On Bayesian Consistency

Yang Xing and Bo Ranneby

Research Report
Centre of Biostochastics

Swedish University of Agricultural Sciences

Report 2008:9
ISSN 1651-8543

On Bayesian consistency

Yang Xing[1] and Bo Ranneby

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

Abstract

We establish a sufficient condition ensuring strong Hellinger consistency of posterior distributions. We also prove a strong Hellinger consistency theorem for the pseudoposterior distributions based on the likelihood ratio with power $0 < \alpha < 1$, which are introduced by Walker and Hjort (2001). Our result is an extension of their theorem for $\alpha = 1/2$.

Keywords: Hellinger consistency, posterior distribution, nonparametric model.

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

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

1 Introduction

Recent years have seen a remarkable development in the area of Bayesian nonparametric inference. One of key points to this area is to establish consistency of posterior distributions with respect to a suitable metric, usually the Hellinger metric. Early works on consistency of posterior distributions were concerned with weak consistency. Freedman (1963) and Diaconis and Freedman (1986) have demonstrated that a prior having positive mass on all weak neighborhoods of the true density is not necessarily weakly consistent. A sufficient condition for weak consistency was suggested by Schwartz (1965). Schwartz (1965) proved that, if the true density 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. Schwartz's theorem provides a powerful tool in establishing weak posterior consistency, see, for example, Barron (1999). However, it seems not to be useful for establishing strong posterior consistency. In many applications like density estimation it is natural to ask for strong consistency of Bayesian procedures. Recent attention has switched to studying strong posterior consistency. It is known that the condition of the true density being in the Kullback-Leibler support cannot 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 the true density f_0. Some additional restrictions must be added to obtain the strong Hellinger consistency of posterior distributions. Barron et al. (1999), Ghosal et al. (1999) and Walker (2004) have obtained important results in this direction. In the present paper we give a sufficient condition of the strong Hellinger consistency of posterior distributions. As an application we give a new proof of the sufficient condition of Barron et al. (1999) by means of the upper bracketing metric entropy. Since the existing sufficient conditions of the strong Hellinger posterior consistency seem to be quite difficult to be verified, Walker and Hjort (2001) introduced one kind of pseudoposterior distributions based on the likelihood ratio with power α. Under the unique condition of the true density belonging to the Kullback-Leibler support of the prior distribution, they obtained a strong consistency theorem for the pseudoposterior distributions with respect to a related metric $H_\alpha(f,g) = \big(1 - \int g^\alpha f^{1-\alpha}\big)^{1/2}$, which agrees nicely with the Hellinger metric only in the case of $\alpha = 1/2$. Another main aim of this paper is to prove strong consistency of pseudoposterior distributions with respect to the Hellinger metric for all $0 < \alpha < 1$.

We consider a family of probability measures dominated by a σ-finite measure μ in a Polish space $\mathbb{X}$ endowed with a σ-algebra $\mathcal{X}$. Assume that

$(X_1, X_2, \ldots, X_n)$ is an independent identically distributed sample of random variables, taking values in $\mathbb{X}$ and having a common probability density function f_0 with respect to μ. For two density functions f and g we denote the Hellinger distance

$$H(f,g) = \Big(\int_{\mathbb{X}} \big(\sqrt{f(x)} - \sqrt{g(x)}\ \big)^2 \mu(dx) \Big)^{1/2}$$

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 density functions is separable with respect to the Hellinger metric and assume that $\mathcal{F}$ is the Borel σ-algebra of $\mathbb{F}$. Denote $A_\varepsilon = \big\{ f : H(f_0, f) \geq \varepsilon \big\}$ and $N_\delta = \big\{ f : K(f_0, f) < \delta \big\}$. The density function f_0 is said to be in the Kullback-Leibler support of Π if $\Pi\big(N_\delta\big) > 0$ for all $\delta > 0$. Let Π be a prior distribution on $\mathbb{F}$. It is known that the posterior distribution Π_n given $X_1, X_2, \ldots, X_n$ has the following expression

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

for all measurable subsets A in $\mathbb{F}$, where $R_n(f) = \prod_{i=1}^n \big\{ f(X_i)/f_0(X_i) \big\}$ stands for the likelihood ratio.

The posterior consistency relies on how the likelihood ratio behaves as the sample size increases to infinity. It seems that the size of likelihood ratio plays a crucial role in determining Bayesian consistency. In this paper we use the quantity $R_n(f)^\alpha$, the likelihood ratio with power α, to study strong consistency of posterior distributions. Different approaches are used to deal with the likelihood ratio of power $\alpha > 1$ and $\alpha < 1$, respectively. In the case $\alpha > 1$ we establish a sufficient condition of the strong posterior consistency, whereas for $\alpha < 1$ we discuss strong consistency of pseudoposterior distributions introduced by Walker and Hjort (2001). For any α in $(0, 1)$ we prove strong Hellinger consistency of the pseudoposterior distributions. This extends a result of Walker and Hjort (2001).

