric posterior distributions

Yang Xing

Convergence rates of nonparametric posterior distributions

Yang Xing[1]

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

Abstract

We study the asymptotic behavior of posterior distributions. We present general posterior convergence rate theorems, which strengthen several results on posterior convergence rates provided by Ghosal and Van der Vaart (2000), Shen and Wasserman (2001) and Walker, Lijor and Prunster (2007). Our main tools are the Hausdorff α-entropy introduced by Xing and Ranneby (2008) and a new notion of prior concentration, which is a slight improvement of the usual prior concentration provided by Ghosal and Van der Vaart (2000). We apply our results to several statistical models and obtain some improvements on current results in these examples.

---

[1]E-mail address: yang.xing@sekon.slu.se

1   Introduction

Recently, a major theoretical advance has occurred in the theory of Bayesian consistency for infinite-dimensional models. Schwartz (1965) first proved that, if the true density function is in the Kullback-Leibler support of the prior distribution, then the sequence of posterior distributions accumulates in all weak neighborhoods of the true density function. It is known that the condition of positivity of prior mass on each Kullback-Leibler neighborhood in Schwartz's theorem is not a necessary condition. When one considers problems of density estimation, it is natural to ask for the strong consistency of Bayesian procedures. Sufficient conditions for the strong Hellinger consistency and for evaluating consistency rates have been currently developed by many authors. In this paper we study the problem of determining whether the posterior distributions accumulate in Hellinger neighborhoods of the true density function. The rate of this convergence can be measured by the size of the smallest shrinking Hellinger balls around the true density function on which posterior masses tend to zero as the sample size increases to infinity. By the fundamental works of Ghosal, Ghosh and Van der Vaart (2000) and Shen and Wasserman (2001), we know that the convergence rate of posterior distributions is completely determined by two quantities: the rate of the metric entropy and the prior concentration rate. Roughly speaking, the rate of the metric entropy describes how large the model is, and the prior concentration rate depends on prior masses near the true distribution. Since the true distribution is unknown, the later assumption actually requires that the prior distribution spreads its mass "uniformly" over the whole density space. Another elegant approach for determination of the convergence rate was provided by Walker (2004), who obtained a sufficient condition for strong consistency by using summability of square root of prior probability instead of the metric entropy method. In this paper, in dealing with the rate of metric entropies we shall apply the Hausdorff α-entropy introduced by Xing and Ranneby (2008), which is much smaller than widely used metric entropies and the bracketing entropy. For some important prior distributions of statistical models the Hausdorff α-entropies of all sieves are uniformly bounded, whereas it is generally impossible to get uniform boundedness of metric entropies of large sieves. The application of the Hausdorff α-entropy leads refinements of several theorems on posterior convergence rates, for instance, the well known assumptions on metric entropies and summability of square root of prior probability have been weakened, which particularly yields that Theorem 5 of Ghosal et al. (2007b) is strengthened into Corollary 3 of this paper. To handle the prior concentration rate, we shall apply a new notion of prior concentration. Our approach

is a slight improvement of the prior concentration provided by Ghosal et al. (2000), and moreover the proof of Lemma 1 on which the approach bases is quite simple. Finally, to get posterior convergence at the optimal rate $1/\sqrt{n}$, we give an extension of Ghosal et al. (2000, Theorem 2.4), in which the universal testing constant has been replaced by any fixed constant.

An outline of this paper is as follows. In Section 2 we define the Hausdorff α-entropy with respect to a given prior and then present general theorems for the determination of posterior convergence rates. We also give a new approach to compute concentration rates. In Section 3 we apply our results to Bernstein polynomial priors, priors based on uniform distribution, log spline models and finite-dimensional models, which leads some improvements on known results for these models. The proofs of the main results are contained in Section 4.

2   Notations and theorems

We consider a family of probability measures dominated by a σ-finite measure μ in $\mathbb{X}$, a Polish space endowed with a σ-algebra $\mathcal{X}$. 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 measure μ. Denote by F_0^∞ the infinite product distribution of the probability distribution F_0 associated with f_0. For two probability densities 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 $\mathbb{F}$. Given a prior distribution Π on $\mathbb{F}$, the posterior distribution Π_n is a random probability measure with the following expression

$$\Pi_n(A) = \Pi\big(A \,|\, X_1, X_2, \ldots, X_n\big) = \frac{\int_A \prod_{i=1}^n f(X_i) \, \Pi(df)}{\int_{\mathbb{F}} \prod_{i=1}^n f(X_i) \, \Pi(df)} = \frac{\int_A R_n(f) \, \Pi(df)}{\int_{\mathbb{F}} R_n(f) \, \Pi(df)}$$

2

for all measurable subsets $A \subset \mathbb{F}$, where $R_n(f) = \prod_{i=1}^{n} \{f(X_i)/f_0(X_i)\}$ is the likelihood ratio. In other words, the posterior distribution Π_n is the conditional distribution of Π given the observations $X_1, X_2, \ldots, X_n$. If the posterior distribution Π_n concentrates on arbitrarily small neighborhoods of the true density function f_0 almost surely or in probability, then it is said to be consistent at f_0 almost surely and in probability respectively. Throughout this paper, almost sure convergence and convergence in probability should be understood as to be with respect to the infinite product distribution F_0^∞ of F_0.

Our aim of this article is to present general theorems on posterior convergence rates at f_0. By the posterior convergence rate theorems of Ghosal et al. (2000), we know that the prior concentration rate and the rate of metric entropy both completely determine the convergence rate of posterior distributions. More specifically, a key inequality to determine almost sure convergence rates of posterior distributions is that for each $\varepsilon > 0$,

$$\int_\mathbb{F} R_n(f)\,\Pi(df) \geq e^{-3\,n\,\varepsilon^2}\,\Pi\big(\,f : H(f_0, f)^2\,||f_0/f||_\infty < \varepsilon^2\,\big)$$

almost surely for all sufficiently large n, where $||g||_\infty$ stands for the supremum norm of the function g on $\mathbb{X}$. This inequality was obtained by Ghosal et al. (2000, Lemma 8.4) under mild assumptions. It appears almost in all of papers handling strong convergence rates of posterior distributions. The reason is that in order to get the convergence rate of posterior distributions one needs to find a suitable lower bound for the denominator in the expression of posterior distributions. This is successfully done in Ghosal et al. (2000), who suggested that the prior Π puts sufficiently amount of mass around the true density function f_0 in the sense: $\Pi\big(\,f : H(f_0, f)^2\,||f_0/f||_\infty < \tilde{\varepsilon}_n^2\,\big) \geq e^{-n\,\tilde{\varepsilon}_n^2\,c}$ for some fixed constant c. Such a sequence $\{\tilde{\varepsilon}_n\}$ is referred to as the concentration rate of the prior Π around f_0. Here we give a slightly stronger result. We introduce a modification of the Hellinger distance

$$H_*(f_0, f) = \left(\int_\mathbb{X} \big(\sqrt{f_0(x)} - \sqrt{f(x)}\,\big)^2 \Big(\frac{2}{3} \sqrt{\frac{f_0(x)}{f(x)}} + \frac{1}{3} \Big)\,\mu(dx) \right)^{\frac{1}{2}}.$$

It is clear that the inequality $||f_0/f||_\infty \geq 1$ holds for all density functions f and f_0 such that the supremum is well-defined, and the quality holds if and only if $f = f_0$ almost surely. Observe also that $H_*(f_0, f) \neq H_*(f, f_0)$ and $3^{-1/2}\,H(f_0, f) \leq H_*(f_0, f)$. Moreover, we have

$$\begin{aligned} H_*(f_0, f) &\leq H(f_0, f)\,||\frac{2}{3}\,\sqrt{f_0/f} + \frac{1}{3}||_\infty^{1/2} \\ &\leq H(f_0, f)\,||f_0/f||_\infty^{1/4} \leq H(f_0, f)\,||f_0/f||_\infty^{1/2} \end{aligned}$$

which yields

$$\left\{f \in \mathbb{F} : H_*(f_0, f) \leq \tilde{\varepsilon}_n\right\} \quad \supset \quad \left\{f \in \mathbb{F} : H(f_0, f)^2 \sqrt{||f_0/f||_\infty} < \tilde{\varepsilon}_n^2 \right\}$$
$$\supset \quad \left\{f \in \mathbb{F} : H(f_0, f)^2 \, ||f_0/f||_\infty < \tilde{\varepsilon}_n^2 \right\}.$$

The following simple lemma shows that, for $\tilde{\varepsilon}_n$ to be a prior concentration rate, it is enough to assume $\Pi\big(W_{\tilde{\varepsilon}_n}\big) \geq e^{-n\,\tilde{\varepsilon}_n^2\,c_3}$, where we have used the notation

$$W_\varepsilon = \left\{f \in \mathbb{F} : H_*(f_0, f) \leq \varepsilon\right\}.$$

Lemma 1. *Let $\varepsilon > 0$ and $c > 0$. Then the inequality*

$$F_0^\infty \left(\, \int_{\mathbb{F}} R_n(f)\,\Pi(df) \leq e^{-n\,\varepsilon^2\,(3+2c)}\,\Pi\big(W_\varepsilon\big) \right) \leq e^{-n\,\varepsilon^2\,c}$$

holds for all n.

Lemma 1 provides a useful approach to compute prior concentration rates, particularly for models in which rate of convergence is governed by the prior concentration rate. It leads a simplification of the proof of Theorem 2.2 of Ghosal et al. (2000). Furthermore, we shall present general posterior convergence rate theorems in which the well known assumptions on metric entropies and summability of square root of prior probability are also weakened. We shall apply the Hausdorff α-entropy $J(\delta, \mathcal{G}, \alpha)$ introduced by Xing and Ranneby (2008). Denote by L_μ the space of all nonnegative integrable functions with the norm $||f||_1 = \int_{\mathbb{X}} f(x)\,\mu(dx)$. Write $\log 0 = -\infty$.

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

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

where the infimum is taken over all coverings $\{B_1, B_2, \ldots, B_N\}$ of $\mathcal{G}$, where N may take the value ∞, such that each B_j is contained in some Hellinger ball $\{f : H(f_j, f) < \delta\}$ of radius δ and center at $f_j \in L_\mu$.

