Bayesian Analysis (2015) **10**, Number 4, pp. 759–789

On Posterior Concentration in Misspecified Models

R. V. Ramamoorthi[*] Karthik Sriram[†], and Ryan Martin[‡]

Abstract. We investigate the asymptotic behavior of Bayesian posterior distributions under independent and identically distributed (*i.i.d.*) misspecified models. More specifically, we study the concentration of the posterior distribution on neighborhoods of $f^\star$, the density that is closest in the Kullback–Leibler sense to the true model f_0. We note, through examples, the need for assumptions beyond the usual Kullback–Leibler support assumption. We then investigate consistency with respect to a general metric under three assumptions, each based on a notion of divergence measure, and then apply these to a weighted L_1-metric in convex models and non-convex models.

Although a few results on this topic are available, we believe that these are somewhat inaccessible due, in part, to the technicalities and the subtle differences compared to the more familiar well-specified model case. One of our goals is to make some of the available results, especially that of Kleijn and van der Vaart (2006), more accessible. Unlike their paper, our approach does not require construction of test sequences. We also discuss a preliminary extension of the *i.i.d.* results to the independent but not identically distributed (*i.n.i.d.*) case.

MSC 2010: Primary 62C10; secondary 62C10.

Keywords: Bayesian, consistency, misspecified, Kullback–Leibler.

1 Introduction

Let $\mathbb{F}_0$ be a family of densities with respect to a σ-finite measure on a measure space. The object of study is the posterior distribution arising out of the model which consists of a prior distribution Π on $\mathbb{F}_0$ and for any given $f \in \mathbb{F}_0$, $Y_{1:n} = (Y_1, Y_2, \ldots, Y_n)$ are independent and identically distributed (*i.i.d.*) as f. We investigate the behavior of the posterior distribution when the "true" model f_0 is not necessarily in $\mathbb{F}_0$. The posterior is typically expected to concentrate around a density $f^\star$ in $\mathbb{F}_0$ that minimizes the Kullback–Leibler divergence from f_0.

An early investigation of this set up goes back to Berk (1966). An extensive study of parametric model appears in Bunke and Milhaud (1998). Lee and MacEachern (2011) investigate concentration of the posterior and its behavior in testing problems when the prior is on an exponential model. The infinite-dimensional nonparametric case has been studied by Kleijn and van der Vaart (2006) and De Blasi and Walker (2013) for the *i.i.d.* case, and Shalizi (2009) for the non-*i.i.d.* case.

[*]Michigan State University, East Lansing, Michigan, USA, ramamoor@stt.msu.edu
[†]Indian Institute of Management Ahmedabad, India, karthiks@iimahd.ernet.in
[‡]University of Illinois at Chicago, USA, rgmartin@uic.edu

 DOI: 10.1214/15-BA941

For the nonparametric $i.i.d.$ case studied in this note, Kleijn and van der Vaart (2006) is the basic paper. The standard approach to this problem is to first identify sets whose posterior probability goes to 0 and then relate these to the topology of interest. In their paper, Kleijn and van der Vaart develop both these aspects together and, in addition to consistency, also develop rates. De Blasi and Walker (2013) take a somewhat different route towards providing sufficient conditions, specifically for Hellinger-consistency.

Our starting point is Kleijn and van der Vaart (2006). To help motivate our approach, we first summarize key steps in their work. Let E_0 denote expectation with respect to f_0.

- They start with the Kullback–Leibler support assumption on the prior Π, i.e.,

$$\Pi\left(f : \mathsf{E}_0 \log \frac{f^\star}{f} < \varepsilon\right) > 0, \text{ for any } \varepsilon > 0. \tag{1}$$

- They cover the sets $S_j = \{f : \ j\varepsilon \leq d(f, f^\star) < (j+1)\varepsilon\}$, for $j \geq 1$, by convex sets A satisfying:

$$\sup_{f\in A} \inf_{0\leq\alpha\leq 1} \mathsf{E}_0 \left(\frac{f}{f^\star}\right)^\alpha < e^{-j^2\varepsilon^2/4}. \tag{2}$$

- Then they show that, if the assumption in (1) is satisfied, posterior probability of sets satisfying (2) goes to 0 by constructing a sequence of exponential tests for a testing problem that involves non-probability measures. Then, based on the number of sets satisfying (2) required to cover S_j, they develop a notion of entropy for testing problems of a set S_j. When such entropy can be controlled suitably, it is shown that posterior probability of $\{d(f, f^\star) \geq \varepsilon\} = \cup_{j\geq 1} S_j$ goes to 0.

In this paper, we first provide a simple proof to show that, if the assumption in (1) is satisfied, then probability of sets satisfying (2) goes to 0. Our proof does not involve testing problems. We further observe that for a given convex set, the condition in (2) is, in fact, equivalent to a simpler condition based on Kullback–Leibler divergence.

Consistency is related to the topology on the space of densities, usually the weak topology or the Hellinger-metric topology. Towards this, we give two examples in the appendix that point out the need for additional assumptions beyond requiring that $f^\star$ be in the topological support or Kullback–Leibler support of Π. In order to gain insight, we first study consistency with respect to a general metric under a set of three assumptions, each based on a notion of divergence. The first assumption is based on Kullback–Leibler divergence, the second is based on (2), and the third is based on a relatively simpler notion. We show that for a weighted L_1-metric, such assumptions hold in convex models or when the specified family is compact. The first assumption mainly works for compact and convex families. The second assumption along with an appropriate metric entropy condition gives consistency for convex families. As a consequence, we derive a consistency result (Theorem 4), which is analogous to Kleijn and van der Vaart (2006). The third assumption is useful for non-convex (e.g., parametric) models. In this case, we circumvent the convexity requirement by making a continuity assumption on the likelihood ratio and show posterior consistency under an appropriate prior-summability or metric-entropy condition. As a particular consequence, Theorem 5 gives Hellinger consistency analogous to De Blasi and Walker (2013).

We believe that our methods are simple and transparent, and provide useful insights on the requirements for the metric, while also making some of the results in Kleijn and van der Vaart (2006) more accessible. As another small difference, we note that our consistency results are presented in the '*almost sure*' sense, as compared to convergence of means. We also look at one immediate extension of the *i.i.d.* results to independent but not identically distributed (*i.n.i.d.*) models. We note that our study in the *i.n.i.d.* case is preliminary and is presented as an initial approach.

The remainder of the paper is organized as follows. Section 2 sets our notation and provides some basic results. Section 3 presents the consistency results for a general metric. Section 4 presents some important results specific to L_1 and weak topologies. Section 5 contains examples to demonstrate the application of our results. Finally, Section 6 discusses an extension of the *i.i.d.* results to the *i.n.i.d.* case. In the interest of flow, supporting results and details of some proofs are included in the appendix.

2 Notations and preliminary results

2.1 Notations

Let $Y_{1:n} = (Y_1, Y_2, \ldots, Y_n)$ be an *i.i.d.* sample from an unknown "true" density f_0 with respect to a σ-finite measure μ on a measure space $(\mathbb{Y}, \mathscr{Y})$. $\mathbb{F}_0$ is a family of density functions specified to model $Y_{1:n}$. f_0 is not necessarily in $\mathbb{F}_0$. Let Π be a prior on $\mathbb{F}_0$. We let P_0 and E_0 denote probability and expectation with respect to f_0. When talking about joint distribution of finite or infinite *i.i.d.* sequences with respect to P_0, we will omit the superscript in P_0^n or P_0^∞.

It is well known that the posterior typically concentrates around a density that minimizes the Kullback–Leibler divergence from f_0, given by

$$K(f_0, f) := \mathsf{E}_0 \log \frac{f_0}{f} = \int \log \frac{f_0}{f} f_0 d\mu.$$

Accordingly, we assume that there is a fixed unique $f^\star \in \mathbb{F}_0$ such that

$$K(f_0, f^\star) = \inf_{f \in \mathbb{F}_0} K(f_0, f).$$

For any density f, we define

$$K^\star(f_0, f) := K(f_0, f) - K(f_0, f^\star).$$

We assume throughout that $\int (f/f^\star) f_0 \, d\mu < \infty$ for all $f \in \mathbb{F}_0$, and also $\int (f_0/f^\star) \, d\mu < \infty$. The latter condition is useful since we will later (Section 4) consider weak and L_1 topologies with respect to the measure μ_0, where $d\mu_0 = (f_0/f^\star) \, d\mu$. The weighted L_1

metric, $L_1(\mu_0)$, appears to be a natural choice in misspecified models, as opposed to the usual $L_1(\mu)$ metric.