2 Strong posterior consistency

The main work of this section is to establish a sufficient condition of Bayesian consistency by means of $R_n(f)^\alpha$ with $\alpha > 1$.

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 positive constants α, c_1, c_2, c_3, c_4 and a sequence $\{\mathcal{F}_n\}_{n=1}^{\infty}$ of subsets of $\mathbb{F}$ such that $\alpha > 1$, $c_1 < 1 - \frac{1}{\alpha}$ and for all large n,*

(i) $E\Big(\int_{A_\varepsilon \cap \mathcal{F}_n} R_n(f)^\alpha \, \Pi(df)\Big)^{\frac{1}{\alpha}} \le c_2 \, e^{n\, c_1\, \varepsilon^2}$, *where E stands for the expectation with respect to the distribution of $X_1, X_2, \ldots, X_n$;*

(ii) $\Pi\big(A_\varepsilon \setminus \mathcal{F}_n\big) < c_4 \, e^{-n\, c_3}$.

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

Note that Theorem 1 fails when $\alpha \le 1$, since Jensen's inequality yields that for $\alpha \le 1$ the left-hand side of the inequality of condition (i) does not exceed one.

Proof. Write $\Pi_n(A_\varepsilon) = \Pi_n(A_\varepsilon \cap \mathcal{F}_n) + \Pi_n(A_\varepsilon \setminus \mathcal{F}_n)$. From Lemma 5 of Barron et al. (1999) and condition (ii) it turns out that the second term $\Pi_n(A_\varepsilon \setminus \mathcal{F}_n) \to 0$ almost surely as $n \to \infty$. It suffices to prove that the first term also tends to zero almost surely.

We first give an estimation of the numerator of $\Pi_n(A_\varepsilon \cap \mathcal{F}_n)$. Let $q = 2\alpha - 1 > 1$ and $p = (2\alpha - 1)/2(\alpha - 1) > 1$. Then $1/p + 1/q = 1$ and hence by Hölder's inequality we get that for any given $\varepsilon > 0$,

$$
\begin{aligned}
&\int_{A_\varepsilon \cap \mathcal{F}_n} R_n(f) \, \Pi(df) \\
&= \int_{A_\varepsilon \cap \mathcal{F}_n} R_n(f)^{\frac{1}{2p}} R_n(f)^{1 - \frac{1}{2p}} \, \Pi(df) \\
&\le \Big(\int_{A_\varepsilon \cap \mathcal{F}_n} R_n(f)^{\frac{1}{2}} \, \Pi(df)\Big)^{\frac{1}{p}} \Big(\int_{A_\varepsilon \cap \mathcal{F}_n} R_n(f)^{q(1 - \frac{1}{2p})} \, \Pi(df)\Big)^{\frac{1}{q}} \\
&= \Big(\int_{A_\varepsilon \cap \mathcal{F}_n} R_n(f)^{\frac{1}{2}} \, \Pi(df)\Big)^{\frac{1}{p}} \Big(\int_{A_\varepsilon \cap \mathcal{F}_n} R_n(f)^{\alpha} \, \Pi(df)\Big)^{\frac{1}{q}}
\end{aligned}
$$

Since $c_1 < 1 - \frac{1}{\alpha}$ the inequality $\frac{1}{2p} > \frac{c_1 \alpha}{q}$ holds. Take a constant b with $0 < b < \frac{1}{2p} - \frac{c_1 \alpha}{q}$. It follows then from Hölder's inequality and Fubini's theorem that for all n large enough,

$$
\begin{aligned}
&F_0^\infty\Big\{\int_{A_\varepsilon \cap \mathcal{F}_n} R_n(f) \, \Pi(df) \ge e^{-n\, b\, \varepsilon^2}\Big\} \\
&= F_0^\infty\Big\{e^{\frac{n\, b\, \varepsilon^2}{\alpha}} \Big(\int_{A_\varepsilon \cap \mathcal{F}_n} R_n(f) \, \Pi(df)\Big)^{\frac{1}{\alpha}} \ge 1\Big\}
\end{aligned}
$$