Note that the infimum can be equivalently taken over all partitions $\{P_1, P_2, \ldots, P_N\}$ of $\mathcal{G}$ such that the Hellinger radius of each subset P_j does

not exceed δ. It was proved in Xing et al.(2008, Lemma 1) that the Hausdorff α-entropy $J(\delta, \mathcal{G}, \alpha)$ is an increasing subadditive function of $\mathcal{G}$ and satisfies $J(\delta, \mathcal{G}, \alpha) \leq \log N(\delta, \mathcal{G})$ for all $\alpha \geq 0$, where $N(\delta, \mathcal{G})$ stands for the minimal number of Hellinger balls of radius δ needed to cover $\mathcal{G}$. For $0 \leq \alpha \leq 1$ and each $\mathcal{G} \subset \mathbb{F}$, we also obtained the following useful inequality

$$\Pi(\mathcal{G})^\alpha \leq e^{J(\delta, \mathcal{G}, \alpha)} \leq \Pi(\mathcal{G})^\alpha \, N(\delta, \mathcal{G})^{1-\alpha}.$$

Our first result in this paper is the following general theorem on posterior strong convergence rates. Denote

$$A_\varepsilon = \big\{ f : \, H(f_0, f) \geq \varepsilon \big\}.$$

Theorem 1. *Let* $\{\bar{\varepsilon}_n\}_{n=1}^\infty$ *and* $\{\tilde{\varepsilon}_n\}_{n=1}^\infty$ *be positive sequences such that* $n \min(\bar{\varepsilon}_n^2, \tilde{\varepsilon}_n^2) \to \infty$ *as* $n \to \infty$. *Suppose that there exist constants* $c_1 > 0$, $c_2 > 0$, $c_3 \geq 0$, $0 \leq \alpha < 1$ *and a sequence* $\{\mathcal{G}_n\}_{n=1}^\infty$ *of subsets on* $\mathbb{F}$ *such that* $\sum_{n=1}^\infty e^{-n \tilde{\varepsilon}_n^2 c_2} < \infty$ *and*

(1) $\sum_{n=1}^\infty e^{J(\bar{\varepsilon}_n, \mathcal{G}_n, \alpha) - n \bar{\varepsilon}_n^2 c_1} < \infty$,

(2) $\sum_{n=1}^\infty e^{n \tilde{\varepsilon}_n^2 (3 + 3c_2 + c_3)} \, \Pi(A_{\tilde{\varepsilon}_n} \setminus \mathcal{G}_n) < \infty$,

(3) $\Pi\big(f : H_*(f_0, f) \leq \tilde{\varepsilon}_n \big) \geq e^{-n \tilde{\varepsilon}_n^2 c_3}$.

Then for $\varepsilon_n = \max(\bar{\varepsilon}_n, \tilde{\varepsilon}_n)$ *and each* $r > 2 + \sqrt{\frac{2(3\alpha + 2\alpha c_2 + \alpha c_3 + c_1)}{1-\alpha}}$, *we have* $\Pi_n\big(A_{r\,\varepsilon_n}\big) \to 0$ *almost surely as* $n \to \infty$.

As direct applications we have

Corollary 1. *Let* $c_1 \geq 0$, $c_2 > 0$, $c_3 \geq 0$ *and* $0 \leq \alpha < 1$. *Suppose that* $\{\varepsilon_n\}_{n=1}^\infty$ *is a positive sequence satisfying* $\sum_{n=1}^\infty e^{-n \varepsilon_n^2 c_2} < \infty$ *and suppose that there exists a sequence* $\{\mathcal{G}_n\}_{n=1}^\infty$ *of subsets on* $\mathbb{F}$ *such that*

(1) $J(\varepsilon_n, \mathcal{G}_n, \alpha) \leq n \varepsilon_n^2 c_1$,

(2) $\Pi(\mathbb{F} \setminus \mathcal{G}_n) \leq e^{-n \varepsilon_n^2 (3 + 3c_2 + c_3)}$,

(3) $\Pi\big(f : H_*(f_0, f) \leq \varepsilon_n \big) \geq e^{-n \varepsilon_n^2 c_3}$.

Then for each $r > 2 + \sqrt{\frac{2(3\alpha + 2\alpha c_2 + \alpha c_3 + c_1 + c_2)}{1-\alpha}}$, *we have that* $\Pi_n\big(A_{r\,\varepsilon_n}\big) \to 0$ *almost surely as* $n \to \infty$.

Proof. It is clear that all conditions of Theorem 1 are fulfilled if we let $\varepsilon_n = \bar{\varepsilon}_n = \tilde{\varepsilon}_n$ and replace the c_1 in Theorem 1 by $c_1 + c_2$ of Corollary 1, and the proof is complete. □

Corollary 1 extends Theorem 2.2 of Ghosal et al. (2000), in which they have stronger conditions than (1) and (3) of Corollary 1. It is probably worth mentioning that for several important prior distributions of infinite-dimensional statistical models, the quantities $J(\varepsilon_n, \mathcal{G}_n, \alpha)$ are uniformly bounded for all n and hence condition (1) of Corollary 1 is trivially fulfilled, whereas general metric entropies of the sieve $\mathcal{G}_n$ grow to infinity as the sample size increases. A slightly different version of Theorem 1 is the following consequence, which is in fact an extension of Proposition 1 of Walker et al. (2007).

Corollary 2. *Let $\{\varepsilon_n\}_{n=1}^{\infty}$ be a positive sequence such that $n\,\varepsilon_n^2 \to \infty$ as $n \to \infty$. Suppose that there exist constants $c_1 > 0$, $c_2 > 0$, $c_3 \geq 0$, $0 \leq \alpha < 1$ and a sequence $\{\cup_{j=1}^{\infty}\mathcal{G}_{nj}\}_{n=1}^{\infty}$ with $\mathcal{G}_{nj} \subset \mathbb{F}$ such that $\sum_{n=1}^{\infty} e^{-n\,\varepsilon_n^2\,c_2} < \infty$ and*

(1) $\displaystyle\sum_{n=1}^{\infty}\sum_{j=1}^{\infty} N(\varepsilon_n, \mathcal{G}_{nj})^{1-\alpha}\,\Pi(\mathcal{G}_{nj})^{\alpha}\,e^{-n\,\varepsilon_n^2\,c_1} < \infty;$

(2) $\displaystyle\sum_{n=1}^{\infty} e^{n\,\varepsilon_n^2\,(3+3c_2+c_3)}\,\Pi\big(A_{\varepsilon_n} \setminus \cup_{j=1}^{\infty}\mathcal{G}_{nj}\big) < \infty,$

(3) $\Pi\big(f : H_*(f_0, f) \leq \varepsilon_n\,\big) \geq e^{-n\,\varepsilon_n^2\,c_3}.$

Then for each $r > 2 + \sqrt{\dfrac{2(3\alpha + 2\alpha c_2 + \alpha c_3 + c_1)}{1-\alpha}}$, we have that $\Pi_n\big(A_{r\,\varepsilon_n}\big) \to 0$ almost surely as $n \to \infty$.

Proof. Let $\mathcal{G}_n = \cup_{j=1}^{\infty}\mathcal{G}_{nj}$. We only need to verify condition (1) of Theorem 1 for such a sieve $\mathcal{G}_n$. By Lemma 1 of Xing and Ranneby (2008) we have

$$\sum_{n=1}^{\infty} e^{J(\varepsilon_n, \mathcal{G}_n, \alpha) - n\,\varepsilon_n^2\,c_1} \leq \sum_{n=1}^{\infty}\sum_{j=1}^{\infty} e^{J(\varepsilon_n, \mathcal{G}_{nj}, \alpha) - n\,\varepsilon_n^2\,c_1}$$

$$\leq \sum_{n=1}^{\infty}\sum_{j=1}^{\infty} N(\varepsilon_n, \mathcal{G}_{nj})^{1-\alpha}\,\Pi(\mathcal{G}_{nj})^{\alpha}\,e^{-n\,\varepsilon_n^2\,c_1} < \infty.$$

Corollary 2 then follows from Theorem 1. □

The assertion of Theorem 1 is an almost sure statement that the posterior distributions outside a Hellinger ball with a multiple of ε_n as radius converge

to zero almost surely. Now we give an in-probability assertion under weaker conditions. Denote

$$V(f, g) = \int_{\mathbb{X}} f(x) \left(\log \frac{f(x)}{g(x)} \right)^2 \mu(dx).$$

Theorem 2. *Let $\{\bar{\varepsilon}_n\}_{n=1}^{\infty}$ and $\{\tilde{\varepsilon}_n\}_{n=1}^{\infty}$ be positive sequences such that $n \min(\bar{\varepsilon}_n^2, \tilde{\varepsilon}_n^2) \to \infty$ as $n \to \infty$. Suppose that there exist constants $c_1 > 0$, $c_2 \geq 0$, $0 \leq \alpha < 1$ and a sequence $\{\mathcal{G}_n\}_{n=1}^{\infty}$ of subsets on $\mathbb{F}$ such that*

(1) $J(\bar{\varepsilon}_n, \mathcal{G}_n, \alpha) - n \bar{\varepsilon}_n^2 c_1 \longrightarrow -\infty, \quad$ *as $n \to \infty$,*

(2) $e^{n \tilde{\varepsilon}_n^2 (2+c_2)} \Pi(A_{\tilde{\varepsilon}_n} \setminus \mathcal{G}_n) \longrightarrow 0, \quad$ *as $n \to \infty$,*

(3) $\Pi\big(f : K(f_0, f) < \tilde{\varepsilon}_n^2 \text{ and } V(f_0, f) < \tilde{\varepsilon}_n^2 \big) \geq e^{-n \tilde{\varepsilon}_n^2 c_2}.$

Then for $\varepsilon_n = \max(\bar{\varepsilon}_n, \tilde{\varepsilon}_n)$ and each $r > 2 + \sqrt{\frac{2(2\alpha + \alpha c_2 + c_1)}{1-\alpha}}$, we have that $\Pi_n\big(A_{r \varepsilon_n}\big) \to 0$ in probability as $n \to \infty$.

Observe that the conditions (2) of Theorem 1 and Theorem 2 are only used to ensure that $\Pi_n(A_{\varepsilon_n} \setminus \mathcal{G}_n) \to 0$ as $n \to \infty$. So one can replace the conditions (2) of these theorems by $\Pi_n(A_{\varepsilon_n} \setminus \mathcal{G}_n) \to 0$ as $n \to \infty$ almost surely and in probability respectively. Theorem 2 is an extended version of Theorem 2.1 of Ghosal et al.(2001) and Theorem 1 of Walker et all.(2007). Furthermore, as a consequence of Theorem 2 we obtain the following slight improvement of Theorem 5 of Ghosal et al.(2007b).

Corollary 3. *Let $\{\varepsilon_n\}_{n=1}^{\infty}$ be a positive sequence such that $n \varepsilon_n^2 \to \infty$ as $n \to \infty$. Suppose that there exist constants $c_1 > 0$, $c_2 \geq 0$, $0 \leq \alpha < 1$ and a sequence $\{\cup_{j=1}^{\infty} \mathcal{G}_{nj}\}_{n=1}^{\infty}$ with $\mathcal{G}_{nj} \subset \mathbb{F}$. If*

(1) $e^{-n\varepsilon_n^2 c_1} \sum_{j=1}^{\infty} N(\varepsilon_n, \mathcal{G}_{nj})^{1-\alpha} \Pi(\mathcal{G}_{nj})^{\alpha} \longrightarrow 0, \quad$ *as $n \to \infty$;*

(2) $e^{n \varepsilon_n^2 (2+c_2)} \Pi(A_{\varepsilon_n} \setminus \cup_{j=1}^{\infty} \mathcal{G}_{nj}) \longrightarrow 0, \quad$ *as $n \to \infty$;*

(3) $\Pi\big(f : K(f_0, f) < \varepsilon_n^2 \text{ and } V(f_0, f) < \varepsilon_n^2 \big) \geq e^{-n \varepsilon_n^2 c_2},$

then for each $r > 2 + \sqrt{\frac{2(2\alpha + \alpha c_2 + c_1)}{1-\alpha}}$, we have that $\Pi_n\big(A_{r \varepsilon_n}\big) \to 0$ in probability as $n \to \infty$.