Let $\langle \mathbb{F}_0 \rangle$ be the smallest convex set containing $\mathbb{F}_0$. In this note, a convex set is one that is closed under mixtures. That is, a general subset $A \subseteq \langle \mathbb{F}_0 \rangle$ is called convex if, for any probability measure ν on A, the mixture $\hat{f}_\nu := \int_A f \, \nu(df)$ belongs to A. It is convenient to extend Π to $\langle \mathbb{F}_0 \rangle$ by defining $\Pi(A) := \Pi(A \cap \mathbb{F}_0)$, for any measurable subset A of $\langle \mathbb{F}_0 \rangle$. Note that we do not assume that $f^\star$ minimizes the Kullback–Leibler divergence in $\langle \mathbb{F}_0 \rangle$. In addition, we define:

$$
\begin{aligned}
h^\star_\alpha(f_0, f) &:= \mathsf{E}_0 \left(f/f^\star \right)^\alpha = \int \left(f/f^\star \right)^\alpha f_0 d\mu, \\
f^{(n)} &:= f^{(n)}(y_1, y_2, \ldots, y_n) := \prod_{i=1}^n f(y_i), \\
\hat{f}^{(n)}_\nu &:= \int f^{(n)} d\nu(f), \text{ where } \nu \text{ is a probability measure on } \langle \mathbb{F}_0 \rangle, \\
\hat{f}^{(n)}_A &:= \hat{f}^{(n)}_{\Pi_A}, \text{ where } \Pi_A(\cdot) := \Pi(A \ \cap \ \cdot \ )/\Pi(A), \\
h^\star_\alpha(f^{(n)}_0, f^{(n)}) &:= \mathsf{E}_0 \left(f^{(n)}/f^{\star(n)} \right)^\alpha = \int \left(f^{(n)}/f^{\star(n)} \right)^\alpha f^{(n)}_0 d\mu.
\end{aligned}
$$

Finally, we write down the formula for the posterior distribution as

$$
\Pi(A|Y_{1:n}) = \Pi(A) \, \frac{\hat{f}^{(n)}_A(Y_1, \ldots, Y_n)/f^{\star(n)}(Y_1, \ldots, Y_n)}{\hat{f}^{(n)}_\Pi(Y_1, \ldots, Y_n)/f^{\star(n)}(Y_1, \ldots, Y_n)}. \tag{3}
$$

2.2 Preliminary results

We start with

Assumption 1. $\forall \, \varepsilon > 0, \, \Pi \left(f : \ K^\star(f_0, f) < \varepsilon \right) > 0.$

The following proposition, which helps handle the denominator of (3), is the main consequence of Assumption 1.

Proposition 1. *If Assumption 1 holds, then, for any $\beta > 0$,*

$$
\liminf_{n \to \infty} e^{n\beta} \cdot \hat{f}^{(n)}_\Pi(Y_1, \ldots, Y_n)/f^{\star(n)}(Y_1, \ldots, Y_n) = \infty \qquad \mathsf{P}_0\text{-}a.s.
$$

The proof of the proposition is along the lines of Lemma 4.4.1 in Ghosh and Ramamoorthi (2003).

In view of Proposition 1, $\Pi(A|Y_{1:n}) \to 0$ P_0-a.s., if it can be ensured that for some $\beta_0 > 0$,

$$
e^{n\beta_0} \cdot \Pi(A) \cdot \frac{\hat{f}^{(n)}_A(Y_{1:n})}{f^{\star(n)}(Y_{1:n})} \to 0 \qquad \mathsf{P}_0\text{-a.s.} \ldots \tag{4}
$$

Towards handling (4), we work with three notions of divergence of f from $f^\star$:

(i) $K^\star(f_0, f)$, based on Kullback–Leibler divergence,

(ii) $(1 - \inf_{0\le\alpha\le1} h^\star_\alpha(f_0, f))$, based on Kleijn and van der Vaart (2006) and

(iii) $(1 - h^\star_{\alpha_0}(f_0, f))$ for some $0 < \alpha_0 < 1$, a notion relatively simpler than (ii).

The proposition below describes the relationship between these. The second condition in the proposition was introduced by Kleijn and van der Vaart (2006).

Proposition 2. *Consider the following three conditions for a subset A:*

(i) For some $\epsilon > 0$, $\inf_{f\in A} K^\star(f_0, f) > \epsilon$.

(ii) For some $\delta > 0$, $\sup_{f\in A} \inf_{0\le\alpha\le1} h^\star_\alpha(f_0, f) < e^{-\delta}$.

(iii) For some $0 < \alpha_0 < 1$ and $\eta > 0$, $\sup_{f\in A} h^\star_{\alpha_0}(f_0, f) < e^{-\eta}$.

For any set A, $(iii) \Rightarrow (ii) \Rightarrow (i)$. Further, if the set A is convex, then they are all equivalent.

The proof of Proposition 2 uses the minimax theorem and is provided in Appendix A. The easy proposition below plays a central role.

Proposition 3. *Suppose $A \subset \langle \mathbb{F}_0 \rangle$ is convex. If for some $0 < \alpha < 1$ and $\delta > 0$,*

$$\sup_{f\in A} h^\star_\alpha(f_0, f) \le e^{-\delta},$$

then for any probability measure ν on A,

$$h^\star_\alpha(f_0^{(n)}, \hat{f}_\nu^{(n)}) \le e^{-n\delta}.$$

Proof. The result follows by the use of convexity and induction. Here is an outline. When $n = 1$, the claim holds by convexity of A.
When $n = 2$, $f_\nu^{(2)}(y_1, y_2)$ is the marginal density of Y_1, Y_2 under the model: $Y_1, Y_2 | f \overset{iid}{\sim} f$ and $f \sim \nu(\cdot)$. Write

$$\left[\frac{f_\nu^{(2)}(y_1, y_2)}{f^{\star(2)}(y_1, y_2)}\right]^\alpha = \left[\frac{f_\nu(y_1|y_2)}{f^\star(y_1)}\right]^\alpha \left[\frac{f_\nu(y_2)}{f^\star(y_2)}\right]^\alpha$$

where the first term inside the brackets on the right-hand side, $f_\nu(y_1|y_2)$, is the conditional density of Y_1 given Y_2, and the second term, $f_\nu(y_2)$, is the marginal density of Y_2, obtained from the joint density $f_\nu^{(2)}(y_1, y_2)$. By convexity of A, for all y_2, $f_\nu(\cdot|y_2) \in A$. Hence, we have

$$\mathsf{E}_0\left[\left(\frac{f_\nu(y_1|y_2)}{f^\star(y_1)}\right)^\alpha \Big| y_2\right] \le\; e^{-\delta}.$$

Since $f_\nu(y_2) \in A$, $\mathsf{E}_0[\frac{f_\nu(y_2)}{f^\star(y_2)}]^\alpha \leq e^{-\delta}$. Therefore,

$$\mathsf{E}_0\left[\frac{f_\nu(y_1,y_2)}{f^\star(y_1,y_2)}\right]^\alpha \leq e^{-2\delta}.$$

A similar induction argument for general n completes the proof. □

Theorem 1. *Suppose Assumption 1 holds. If $A \subset \langle \mathbb{F}_0 \rangle$ is convex and satisfies (i) (or equivalently, (ii) or (iii)) of Proposition 2, then*

$$\Pi(A|Y_{1:n}) \to 0 \qquad \mathsf{P}_0\text{-}a.s.$$

Proof. Suppose $\sup_{f\in A} h^\star_\alpha(f_0,f) = \sup_{f\in A} \mathsf{E}_0(f/f^\star)^\alpha < e^{-\delta}$ for some $0<\alpha<1$ and $\delta>0$. Then, by Proposition 3, $h^\star_\alpha(f_0^{(n)}, \hat{f}_A^{(n)}) \leq e^{-n\delta}$. Let $2\beta_0 < \delta$. Then

$$\begin{aligned}
&\mathsf{P}_0\left(\int_A \prod_{i=1}^n \frac{f(Y_i)}{f^\star(Y_i)} d\Pi(f) > e^{-2n\beta_0}\right)\\
&= \mathsf{P}_0\left(\left(\int_A \prod_{i=1}^n \Pi(A)\frac{f(Y_i)}{f^\star(Y_i)} d\Pi_A(f)\right)^\alpha > e^{-2n\alpha\beta_0}\right)\\
&\leq \Pi^\alpha(A)\cdot h^\star_\alpha(f_0^{(n)}, f_A^{(n)})\cdot e^{2n\beta_0\alpha}
\end{aligned}$$

Hence

$$\mathsf{P}_0\left\{\int_A \prod_{i=1}^n \frac{f(Y_i)}{f^\star(Y_i)} d\Pi(f) > e^{-2n\beta_0}\right\} \leq e^{-n(\delta-2\beta_0)}.$$

Since the expression on the right-hand side is summable, we observe by using Borel–Cantelli lemma that (4) is satisfied. This observation, along with Proposition 1, gives the result. □

3 Consistency with general metric

Consistency requires the posterior to concentrate on neighborhoods of $f^\star$ with respect to some metric d. In developing conditions for consistency with respect to d, we encounter a few issues.

First, a necessary condition is that $f^\star$ be in the topological support of Π with respect to this metric. Assumption 1 by itself does not ensure this. We present two examples in Appendix B to illustrate this and point out the need for stronger assumptions. The first example shows that while the presence of $f^\star$ in the L_1 support of Π is necessary for consistency, this is not automatically guaranteed by the positivity of Kullback–Leibler neighborhoods specified in Assumption 1. The next example demonstrates that the presence of $f^\star$ in the L_1 support and Assumption 1 by themselves are not enough to ensure consistency.

Second, since the complement of a d-neighborhood is not convex in general, the equivalence in Proposition 2 is inapplicable. One approach is to suitably cover the complement by d-balls. This in turn requires that each ball satisfy one of the three conditions in Proposition 2. Motivated by these, we investigate consequences of each of the following set of assumptions.

Assumption 2a. Every neighborhood $U = \{f \in \langle \mathbb{F}_0 \rangle : d(f, f^\star) < \epsilon\}$ contains a set of the form $\{f \in \langle \mathbb{F}_0 \rangle :\ K^\star(f_0, f) < \delta\}$ for some $\delta > 0$.