$$
\begin{aligned}
&\le\ e^{\frac{n\,b\,\varepsilon^2}{\alpha}}\,E\Big(\int_{A_\varepsilon\cap\mathcal{F}_n} R_n(f)\,\Pi(df)\Big)^{\frac{1}{\alpha}}\\
&\le\ e^{\frac{n\,b\,\varepsilon^2}{\alpha}}\left(E\Big(\int_{A_\varepsilon\cap\mathcal{F}_n} R_n(f)^{\frac{1}{2}}\,\Pi(df)\Big)^{\frac{1}{\alpha}}\right)^{\frac{1}{p}}\left(E\Big(\int_{A_\varepsilon\cap\mathcal{F}_n} R_n(f)^{\alpha}\,\Pi(df)\Big)^{\frac{1}{\alpha}}\right)^{\frac{1}{q}}\\
&\le\ e^{\frac{n\,b\,\varepsilon^2}{\alpha}}\left(E\int_{A_\varepsilon\cap\mathcal{F}_n} R_n(f)^{\frac{1}{2}}\,\Pi(df)\right)^{\frac{1}{p\,\alpha}}\left(E\Big(\int_{A_\varepsilon\cap\mathcal{F}_n} R_n(f)^{\alpha}\,\Pi(df)\Big)^{\frac{1}{\alpha}}\right)^{\frac{1}{q}}\\
&\le\ c_2^{\frac{1}{q}}\,e^{\frac{n\,b\,\varepsilon^2}{\alpha}+\frac{n\,c_1\,\varepsilon^2}{q}}\left(E\int_{A_\varepsilon} R_n(f)^{\frac{1}{2}}\,\Pi(df)\right)^{\frac{1}{p\,\alpha}}\\
&=\ c_2^{\frac{1}{q}}\,e^{\frac{n\,b\,\varepsilon^2}{\alpha}+\frac{n\,c_1\,\varepsilon^2}{q}}\left(\int_{A_\varepsilon}\Big(\int\sqrt{f(x)f_0(x)}\,\mu(dx)\Big)^n\,\Pi(df)\right)^{\frac{1}{p\,\alpha}}\\
&=\ c_2^{\frac{1}{q}}\,e^{\frac{n\,b\,\varepsilon^2}{\alpha}+\frac{n\,c_1\,\varepsilon^2}{q}}\left(\int_{A_\varepsilon}\big(1-\frac{1}{2}H(f_0,f)^2\big)^n\,\Pi(df)\right)^{\frac{1}{p\,\alpha}}\\
&\le\ c_2^{\frac{1}{q}}\,e^{\frac{n\,b\,\varepsilon^2}{\alpha}+\frac{n\,c_1\,\varepsilon^2}{q}}\,(1-\frac{\varepsilon^2}{2})^{\frac{n}{p\,\alpha}}\\
&=\ c_2^{\frac{1}{q}}\,e^{\frac{n\,b\,\varepsilon^2}{\alpha}+\frac{n\,c_1\,\varepsilon^2}{q}}\,e^{\frac{n}{p\,\alpha}\log(1-\frac{\varepsilon^2}{2})}\le c_2^{\frac{1}{q}}\,e^{n\,\varepsilon^2\,(\frac{b}{\alpha}+\frac{c_1}{q}-\frac{1}{2p\,\alpha})}.
\end{aligned}
$$

From $\frac{b}{\alpha}+\frac{c_1}{q}-\frac{1}{2p\,\alpha}<0$ it turns out that $\sum_{n=1}^{\infty} e^{n\,\varepsilon^2\,(\frac{b}{\alpha}+\frac{c_1}{q}-\frac{1}{2p\,\alpha})}<\infty$. By the first Borel-Cantelli Lemma, the inequality

$$\int_{A_\varepsilon\cap\mathcal{F}_n} R_n(f)\,\Pi(df)\le e^{-n\,b\,\varepsilon^2}$$

holds almost surely for all n large enough.

To estimate the denominator of $\Pi_n(A_\varepsilon\cap\mathcal{F}_n)$ it is enough to apply the known result that for each given $\delta>0$ the denominator is bounded below by $e^{-n\,\delta}$ almost surely for large n, see Lemma 4 of Barron et al. (1999). Choosing $\delta=b\,\varepsilon^2/2$ and applying the above estimations for both the numerator and the denominator, we obtain that

$$\Pi_n(A_\varepsilon\cap\mathcal{F}_n)\le e^{-n\,b\,\varepsilon^2+n\,\delta}=e^{-\frac{n\,b\,\varepsilon^2}{2}}$$

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

Sometimes it is convenient to find a suitable covering of $\mathbb{F}$ in order to obtain consistent posteriors. Here we present a result in this direction. Given $\varepsilon>0$, we consider a partition of A_ε defined by

$$A_{nj} = \left\{ f \in A_\varepsilon : \ e^{n\,a_{j-1}} \le R_n(f) < e^{n\,a_j} \right\} \qquad \text{for} \quad j = 1, \ldots,$$

where $a_0 = -\infty$ and $\{a_j\}_1^\infty$ is a sequence increasing to ∞.