7

Proof. For $\mathcal{G}_n = \cup_{j=1}^\infty \mathcal{G}_{nj}$, by Lemma 1 of Xing and Ranneby (2008) we get

$$e^{J(\varepsilon_n, \mathcal{G}_n, \alpha) - n\varepsilon_n^2 c_1} \leq \sum_{j=1}^\infty e^{J(\varepsilon_n, \mathcal{G}_{nj}, \alpha) - n\varepsilon_n^2 c_1} \leq e^{-n\varepsilon_n^2 c_1} \sum_{j=1}^\infty N(\varepsilon_n, \mathcal{G}_{nj})^{1-\alpha} \Pi(\mathcal{G}_{nj})^\alpha$$

which tends to zero as $n \to \infty$ and condition (1) of Theorem 2 holds. Then by Theorem 2 we conclude the proof. $\qquad\square$

The above theorems cannot yield a convergence rate $1/\sqrt{n}$ because of the assumption $n\varepsilon_n^2 \to \infty$. Particularly, these theorems cannot well serve finite-dimensional models. Ghosal et al. (2000, 2007a) have obtained a nice theorem to handle such models. Denote

$$B_{\varepsilon_n^2} = \big\{ f : K(f_0, f) < \varepsilon_n^2 \text{ and } V(f_0, f) < \varepsilon_n^2 \big\}.$$

Now our result is

Theorem 3. *Let $\{\varepsilon_n\}_{n=1}^\infty$ be a positive sequence such that $n\varepsilon_n^2$ are uniformly bounded away from zero, i.e., there exists a constant $c_0 > 0$ such that $n\varepsilon_n^2 \geq c_0$ for all n . Suppose that there exist constants $0 < \alpha < 1$, $c_1 < \frac{1-\alpha}{18}$ and a sequence $\{\mathcal{G}_n\}_{n=1}^\infty$ of subsets on $\mathbb{F}$ such that*

(1) $\dfrac{e^{2n\varepsilon_n^2} \Pi(A_{\varepsilon_n} \backslash \mathcal{G}_n)}{\Pi(B_{\varepsilon_n^2})} \longrightarrow 0, \quad \text{as } n \to \infty,$

(2) $\exp\!\Big(J\big(\tfrac{j\varepsilon_n}{3}, \{f \in \mathcal{G}_n : j\varepsilon_n \leq H(f_0, f) < 2j\varepsilon_n\}, \alpha\big)\Big) \leq e^{c_1 j^2 n\varepsilon_n^2} \Pi(B_{\varepsilon_n^2})^\alpha$
for all sufficiently large positive integers j and n.

Then for each $r_n \to \infty$ we have that $\Pi_n\big(A_{r_n \varepsilon_n}\big) \to 0$ in probability as $n \to \infty$.

Remark. Here we adopt the convention that if the denominator of a quotient equals zero then the numerator must also be zero. Hence, Theorem 3 is still true even when $\Pi(B_{\varepsilon_n^2}) = 0$ for some n.

Corollary 4. *Let $\{\varepsilon_n\}_{n=1}^\infty$ be a positive sequence such that $n\varepsilon_n^2$ are uniformly bounded away from zero. Suppose that there exist constants c_1, c_2 and a sequence $\{\mathcal{G}_n\}_{n=1}^\infty$ of subsets on $\mathbb{F}$ such that*

(1) $\log N\big(\tfrac{j\varepsilon_n}{3}, \{f \in \mathcal{G}_n : j\varepsilon_n \leq H(f_0, f) < 2j\varepsilon_n\}\big) \leq c_1 n\varepsilon_n^2$ *for all integers j and n large enough,*

(2) $\dfrac{e^{2n\varepsilon_n^2}\,\Pi(A_{\varepsilon_n}\backslash\mathcal{G}_n)}{\Pi(B_{\varepsilon_n^2})}\longrightarrow 0,\quad$ as $n\to\infty,$

(3) $\dfrac{\Pi\big(f\in\mathcal{G}_n:j\varepsilon_n\le H(f_0,f)<2j\varepsilon_n\big)}{\Pi(B_{\varepsilon_n^2})}\le e^{c_2 j^2 n\varepsilon_n^2}$ *for all large integers j and n.*

Then for each $r_n\to\infty$ we have that $\Pi_n\big(A_{r_n\,\varepsilon_n}\big)\to 0$ in probability as $n\to\infty$.

Corollary 4 is a slightly stronger version of Theorem 2.4 of Ghosal et al. (2000). A notable improvement in Corollary 4 is that we have no restriction on the constant c_2, whereas their constant c_2 equals half of some universal testing constant.

Proof of Corollary 4. We only need to check condition (2) of Theorem 3. It follows from Lemma 1 of Xing and Ranneby (2008) and conditions (1) and (3) that

$$
\begin{aligned}
J\big(&\frac{j\varepsilon_n}{3},\{f\in\mathcal{G}_n:j\varepsilon_n\le H(f_0,f)<2j\varepsilon_n\},\alpha\big)\\
&\le \alpha\log\Pi\big(f\in\mathcal{G}_n:j\varepsilon_n\le H(f_0,f)<2j\varepsilon_n\big)\\
&\quad +(1-\alpha)\log N\big(\frac{j\varepsilon_n}{3},\{f\in\mathcal{G}_n:j\varepsilon_n\le H(f_0,f)<2j\varepsilon_n\}\big)\\
&\le \alpha\log\big(e^{c_2 j^2 n\varepsilon_n^2}\Pi(B_{\varepsilon_n^2})\big)+(1-\alpha)c_1 n\varepsilon_n^2\\
&= \big(\alpha c_2+\frac{(1-\alpha)c_1}{j^2}\big)j^2 n\varepsilon_n^2+\alpha\log\Pi(B_{\varepsilon_n^2}).
\end{aligned}
$$

Taking a small α in $(0,1)$ and then letting j be large enough, we have that $\alpha c_2+\frac{(1-\alpha)c_1}{j^2}<\frac{1-\alpha}{18}$ and hence condition (2) of Theorem 3 is fulfilled. The proof of Corollary 4 is complete. $\qquad\square$

We will conclude this section by presenting an analogue of Theorem 3, which gives an almost sure assertion under stronger conditions.

Theorem 4. *Let $\{\varepsilon_n\}_{n=1}^{\infty}$ be a positive sequence such that there exists a constant $c_0>0$ such that $n\varepsilon_n^2\ge c_0\log n$ for all large n . Suppose that there exist constants $0<\alpha<1$, $c_1<\frac{1-\alpha}{18}$, $c_2>\frac{1}{c_0}$ and a sequence $\{\mathcal{G}_n\}_{n=1}^{\infty}$ of subsets on $\mathbb{F}$ such that*

(1) $\displaystyle\sum_{n=1}^{\infty}\frac{e^{n\,\varepsilon_n^2\,(3+2c_2)}\,\Pi(A_{\varepsilon_n}\backslash\mathcal{G}_n)}{\Pi(W_{\varepsilon_n})}<\infty,$

9

(2) $\exp\left(J\left(\frac{j\varepsilon_n}{3}, \{f \in \mathcal{G}_n : j\varepsilon_n \leq H(f_0, f) < 2j\varepsilon_n\}, \alpha\right)\right) \leq e^{c_1 j^2 n\varepsilon_n^2} \, \Pi(W_{\varepsilon_n})^\alpha$
    for all sufficiently large positive integers j and n.

Then for each r large enough we have that $\Pi_n\left(A_{r\,\varepsilon_n}\right) \to 0$ almost surely as $n \to \infty$.

Completely following the proof of Corollary 4, we have the following consequence of Theorem 4.

Corollary 5. *Let $\{\varepsilon_n\}_{n=1}^\infty$ be a positive sequence such that there exists a constant $c_0 > 0$ such that $n\,\varepsilon_n^2 \geq c_0 \log n$ for all large n . Suppose that there exist constants c_1, $c_2 > \frac{1}{c_0}$, c_3 and a sequence $\{\mathcal{G}_n\}_{n=1}^\infty$ of subsets on $\mathbb{F}$ such that*

(1) $\log N\left(\frac{j\varepsilon_n}{3}, \{f \in \mathcal{G}_n : j\varepsilon_n \leq H(f_0, f) < 2j\varepsilon_n\}\right) \leq c_1 n\varepsilon_n^2$ *for all integers j and n large enough,*

(2) $\sum_{n=1}^\infty \frac{e^{n\,\varepsilon_n^2\,(3+2c_2)} \, \Pi(A_{\varepsilon_n} \backslash \mathcal{G}_n)}{\Pi(W_{\varepsilon_n})} < \infty,$

(3) $\frac{\Pi\left(f \in \mathcal{G}_n : j\varepsilon_n \leq H(f_0, f) < 2j\varepsilon_n\right)}{\Pi(W_{\varepsilon_n})} \leq e^{c_3 j^2 n\varepsilon_n^2}$ *for all large integers j and n.*

Then for each r large enough we have that $\Pi_n\left(A_{r\,\varepsilon_n}\right) \to 0$ almost surely as $n \to \infty$.

3   Illustrations

In this section we apply our theorems to Bernstein polynomial priors, priors based on uniform distribution, log spline models and finite-dimensional models. This leads some improvements on known results for these models.

3.1   Bernstein polynomial prior

A Bernstein polynomial prior is a probability measure on the space of continuous probability distribution functions on $[0, 1]$. Petrone (1999) introduced the Bernstein polynomial prior Π by putting a prior distribution on the class of Bernstein densities in $[0, 1]$ in the following way:

$$b(x; k, F) = \sum_{j=1}^{k} \left(F(j/k) - F((j-1)/k)\right) \beta(x; j, k - j + 1),$$

where $\beta(x; a, b)$ stands for the beta density $\beta(x; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}$, k has a distribution $\rho(\cdot)$, F is a random distribution independent of ρ. In other words, if B_j stands for the set of all Bernstein densities of order j, then $\Pi(\cdot) = \sum_{j=1}^{\infty} \rho(j) \Pi_{B_j}(\cdot)$, where the probability measure Π_{B_j} is the normalized restriction of Π on B_j. We refer to Petrone and Wasserman (2002) for a detailed description of the Bernstein polynomial prior, in which consistency of the posterior distribution for the Bernstein polynomial prior is discussed. Rates of convergence have been established under suitable tail conditions on ρ by Ghosal (2001) and Walker et al. (2007), where the convergence is understood as convergence in F_0^{∞}-probability. Ghosal (2001, Theorem 2.3) proved that the prior concentration rate is $(\log n)^{1/3}/n^{1/3}$ and the entropy rate is $(\log n)^{5/6}/n^{1/3}$ under the tail assumption $\rho(j) \approx e^{-cj}$ for all j, which yields the convergence rate $(\log n)^{5/6}/n^{1/3}$. Walker et al. (2007) obtained the entropy rate $(\log n)^{1/3}/n^{1/3}$ under the lighter tail condition $\rho(j) \leq e^{-4j \log j} = (1/j^j)^4$ for all j, which leads the convergence rate $(\log n)^{1/3}/n^{1/3}$. In the following we establish the entropy rate $1/n^{\gamma}$ under the tail condition $\rho(j) \leq (1/j^j)^{c_0}$ for all j, where c_0 is any fixed positive constant and γ is any fixed constant strictly less than $1/2$. Hence we also get the convergence rate $(\log n)^{1/3}/n^{1/3}$ under the weaker condition $\rho(j) \leq (1/j^j)^{c_0}$. This improves the result of Walker et al. (2007).