Assumption 2b. Every neighborhood $U = \{f \in \langle \mathbb{F}_0 \rangle : d(f, f^\star) < \epsilon\}$ contains a set of the form $\{f \in \langle \mathbb{F}_0 \rangle : \inf_{0 \le \alpha \le 1} h^\star_\alpha(f_0, f) > e^{-\delta}\}$ for some $\delta > 0$.

Assumption 2c. Every neighborhood $U = \{f \in \mathbb{F}_0 : d(f, f^\star) < \epsilon\}$ contains a set of the form $\{f \in \mathbb{F}_0 : h^\star_{\alpha_0}(f_0, f) > e^{-\delta}\}$ for some $0 < \alpha_0 < 1$ and $\delta > 0$.

We note that Assumptions 2a and 2b are stated in terms of the convexification $\langle \mathbb{F}_0 \rangle$ of $\mathbb{F}_0$, whereas Assumption 2c is stated in terms of $\mathbb{F}_0$. The presence of $\langle \mathbb{F}_0 \rangle$ makes it hard to verify the first two assumptions in non-convex models. However, we study the consequences of these assumptions in a general metric space because of the insight it provides into the requirements on the metric for consistency and shows the usefulness of Assumption 2c in non-convex models. In the next section, we discuss sufficient conditions for Assumptions 2a, 2b, and 2c, with respect to L_1 and weak topologies.

For the rest of this section, we study posterior consistency based on each of these assumptions. We find that Assumption 2a is mainly useful when $\mathbb{F}_0$ is convex and compact, Assumption 2b is useful when $\mathbb{F}_0$ is convex but may not be compact, and Assumption 2c is useful when the family is neither convex nor compact.

The following theorem based on Assumption 2a is an easy consequence of Theorem 1.

Theorem 2. *Let d be a metric such that d-balls in $\langle \mathbb{F}_0 \rangle$ are convex sets and $\mathbb{F}_0$ is compact with respect to d. Let $U = \{f \in \langle \mathbb{F}_0 \rangle : d(f, f^\star) < \epsilon\}$. Suppose Assumptions 1 and 2a hold. Then*

$$\Pi(U^c | Y_{1:n}) \to 1 \qquad \mathsf{P}_0\text{-}a.s.$$

Proof. By Assumption 2a, let $\{f \in \langle \mathbb{F}_0 \rangle : K^\star(f_0, f) < \delta\} \subset \{f \in \langle \mathbb{F}_0 \rangle : d(f, f^\star) < \epsilon/2\}$. Since $U^c \cap \mathbb{F}_0$ is compact, it can be covered by $B_1, B_2, \ldots, B_k$ all contained in $\langle \mathbb{F}_0 \rangle$ with $B_i = \{f \in \langle \mathbb{F}_0 \rangle : d(f, f_i) < \epsilon/3\}$ for some $f_1, f_2, \ldots, f_k \in \mathbb{F}_0$. Each of these balls is convex and disjoint from $\{f \in \langle \mathbb{F}_0 \rangle : d(f, f^\star) < \epsilon/2\}$. Assumption 2a ensures that each B_i satisfies property (i) of Proposition 2. Since there are finitely many such sets, the result follows using Theorem 1. □

In the proof of Proposition 2 provided in Appendix A, it is clear that the choice of α_0 and η made while establishing equivalence of conditions depends on the specific set A. Hence, it does not appear that the approach based on Assumption 2a can be carried easily beyond convex and compact families. Below, we take an approach based on Assumption 2b, which is more in line with Kleijn and van der Vaart (2006). Theorem 1 derives posterior consistency for convex sets. Since complement of a d-neighborhoods will not be convex in general, the approach here is to cover with a suitable number of convex sets with diminishing posterior probabilities. Towards this end, we make the following assumption.

Assumption 3. There exist subsets $\{V_n, W_n\}_{n\geq 1}$ such that $\mathbb{F}_0 \subseteq V_n \cup W_n$ and

(a) $\Pi(W_n) < e^{-n\Delta}$ for some $\Delta > 2\mathsf{E}_0 \log \frac{f_0}{f_*}$.

(b) For every $\epsilon > 0$, V_n can be covered by J_n d-balls in $\langle \mathbb{F}_0 \rangle$ of radius less than ϵ, where J_n is a polynomial in n, i.e., for some $r > 0$, $J_n \leq an^r$.

The simple lemma below will be useful to derive the results that follow.

Lemma 1. *Let $T_i, i = 1, 2, \ldots, k$ be non-negative random variables. Then*

$$\mathsf{P}\left(\sum_{i=1}^{k} T_i > e^{-\epsilon}\right) \leq e^{\epsilon} \sum_{i=1}^{k} \inf_{0\leq\alpha\leq 1} \mathsf{E} T_i^{\alpha}.$$

Proof. The result follows since, $\mathsf{P}(\sum_{i=1}^{k} T_i > e^{-\epsilon}) = \mathsf{P}(\sum_{i=1}^{k} \min(T_i, 1) > e^{-\epsilon})$ and $\min(T_i, 1) \leq T_i^{\alpha_i}$ for any $0 < \alpha_i < 1$. Consequently,

$$\mathsf{P}\left(\sum_{i=1}^{k} T_i > e^{-\epsilon}\right) \leq \mathsf{P}\left(\sum_{i=1}^{k} T_i^{\alpha_i} > e^{-\epsilon}\right) \leq e^{\epsilon} \sum_{i=1}^{k} \mathsf{E} T_i^{\alpha_i}.$$

Taking the infimum over $\alpha_1, \alpha_2, \ldots, \alpha_k$ on both sides, we get the result. □

We now derive posterior consistency under Assumptions 2b and 3, followed by a result that is analogous to the posterior consistency result in Kleijn and van der Vaart (2006).

Theorem 3. *Let metric d be such that d-balls in $\langle \mathbb{F}_0 \rangle$ are convex sets and let $U_\epsilon = \{f \in \langle \mathbb{F}_0 \rangle : d(f, f^\star) < \epsilon\}$. If Assumptions 1, 2b and 3 hold, then*

$$\Pi(U_\epsilon^c | Y_{1:n}) \to 0 \qquad \mathsf{P}_0\text{-}a.s.$$

Proof. Let $U_{\epsilon/2} = \{f \in \langle \mathbb{F}_0 \rangle : d(f^\star, f) < \epsilon/2\}$, and as guaranteed by Assumption 2b, let

$$\{f \in \langle \mathbb{F}_0 \rangle : \inf_{0\leq\alpha\leq 1} h_\alpha^\star(f_0, f) > e^{-\delta}\} \subset U_{\epsilon/2}.$$

Let $A_1, A_2, \ldots, A_{J_n}$ be open d-balls of radius $\epsilon/3$ that cover $U_\epsilon^c \cap V_n$. Then,

$$\begin{aligned}\mathsf{P}_0\left(\int_{U_\epsilon^c \cap V_n} f^{(n)}/f^{\star(n)} d\Pi > e^{-n\beta}\right) &\leq \mathsf{P}_0\left(\sum_{i=1}^{J_n} \int_{A_i} f^{(n)}/f^{\star(n)} d\Pi > e^{-n\beta}\right) \\ &\leq \mathsf{P}_0\left(\sum_{i=1}^{J_n} \int_{A_i} f^{(n)}/f^{\star(n)} d\Pi_i > e^{-n\beta}\right),\end{aligned}$$

where Π_i is Π normalized to 1 on A_i. Set $T_i = \int_{A_i} f^{(n)}/f^{\star(n)} d\Pi_i$ so that the expression on the right-hand side of the last inequality can be written as:

$$\mathsf{P}_0\left(\sum_{i=1}^{J_n} T_i > e^{-n\beta}\right).$$

By Lemma 1,

$$\mathsf{P}_0\left(\sum_{i=1}^{J_n} T_i > e^{-n\beta}\right) \leq e^{n\beta} \sum_i \inf_{0\leq\alpha\leq 1} \mathsf{E}_0 T_i^\alpha.$$

Since A_i does not intersect $U_{\epsilon/2}$, $\inf_{0\leq\alpha\leq 1} \mathsf{E}_0 (f/f^\star)^\alpha < e^{-\delta}$. By Proposition 3, for each i, $\inf_{0\leq\alpha\leq 1} \mathsf{E} T_i^\alpha < e^{-n\delta}$, so that

$$\mathsf{P}_0\left(\int_{U_\epsilon^c \cap V_n} f^{(n)}/f^{\star(n)} d\Pi > e^{-n\beta}\right) < e^{n\beta} n^r e^{-n\delta}.$$

A choice of small enough β and an application of Borel–Cantelli lemma with $\beta_0 < \beta$ gives

$$e^{n\beta_0} \int_{U_\epsilon^c \cap V_n} f^{(n)}/f^{\star(n)} d\Pi \to 0 \qquad \mathsf{P}_0\text{-a.s.}$$

This, along with Proposition 1, gives $\Pi(U_\epsilon^c \cap V_n | Y_{1:n}) \to 0 \quad \mathsf{P}_0$-a.s.