Corollary 1. *Suppose that f_0 is in the Kullback-Leibler support of Π and suppose that there exist constants $\alpha > 1$, $c < 1 - \frac{1}{\alpha}$ and a positive series $\sum_{j=1}^\infty d_j < \infty$ such that*

$$\Pi(A_{nj})^{\frac{1}{\alpha}} \le d_j\, e^{-n(a_j - c\varepsilon^2)}$$

for all j and all but finitely many n. Then $\Pi_n(A_\varepsilon) \to 0$ almost surely as $n \to \infty$.

Proof. It is no restriction to assume that $c > 0$. From the inequality $(x+y)^{\frac{1}{\alpha}} \le x^{\frac{1}{\alpha}} + y^{\frac{1}{\alpha}}$ for all $x, y \ge 0$ it turns out that

$$\begin{aligned} E\left(\int_{A_\varepsilon} R_n(f)^\alpha \, \Pi(df) \right)^{\frac{1}{\alpha}} &\le E \sum_{j=1}^\infty \left(\int_{A_{nj}} R_n(f)^\alpha \, \Pi(df) \right)^{\frac{1}{\alpha}} \\ &= \sum_{j=1}^\infty E \left(\int_{A_{nj}} R_n(f)^\alpha \, \Pi(df) \right)^{\frac{1}{\alpha}} \\ &\le \sum_{j=1}^\infty e^{n\,a_j}\, \Pi(A_{nj})^{\frac{1}{\alpha}} \ \le \ e^{nc\varepsilon^2} \sum_{j=1}^\infty d_j, \end{aligned}$$

which by Theorem 1 concludes the proof. □

Corollary 1 is an analogue of Theorem 3 in Walker et al. (2005).

3 Log strong pseudoposterior consistency

In this section we study one class of pseudoposterior distributions introduced by Walker and Hjort (2001). We extend their result on the strong consistency of pseudoposterior distributions. We also use a slight modification of the pseudoposteriors of Walker and Hjort to deal with strong consistency of posterior distributions.

Wasserman (2000) introduced a psuedolikelihood function-data-dependent prior to obtain asymptotic properties for mixture models. Later, Walker and

Hjort (2001) used a different psuedolikelihood function to establish a strong consistency theorem. Given $0 \leq \alpha \leq 1$, they defined the pseudoposterior distribution Q_n^α based on Π by

$$Q_n^\alpha(df) = \frac{R_n(f)^\alpha\, \Pi(df)}{\int_{\mathbb{F}} R_n(f)^\alpha\, \Pi(df)},$$

which can be considered as we are using the data-dependent prior $\Pi(df) \Big/ \prod_{i=1}^{n} f^{1-\alpha}(x_i)$. Clearly, $Q_n^0 = \Pi$ and $Q_n^1 = \Pi_n$. So for $\alpha = 1$ the condition of the true density f_0 being in the Kullback-Leibler support cannot guarantee strong Hellinger consistency of Q_n^α. However, when $0 < \alpha < 1$, under this unique condition Walker and Hjort (2001) proved the strong consistency of pseudoposteriors with respect to the metric defined by $H_\alpha(f, f_0) = \big(1 - \int f_0^\alpha f^{1-\alpha}\big)^{1/2}$ which, in the particular case of $\alpha = 1/2$, agrees nicely with the Hellinger distance. Hence, for $\alpha = 1/2$, they obtained strong Hellinger consistent pseudoposteriors. Now for all $0 < \alpha < 1$ we prove the strong Hellinger consistency of the pseudoposteriors.

Theorem 2. *Let $0 < \alpha < 1$. If f_0 is in the Kullback-Leibler support of Π then for each $\varepsilon > 0$ we have that $Q_n^\alpha(A_\varepsilon) \to 0$ almost surely as $n \to \infty$.*

Proof. Given $\varepsilon > 0$ and $b > 0$, we have

$$F_0^\infty\Big\{\int_{A_\varepsilon} R_n(f)^\alpha\, \Pi(df) \geq e^{-n\,b\,\varepsilon^2}\Big\} \leq e^{n\,b\,\varepsilon^2}\, E\int_{A_\varepsilon} R_n(f)^\alpha\, \Pi(df).$$