From Ghosal (2001) it follows that there exists an absolute constant $c > 0$ such that $N(\varepsilon_n, B_j) \leq (c/\varepsilon_n)^j$ for all j. Given $\gamma < 1/2$, choose $\mathcal{G}_{nj} = B_j$ and $\varepsilon_n = 1/n^{\gamma}$. Take $0 < \alpha < 1$ such that $\gamma(2 + 1/d) < 1$ with $d = c_0 \alpha/(1 - \alpha)$. To verify condition (1) of Corollary 3, by $\rho(j) \leq (1/j^j)^{c_0}$ we obtain that

$$e^{-n\varepsilon_n^2 c_1} \sum_{j=1}^{\infty} N(\varepsilon_n, B_j)^{1-\alpha} \Pi(B_j)^{\alpha}$$

$$\leq e^{-n\varepsilon_n^2 c_1} \sum_{j=1}^{\infty} \left(\frac{c}{\varepsilon_n}\right)^{j(1-\alpha)} \rho(j)^{\alpha}$$

$$\leq e^{-n\varepsilon_n^2 c_1} \sum_{j=1}^{\infty} \left(\frac{c}{\varepsilon_n j^{\frac{c_0 \alpha}{1-\alpha}}}\right)^{j(1-\alpha)}$$

$$\leq e^{-c_1 n^{1-2\gamma}} \sum_{1 \leq j \leq (2cn^{\gamma})^{1/d}} \left(\frac{cn^{\gamma}}{j^d}\right)^{j(1-\alpha)} + \sum_{j \geq (2cn^{\gamma})^{1/d}} \left(\frac{1}{2}\right)^{j(1-\alpha)},$$

where the last sum on the right hand side tends to zero as $n \to \infty$. To estimate the first term, note that for $g(t) = (cn^{\gamma}/t^d)^t$ we have that $g'(t) = (cn^{\gamma}/t^d)^t \left(\log cn^{\gamma}/t^d - d\right) = 0$ is equivalent to $t = c^{1/d} v^{\gamma/d} e^{-1}$. This implies

11

$g(j) \leq e^{d_1 n^{\gamma/d}}$ for all j, where d_1 stands for the constant $dc^{1/d}e^{-1}$. Therefore, by the inequality $x \leq e^x$ for $x \geq 0$ we get

$$e^{-c_1 n^{1-2\gamma}} \sum_{1 \leq j \leq (2cn^\gamma)^{1/d}} \left(\frac{cn^\gamma}{j^d}\right)^{j(1-\alpha)}$$

$$\leq e^{-c_1 n^{1-2\gamma}} (2cn^\gamma)^{1/d} e^{d_1(1-\alpha)n^{\gamma/d}}$$

$$\leq e^{-c_1 n^{1-2\gamma}+(2^{1/d}c^{1/d}+d_1-d_1\alpha)n^{\gamma/d}} \longrightarrow 0, \quad \text{as} \quad j \to \infty,$$

since $1 - 2\gamma > \gamma/d$, and hence condition (1) of Corollary 3 holds. Condition (2) of Corollary 3 is trivially fulfilled and hence by Corollary 3 we obtain that the entropy rate is at least $1/n^\gamma$ for any given $\gamma < 1/2$.

3.2 Prior based on uniform distribution

Ghosal et al. (1997) established posterior consistency for prior distributions based on uniform distributions of finite subsets. Priors based on discrete uniform distributions were further studied in Ghosal et al. (2000), in which they used the bracketing entropy as a tool to compute the convergence rate of posterior distributions. As an application of Theorem 1, we now give a slight extension. Given $\varepsilon_n > 0$, assume that there exist density functions $f_1, \ldots, f_{N_n}$ such that all sets $\{f : H_*(f, f_j) \leq 3^{-1/2}\varepsilon_n\}$ form a covering of $\mathbb{F}$. Denote by μ_n the uniform discrete probability measure on the set $\{f_1, \ldots, f_{N_n}\}$. Define a prior distribution $\Pi = \sum_{j=1}^{\infty} a_j \mu_j$ for a given sequence a_j with $a_j > 0$ and $\sum_{j=1}^{\infty} a_j = 1$.

Theorem 5. If $\log N_n + \log n + \log \frac{1}{a_n} = O(n\varepsilon_n^2)$ as $n \to \infty$, then the posterior distributions Π_n converge almost surely at least at the rate ε_n, that is, $\Pi_n(f : H(f_0, f) \geq r\varepsilon_n) \longrightarrow 0$ as $n \to \infty$ almost surely for any given sufficiently large r.

Proof. From $\log n = O(n\varepsilon_n^2)$ it follows that $\sum_{n=1}^{\infty} e^{-n\varepsilon_n^2 c} < \infty$ for all large $c > 0$. For any f with $H_*(f, f_j) \leq 3^{-1/2}\varepsilon_n$ we have $H(f, f_j) \leq 3^{1/2}H_*(f, f_j) \leq \varepsilon_n$. Hence we obtain that $\{f : H_*(f, f_j) \leq 3^{-1/2}\varepsilon_n\} \subset \{f : H(f, f_j) < \varepsilon_n\}$ and then $J(\varepsilon_n, \mathbb{F}, 0) \leq \log N_n = O(n\varepsilon_n^2)$, which implies condition (1) of Theorem 1. Condition (2) is trivially fulfilled. Condition (3) follows from the fact that $\Pi(f : H_*(f_0, f) \leq \varepsilon_n) \geq \Pi(f : H_*(f_0, f) \leq 3^{-1/2}\varepsilon_n) \geq a_n/N_n \geq e^{-n\varepsilon_n^2 c_3}$ for some $c_3 > 0$, since $\{f : H_*(f_0, f) \leq 3^{-1/2}\varepsilon_n\}$ contains at least some density function of the set $\{f_1, \ldots, f_{N_n}\}$. The proof of Theorem 5 is complete. $\square$

It seems to be unusual to find a covering of the density space with covering sets of type $\{f : H_*(f, f_j) \leq c\,\varepsilon_n\}$. The most widely used norm for continuous functions should be the supremum norm. In fact, one can easily construct a new covering $\cup_{j=1}^{N_n}\{f : H_*(f, f_j) \leq c\,\varepsilon_n\}$ of $\mathbb{F}$ in terms of a given covering $\cup_{j=1}^{N_n}\{f : ||\sqrt{f} - \sqrt{g_j}||_\infty \leq \varepsilon_n\}$ with nonnegative bounded functions g_j (not necessarily density functions), as shown in the following: Take $f_j(x) = (\sqrt{g_j(x)} + \varepsilon_n)^2 / \int_{\mathbb{X}} (\sqrt{g_j(x)} + \varepsilon_n)^2\, \mu(dx)$, where we assume that μ is a probability measure on $\mathbb{X}$ and $\varepsilon_n \leq 1$ for all n. Then for each f with $||\sqrt{f} - \sqrt{g_j}||_\infty \leq \varepsilon_n$, that is, $\sqrt{g_j(x)} - \varepsilon_n \leq \sqrt{f(x)} \leq \sqrt{g_j(x)} + \varepsilon_n$ on $\mathbb{X}$, we have

$$
\frac{f(x)}{f_j(x)} = \frac{f(x)}{(\sqrt{g_j(x)} + \varepsilon_n)^2} \int_{\mathbb{X}} \left(\sqrt{g_j(x)} + \varepsilon_n\right)^2 \mu(dx) \leq \int_{\mathbb{X}} \left(\sqrt{f_j^*(x)} + 2\,\varepsilon_n\right)^2 \mu(dx)
$$

$$
= 1 + 4\,\varepsilon_n^2 + 4\,\varepsilon_n \int_{\mathbb{X}} \sqrt{f_j^*(x)}\, \mu(dx) \leq (1 + 2\,\varepsilon_n)^2,
$$

where f_j^* is some density function in $\{f : ||\sqrt{f} - \sqrt{g_j}||_\infty \leq \varepsilon_n\}$ and the last inequality follows from $||\sqrt{f_*}||_1 \leq ||\sqrt{f_*}||_2 = 1$. This implies that

$$
\begin{aligned}
H_*(f, f_j) &\leq \left(\frac{2}{3}(1 + 2\,\varepsilon_n) + \frac{1}{3}\right)^{\frac{1}{2}} H(f, f_j) \leq 2\,H(f, f_j) \\
&\leq 2\,H\big(f, (\sqrt{g_j} + \varepsilon_n)^2\big) + 2\,H\big((\sqrt{g_j} + \varepsilon_n)^2, f_j\big) \\
&\leq 4\,\varepsilon_n + 2\left(\left(\int_{\mathbb{X}} \left(\sqrt{g_j(x)} + \varepsilon_n\right)^2 \mu(dx)\right)^{\frac{1}{2}} - 1\right) \\
&\leq 4\,\varepsilon_n + 2\big((1 + 2\,\varepsilon_n) - 1\big) = 8\,\varepsilon_n.
\end{aligned}
$$

Therefore, we have

$$
\bigcup_{j=1}^{N_n} \{f : H_*(f, f_j) \leq 8\,\varepsilon_n\} \supset \bigcup_{j=1}^{N_n} \{f : ||f - g_j||_\infty \leq \varepsilon_n\} \supset \mathbb{F}.
$$

Observe that the numbers of covering subsets in both type coverings are equal.