As for W_n, first an argument in the lines of Lemma 4.4.1 of Ghosh and Ramamoorthi (2003) can be used to conclude that for any $\beta > \mathsf{E}_0 \log \frac{f_0}{f^\star}$,

$$\liminf_{n\to\infty} e^{n\beta} \int_{\langle \mathbb{F}_0 \rangle} f^{(n)}/f_0^{(n)} d\Pi = \infty \qquad \mathsf{P}_0\text{-a.s.} \tag{5}$$

Then, for $\Delta > 2\mathsf{E}_0 \log \frac{f_0}{f^\star}$, an application of Markov's inequality gives

$$\begin{aligned} &\mathsf{P}_0\left(\int_{W_n} f^{(n)}/f_0^{(n)} d\Pi > e^{-n\frac{\Delta}{2}}\right) \\ &\leq e^{n\frac{\Delta}{2}} \cdot \int_{W_n} \mathsf{E}_0\left(f^{(n)}/f_0^{(n)}\right) d\Pi = e^{n\frac{\Delta}{2}} \Pi(W_n) \leq e^{-n\frac{\Delta}{2}}. \end{aligned} \tag{6}$$

Equations (5) and (6) together imply that $\Pi(W_n | Y_{1:n}) \to 0 \quad \mathsf{P}_0$-a.s. □

The approach taken in Theorem 3 can be adapted to derive a result that is analogous to Corollary 2.1 of Kleijn and van der Vaart (2006). The entropy condition in their paper assumes that each set $S_j = \{f \in \mathbb{F}_0 : j\epsilon \leq d(f, f^\star) < (j+1)\epsilon\}$ can be covered by N_j convex sets B_k with the property $\sup_{f\in B_k} \inf_{0\leq\alpha\leq 1} h_\alpha^\star(f_0, f^\star) < e^{-j^2\epsilon^2/4}$. In our approach, this corresponds to a stronger version of Assumption 2b as stated in the theorem below. If $\sup_{j\geq 1} N_j < \infty$, then they show that $\mathsf{E}_0[\Pi(d(f, f^\star) \geq \epsilon | Y_{1:n})] \to 0$. An analogous result using our approach is as follows.

Theorem 4. *Let metric d be such that d-balls in $\langle \mathbb{F}_0 \rangle$ are convex sets. Suppose Assumption 1 and a stronger version of Assumption 2b (where δ depends on ϵ) hold, i.e.,*

> *For any $\epsilon > 0$, $U_\epsilon = \{f \in \langle \mathbb{F}_0 \rangle : d(f, f^\star) < \epsilon\}$ contains a set of the form $\{f \in \langle \mathbb{F}_0 \rangle : \inf_{0 \le \alpha \le 1} h^\star_\alpha(f_0, f) > e^{-\epsilon^2}\}$.*

Let N_j be the minimum number of d-balls in $\langle \mathbb{F}_0 \rangle$ of radius $j\epsilon/3$ that cover the set $S_j = \{f \in \mathbb{F}_0 : j\epsilon \le d(f, f^\star) < (j+1)\epsilon\}$. If $\sup_{j \ge 1} N_j < \infty$, then

$$\Pi(d(f, f^\star) > \epsilon | Y_{1:n}) \to 0 \qquad \mathsf{P}_0\text{-}a.s.$$

Proof. Along the lines of Lemma 1, we get

$$\mathsf{P}_0 \left(\int_{U_\epsilon^c} f^{(n)}/f^{\star(n)} d\Pi > e^{-n\beta} \right) \le e^{n\beta} \cdot \sum_{j=1}^{\infty} \inf_{0 \le \alpha \le 1} \mathsf{E}_0 \left(\int_{S_j} f^{(n)}/f^{\star(n)} d\Pi_j \right)^\alpha .$$

Note that for any f_1 such that $d(f, f_1) \ge j\epsilon$, $B_{f_1} = \{f : \ d(f, f_1) < j\epsilon/3\}$ does not intersect with $U_{j\epsilon/2}$. Hence $\sup_{f \in B_{f_1}} \inf_{0 \le \alpha \le 1} \mathsf{E}_0 (f/f^\star)^\alpha < e^{-j^2\epsilon^2/4}$.

Using Proposition 3 and the fact that S_j can be covered by N_j convex sets of the form B_{f_1}, we get

$$\mathsf{P}_0 \left(\int_{U_\epsilon^c} \frac{f^{(n)}}{f^{\star(n)}} d\Pi > e^{-n\beta} \right) \le e^{n\beta} \cdot \sum_{j=1}^{\infty} N_j \cdot e^{-nj^2\epsilon^2/4} \le e^{n\beta} \cdot \sup_{j \ge 1} N_j \cdot \frac{e^{-n\epsilon^2/4}}{1 - e^{-n\epsilon^2/4}}.$$

A suitable choice of small enough β and an application of Borel–Cantelli lemma, gives that for $\beta_0 < \beta$, $e^{n\beta_0} \int_{U_\epsilon^c} f^{(n)}/f^{\star(n)} d\Pi \to 0 \quad \mathsf{P}_0$-a.s. This, along with Assumption 1, gives the result. □

As noted earlier, Assumptions 2a and 2b are stated in terms of $\langle \mathbb{F}_0 \rangle$, which makes them difficult to verify for non-convex models. Assumption 2c helps handle the case of non-convex families. We now derive consistency results under Assumption 2c and the following continuity assumption.

> **Assumption 4.** For any $f_1, f_2 \in \mathbb{F}_0$ and for some monotonically increasing function $\eta(\cdot)$ with $\eta(0) = 0$ we have
>
> $$\mathsf{E}_0 \left| \frac{f_1}{f^\star} - \frac{f_2}{f^\star} \right| \le \eta(d(f_1, f_2)).$$

Note that Assumption 4 is satisfied by $d = L_1(\mu_0)$, in which case $\eta(\cdot)$ is just the identity function. If $f_0/f^\star \in L_\infty(\mu)$ then the assumption is satisfied by $L_1(\mu)$. Also if $f_0/f^\star \in L_2(\mu)$, an application of Cauchy–Schwartz inequality shows that it is satisfied for $d = L_2(\mu)$.

Towards deriving consistency, the next lemma shows that if Assumptions 2c and 4 hold then U_ϵ^c can be covered by d-balls whose posterior probabilities diminish to zero with increasing n.

Lemma 2. *Let $U^c = \{f \in \mathbb{F}_0 : d(f, f^\star) \geq \epsilon\}$. Suppose Assumptions 2c and 4 hold with respect to U and a metric d. Let α_0, δ be as in Assumption 2c and $\eta(\cdot)$ as in Assumption 4. Then, for any $f_1 \in U^c$ there exists an open ball $B(f_1, r)$ around f_1, with the radius r depending only on δ, α_0 and $\eta(\cdot)$, such that*

$$\mathsf{E}_0 \left(\int_{B(f_1,r)} f^{(n)}/f^{\star(n)} d\Pi(f) \right)^{\alpha_0} \leq e^{-n\frac{\delta}{2}} \cdot \Pi\left(B\left(f_1, r\right)\right)^{\alpha_0}.$$

Proof. Let $r = \eta^{-1}((e^{-\frac{\delta}{2}} - e^{-\delta})^{\frac{1}{\alpha_0}})$ and $\nu(\cdot)$ be any probability measure on $\mathbb{F}_0$. Since $0 < \alpha_0 < 1$,

$$\begin{aligned}
&\mathsf{E}_0 \left(\int_{B(f_1,r)} f/f^\star \; d\nu(f) \right)^{\alpha_0} \\
&\leq \mathsf{E}_0 \left(\int_{B(f_1,r)} \left| \frac{f}{f^\star} - \frac{f_1}{f^\star} \right| d\nu(f) \right)^{\alpha_0} + \mathsf{E}_0 \left(\int_{B(f_1,r)} \frac{f_1}{f^\star} d\nu(f) \right)^{\alpha_0} \\
&\text{(then by Jensen's inequality)} \\
&\leq \left(\int_{B(f_1,r)} \mathsf{E}_0 \left| \frac{f}{f^\star} - \frac{f_1}{f^\star} \right| d\nu(f) \right)^{\alpha_0} + \mathsf{E}_0 \left(\frac{f_1}{f^\star} \right)^{\alpha_0} \nu(B\left(f_1, r\right))^{\alpha_0}.
\end{aligned}$$

By Assumption 4, the first term of the above inequality satisfies

$$\left(\int_{B(f_1,r)} \mathsf{E}_0 \left| \frac{f}{f^\star} - \frac{f_1}{f^\star} \right| d\nu(f) \right)^{\alpha_0} \leq (e^{-\frac{\delta}{2}} - e^{-\delta}) \cdot \nu(B\left(f_1, r\right))^{\alpha_0}.$$

By Assumption 2c, the second term of the inequality is bounded as

$$\mathsf{E}_0 \left(\frac{f_1}{f^\star} \right)^{\alpha_0} \cdot \nu(B\left(f_1, r\right))^{\alpha_0} < e^{-\delta} \cdot \nu(B\left(f_1, r\right))^{\alpha_0}.$$

Therefore, it follows that for any probability measure $\nu(\cdot)$ on $\mathbb{F}_0$ we have

$$\mathsf{E}_0 \left(\int_{B(f_1,r)} f/f^\star \; d\nu(f) \right)^{\alpha_0} \leq e^{-\frac{\delta}{2}} \cdot \nu(B\left(f_1, r\right))^{\alpha_0}. \tag{7}$$

Equation (7) is the result for $n = 1$. An induction argument on n along the lines of Proposition 3 can now be used to obtain the result. To see this for $n = 2$, note that, as in the proof of Proposition 3, we can write

$$\left[\frac{f_\nu^{(2)}(y_1, y_2)}{f^{\star(2)}(y_1, y_2)} \right]^{\alpha_0} = \left[\frac{f_\nu(y_1|y_2)}{f^\star(y_1)} \right]^{\alpha_0} \left[\frac{f_\nu(y_2)}{f^\star(y_2)} \right]^{\alpha_0}.$$