From Hölder's inequality and Fubini's theorem it turns out that

$$\begin{aligned}
& E\int_{A_\varepsilon} R_n(f)^\alpha\, \Pi(df) \\
&= E\int_{A_\varepsilon} R_n(f)^{\frac{\alpha}{2}} R_n(f)^{\frac{\alpha}{2}}\, \Pi(df) \\
&\leq E\left(\left(\int_{A_\varepsilon} R_n(f)^{\frac{\alpha}{2}\cdot\frac{2}{2-\alpha}}\, \Pi(df)\right)^{\frac{2-\alpha}{2}} \left(\int_{A_\varepsilon} R_n(f)^{\frac{\alpha}{2}\cdot\frac{2}{\alpha}}\, \Pi(df)\right)^{\frac{\alpha}{2}}\right) \\
&\leq \left(E\int_{A_\varepsilon} R_n(f)^{\frac{\alpha}{2-\alpha}}\, \Pi(df)\right)^{\frac{2-\alpha}{2}} \left(E\int_{A_\varepsilon} R_n(f)\, \Pi(df)\right)^{\frac{\alpha}{2}} \\
&= \left(E\int_{A_\varepsilon} R_n(f)^{\frac{\alpha}{2-\alpha}}\, \Pi(df)\right)^{\frac{2-\alpha}{2}} \left(\int_{A_\varepsilon} E\big(R_n(f)\big)\, \Pi(df)\right)^{\frac{\alpha}{2}} \\
&\leq \left(E\int_{A_\varepsilon} R_n(f)^{\frac{\alpha}{2-\alpha}}\, \Pi(df)\right)^{\frac{2-\alpha}{2}}.
\end{aligned}$$

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

$$E\int_{A_\varepsilon} R_n(f)^\alpha\,\Pi(df) \le \Big(E\int_{A_\varepsilon} R_n(f)^{\frac{\alpha}{2^m(1-\alpha)+\alpha}}\,\Pi(df)\Big)^{\frac{2^m(1-\alpha)+\alpha}{2^m}},$$

which by Hölder's inequality does not exceed

$$\begin{aligned}\Big(E\int_{A_\varepsilon} R_n(f)^{\frac{1}{2}}\,\Pi(df)\Big)^{\alpha\,2^{1-m}} &= \Big(\int_{A_\varepsilon}\Big(\int\sqrt{f(x)f_0(x)}\,\mu(dx)\Big)^n\,\Pi(df)\Big)^{\alpha\,2^{1-m}}\\ &\le (1-\frac{\varepsilon^2}{2})^{n\,\alpha\,2^{1-m}} \le e^{-n\varepsilon^2\alpha\,2^{-m}}.\end{aligned}$$

Hence we have obtained

$$F_0^\infty\Big\{\int_{A_\varepsilon} R_n(f)^\alpha\,\Pi(df) \ge e^{-n\,b\,\varepsilon^2}\Big\} \le e^{n\,\varepsilon^2(b-\alpha\,2^{-m})} \qquad \text{for all} \quad n.$$

For any positive constant $b < \frac{1-\alpha}{2}$ we have that $b-\alpha\,2^{-m} < 0$. It then follows from the first Borel-Cantelli Lemma that

$$\int_{A_\varepsilon} R_n(f)^\alpha\,\Pi(df) \le e^{-n\,b\,\varepsilon^2}$$

holds almost surely for all n large enough.

On the other hand, using the same argument as the proof of Lemma 4 of Barron et al. (1999), we get that for each given $\delta > 0$,

$$\int_{\mathbb{F}} R_n(f)^\alpha\,\Pi(df) \ge e^{-n\,\alpha\,\delta}$$

almost surely for all n large enough. Taking $b = \frac{1-\alpha}{4}$ and $\delta = \frac{(1-\alpha)\,\varepsilon^2}{5\,\alpha}$, we get that

$$Q_n^\alpha(A_\varepsilon) \le e^{-n\,b\,\varepsilon^2+n\,\alpha\,\delta} = e^{-n\,(1-\alpha)\,\varepsilon^2/20}$$

almost surely for large n. The proof of Theorem 2 is complete. □

We will finish this section by presenting a simple fact on a modification of Q_n^α. Denote

$$\widetilde{Q}_n^\alpha(df) = \frac{R_n(f)^\alpha\,\Pi(df)}{\Big(\int_{\mathbb{F}} R_n(f)\,\Pi(df)\Big)^\alpha} \qquad \text{for} \quad 0<\alpha<1.$$

By the proof of Theorem 2 we get

Theorem 3. *Let $0 < \alpha < 1$. If f_0 is in the Kullback-Leibler support of Π then for each $\varepsilon > 0$ we have that $\widetilde{Q}_n^\alpha(A_\varepsilon) \to 0$ almost surely as $n \to \infty$.*

The pseudoposteriors $\widetilde{Q}_n^\alpha$ can be used to characterize the Hellinger consistency of the posterior distributions Π_n.