For models with $H_*(f, g)$ controlled by a constant multiple of the Hellinger metric $H(f, g)$ such as the exponential family and a model with uniformly bounded supremum norm $||f/g||_\infty$ for all density functions f and g, it is not necessary to assume that the probability measures μ_n constructed above concentrate on a finite number of points. Here we give an extension of Theorem 5. Let $c_0 \geq 1$ and $\mathbb{F}_{c_0}$ be a subfamily of $\mathbb{F}$ such that $H_*(f, g) \leq c_0\, H(f, g)$ for all $f, g \in \mathbb{F}_{c_0}$. Given $\varepsilon_n > 0$, let $\{P_1, \ldots, P_{K_n}\}$ be a partition of $\mathbb{F}_{c_0}$ such that for each P_i there exists f_i in $\mathbb{F}_{c_0}$ with $P_i \subset \{f : H(f_i, f) \leq \varepsilon_n / 2\,c_0\}$. Take

any probability measure $\bar{\mu}_n$ on $\mathbb{F}_{c_0}$ with $\bar{\mu}_n(P_i) = 1/K_n$ for $i = 1, 2, \ldots, K_n$. Define then a prior distribution $\bar{\Pi} = \sum_{j=1}^{\infty} a_j \bar{\mu}_j$ for a given sequence a_j with $a_j > 0$ and $\sum_{j=1}^{\infty} a_j = 1$. Now we have

Theorem 6. *Let $f_0 \in \mathbb{F}_{c_0}$. If $\log K_n + \log n + \log \frac{1}{a_n} = O(n\,\varepsilon_n^2)$ as $n \to \infty$, then the posterior distributions $\bar{\Pi}_n$ converge at least at the rate ε_n almost surely.*

Proof. By the proof of Theorem 5, we only need to verify condition (3) of Theorem 1. Take $f_{i_0} \in \mathbb{F}_{c_0}$ such that $H(f_0, f_{i_0}) \leq \varepsilon_n/2\,c_0$. Then, for all $f \in \mathbb{F}_{c_0}$ with $H(f_{i_0}, f) \leq \varepsilon_n/2\,c_0$, we have that $H_*(f_0, f) \leq c_0\,H(f_0, f) \leq c_0\,H(f_0, f_{i_0}) + c_0\,H(f_{i_0}, f) \leq \varepsilon_n$. Hence we get that $\Pi(f : H_*(f_0, f) \leq \varepsilon_n) \geq \Pi(P_{i_0}) \geq a_n/K_n \geq e^{-n\varepsilon_n^2 c_3}$ for some $c_3 > 0$, and the proof of Theorem 6 is complete. $\square$

Observe that, given a covering $\{O_1, O_2, \ldots, O_{K_n}\}$ of $\mathbb{F}_{c_0}$, one can easily construct a partition $\{P_1, P_2, \ldots, P_{K_n}\}$ of $\mathbb{F}_{c_0}$ in the following way: $P_1 = O_1 \cap \mathbb{F}_{c_0}$ and $P_i = (O_i - \cup_{l=1}^{i-1} P_l) \cap \mathbb{F}_{c_0}$ for $i = 2, 3, \ldots, K_n$.

Example (Exponential families). We consider the exponential family of all density functions of the form $e^{h(x)}$, where the function $h(x)$ belongs to a fixed bounded subset in the Sobolev space $C^p[0, 1]$ with $p > 0$. A subclass of this family has been recently studied by Scricciolo (2006). Following a result of Kolmogorov and Tihomirov (1959), we know that the ε-entropy of this family with respect to the norm $\| \cdot \|_\infty$ equals $O(\varepsilon^{-1/p})$. Thus, using the above argument we get that $\log K_n = O(n\,\varepsilon_n^2)$ for $\varepsilon_n = n^{-p/(2p+1)}$ and hence by Theorem 6 the posterior distributions constructed above converge at the rate $\varepsilon_n = n^{-p/(2p+1)}$, which is known to be the optimal rate of convergence in the minimax sense under the Hellinger loss.

3.3 Log spline models

Log spline models for density estimation have been studied, among others, by Stone (1990) and Ghosal et al. (2000). Let $\big[(k-1)/K_n, k/K_n\big)$ with $k = 1, 2, \ldots, K_n$ be a partition of the interval $[0, 1)$. The space of splines of order q relative to this partition is the set of all functions $f : [0, 1) \to \mathbb{R}$ such that f is $q - 2$ times continuously differentiable on $[0, 1)$ and the restriction of f on each $\big[(k-1)/K_n, k/K_n\big)$ is a polynomial of degree strictly less then q.

Let $J_n = q + K_n - 1$. This space of splines is a J_n-dimensional vector space with a B-spline basis $B_1(x), B_2(x), \ldots, B_{J_n}(x)$, see Ghosal et al. (2000) for the details of such a basis. Let $\mathbb{F}$ be the set of all density functions in $C^\alpha[0,1]$. Assume that the true density function $f_0(x)$ is bounded away from zero and infinity. We consider the J_n-dimensional exponential subfamily of $C^\alpha[0,1]$ of the form

$$f_\theta(x) = \exp\Big(\sum_{j=1}^{J_n} \theta_j B_j(x) - c(\theta) \Big),$$

where $\theta = (\theta_1, \theta_2, \ldots, \theta_{J_n}) \in \Theta_0 = \{(\theta_1, \theta_2, \ldots, \theta_{J_n}) \in \mathbb{R}^{J_n} : \sum_{j=1}^{J_n} \theta_j = 0\}$ and the constant $c(\theta)$ is chosen such that $f_\theta(x)$ is a density function in $[0,1]$. Each prior on Θ_0 induces naturally a prior on $\mathbb{F}$. Let $||\theta||_\infty = \max_j |\theta_j|$ be the infinity norm on Θ_0. Assume that $a_1 n^{1/(2\alpha+1)} \leq K_n \leq a_2 n^{1/(2\alpha+1)}$ for two fixed positive constants a_1 and a_2. Assume that the prior Π for Θ_0 is supported on $[-M, M]^{J_n}$ for some $M \geq 1$ and has a density function with respect to the Lebesgue measure on Θ_0, which is bounded below by $a_3^{J_n}$ and above by $a_4^{J_n}$. Take a constant $d > 0$ such that $d\,||\theta||_\infty \leq ||\log f_\theta(x)||_\infty$ for all $\theta \in \Theta_0$. Ghosal et al. (2000, Theorem 4.5) proved that, if $f_0 \in C^\alpha[0,1]$ with $q \geq \alpha \geq 1/2$ and $||\log f_0(x)||_\infty \leq d\,M/2$, the posteriors Π_n converge in probability at the rate $n^{-\alpha/(2\alpha+1)}$. Using Corollary 5 we now get that under the same assumptions as in Ghosal et al. (2000, Theorem 4.5), the posteriors Π_n are in fact convergent almost surely at the rate $\varepsilon_n = n^{-\alpha/(2\alpha+1)}$. To see this, take $\mathcal{G}_n = \mathbb{F}$. Clearly, $n\varepsilon_n^2 \geq \log n$ for all large n. Condition (1) of Corollary 5 has been verified by Ghosal et al. (2000) and condition (2) is trivially fulfilled. Condition (3) follows also from the proof of Theorem 4.5 in Ghosal et al. (2000), since the inequality $H_*(f_0, f_\theta) \leq H(f_0, f_\theta)\,\big||f_0/f_\theta\big|\big|_\infty^{1/2}$ holds for all $\theta \in \Theta_0$.

3.4 Finite-dimensional models

Let $\beta > 0$ and let Θ be a bounded subset in $\mathbb{R}^d$ with the Euclidean norm $||\cdot||$. Denote by $\mathbb{F}$ the family of all density functions f_θ with the parameter θ in Θ satisfying

$$a_1\,||\theta_1 - \theta_2||^\beta \leq H(f_{\theta_1}, f_{\theta_2}) \leq \sqrt{3}\,H_*(f_{\theta_1}, f_{\theta_2}) \leq a_2\,||\theta_1 - \theta_2||^\beta$$

for all $\theta_1, \theta_2 \in \Theta$, where a_1 and a_2 are two fixed positive constants. Assume that the true value θ_0 is in Θ and that the density function of the prior distribution Π with respect to the Lebesgue measure on Θ is uniformly bounded away from zero and infinity. Under slightly weaker conditions, Ghosal et al.

15

(2000) proved that the posterior distributions Π_n converge in probability at the rate $1/\sqrt{n}$. Now we give an almost sure assertion for this model.

Theorem 7. *Under the above assumptions, the posterior distributions Π_n converge almost surely at least at the rate $\sqrt{\log n}/\sqrt{n}$.*

Proof. We shall apply Corollary 5 for $\mathcal{G}_n = \mathbb{F}$. Clearly, $n\varepsilon_n^2 = \log n$ for $\varepsilon_n = \sqrt{\log n}/\sqrt{n}$. Condition (1) has been verified in the proof of Theorem 5.1 of Ghosal et al.(2000). Condition (2) is trivially fulfilled. Using $H_*(f_{\theta_0}, f_\theta) \leq 3^{-1/2} a_2 \|\theta_0 - \theta\|^\beta$, we have that $\Pi(W_{\varepsilon_n}) \geq \Pi\big(\theta : \|\theta - \theta_0\| \leq (\sqrt{3}\,\varepsilon_n/a_2)^{1/\beta}\big)$. Hence, the verification of condition (3) follows from the same lines as the proof of Ghosal et al. (2000, Theorem 5) and then by Corollary 5 we conclude the proof of Theorem 7. $\qquad\square$

4 Lemmas and Proofs

In this section we give proofs of our lemmas and theorems. For simplicity of notations, we assume throughout this section that $\varepsilon_n = \bar{\varepsilon}_n = \tilde{\varepsilon}_n$.

Proof of Lemma 1. It is no restriction to assume that $\Pi\big(W_\varepsilon\big) > 0$. Using Jensen's inequality for the convex function $x^{-1/2}$ for $x > 0$ and Chebyshev's inequality, we obtain that

$$
\begin{aligned}
F_0^\infty\Big(\int_{\mathbb{F}} & R_n(f)\,\Pi(df) \leq e^{-n\,\varepsilon^2\,(3+2c)}\,\Pi\big(W_\varepsilon\big) \Big) \\
&\leq F_0^\infty\Big(\int_{W_\varepsilon} R_n(f)\,\Pi(df) \leq e^{-n\,\varepsilon^2\,(3+2c)}\,\Pi\big(W_\varepsilon\big) \Big) \\
&= F_0^\infty\Big(e^{n\,\varepsilon^2\,(\frac{3}{2}+c)} \leq \Big(\frac{1}{\Pi(W_\varepsilon)} \int_{W_\varepsilon} R_n(f)\,\Pi(df) \Big)^{-\frac{1}{2}} \Big) \\
&\leq F_0^\infty\Big(e^{n\,\varepsilon^2\,(\frac{3}{2}+c)} \leq \frac{1}{\Pi(W_\varepsilon)} \int_{W_\varepsilon} R_n(f)^{-\frac{1}{2}}\,\Pi(df) \Big) \\
&\leq e^{-n\,\varepsilon^2\,(\frac{3}{2}+c)} \frac{1}{\Pi(W_\varepsilon)}\, E \int_{W_\varepsilon} R_n(f)^{-\frac{1}{2}}\,\Pi(df) \\
&= e^{-n\,\varepsilon^2\,(\frac{3}{2}+c)} \frac{1}{\Pi(W_\varepsilon)} \int_{W_\varepsilon} \Big(E\sqrt{\frac{f_0(X_1)}{f(X_1)}} \Big)^n \Pi(df).
\end{aligned}
$$

16

On the other hand, we have

$$E\sqrt{\frac{f_0(X_1)}{f(X_1)}} = 1 + E\frac{\sqrt{f_0(X_1)} - \sqrt{f(X_1)}}{\sqrt{f(X_1)}}$$

$$= 1 + \int_{\mathbb{X}} \frac{\sqrt{f_0(x)} - \sqrt{f(x)}}{\sqrt{f(x)}} \left(\sqrt{f_0(x)} - \sqrt{f(x)} + \sqrt{f(x)} \right) \sqrt{f_0(x)}\, \mu(dx)$$

$$= 1 + \int_{\mathbb{X}} \left(\sqrt{f_0(x)} - \sqrt{f(x)} \right)^2 \frac{\sqrt{f_0(x)}}{\sqrt{f(x)}}\, \mu(dx)$$

$$\quad + \int_{\mathbb{X}} \left(f_0(x) - \sqrt{f(x)\, f_0(x)} \right) \mu(dx)$$

$$= 1 + \int_{\mathbb{X}} \left(\sqrt{f_0(x)} - \sqrt{f(x)} \right)^2 \frac{\sqrt{f_0(x)}}{\sqrt{f(x)}}\, \mu(dx)$$

$$\quad + \frac{1}{2} \int_{\mathbb{X}} \left(\sqrt{f_0(x)} - \sqrt{f(x)} \right)^2 \mu(dx)$$

$$= 1 + \frac{3}{2} H_*(f_0, f)^2 \le e^{\frac{3}{2} H_*(f_0,f)^2} \le e^{\frac{3}{2}\varepsilon^2},$$

where the last inequality holds when $f \in W_\varepsilon$. Hence we get

$$F_0^\infty \left(\int_{\mathbb{F}} R_n(f)\, \Pi(df) \le e^{-n\varepsilon^2\,(3+2c)}\, \Pi\!\left(W_\varepsilon\right) \right)$$

$$\le e^{-n\varepsilon^2\,(\frac{3}{2}+c)} \frac{1}{\Pi(W_\varepsilon)} \int_{W_\varepsilon} e^{\frac{3}{2}n\varepsilon^2}\, \Pi(df) = e^{-n\varepsilon^2\,c}.$$

The proof of Lemma 1 is complete. □

In the proof of Theorem 1 we use the following Lemma, which is similar to Lemma 5 of Barron et al. (1999).