Now, by (7), $\mathsf{E}(\frac{f_\nu(y_2)}{f^\star(y_2)})^{\alpha_0} \leq e^{-\frac{\delta}{2}} \cdot \nu(B(f_1, r))^{\alpha_0}$. Further, since (7) holds for any probability measure, taking the measure $\frac{f(y_2)}{f_\nu(y_2)} d\nu$, we get

$$\mathsf{E}_0 \left[\left(\frac{f_\nu(y_1|y_2)}{f^\star(y_1)} \right)^{\alpha_0} \Big| y_2 \right] \leq \; e^{-\frac{\delta}{2}}.$$

It therefore follows that for any probability measure ν on $\mathbb{F}_0$,

$$\mathsf{E}_0\left[\frac{f_\nu^{(2)}(y_1,y_2)}{f^{\star(2)}(y_1,y_2)}\right]^{\alpha_0} \leq\; e^{-2\cdot\frac{\delta}{2}}\cdot\nu(B\,(f_1,r))^{\alpha_0}.$$

A similar induction argument for general n completes the proof. □

To ensure that the total posterior probability of U^c goes to zero, we need to be able to cover it with sets of the form $B(f_1,r)$ that satisfy a prior-summability assumption as in De Blasi and Walker (2013) or metric entropy assumption as in Kleijn and van der Vaart (2006). The following theorem is an immediate consequence of Lemma 2.

Theorem 5. *Let $U^c=\{f\in\mathbb{F}_0 : d(f,f^\star)\geq\epsilon\}$. Suppose Assumptions 1, 2c and 4 hold. Let α_0 be as in Assumption 2c and $\eta(\cdot)$ be as in Assumption 4. Suppose for any given $r>0$, $r_{\alpha_0}=\eta^{-1}(r^{\frac{1}{\alpha_0}})$ and $\{B(f_j,r_{\alpha_0}),\ j\geq 1\}$ is an open cover of U^c such that one of the following two conditions (a) or (b) holds:*

(a) $\sum_j \Pi(B(f_j,r_{\alpha_0}))^{\alpha_0}<\infty$.

(b) Assumption 3 holds.

Then, $\Pi(U^c|Y_{1:n})\to 0\quad \mathsf{P}_0$*-a.s.*

Proof. If condition (a) holds, then since $0<\alpha_0<1$, we get

$$\begin{aligned}\mathsf{P}_0\left(\int_{U^c} f^{(n)}/f^{\star(n)}d\Pi > e^{-n\beta}\right) &\leq e^{n\beta}\sum_{j\geq 1}\mathsf{E}_0\left(\int_{B(f_j,r_{\alpha_0})} f^{(n)}/f^{\star(n)}d\Pi\right)^{\alpha_0}\\ &\leq e^{n\beta}\cdot e^{-n\delta/2}\cdot\sum_{j\geq 1}\Pi(B\,(f_1,r))^{\alpha_0}.\end{aligned}$$

A suitable β and Borel–Cantelli Lemma give that for $\beta_0<\beta$, $e^{n\beta_0}\int_{U^c} f^{(n)}/f^{\star(n)}d\Pi\to 0\ \ \mathsf{P}_0$-a.s. This, along with Proposition 1, gives $\Pi(U^c|Y_{1:n})\to\ 0\ \ \mathsf{P}_0$-a.s.
If condition (b) holds then the proof is along the same lines as for Theorem 3. □

Remark 1. As noted earlier, Assumption 4 automatically holds for $d=L_1(\mu_0)$. In that case, the function $\eta(\cdot)$ is just the identity function, and the result based on condition (a) of Theorem 5 is analogous to Corollary 1 of De Blasi and Walker (2013).

Remark 2. Theorem 5 can be easily applied to *i.i.d.* parametric models, i.e., when $\mathbb{F}_0=\{f_\theta:\ \theta\in\Theta\}$. Let $f^\star=f_{\theta^*}$ for some $\theta^*\in\Theta$ be the minimizer of Kullback–Leibler divergence from f_0. It is easy to see that Assumption 1 is ensured as long as the prior Π assigns a positive probability to every open d-neighborhood of θ^* and $\mathsf{E}_0\log\frac{f_{\theta^*}}{f_\theta}$ is continuous in θ. Further, note that Theorems 7 and 8 in Section 4 provide sufficient conditions for Assumption 2c to hold with respect to $L_1(\mu_0)$. However, to apply the results for the parametric model, we need the assumption to hold with respect to the metric d. This can be easily ensured if in addition to conditions of Theorems 7 or 8, it can be verified that for some monotonically increasing function $\zeta(\cdot)$, with $\zeta(0) = 0$:

$$\int |f_\theta - f_{\theta^*}| d\mu_0 \geq \zeta(d(\theta, \theta^*)), \ \forall \ \theta \in \Theta.$$

Then, by defining the metric on $\mathbb{F}_0$ as $d(f_{\theta_1}, f_{\theta_2}) := d(\theta_1, \theta_2)$, and using Assumption 4, Theorem 5 can be applied. We provide examples of this approach to parametric models in Section 5.

Remark 3. Assumption 4 is a continuity condition on $\mathsf{E}_0(\frac{f}{f^\star})$. It is possible to work with an an alternative condition that assumes continuity of $\mathsf{E}_0(\frac{f}{f_1})$ for any $f_1 \in \mathbb{F}_0$. In particular, if $f_1 \in U^c$, by Assumption 2c, $\mathsf{E}_0(\frac{f_1}{f^\star})^{\alpha_0} < e^{-\delta}$. Then, the conclusion analogous to that of Lemma 2, which is crucial for Theorem 5, can be obtained by defining an open set $B_1 := \{\theta \in \Theta : \ E[\frac{f_1}{f^\star}] < e^{\frac{\delta}{2}}\}$. Then for $\alpha = \frac{\alpha_0}{2}$ and any probability measure $\nu(\cdot)$ on B_1, it can be shown (by using Cauchy–Schwartz and Jensen's inequality) that

$$\begin{aligned}
\mathsf{E}_0\left[\left(\int_{B_1} \frac{f}{f^\star} d\nu\right)^\alpha\right] &= E\left[\left(\frac{f_1}{f^\star}\right)^\alpha \cdot \left(\int_{B_1} \frac{f}{f_1} d\nu\right)^\alpha\right] \\
&\leq \left(\mathsf{E}_0\left[\left(\frac{f_1}{f^\star}\right)^{2\alpha}\right]\right)^{\frac{1}{2}} \cdot \left(\int_{B_1} \mathsf{E}_0\left[\frac{f}{f_1}\right] d\nu\right)^\alpha < e^{-\alpha\frac{\delta}{2}}.
\end{aligned}$$

Posterior consistency can then be derived if U^c can be covered by suitably many sets of the form B_1, e.g., when $\mathbb{F}_0$ is compact. We work with such an assumption in Section 6 (Assumption D) while extending results to the *i.n.i.d.* case.

4 Weak and L_1 consistency

Assumptions 2a, 2b and 2c are crucial for Theorems 2, 3 and 5, respectively. These, we feel, are in general hard to verify. Here, we focus on specific topologies and discuss cases where these assumptions hold.

Recall $d\mu_0 = (f_0/f^\star)\, d\mu$. Our interest is in two topologies on $\langle \mathbb{F}_0 \rangle$. First, the weak topology on $L_1(\mu_0)$ induced by $L_\infty(\mu_0)$. The basic open neighborhoods of $f^\star$ here are finite intersections of sets of the form

$$\left\{g \in L_1(\mu_0) : |\int \varphi_k g d\mu_0 - \int \varphi f^\star d\mu_0| < \epsilon_k, \qquad \varphi_k \in L_\infty(\mu_0)\right\}.$$

We will refer to this as the μ_0-weak topology. The other topology is the L_1 topology which yields neighborhoods of the form $\{g : \int |g - f^\star| d\mu_0 < \epsilon\}$. Of interest are also the usual weak and total variation topologies on densities. These correspond to μ-weak topology and $L_1(\mu)$ topology. In the context of consistency, our interest is in the concentration of the posterior in neighborhoods of $f^\star$. We formally define

Definition 1. The sequence of posterior distributions $\{\Pi(\cdot|Y_1, Y_2, \ldots, Y_n)\}_{n\geq 1}$ is said to be μ_0-weakly consistent if, for any μ_0-weak neighborhood U of $f^\star$,

$$\Pi(U|Y_{1:n}) \to 1 \qquad \mathsf{P}_0\text{-a.s.}$$

We will now show in the theorem below that when $\mathbb{F}_0$ itself is convex, Assumptions 2a, 2b and 2c are ensured with respect to weak and $L_1(\mu_0)$ topologies.