Theorem 4. *Let $0 < \alpha_0 < 1$ and $\varepsilon > 0$. Suppose that f_0 is in the Kullback-Leibler support of Π. Then $\Pi_n(A_\varepsilon) \to 0$ almost surely as $n \to \infty$ if and only if $\max\limits_{\alpha_0 \le \alpha \le 1} \widetilde{Q}_n^\alpha(A_\varepsilon) \to 0$ almost surely, that is,*

$$F_0^\infty\Big\{\bigcap_{l=1}^\infty \bigcup_{k=1}^\infty \bigcap_{n=k}^\infty \Big\{ \sup_{\alpha_0 \le \alpha \le 1} \widetilde{Q}_n^\alpha(A_\varepsilon) \le \frac{1}{l} \Big\}\Big\} = 1.$$

Proof. The "if " part is trivial. The "only if " part follows from Hölder's inequality, as shown in the following

$$\widetilde{Q}_n^\alpha(A_\varepsilon) = \frac{\int_{A_\varepsilon} R_n(f)^\alpha \, \Pi(df)}{\Big(\int_{\mathbb{F}} R_n(f)\,\Pi(df)\Big)^\alpha} \le \frac{\Big(\int_{A_\varepsilon} R_n(f)\,\Pi(df)\Big)^\alpha \, \big(\Pi(A_\varepsilon)\big)^{1-\alpha}}{\Big(\int_{\mathbb{F}} R_n(f)\,\Pi(df)\Big)^\alpha}$$
$$\le \Pi_n(A_\varepsilon)^\alpha \le \Pi_n(A_\varepsilon)^{\alpha_0}$$

for all $\alpha_0 \le \alpha \le 1$. The proof of Theorem 4 is complete. □

4 Applications

In this section we discuss some applications of our theorems.

4.1 Upper bracketing metric entropy

Barron et al. (1999) provided a very useful general theorem on strong Hellinger consistency, in which they used the upper bracketing metric entropy. Let L_μ be the space of all nonnegative integrable functions with respect to a measure μ on $\mathbb{X}$. Given a subset $\mathcal{G}$ of $\mathbb{F}$ and $\delta > 0$, the upper bracketing metric entropy $\mathcal{H}(\delta, \mathcal{G})$ 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) $\int_{\mathbb{X}} f_j(x)\,\mu(dx) \le 1 + \delta$ for all j; (b) For each $f \in \mathcal{G}$ there exists some f_j with $f \le f_j$. Now as an application of Theorem 1, we give a new proof of the following result of Barron et al. (1999, Theorem 1).

Proposition 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 $0 < \delta \leq \frac{\varepsilon^2}{2}$, c_1, $c_2 > 0$, $0 < c_3 < \frac{\varepsilon^2}{2}$, and $\mathcal{F}_n \subset \mathbb{F}$ such that for all large n,*

(i) $\mathcal{H}(\delta, \mathcal{F}_n) < n\, c_3$;

(ii) $\Pi\big(\mathbb{F} \setminus \mathcal{F}_n\big) < c_2\, e^{-n\, c_1}$.

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

Proof. We only need verify condition (i) of Theorem 1. Take a large α with $\alpha > 1$ and $1 - \frac{1}{\alpha} > \frac{c_3}{\varepsilon^2} + \frac{1}{2}$. For each $\mathcal{F}_n$, we take functions $f_1, f_2, \ldots, f_{N_n}$ in L_μ such that $N_n < e^{n\, c_3}$, $\int_{\mathbb{X}} f_j(x)\, \mu(dx) \leq 1 + \varepsilon$ for all j and each $f \in \mathcal{F}_n$ does not exceed some f_j. We can then make a partition $A_\varepsilon \cap \mathcal{F}_n = \cup_{j=1}^{N_n} B_j$ such that $f \leq f_j$ for all $f \in B_j$. Hence we have