Lemma 2. *Let $c_2 > 0$ and $c_3 \ge 0$. Let $\{\varepsilon_n\}_{n=1}^\infty$ be a positive sequence such that $\Pi(W_{\varepsilon_n}) \ge e^{-n\varepsilon_n^2 c_3}$ for all n and $\sum_{n=1}^\infty e^{-n\varepsilon_n^2 c_2} < \infty$. If a sequence $\{D_n\}_{n=1}^\infty$ of subsets in $\mathbb{F}$ satisfies $\sum_{n=1}^\infty e^{n\varepsilon_n^2 (3+3c_2+c_3)} \Pi(D_n) < \infty$, then $\Pi_n(D_n) \to 0$ almost surely as $n \to \infty$.*

17

Proof. From Chebyshev's inequality and Fubini's theorem it turns out that

$$F_0^\infty \Big\{ \int_{D_n} R_n(f)\,\Pi(df) \geq e^{-n\,\varepsilon_n^2\,(3+3c_2+c_3)} \Big\}$$

$$\leq e^{n\,\varepsilon_n^2\,(3+3c_2+c_3)}\, E \int_{D_n} R_n(f)\,\Pi(df)$$

$$= e^{n\,\varepsilon_n^2\,(3+3c_2+c_3)} \int_{D_n} E\, R_n(f)\,\Pi(df)$$

$$= e^{n\,\varepsilon_n^2\,(3+3c_2+c_3)}\, \Pi(D_n)$$

for all n. Hence by the first Borel-Cantelli Lemma we get that

$$\int_{D_n} R_n(f)\,\Pi(df) \leq e^{-n\,\varepsilon_n^2\,(3+3c_2+c_3)}$$

almost surely for all n large enough. On the other hand, Lemma 1 and the first Borel-Cantelli Lemma yield that

$$\int_{\mathbb{F}} R_n(f)\,\Pi(df) \geq \Pi(W_{\varepsilon_n})\,e^{-n\varepsilon_n^2(3+2c_2)} \geq e^{-n\,\varepsilon_n^2\,(3+2c_2+c_3)}$$

almost surely for all n. Therefore, we obtain that with probability one,

$$\Pi_n(D_n) = \frac{\int_{D_n} R_n(f)\,\Pi(df)}{\int_{\mathbb{F}} R_n(f)\,\Pi(df)} \leq e^{-n\,\varepsilon_n^2\,c_2},$$

which tends to zero as $n \to \infty$ and the proof of Lemma 2 is complete. $\square$

Proof of Theorem 1. It is clear that if condition (1) holds for some $\alpha = \alpha_0$ then it also holds for any $\alpha \geq \alpha_0$. So we may assume that $0 < \alpha < 1$. Given $r > 2 + \sqrt{\frac{2(3\alpha+2\alpha c_2+\alpha c_3+c_1)}{1-\alpha}}$, we have

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

It then follows from Lemma 2 that $\Pi_n\big(A_{\varepsilon_n} \setminus \mathcal{G}_n\big) \to 0$ almost surely as $n \to \infty$. So it suffices to prove that $\Pi_n\big(\mathcal{G}_n \cap A_{r\varepsilon_n}\big) \to 0$ almost surely as $n \to \infty$. By the definition of $J(\varepsilon_n, \mathcal{G}_n, \alpha)$, for each fixed n there exist functions $f_1, f_2, \ldots, f_N$ in L_μ such that $\mathcal{G}_n \cap A_{r\varepsilon_n} \subset \bigcup_{j=1}^N B_j$, where $B_j = \mathcal{G}_n \cap A_{r\varepsilon_n} \cap \{f : H(f_j, f) < \varepsilon_n\}$ and $\sum_{j=1}^N \Pi(B_j)^\alpha \leq 2\,e^{J(\varepsilon_n, \mathcal{G}_n, \alpha)}$. It is no restriction to assume that all the sets B_j are disjoint and nonempty. Taking a $f_j^\star \in B_j$ we get that $H(f_j, f_0) \geq H(f_j^\star, f_0) - H(f_j^\star, f_j) \geq (r-1)\,\varepsilon_n$. Now for each B_j we have

$$\int_{B_j} R_n(f)\,\Pi(df) = \Pi(B_j) \prod_{k=0}^{n-1} \frac{\int_{B_j} R_{k+1}(f)\,\Pi(df)}{\int_{B_j} R_k(f)\,\Pi(df)} = \Pi(B_j) \prod_{k=0}^{n-1} \frac{f_{kB_j}(X_{k+1})}{f_0(X_{k+1})},$$

18

where $f_{kB_j}(x) = \int_{B_j} f(x)\, R_k(f)\, \Pi(df) \big/ \int_{B_j} R_k(f)\, \Pi(df)$ and $R_0(f) = 1$. The function f_{kB_j} was introduced by Walker (2004) and can be considered as the predictive density of f with a normalized posterior distribution, restricted on the set B_j. Clearly, Jensen's inequality yields that $H(f_{kB_j}, f_j)^2 \leq \varepsilon_n^2$ for each k. Hence $H(f_{kB_j}, f_0) \geq H(f_j, f_0) - H(f_j, f_{kB_j}) \geq (r-2)\,\varepsilon_n > 0$. Since $\sum_{n=1}^{\infty} e^{-n\,\varepsilon_n^2\, c_2} < \infty$, it turns out from Lemma 1 and the first Borel-Cantelli Lemma that $\int_{\mathbb{F}} R_n(f)\, \Pi(df) \geq e^{-n\,\varepsilon_n^2\,(3+2c_2)}\, \Pi\big(W_{\varepsilon_n}\big)$ almost surely for all n large enough. Hence, by condition (3) we obtain that

$$\Pi_n\big(\mathcal{G}_n \cap A_{r\varepsilon_n}\big) \leq \big(\Pi_n(\mathcal{G}_n \cap A_{r\varepsilon_n})\big)^{\alpha} \leq \big(\sum_{j=1}^{N} \Pi_n(B_j) \big)^{\alpha} \leq \sum_{j=1}^{N} \Pi_n(B_j)^{\alpha}$$

$$= \frac{\sum_{j=1}^{N} \Pi(B_j)^{\alpha} \prod_{k=0}^{n-1} \frac{f_{kB_j}(X_{k+1})^{\alpha}}{f_0(X_{k+1})^{\alpha}}}{\left(\int_{\mathbb{F}} R_n(f)\, \Pi(df) \right)^{\alpha}} \leq \frac{\sum_{j=1}^{N} \Pi(B_j)^{\alpha} \prod_{k=0}^{n-1} \frac{f_{kB_j}(X_{k+1})^{\alpha}}{f_0(X_{k+1})^{\alpha}}}{\left(\Pi(W_{\varepsilon_n})\, e^{-n\varepsilon_n^2(3+2c_2)} \right)^{\alpha}}$$

$$\leq e^{n\,\varepsilon_n^2\,(3+2c_2+c_3)\,\alpha} \sum_{j=1}^{N} \Pi(B_j)^{\alpha} \prod_{k=0}^{n-1} \frac{f_{kB_j}(X_{k+1})^{\alpha}}{f_0(X_{k+1})^{\alpha}}$$

almost surely for all n large enough. Since $r > 2 + \sqrt{\frac{2(3\alpha + 2\alpha c_2 + \alpha c_3 + c_1)}{1-\alpha}}$, the inequality $(3 + 2c_2 + c_3)\,\alpha < \frac{1}{2}(r-2)^2\,(1-\alpha) - c_1$ holds. Take a constant b with $(3 + 2c_2 + c_3)\,\alpha < b < \frac{1}{2}(r-2)^2\,(1-\alpha) - c_1$. Denote $\mathcal{F}_k = \sigma\{X_1, X_2, \ldots, X_k\}$. Then we have

$$F_0^{\infty}\Big\{ \sum_{j=1}^{N} \Pi(B_j)^{\alpha} \prod_{k=0}^{n-1} \frac{f_{kB_j}(X_{k+1})^{\alpha}}{f_0(X_{k+1})^{\alpha}} \geq e^{-n\,\varepsilon_n^2\, b} \Big\}$$

$$\leq e^{n\,\varepsilon_n^2\, b}\, E\left(\sum_{j=1}^{N} \Pi(B_j)^{\alpha} \prod_{k=0}^{n-1} \frac{f_{kB_j}(X_{k+1})^{\alpha}}{f_0(X_{k+1})^{\alpha}} \right)$$

$$= e^{n\,\varepsilon_n^2\, b} \sum_{j=1}^{N} \Pi(B_j)^{\alpha}\, E\left(\prod_{k=0}^{n-1} \frac{f_{kB_j}(X_{k+1})^{\alpha}}{f_0(X_{k+1})^{\alpha}} \right)$$

$$\leq 2\, e^{J(\varepsilon_n, \mathcal{G}_n, \alpha) + n\,\varepsilon_n^2\, b} \max_{1 \leq j \leq N} E\left(\prod_{k=0}^{n-1} \frac{f_{kB_j}(X_{k+1})^{\alpha}}{f_0(X_{k+1})^{\alpha}} \right),$$

where

$$E\left(\prod_{k=0}^{n-1} \frac{f_{kB_j}(X_{k+1})^{\alpha}}{f_0(X_{k+1})^{\alpha}} \right) = E\left(E\Big(\prod_{k=0}^{n-1} \frac{f_{kB_j}(X_{k+1})^{\alpha}}{f_0(X_{k+1})^{\alpha}} \,\Big|\, \mathcal{F}_{n-1} \Big) \right)$$