Theorem 6. *If $\mathbb{F}_0$ is convex then Assumptions 2a, 2b and 2c hold both with respect to the $L_1(\mu_0)$ topology and the μ_0-weak topology.*

Proof. First, using the Cauchy–Schwartz inequality, we get

$$\begin{aligned}\int |f^\star - f| d\mu_0 &= \int \left|1 - \frac{f}{f^\star}\right| f_0 d\mu = \int \left|\left(1 - \sqrt{\frac{f}{f^\star}}\right)\right| \cdot \left(1 + \sqrt{\frac{f}{f^\star}}\right) f_0 d\mu \\ &\leq \left(\int \left|1 - \sqrt{\frac{f}{f^\star}}\right|^2 f_0 d\mu\right)^{\frac{1}{2}} \cdot \left(\int \left(1 + \sqrt{\frac{f}{f^\star}}\right)^2 f_0 d\mu\right)^{\frac{1}{2}}.\end{aligned}$$

Since $\mathbb{F}_0$ is convex, by Lemma 2.3 of Kleijn and van der Vaart (2006), $\mathsf{E}_0 \frac{f}{f^\star} \leq 1$. Hence, the second term in the above inequality is bounded because

$$\mathsf{E}_0 \left(1 + \sqrt{\frac{f}{f^\star}}\right)^2 = 2\left(1 + \mathsf{E}_0 \frac{f}{f^\star}\right) \leq 4.$$

Similarly, for the first term,

$$\begin{aligned}\mathsf{E}_0 \left(1 - \sqrt{\frac{f}{f^\star}}\right)^2 &= \mathsf{E}_0 \left(1 - \sqrt{\frac{f}{f^\star}}\right)^2 = 1 + \mathsf{E}_0 \frac{f}{f^\star} - 2\mathsf{E}_0 \sqrt{\frac{f}{f^\star}} \\ &\leq 2\left(1 - \mathsf{E}_0 \sqrt{\frac{f}{f^\star}}\right) = \frac{1 - h^\star_{\frac{1}{2}}(f_0, f)}{\frac{1}{2}} \leq K^*(f_0, f).\end{aligned}$$

Thus using Lemma 5, we get

$$\int |f^\star - f| d\mu_0 \leq 2 \cdot \sqrt{\frac{1 - h^\star_\alpha(f_0, f)}{\alpha}} (\text{with } \alpha = 0.5) \leq 2\sqrt{K^\star(f_0, f)}.$$

The last inequality ensures that Assumption 2a, 2b and 2c hold with respect to $L_1(\mu_0)$. Since every weak neighborhood contains an L_1-neighborhood, assumptions hold with respect to the μ_0-weak topology as well. □

Remark 4 (**μ_0-Weak Consistency**). By Theorem 1, Assumption 2a along with Assumption 1 ensures μ_0-weak consistency. This is because the complement of a weak neighborhood is a finite union of convex sets. Further, by Theorem 6, if $\mathbb{F}_0$ is convex, Assumption 1 is enough to ensure μ_0-weak consistency.

When $\mathbb{F}_0$ is not convex, Assumptions 2a and 2b are not easy to verify. In that case, it may be easier to work with Assumption 2c. Next, we derive two results with sufficient conditions for Assumption 2c to hold with respect to the $L_1(\mu_0)$ metric. The first simpler result below is obtained when $\mathbb{F}_0$ is $L_1(\mu_0)$ compact.

Theorem 7. *If $\mathbb{F}_0$ is $L_1(\mu_0)$-compact, then Assumption 2c holds with respect to $d = L_1(\mu_0)$.*

Proof. Suppose $f_1 \in \mathbb{F}_0$ is such that $d(f_1, f^\star) \geq \epsilon$. Since $f^\star$ is assumed to be unique, $\exists\ \eta > 0$ such that $K^\star(f_0, f_1) > \eta$ and further by Lemma 5, $\exists\ 0 < \alpha < 1$ such that $h^\star_\alpha(f_0, f_1) < 1 - \alpha\eta < e^{-\alpha\eta}$. Here α and η may depend on f_1. Now, let $B_{f_1} = \{f \in \mathbb{F}_0 : d(f, f_1) < r_\alpha\}$, where $r_\alpha \leq (e^{-\alpha\eta/2} - e^{-\alpha\eta})^{\frac{1}{\alpha}}$. For $f \in B_{f_1}$, since $0 < \alpha < 1$,

$$
\begin{aligned}
h^\star_\alpha(f_0, f) &\leq h^\star_\alpha(f_0, f_1) + \int \left| \left(\frac{f}{f^\star}\right)^\alpha - \left(\frac{f_1}{f^\star}\right)^\alpha \right| f_0 d\mu \\
&\leq h^\star_\alpha(f_0, f_1) + \int \left| \frac{f}{f^\star} - \frac{f_1}{f^\star} \right|^\alpha f_0 d\mu \\
&\leq h^\star_\alpha(f_0, f_1) + \left(\int |f - f_1|\, d\mu_0 \right)^\alpha \leq e^{-\delta}, \text{ where } \delta = \alpha\eta/2.
\end{aligned}
$$

The last step uses Jensen's inequality. We have essentially shown that if $d(f_1, f^\star) \geq \epsilon$ there is an open $L_1(\mu_0)$-ball B_{f_1} around f_1 and $\exists\, 0 < \alpha < 1$, $\delta > 0$ such that $h^\star_\alpha(f_0, f) \leq e^{-\delta}$, for all $f \in B_{f_1}$. Then, as noted in proof of Proposition 2 in Appendix A, we would also have $h^\star_{\alpha'}(f_0, f) < e^{-\delta}$ for all $\alpha' < \alpha$. Since $\mathbb{F}_0$ is $L_1(\mu_0)$-compact, $\{f : d(f, f^\star) \geq \epsilon\}$ can be covered by finitely many such balls $B_{f_1}, B_{f_2}, \ldots, B_{f_k}$, thus obtaining an α and δ corresponding to each ball. The result is obtained by choosing the minimum of these finitely many α's and δ's and noting that $\{f \in \mathbb{F}_0 : d(f, f^\star) \geq \epsilon\} \subseteq \cup_{i=1}^k B_{f_i} \subseteq \{f \in \mathbb{F}_0 : h^\star_\alpha(f_0, f^\star) \leq e^{-\delta}\}$. □

The following theorem and the corollary give sufficient conditions for Assumption 2c to hold with respect to $L_1(\mu_0)$, when $\mathbb{F}_0$ is neither convex nor compact.

Theorem 8. *If $\exists\ 0 < \alpha_0 < 1$ such that $\sup_{f \in \mathbb{F}_0} \mathsf{E}_0(\frac{f}{f^\star})^{\alpha_0} \leq 1$ and suppose $\sup_{f \in \mathbb{F}_0} \mathsf{E}_0(\frac{f}{f^\star})^2 < \infty$. Then Assumption 2c holds with respect to $d = L_1(\mu_0)$.*

Proof. Without loss of generality, assume $\alpha_0 = \frac{1}{2^{K-1}}$ for some $K > 1$. Define $a := (\frac{f}{f^\star})^{\frac{1}{2^K}}$. Then

$$
\begin{aligned}
\mathsf{E}_0 \left| \frac{f}{f^\star} - 1 \right| &= \mathsf{E}_0 \left| a^{2^k} - 1 \right| = \mathsf{E}_0 \left[|a - 1| \cdot \left| 1 + a^2 + a^3 + \cdots + a^{2^k - 1} \right| \right] \\
&\leq \mathsf{E}_0 \left(|a - 1|^2 \right)^{\frac{1}{2}} \cdot \left(\mathsf{E}_0 \left| 1 + a^2 + a^3 + \cdots + a^{2^k - 1} \right|^2 \right)^{\frac{1}{2}}.
\end{aligned}
$$

For the first term on the right-hand side of the above inequality, we have

$$\mathsf{E}_0 \left|a-1\right|^2 = \mathsf{E}_0 \left(\frac{f}{f^\star}\right)^{\frac{1}{2^{k-1}}} + 1 - 2\mathsf{E}_0 \left(\frac{f}{f^\star}\right)^{\frac{1}{2^k}} \leq 2\left(1 - \mathsf{E}_0 \left(\frac{f}{f^\star}\right)^{\frac{1}{2^k}}\right).$$

Note that every term within the expansion $\mathsf{E}_0|1+a^2+a^3+\cdots+a^{2^k-1}|^2$ is of the form $\mathsf{E}_0 a^l = \mathsf{E}_0(\frac{f}{f^\star})^{\frac{l}{2^k}}$, where $l \leq 2^{k+1}-2$. In particular, the second term is bounded by some constant multiple of $\sqrt{\sup_{f\in\mathbb{F}_0} \mathsf{E}_0(\frac{f}{f^\star})^2}$. Hence the result follows. □

Corollary 1. *If the log-likelihood ratio* $\log \frac{f}{f^\star}$ *is uniformly bounded, then Assumption 2c holds with respect to* $d = L_1(\mu_0)$*.*

Proof. By uniform boundedness, $\exists\ \alpha$ that does not depend on f such that $\alpha \cdot |\log \frac{f}{f^\star}| < \frac{1}{2}$. We note that when $t < 1, e^t < \frac{1}{1-t}$. In particular, for $t < 1/2$, $e^t - 1 < t/(1-t) < 2t$. Applying this inequality, we get

$$e^{\alpha\cdot\log\frac{f}{f^\star}} - 1 < -2\alpha \cdot \log\frac{f^\star}{f}.$$

Therefore, $\mathsf{E}_0(\frac{f}{f^\star})^\alpha < 1 - 2\alpha\cdot\mathsf{E}_0 \log\frac{f^\star}{f}$. Since $0 \leq 2\alpha\cdot\mathsf{E}_0\log\frac{f^\star}{f} < 1$, we have $\mathsf{E}_0(\frac{f}{f^\star})^\alpha \leq 1$. Clearly, uniform boundedness also ensures that $\mathsf{E}_0(\frac{f}{f^\star})^2$ is uniformly bounded. Hence Theorem 8 implies that Assumption 2c holds. □

Remark 5 ($L_1(\mu_0)$**-consistency**)*.* Clearly, Theorem 3 of the previous section can be applied for $d = L_1(\mu_0)$, along with the sufficient conditions presented in this section for verifying Assumptions 2b or 2c. In particular, we can conclude by Theorems 5 and 7 that, if $\mathbb{F}_0$ is $L_1(\mu_0)$ compact, Assumption 1 is enough to ensure $L_1(\mu_0)$-consistency. This is because Assumption 4 is automatically satisfied by $L_1(\mu_0)$, the entropy condition will hold by compactness and Assumption 2c holds due to compactness by Theorem 7.