$$
\begin{aligned}
& E\Big(\int_{A_\varepsilon \cap \mathcal{F}_n} R_n(f)^\alpha\ \Pi(df)\Big)^{\frac{1}{\alpha}} \\
= \ & \int_{\mathbb{X}^n} \Big(\int_{A_\varepsilon \cap \mathcal{F}_n} \prod_{i=1}^n f(x_i)^\alpha\ \Pi(df)\Big)^{\frac{1}{\alpha}} \mu(dx_1)\ldots\mu(dx_n) \\
= \ & \int_{\mathbb{X}^n} \Big(\sum_{j=1}^{N_n} \int_{B_j} \prod_{i=1}^n f(x_i)^\alpha\ \Pi(df)\Big)^{\frac{1}{\alpha}} \mu(dx_1)\ldots\mu(dx_n) \\
\leq \ & \sum_{j=1}^{N_n} \int_{\mathbb{X}^n} \Big(\int_{B_j} \prod_{i=1}^n f(x_i)^\alpha\ \Pi(df)\Big)^{\frac{1}{\alpha}} \mu(dx_1)\ldots\mu(dx_n) \\
\leq \ & \sum_{j=1}^{N_n} \int_{\mathbb{X}^n} \Big(\int_{B_j} \prod_{i=1}^n f_j(x_i)^\alpha\ \Pi(df)\Big)^{\frac{1}{\alpha}} \mu(dx_1)\ldots\mu(dx_n) \\
\leq \ & \sum_{j=1}^{N_n} \int_{\mathbb{X}^n} \prod_{i=1}^n f_j(x_i)\ \mu(dx_1)\ldots\mu(dx_n) \\
\leq \ & N_n\, (1 + \frac{\varepsilon^2}{2})^n \leq e^{n\, (\frac{c_3}{\varepsilon^2} + \frac{1}{2})\, \varepsilon^2}.
\end{aligned}
$$

Thus we have obtained condition (i) of Theorem 1 and the proof of Proposition 1 is complete. □

Proposition 1 has been widely applied in many statistical models, we refer to Barron et al. (1999) for the details.

4.2 Pseudo-Bayes estimator

Convergence of the pseudoposteriors Q_n^α implies the existence of pseudo-Bayes estimators. A useful pseudo-Bayes estimator based on the Q_n^α is given by

$$f_n^\alpha(x) = \int_{\mathbb{F}} f(x)\, Q_n^\alpha(df).$$

Given $\varepsilon > 0$, from convexity of the squared Hellinger distance and Jensen's inequality it turns out that

$$\begin{aligned} H^2(f_n^\alpha, f_0) &\leq \int_{\mathbb{F}} H^2(f, f_0)\, Q_n^\alpha(df) \\ &\leq \int_{A_\varepsilon} 2\, Q_n^\alpha(df) + \int_{\mathbb{F}\setminus A_\varepsilon} \varepsilon^2\, Q_n^\alpha(df) \\ &\leq 2\, Q_n^\alpha(A_\varepsilon) + \varepsilon^2, \end{aligned}$$

where, by Theorem 2, the first term in the last sum tends to zero almost surely as $n \to \infty$. This yields almost sure convergence of the pseudo-Bayes estimator f_n^α to the true density f_0 with respect to the Hellinger distance.

4.3 Parametric family

Let $\mathbb{F} = \big\{ f_\theta(x) \,:\; \theta \in \Theta \big\}$ be a class of density functions with respect a dominating σ-finite measure μ on $\mathbb{X}$. Consider a sample $X_1, X_2 \ldots, X_n$ of i.i.d. observations from $f_{\theta_0}(x)$ with $\theta_0 \in \Theta$. The maximum likelihood estimator (MLE) $\hat{\theta}_n$ is defined by

$$\hat{\theta}_n = \arg\max\nolimits_{\theta\in\Theta} L_n(\theta),$$

where $L_n(\theta) = \prod\limits_{i=1}^{n} f_\theta(X_i)$ is the likelihood function. We assume that such a $\hat{\theta}_n$ exists (not necessary uniquely). Let Π be a prior distribution on Θ. Denote $R_n(\theta) = \prod\limits_{i=1}^{n} f_\theta(X_i)\big/ f_{\theta_0}(X_i)$. Walker and Hjort (2001, Theorem 3) gave the following estimation of posterior distributions: if $\Pi\big(N_\delta\big) > 0$ for all $\delta > 0$ then $\Pi_n(A_\varepsilon) \leq e^{-nc}\, R_n(\hat{\theta}_n)^{\frac{1}{2}}$ almost surely for large n for any $\varepsilon > 0$ and $c < \frac{1}{2}\,\varepsilon^2$. Hence the strong consistency of $\Pi_n(A_\varepsilon)$ is guaranteed if $R_n(\hat{\theta}_n) = o(e^{n2c})$ almost surely as $n \to \infty$. Now we prove that the estimation of Walker and Hjort (2001, Theorem 3) is still true if the square root is replaced by any exponent α with $0 < \alpha < 1$.