19

$$= E\left(\prod_{k=0}^{n-2} \frac{f_{kB_j}(X_{k+1})^\alpha}{f_0(X_{k+1})^\alpha} \; E\left(\frac{f_{n-1B_j}(X_n)^\alpha}{f_0(X_n)^\alpha} \;\middle|\; \mathcal{F}_{n-1} \right) \right).$$

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

$$E\left(\frac{f_{n-1B_j}(X_n)^\alpha}{f_0(X_n)^\alpha} \;\middle|\; \mathcal{F}_{n-1} \right) = E\left(\frac{f_{n-1B_j}(X_n)^{\frac{\alpha}{2}}}{f_0(X_n)^{\frac{\alpha}{2}}} \frac{f_{n-1B_j}(X_n)^{\frac{\alpha}{2}}}{f_0(X_n)^{\frac{\alpha}{2}}} \;\middle|\; \mathcal{F}_{n-1} \right)$$

$$\le E\left(\frac{f_{n-1B_j}(X_n)^{\frac{\alpha}{2}\cdot\frac{2}{2-\alpha}}}{f_0(X_n)^{\frac{\alpha}{2}\cdot\frac{2}{2-\alpha}}} \;\middle|\; \mathcal{F}_{n-1} \right)^{\frac{2-\alpha}{2}} E\left(\frac{f_{n-1B_j}(X_n)^{\frac{\alpha}{2}\cdot\frac{2}{\alpha}}}{f_0(X_n)^{\frac{\alpha}{2}\cdot\frac{2}{\alpha}}} \;\middle|\; \mathcal{F}_{n-1} \right)^{\frac{\alpha}{2}}$$

$$= E\left(\frac{f_{n-1B_j}(X_n)^{\frac{\alpha}{2-\alpha}}}{f_0(X_n)^{\frac{\alpha}{2-\alpha}}} \;\middle|\; \mathcal{F}_{n-1} \right)^{\frac{2-\alpha}{2}}.$$

Take an integer m with $\frac{\alpha}{1-\alpha} \le 2^m < \frac{2\alpha}{1-\alpha}$. Repeating the above procedure $m-1$ more times we obtain that with probability one,

$$E\left(\frac{f_{n-1B_j}(X_n)^\alpha}{f_0(X_n)^\alpha} \;\middle|\; \mathcal{F}_{n-1} \right) \le E\left(\frac{f_{n-1B_j}(X_n)^{\frac{\alpha}{2^m(1-\alpha)+\alpha}}}{f_0(X_n)^{\frac{\alpha}{2^m(1-\alpha)+\alpha}}} \;\middle|\; \mathcal{F}_{n-1} \right)^{\frac{2^m(1-\alpha)+\alpha}{2^m}},$$

which by the conditional Hölder's inequality is less than

$$E\left(\frac{f_{n-1B_j}(X_n)^{\frac{1}{2}}}{f_0(X_n)^{\frac{1}{2}}} \;\middle|\; \mathcal{F}_{n-1} \right)^{\frac{\alpha}{2^{m-1}}}$$

$$= \left(\int \sqrt{f_{n-1B_j}(X_n)\, f_0(X_n)}\, \mu(dX_n) \right)^{\frac{\alpha}{2^{m-1}}}$$

$$= \left(1 - \frac{H(f_{n-1B_j}, f_0)^2}{2} \right)^{\frac{\alpha}{2^{m-1}}} \le \left(1 - \frac{(r-2)^2\, \varepsilon_n^2}{2} \right)^{\frac{\alpha}{2^{m-1}}}$$

$$\le e^{-2^{-m}\,(r-2)^2\,\alpha\,\varepsilon_n^2} \le e^{\frac{1}{2}\,(r-2)^2\,(\alpha-1)\,\varepsilon_n^2}.$$

Hence, with probability one, we have

$$E\left(\prod_{k=0}^{n-1} \frac{f_{kB_j}(X_{k+1})^\alpha}{f_0(X_{k+1})^\alpha} \right) \le e^{\frac{1}{2}\,(r-2)^2\,(\alpha-1)\,\varepsilon_n^2}\, E\left(\prod_{k=0}^{n-2} \frac{f_{kB_j}(X_{k+1})^\alpha}{f_0(X_{k+1})^\alpha} \right).$$

Repeating the same argument $n-1$ times, we obtain that for each j,

$$E\left(\prod_{k=0}^{n-1} \frac{f_{kB_j}(X_{k+1})^\alpha}{f_0(X_{k+1})^\alpha} \right) \le e^{\frac{1}{2}\,(r-2)^2\,(\alpha-1)\,n\,\varepsilon_n^2}.$$

Therefore, we have gotten that for all n,

$$F_0^\infty \Big\{ \sum_{j=1}^N \Pi(B_j)^\alpha \prod_{k=0}^{n-1} \frac{f_{kB_j}(X_{k+1})^\alpha}{f_0(X_{k+1})^\alpha} \geq e^{-n\,\varepsilon_n^2\,b} \Big\}$$

$$\leq 2\, e^{J(\varepsilon_n, \mathcal{G}_n, \alpha) + n\,\varepsilon_n^2 \left(b + \frac{1}{2}\,(r-2)^2\,(\alpha-1) \right)}$$

$$\leq 2\, e^{J(\varepsilon_n, \mathcal{G}_n, \alpha) - n\,\varepsilon_n^2\,c_1}.$$

Thus, together with condition (1), the first Borel-Cantelli Lemma yields that

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

almost surely for all n large enough. Hence we have

$$\Pi_n\big(\mathcal{G}_n \cap A_{r\varepsilon_n}\big) \leq 2\, e^{n\,\varepsilon_n^2\,(3\,\alpha + 2\,\alpha\,c_2 + \alpha\,c_3 - b)},$$

which tends to zero as $n \to \infty$, since $(3 + 2c_2 + c_3)\,\alpha < b$ and $n\,\varepsilon_n^2 \to \infty$ as $n \to \infty$. The proof of Theorem 1 is complete. $\square$

To prove Theorem 2, we need a replacement of Lemma 2 under weaker conditions.

Lemma 3. *Let $c_2 \geq 0$ and let $\{\varepsilon_n\}_{n=1}^\infty$ be a positive sequence such that $\Pi(B_{\varepsilon_n^2}) \geq e^{-n\,\varepsilon_n^2\,c_2}$ for all n. If a sequence $\{D_n\}_{n=1}^\infty$ of subsets in $\mathbb{F}$ satisfies $e^{n\,\varepsilon_n^2\,(2+c_2)}\,\Pi(D_n) \to 0$ as $n \to \infty$, then $\Pi_n(D_n) \to 0$ in probability as $n \to \infty$.*

Proof. From Lemma 1 of Shen et al. (2001) or Lemma 8.1 of Ghosal et al. (2000) it turns out that we have, with probability tending to 1,

$$\Pi_n(D_n) \leq \frac{\int_{D_n} R_n(f)\,\Pi(df)}{\Pi(B_{\varepsilon_n^2})\,e^{-2n\varepsilon_n^2}} \leq e^{n\,\varepsilon_n^2\,(2+c_2)} \int_{D_n} R_n(f)\,\Pi(df).$$

Hence for any given $\delta > 0$ we have that

$$F_0^\infty\{\Pi_n(D_n) \geq \delta\} \;\leq\; F_0^\infty\Big\{ e^{n\,\varepsilon_n^2\,(2+c_2)} \int_{D_n} R_n(f)\,\Pi(df) \geq \delta \Big\} + o(1)$$

$$\leq\; \frac{1}{\delta}\, e^{n\,\varepsilon_n^2\,(2+c_2)}\, E \int_{D_n} R_n(f)\,\Pi(df) + o(1)$$

$$=\; \frac{1}{\delta}\, e^{n\,\varepsilon_n^2\,(2+c_2)}\, \Pi(D_n) + o(1) \longrightarrow 0, \quad \text{as} \quad n \to \infty,$$

21

which concludes the proof of Lemma 3. $\qquad\square$

Proof of Theorem 2. Assume that $0 < \alpha < 1$. The proof of Theorem 2 follows from the same lines as the proof of Theorem 1. By Lemma 3 it suffices to prove that $\Pi_n\big(\mathcal{G}_n \cap A_{r\,\varepsilon_n}\big) \to 0$ in probability as $n \to \infty$. For any given $\delta > 0$, following the proof of Theorem 1 we get

$$
\begin{aligned}
F_0^\infty &\big\{\Pi_n\big(\mathcal{G}_n \cap A_{r\,\varepsilon_n}\big) \geq \delta\big\} \\
&\leq F_0^\infty\bigg\{ e^{n\,\varepsilon_n^2\,(2+c_2)\,\alpha} \sum_{j=1}^N \Pi(B_j)^\alpha \prod_{k=0}^{n-1} \frac{f_{kB_j}(X_{k+1})^\alpha}{f_0(X_{k+1})^\alpha} \geq \delta \bigg\} + \mathrm{o}(1) \\
&\leq \frac{1}{\delta}\, e^{n\,\varepsilon_n^2\,(2+c_2)\,\alpha} \sum_{j=1}^N \Pi(B_j)^\alpha\, E \prod_{k=0}^{n-1} \frac{f_{kB_j}(X_{k+1})^\alpha}{f_0(X_{k+1})^\alpha} + \mathrm{o}(1) \\
&\leq \frac{2}{\delta}\, e^{J(\varepsilon_n,\mathcal{G}_n,\alpha)+n\,\varepsilon_n^2\,(2+c_2)\,\alpha} \max_{1 \leq j \leq N} E \prod_{k=0}^{n-1} \frac{f_{kB_j}(X_{k+1})^\alpha}{f_0(X_{k+1})^\alpha} + \mathrm{o}(1) \\
&\leq \frac{2}{\delta}\, e^{J(\varepsilon_n,\mathcal{G}_n,\alpha)+n\,\varepsilon_n^2\left((2+c_2)\,\alpha+\frac{1}{2}(r-2)^2(\alpha-1)\right)} + \mathrm{o}(1) \\
&\leq \frac{2}{\delta}\, e^{J(\varepsilon_n,\mathcal{G}_n,\alpha)-n\,\varepsilon_n^2\,c_2} + \mathrm{o}(1) \longrightarrow 0 \qquad \text{as} \quad n \to \infty,
\end{aligned}
$$

where the last inequality follows from $r > 2 + \sqrt{\frac{2(2\alpha+\alpha c_2+c_1)}{1-\alpha}}$. The proof of Theorem 2 is complete. $\qquad\square$