5 Examples

5.1 Mixture models

The mixture models discussed in Kleijn and van der Vaart (2006) are covered by our results. In particular, let $y \mapsto f(y|z)$ be a fixed density with respect to μ for each $z \in \mathcal{Z}$, and $f(y,z)$ be jointly measurable. For every probability measure ν on $\mathcal{Z}$ let

$$p_\nu(y) = \int f(y|z)d\nu(z).$$

Let $\mathbf{M}$ be the set of probability measures on $\mathcal{Z}$. Consider the model $\mathbb{F}_0 = \{p_\nu : \nu \in \mathbf{M}\}$. Let f_0 be the "true" distribution and assume that $f^\star \in \mathbb{F}_0$ satisfies $K(f_0, f^\star) = \inf_{\nu\in\mathbf{M}} K(f_0, p_\nu)$. As before set $d\mu_0 = \frac{f_0}{f^\star}d\mu$. Since $\mathbb{F}_0$ is convex by Theorem 6, Assumptions 2a, 2b, and 2c are satisfied. Therefore, as noted in Remark 4, the posterior would be μ_0-weakly consistent, provided the prior satisfies Assumption 1.

Kleijn and van der Vaart (2006) specialize the above model to the case when $z \mapsto f(y|z)$ is continuous for all y and $\mathcal{Z} = [-M, M]$ is compact. Under certain assumptions (including identifiability), they show that there is a unique $f^\star$ that minimizes $K(f_0, f)$. They further argue that $\mathbb{F}_0$ is $L_1(\mu)$ compact. When $\{f(y|z) : z \in [-M, M]\}$ is the normal location family, they show that their assumptions hold for the Dirichlet prior. Since this is a convex family, by Theorem 6, Assumptions 2a, 2b, and 2c hold with respect to $L_1(\mu_0)$. If $f_0/f^\star \in L_\infty(\mu)$, then the map $T : (\mathbb{F}_0, L_1(\mu)) \mapsto (\mathbb{F}_0, L_1(\mu_0))$, defined by $T(f) = f$, is continuous. Therefore, $L_1(\mu)$-compactness implies $L_1(\mu_0)$-compactness. Hence, Theorem 2 implies that $L_1(\mu_0)$-consistency holds. Since $f_0/f^\star \in L_\infty(\mu)$, this also implies that $L_1(\mu)$-consistency holds.

5.2 Normal regression

Consider the family of bivariate densities $\mathbb{F}_0$ of the form $f_\theta(y, x) = \varphi(y - \theta(x))g(x)$ where $\varphi(\cdot)$ is the standard normal density and $\theta \in \Theta$, a class of uniformly bounded continuous functions on the space of X. We assume that the true density f_0 is such that $Y - \theta_0(X) \sim p_0(\cdot)$, a density with mean 0 that does not depend on X. It's easy to see that $f^\star(y, x) = f_{\theta_0}(y, x) = \varphi(y - \theta_0(x))g(x)$. We are interested in posterior consistency with respect to the following metric:

$$d(f_{\theta_1}, f_{\theta_2}) = \sqrt{\mathsf{E}_0(\theta_1(X) - \theta_2(X))^2}.$$

Let $Z = Y - \theta_0(X)$. We assume that $\mathsf{E}_0[e^{M|Z|}] < \infty$, $\forall\ M > 0$. For notational simplicity, we denote $\mu_X := \theta(X) - \theta_0(X)$. Note that

$$\log \frac{f_\theta}{f_{\theta_0}} = Z \cdot \mu_X - \frac{\mu_X^2}{2},$$
$$\mathsf{E}_0 \log \frac{f_\theta}{f_{\theta_0}} = -\mathsf{E}_0 \frac{\mu_X^2}{2} = -\mathsf{E}_0(\theta(X) - \theta_0(X))^2.$$

This immediately ensures that Assumption 1 holds, as long as the prior puts positive mass on d-neighborhoods of θ_0. Towards verifying Assumption 2c we note by using Taylor's approximation for $h_\alpha^\star(f_0, f_\theta)$ as a function of α, at $\alpha = 0$, that for some $\xi \in (0, \alpha)$,

$$\left| \mathsf{E}_0 \left(\frac{f_\theta}{f_{\theta_0}} \right)^\alpha - 1 - \alpha \cdot \mathsf{E}_0 \log \frac{f_\theta}{f_{\theta_0}} \right| \le \frac{\alpha^2}{2} \mathsf{E}_0 \left[\left(\log \frac{f_\theta}{f_{\theta_0}} \right)^2 e^{\xi \log \frac{f_\theta}{f_{\theta_0}}} \right].$$

Since $\mathsf{E}_0[e^{M|Z|}] < \infty$ for any M and $\mu_x = \theta(x) - \theta_0(x)$ is uniformly bounded, the expectation on the right-hand side of the above inequality will be bounded by some large enough constant $C > 0$. Hence, we get

$$\left| \mathsf{E}_0 \left(\frac{f_\theta}{f_{\theta_0}} \right)^\alpha - 1 + \alpha \cdot \mathsf{E}_0 \frac{\mu_X^2}{2} \right| \le \frac{\alpha^2}{2} C.$$

Therefore,

$$\mathsf{E}_0 \left(\frac{f_\theta}{f_{\theta_0}} \right)^\alpha \le 1 - \alpha \mathsf{E}_0(\theta(X) - \theta_0(X))^2 + \frac{\alpha^2}{2} C.$$

For any $0 < \epsilon < 1$, note that if $d(\theta, \theta_0) = \sqrt{\mathsf{E}_0(\theta(X) - \theta_0(X))^2} > \epsilon$, then

$$\mathsf{E}_0 \left(\frac{f_\theta}{f_{\theta_0}} \right)^\alpha \quad \leq \quad 1 - \alpha\epsilon^2 + \frac{\alpha^2}{2} C.$$

In particular, with $\alpha = \frac{\epsilon^2}{C}$, when $d(\theta, \theta_0) > \epsilon$, we have

$$\mathsf{E}_0 \left(\frac{f_\theta}{f_{\theta_0}} \right)^\alpha \leq e^{-\frac{\epsilon^4}{2C}}. \tag{8}$$

The above inequality ensures that Assumption 2c holds with respect to d.

Finally, to verify Assumption 4, first note that since μ_X is uniformly bounded and $\mathsf{E}_0(e^{M|Z|}) < \infty, \ \forall \ M$, for some $C_1 > 0, M_1 > 0$, we have

$$\mathsf{E}_0 \left| \frac{f_{\theta_1}}{f_{\theta_0}} - \frac{f_{\theta_2}}{f_{\theta_0}} \right| = \mathsf{E}_0 \left[\left| \frac{f_{\theta_1}}{f_{\theta_2}} - 1 \right| \cdot \frac{f_{\theta_2}}{f_{\theta_0}} \right] \leq C_1 \cdot \mathsf{E}_0 \left[\left| \frac{f_{\theta_1}}{f_{\theta_2}} - 1 \right| \cdot e^{M_1|Z|} \right].$$

Now, denote $\mu'(X) = \theta_1(X) - \theta_2(X)$. Again, using Taylor's formula and the fact that μ'_X is uniformly bounded, we get that for some $C_2 > 0, M_2 > 0$,

$$\begin{aligned} \left| \left(\frac{f_{\theta_1}}{f_{\theta_2}} \right) - 1 \right| &= \left| e^{\left(Z \cdot \mu'_X - \frac{\mu'^2_X}{2} \right)} - 1 \right| \\ &\leq \sup_{0<\xi<1} \left[|\mu'_X| \cdot \left| Z - \frac{\mu'_X}{2} \right| e^{\xi \cdot \left(Z \cdot \mu'_X - \frac{\mu'^2_X}{2} \right)} \right] \\ &\leq C_2 \cdot \left[|\mu'_X| \cdot \left| Z - \frac{\mu'_X}{2} \right| e^{M_2|Z|} \right]. \end{aligned}$$

Therefore, putting the above two inequalities together, we get for some $M > 0, \ C > 0$,

$$\begin{aligned} \mathsf{E}_0 \left| \frac{f_{\theta_1}}{f_{\theta_0}} - \frac{f_{\theta_2}}{f_{\theta_0}} \right| &\leq C \cdot \left[|\mu'_X| \cdot \left| Z - \frac{\mu'_X}{2} \right| e^{M|Z|} \right] \\ &\leq C \cdot \left(\mathsf{E}_0 \left[|\mu'^2_X| \right] \right)^{\frac{1}{2}} \cdot \left(\mathsf{E}_0 \left[\left| Z - \frac{\mu'_X}{2} \right|^2 e^{2M|Z|} \right] \right)^{\frac{1}{2}}. \end{aligned}$$

The last step uses Cauchy–Schwartz inequality. Since the last term in the above inequality is finite, for some suitably large $K > 0$, we can write

$$\mathsf{E}_0 \left| \frac{f_{\theta_1}}{f_{\theta_0}} - \frac{f_{\theta_2}}{f_{\theta_0}} \right| \leq (\mathsf{E}_0(\theta_1(X) - \theta_2(X))^2)^{\frac{1}{2}} \cdot K,$$

which implies that Assumption 4 holds. Hence Theorem 5 is applicable, as long as the prior-summability (part (a)) or the entropy condition (part (b)) of the theorem holds.