Theorem 5. *Let $0 < \alpha < 1$. If f_0 is in the Kullback-Leibler support of Π then*

$$\Pi_n(A_\varepsilon) \leq e^{-nc}\, R_n(\hat{\theta}_n)^\alpha$$

almost surely for large n for any $\varepsilon > 0$ and $c < \frac{\alpha\,\varepsilon^2}{2}$. Therefore, if furthermore $R_n(\hat{\theta}_n) = \mathrm{o}\big(e^{\frac{n\,c}{\alpha}}\big)$ then $\Pi_n(A_\varepsilon) \to 0$ almost surely as $n \to \infty$.

Proof. It follows from the definition of $\hat{\theta}_n$ that

$$\Pi_n(A_\varepsilon) = \frac{\int_{A_\varepsilon} R_n(\theta)\,\Pi(d\theta)}{\int_\Theta R_n(\theta)\,\Pi(d\theta)} \leq R_n(\hat{\theta}_n)^\alpha\, \frac{\int_{A_\varepsilon} R_n(\theta)^{1-\alpha}\,\Pi(d\theta)}{\int_\Theta R_n(\theta)\,\Pi(d\theta)}.$$

In the proof of Theorem 2 we have obtained that for any $c < \frac{\alpha\,\varepsilon^2}{2}$, the inequality

$$\int_{A_\varepsilon} R_n(f)^{1-\alpha}\,\Pi(df) \leq e^{-n\,c}$$

holds almost surely for all n large enough. On the other hand, by Lemma 4 of Barron et al. (1999) we have that, for any given $\delta > 0$, $\int_\Theta R_n(\theta)\,\Pi(d\theta)$ is bounded below by $e^{-n\,\delta}$ almost surely for large n. Therefore, we have

$$\Pi_n(A_\varepsilon) \leq e^{n\,(\delta - c)}\, R_n(\hat{\theta}_n)^\alpha$$

almost surely for large n. By the arbitrariness of $\delta > 0$ and $c < \frac{\alpha\,\varepsilon^2}{2}$ we have obtained the required inequality for all n large enough, and the proof of Theorem 5 is complete. □

An application of Theorem 1 yields an analogue of Theorem 5 by means of the likelihood functions.

Theorem 6. *Let $\varepsilon > 0$. Suppose that μ is a finite measure on $\mathbb{X}$ and that there exists a constant b such that $b + \log\mu(\mathbb{X}) < \varepsilon^2$. If f_0 is in the Kullback-Leibler support of Π and $L_n(\hat{\theta}_n) = \mathrm{O}\big(e^{n\,b}\big)$ for all large n, then $\Pi_n(A_\varepsilon) \to 0$ almost surely as $n \to \infty$.*

Proof. Take two positive constants c and α such that $b + \log\mu(\mathbb{X}) < c\,\varepsilon^2 < \varepsilon^2$ and $c < 1 - \frac{1}{\alpha}$. Hence we have

$$\begin{aligned} E\Big(\int_{A_\varepsilon} R_n(\theta)^\alpha\;\Pi(d\theta)\Big)^{\frac{1}{\alpha}} &= \int_{\mathbb{X}^n}\Big(\int_{A_\varepsilon} L_n(\theta)^\alpha\;\Pi(d\theta)\Big)^{\frac{1}{\alpha}}\mu(dx_1)\ldots\mu(dx_n) \\ &= \mathrm{O}\big(e^{n\,b}\,\mu(\mathbb{X})^n\big) = \mathrm{O}\big(e^{n\,b + n\,\log\mu(\mathbb{X})}\big) \\ &= \mathrm{O}\big(e^{n\,c\,\varepsilon^2}\big) \qquad \text{as } n \to \infty, \end{aligned}$$

which together with Theorem 1 concludes the proof of Theorem 6. □

5 Discussion

We have used power functions like R^α, where the base R is the likelihood ratio and the exponent α is a fixed constant, to deal with consistency of posterior distributions. Since the size of the likelihood ratio plays a crucial role in determining Bayesian consistency, it is important to find out asymptotic properties of the likelihood ratio as the sample size increases to infinity. The power function is just a special one which is used to describe the size of the likelihood ratio. A natural extension is to use some suitable functions $g(R)$ increasing to infinity instead of R^α.

The uniform convergence of the $\widetilde{Q}_n^\alpha$ as $n \to \infty$ is equivalent to convergence of the posterior distributions Π_n. Hence, it is worth to understand how well the $\widetilde{Q}_n^\alpha$ can approximate Π_n as α increases to one.

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

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

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

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

[8] Walker, S.G. and Hjort, N.L. (2001). On Bayesian consistency. *J. R. Statist. Soc. Ser. B* **63**, 811-821.

[9] Walker, S., Lijoi, A. and Prunster, I. (2005). Data tracking and the understanding of Bayesian consistency. *Biometrika,* **92**, 4, 765-788.

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