Proof of Theorem 3. Since $\Pi_n(A_{r_n\varepsilon_n}) \leq \Pi_n\big(\mathcal{G}_n \cap A_{r_n\varepsilon_n}\big) + \Pi_n\big(A_{\varepsilon_n} \setminus \mathcal{G}_n\big)$, it suffices that the terms on the right hand side both tend to zero in probability. Given $\delta > 0$, the proof of Lemma 3 implies that

$$
F_0^\infty\big\{\Pi_n\big(A_{\varepsilon_n} \setminus \mathcal{G}_n\big) \geq \delta\big\} \leq \frac{e^{2n\varepsilon_n^2}\Pi(A_{\varepsilon_n} \setminus \mathcal{G}_n)}{\delta\, \Pi(B_{\varepsilon_n^2})} + \mathrm{o}(1),
$$

which by condition (1) tends to zero as $n \to \infty$. Assume that $[r_n]$ stands for the largest integer less than or equal to r_n and assume that $D_j = \big\{f \in \mathcal{G}_n : j\varepsilon_n \leq H(f_0,f) < 2j\varepsilon_n\big\}$ (Indeed, D_j is an empty set for $j > \sqrt{2}/\varepsilon_n$ since the Hellinger distance cannot exceed $\sqrt{2}$). Then we have

$$
\begin{aligned}
F_0^\infty\big\{\Pi_n\big(\mathcal{G}_n \cap A_{r_n\varepsilon_n}\big) \geq \delta\big\} &\leq F_0^\infty\bigg\{ \sum_{j=[r_n]}^\infty \Pi_n(D_j) \geq \delta \bigg\} \\
&\leq F_0^\infty\bigg\{ \sum_{j=[r_n]}^\infty \Pi_n(D_j)^\alpha \geq \delta \bigg\} \leq \frac{1}{\delta}\, E \sum_{j=[r_n]}^\infty \Pi_n(D_j)^\alpha \leq \frac{1}{\delta} \sum_{j=[r_n]}^\infty E\, \Pi_n(D_j)^\alpha.
\end{aligned}
$$

Take a partition $\bigcup_{i=1}^{N_j} D_{ji}$ for each D_j such that $D_{ji} \subset \{f : H(f_{ji}, f) < \frac{j\varepsilon_n}{3}\}$ for some f_{ji} in L_μ and

$$\sum_{i=1}^{N_j} \Pi(D_{ji})^\alpha \le 2 \exp\left(J(\frac{j\varepsilon_n}{3}, D_j, \alpha)\right) \le 2e^{c_1 j^2 n \varepsilon_n^2} \Pi(B_{\varepsilon_n^2})^\alpha,$$

where the last inequality follows from condition (2). Using the same argument as the proof of Theorem 1, one can get $H(f_{kD_{ji}}, f_0) \ge j\varepsilon_n/3$ and hence we have with probability tending to 1

$$
\begin{aligned}
F_0^\infty\{\Pi_n(\mathcal{G}_n \cap A_{r_n \varepsilon_n}) \ge \delta\} &\le \frac{1}{\delta} \sum_{j=[r_n]}^{\infty} E\left(\sum_{i=1}^{N_j} \Pi_n(D_{ji})\right)^\alpha \\
&\le \frac{1}{\delta} \sum_{j=[r_n]}^{\infty} \sum_{i=1}^{N_j} E\,\Pi_n(D_{ji})^\alpha \\
&\le \frac{e^{2\alpha n \varepsilon_n^2}}{\delta\,\Pi(B_{\varepsilon_n^2})^\alpha} \sum_{j=[r_n]}^{\infty} \sum_{i=1}^{N_j} \Pi(D_{ji})^\alpha\, e^{\frac{1}{18} j^2 n \varepsilon_n^2 (\alpha-1)} \\
&\le \frac{2}{\delta} \sum_{j=[r_n]}^{\infty} e^{n\varepsilon_n^2(2\alpha + c_1 j^2 + \frac{1}{18} j^2(\alpha-1))} \\
&\le \frac{2}{\delta} \sum_{j=[r_n]}^{\infty} \frac{1}{-n\,\varepsilon_n^2\left(2\alpha + c_1\,j^2 + \frac{1}{18}\,j^2\,(\alpha-1)\right)} \\
&\le \frac{3}{\delta\,n\,\varepsilon_n^2} \sum_{j=[r_n]}^{\infty} \frac{1}{-c_1\,j^2 + \frac{1}{18}\,j^2\,(1-\alpha)} \\
&\le \frac{3}{\delta\,n\,\varepsilon_n^2\left(-c_1 + \frac{1}{18}\,(1-\alpha)\right)} \sum_{j=[r_n]}^{\infty}\left(\frac{1}{j-1} - \frac{1}{j}\right) \\
&= \frac{3}{\delta\,n\,\varepsilon_n^2\left(-c_1 + \frac{1}{18}\,(1-\alpha)\right)([r_n]-1)}
\end{aligned}
$$

which tends to zero as $n \to \infty$, since $n\,\varepsilon_n^2$ are uniformly bounded away from zero and $r_n \to \infty$, where the second inequality follows from $\alpha \in (0,1)$, the third from the proof of Theorem 1, the fifth from the elementary inequality $e^{-x} < \frac{1}{x}$ for $x > 0$ and some of the inequalities only hold for all large n. Thus, we have proved that $\Pi_n(\mathcal{G}_n \cap A_{r_n \varepsilon_n})$ converges to zero in probability. The proof of Theorem 3 is complete. $\qquad\square$

Proof of Theorem 4. From $n\,\varepsilon_n^2 \geq c_0 \log n$ and $c_0\,c_2 > 1$ it turns out that $e^{-n\,\varepsilon_n^2\,c_2} \leq 1/n^{c_0\,c_2}$ and hence, by the first Borel-Cantelli Lemma and Lemma 1, we obtain that

$$\int_{\mathbb{F}} R_n(f)\,\Pi(df) \geq e^{-n\,\varepsilon_n^2\,(3+2c_2)}\,\Pi\big(W_{\varepsilon_n}\big)$$

almost surely for all large n. Then, following the proofs of Lemma 3 and Theorem 3, one can get that for any $\delta > 0$ and $r > 1$,

$$F_0^\infty\big\{\Pi_n\big(A_{r\varepsilon_n}\big) \geq \delta\big\} \leq F_0^\infty\big\{\Pi_n\big(A_{\varepsilon_n}\setminus\mathcal{G}_n\big) \geq \frac{\delta}{2}\big\} + F_0^\infty\big\{\Pi_n\big(\mathcal{G}_n\cap A_{r\varepsilon_n}\big) \geq \frac{\delta}{2}\big\}$$

$$\leq \frac{2\,e^{n\,\varepsilon_n^2\,(3+2c_2)}\,\Pi(A_{\varepsilon_n}\setminus\mathcal{G}_n)}{\delta\,\Pi(W_{\varepsilon_n})} + \frac{4}{\delta}\sum_{j=[r]}^{\infty} e^{n\varepsilon_n^2\left((3+2c_2)\alpha + c_1 j^2 + \frac{1}{18} j^2(\alpha-1)\right)} := a_n + b_n.$$

Condition (1) yields that $\sum_{n=1}^{\infty} a_n < \infty$. On the other hand, since $c_1 < \frac{1-\alpha}{18}$, we have that for all $n \geq 2$ and for all r so large that $(3+2c_2)\alpha + \left(c_1 + \frac{\alpha-1}{18}\right)\left(r^2 - 1\right) \leq -\frac{2}{c_0}$,

$$
\begin{aligned}
b_n &\leq \frac{4\,e^{n\varepsilon_n^2(3+2c_2)\alpha}}{\delta} \sum_{j=[r]^2}^{\infty} e^{n\varepsilon_n^2\left(c_1 + \frac{\alpha-1}{18}\right)j} = \frac{4\,e^{n\varepsilon_n^2\left((3+2c_2)\alpha + \left(c_1+\frac{\alpha-1}{18}\right)[r]^2\right)}}{\delta\left(1 - e^{n\varepsilon_n^2\left(c_1+\frac{\alpha-1}{18}\right)}\right)} \\[2mm]
&\leq \frac{4\,n^{c_0\left((3+2c_2)\alpha + \left(c_1+\frac{\alpha-1}{18}\right)[r]^2\right)}}{\delta\left(1 - n^{c_0\left(c_1+\frac{\alpha-1}{18}\right)}\right)} \leq \frac{4\,n^{c_0\left((3+2c_2)\alpha + \left(c_1+\frac{\alpha-1}{18}\right)(r^2-1)\right)}}{\delta\left(1 - 2^{c_0\left(c_1+\frac{\alpha-1}{18}\right)}\right)} \\[2mm]
&\leq \frac{4\,n^{-2}}{\delta\left(1 - 2^{c_0\left(c_1+\frac{\alpha-1}{18}\right)}\right)}.
\end{aligned}
$$

and hence $\sum_{n=1}^{\infty} b_n < \infty$. Thus, by the first Borel-Cantelli Lemma we obtain that $\Pi_n\big(A_{r\varepsilon_n}\big) \leq \delta$ almost surely for all large n, which concludes the proof of Theorem 4. $\qquad\square$

References

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

[2] Ghosal, S. (2001). Convergence rates for density estimation with Bernstein polynomials. *Ann. Statist.* **29**, 1264-1280.

[3] Ghosal, S., Ghosh, J.K., and Ramamoorthi, R.V. (1997). Non-informative priors via sieves and packing numbers. In *Advances in Statistical Decision Theory*, S. Panchapakeshan and N.Balakrishnan (eds.), 129-140. Birkhäuser, Boston.

[4] Ghosal, S., Ghosh, J.K., and Van der Vaart, A.W. (2000). Convergence distributions. *Ann. Statist.* **28**, 500-531.

[5] Ghosal, S. and Van der Vaart, A.W. (2001).Entropies and rates of convergence for maximum rates of posterior likelihood and Bayes estimation for mixtures of normal densities. *Ann. Statist.* **29**, 1233-1263.

[6] Ghosal, S. and Van der Vaart, A.W. (2007a). Convergence rates of posterior distributions for noniid observations. *Ann. Statist.* **35**, 192-223.

[7] Ghosal, S. and Van der Vaart,, A.W. (2007b). Posterior convergence rates of Dirichlet mixtures at smooth densities. *Ann. Statist.* **35**, 697-723.

[8] Kolmogorov, A.N. and Tihomirov, V.M. (1959). ε-entropy and ε-capacity of sets in function spaces. *Uspekhi Mat. Nauk* **14**, 3-86 [in Russian; English transl. *Amer. Math. Soc. Transl. Ser. 2*, **17**, 277-364 (1961)].

[9] Lijoi, A., Prunster, I., and Walker, S. (2007). On convergence rates for nonparametric posterior distributions. *Aust. N. Z. J. Stat.* **49**(3), 209-219.

[10] Petrone, S. (1999). Random Bernstein polynomials. *Scand. J. Statist.* **26**, 373-393.

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

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

[13] Scricciolo, C. (2006). Convergence rates for Bayesian density estimation of infinite-dimensional exponential families. *Ann. Statist.* **34**, 2897-2920.

[14] Shen, X. and Wasserman, L. (2001). Rates of convergence of posterior distributions. *Ann. Statist.* **29**, 687-714.

[15] Stone, C.J. (1990). Large-sample inference for log-spline models. *Ann. Statist.* **18**, 717-741.

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

[17] Walker, S., Lijoi, A., and Prunster, I. (2007). On rates of convergence for posterior distributions in infinite-dimensional models. *Ann. Statist.* **35**, 738-746.

[18] Xing, Y. and Ranneby, B. (2008). Sufficient conditions for Bayesian consistency. Research Report 2008:6, Centre of Biostochastics, Swedish University of Agricultural Sciences.