5.3 Bayesian quantile regression

Consider the family of bivariate densities $\mathbb{F}_0$ of the form $f(y,x) = \varphi(y-\theta(x))g(x)$ where $\varphi(\cdot)$ is the asymmetric Laplace density given by $\varphi(z) = \tau(1-\tau)e^{-z(\tau - I_{(z\le 0)})}$, $z \in (-\infty,\infty)$ with $I_{(\cdot)}$ being the indicator function, $0<\tau<1$ and $\theta \in \Theta$, a class of uniformly bounded continuous functions on the space of X. It is easy to check that the τth quantile of φ is 0. Hence, this is one particular formulation used for Bayesian quantile regression (see Yu and Moyeed 2001). We assume that the true density is such that $Y - \theta_0(X) \sim p_0(\cdot)$, a density which does not depend on X and whose τth quantile is 0. It's easy to see that $f^\star(y,x) = f_{\theta_0}(y,x) = \varphi(y-\theta_0(x))g(x)$ (see Proposition 1 in Sriram et al. (2013)). We are interested in posterior consistency with respect to the following metric:

$$d(f_{\theta_1}, f_{\theta_2}) = \mathsf{E}_0\left|\theta_1(X) - \theta_2(X)\right|.$$

Let $Z = Y - \theta_0(X)$. It can be seen that (see Lemma 1 of Sriram et al. 2013)

$$\begin{aligned} \left|\log\frac{f_{\theta_1}}{f_{\theta_2}}\right| &\le \left|\theta_1(X) - \theta_2(X)\right|, \qquad (9)\\ \mathsf{E}_0\left|\log\frac{f_{\theta_1}}{f_{\theta_2}}\right| &\le \mathsf{E}_0\left|\theta_1(X) - \theta_2(X)\right|. \end{aligned}$$

This immediately ensures that Assumption 1 holds, as long as the prior puts positive mass on d-neighborhoods of θ_0. Further, since θ are uniformly bounded, the first of the above two inequalities ensures that $\frac{f_{\theta_1}}{f_{\theta_2}}$ is uniformly bounded. By Corollary 1 to Theorem 8, it follows that Assumption 2c will be satisfied with respect to $L_1(\mu_0)$. Further, we argue that Assumption 2c holds with respect to the metric d. To see this, first, it can be checked using the form of asymmetric Laplace density that

$$\left|\frac{f_\theta}{f_{\theta_0}} - 1\right| \ge \begin{cases} \left(1 - e^{-(\theta-\theta_0)(1-\tau)}\right)\cdot I_{(Z\le 0)} & \text{if } \theta-\theta_0 \ge 0,\\ \left(1 - e^{(\theta-\theta_0)\tau}\right)\cdot I_{(Z>0)} & \text{if } \theta-\theta_0 < 0. \end{cases}$$

Since $|\theta(X) - \theta_0(X)|$ is assumed to be uniformly bounded, we can further say that there exists a constant $C_0 > 0$ such that

$$\left|\frac{f_\theta}{f_{\theta_0}} - 1\right| \ge C_0|\theta(X) - \theta_0(X)|\cdot\left(I_{Z\le 0}\cdot I_{\theta-\theta_0\ge 0} + I_{Z>0}\cdot I_{\theta-\theta_0<0}\right).$$

Now, noting that $\mathsf{E}_0[I_{Z\le 0}|X] = \mathsf{P}_0(Z\le 0|X) = \tau$, we get

$$\begin{aligned} \mathsf{E}_0\left|\frac{f_\theta}{f_{\theta_0}} - 1\right| &\ge C_0\mathsf{E}_0\left[|\theta(X)-\theta_0(X)|\cdot(\tau\cdot I_{\theta-\theta_0\ge 0} + (1-\tau)\cdot I_{\theta-\theta_0<0})\right]\\ &\ge C_0\min(\tau,1-\tau)\cdot\mathsf{E}_0|\theta(X)-\theta_0(X)|\\ &= C_0\min(\tau,1-\tau)\cdot d(f_\theta, f_{\theta_0}). \end{aligned}$$

Since we have already argued that Assumption 2c holds for $L_1(\mu_0)$, the above inequality ensures that it also holds with respect to d. Finally, to check Assumption 4, we use (9) and the fact that $\frac{f_{\theta_2(X)}}{f_{\theta_0(X)}}$ is uniformly bounded, to get that for some $C_1 > 0$,

$$\mathsf{E}_0 \left| \frac{f_{\theta_1(X)}}{f_{\theta_0(X)}} - \frac{f_{\theta_2(X)}}{f_{\theta_0(X)}} \right| = \mathsf{E}_0 \left[\left| \frac{f_{\theta_1(X)}}{f_{\theta_2(X)}} - 1 \right| \cdot \frac{f_{\theta_2(X)}}{f_{\theta_0(X)}} \right] \leq C_1 \cdot \mathsf{E}_0 \left| \frac{f_{\theta_1(X)}}{f_{\theta_2(X)}} - 1 \right|.$$

Hence,

$$\mathsf{E}_0 \left| \frac{f_{\theta_1(X)}}{f_{\theta_2(X)}} - 1 \right| \leq \max \left(\mathsf{E}_0 \left| e^{-|\theta_1(X) - \theta_2(X)|} - 1 \right|, \mathsf{E}_0 \left| e^{|\theta_1(X) - \theta_2(X)|} - 1 \right| \right),$$

which by Taylor's formula is

$\leq C \cdot \mathsf{E}_0 \left| \theta_1(X) - \theta_2(X) \right|$ for some constant C.

Therefore, Assumption 4 holds. Hence Theorem 5 is applicable, as long as the prior-summability (part (a)) or the entropy condition (part (b)) of the theorem holds.

6 An extension to $i.n.i.d.$ models

The ideas developed in the previous sections allow us to begin some easy and direct applications to $i.i.d.$ parametric models and to the case of independent but non-identically distributed ($i.n.i.d.$) response. In this section, we outline these ideas. In the interest of flow, the results are presented here and the proofs are deferred to Appendix C.

We will assume that the distribution of the response Y is determined in principle by the knowledge of a covariate vector $\mathbf{X}$. In other words, there exists an unknown "true" density function $f_{0\mathbf{x}}(\cdot)$ with $\mathbf{x} \in \mathcal{X}$, such that $Y|\mathbf{X} = \mathbf{x} \sim f_{0\mathbf{x}}$. So, for the ith observed response Y_i with covariate value $\mathbf{X}_i = \mathbf{x}_i$, $Y_i \sim f_{0\mathbf{x}_i}$. The $\mathbf{X}_i$ could be non-random and hence $Y_1, Y_2, \ldots, Y_n$ are independent but non-identically distributed. $\mathsf{E}_{\mathbf{x}}[\cdot]$ will denote the expectation w.r.t. the density $f_{0\mathbf{x}}$. We will denote by P_x, the probability with respect to f_{0x} and P_0 with respect to the infinite product measure $f_{0\mathbf{x}_1} \times f_{0\mathbf{x}_2} \times \cdots$, and by $\mathsf{E}_0[\cdot]$ the expectation w.r.t. this product measure.

Suppose we have a family of densities $\mathbb{F}_0 = \{f_t : \ t \in [-M, M]\}$. Let Θ be a class of continuous functions from $\mathcal{X}$ to $[-M, M]$. For ease of notation, we write $\theta(\mathbf{x})$ as $\theta_{\mathbf{x}}$. The specified model is that $Y_i \sim f_{\theta_{\mathbf{x}_i}}$, where $\theta \in \Theta$ is the unknown possibly infinite dimensional parameter. First, we make the following assumption with regards to the covariates and the parameter space.

> **Assumption A.** The covariate space $\mathcal{X}$ is compact w.r.t. a norm $\|\cdot\|$ and Θ is a compact subset of continuous functions from $\mathcal{X} \to \mathcal{R}$ endowed with the sup-norm metric, i.e., $d(\theta_1, \theta_2) = \sup_{x \in \mathcal{X}} |\theta_1(x) - \theta_2(x)|$.

A straightforward parametric example for Θ would be any class of smooth functions defined on a compact set $\mathcal{X}$ and parametrized by finitely many parameters, also taking values on some compact set. In this case, sup-norm metric would be equivalent to the Euclidean metric on the finite dimensional parameter space. As an example for a non-parametric class of functions, let $\mathcal{X} = [0, 1]$ and let Θ be the sup-norm closure of the collection of polynomials $\mathcal{S}$ defined on $[0, 1]$ given by $\mathcal{S} := \{\theta \ : \ \theta(x) =$

$\sum_{j=1}^{k} a_j x^j$ for some $k \geq 1$ and such that $a_j \leq \frac{1}{j^3}\}$. It can be checked that this class is equi-uniformly-continuous and uniformly bounded. Hence, Θ will be compact by Arzelà–Ascoli theorem.

Let $\Pi(\cdot)$ be a prior on the parameter space Θ and let θ^* be the minimizer (with respect to θ) of $\mathsf{E}_{\mathbf{x}} \log \frac{f_{0\mathbf{x}}}{f_{\theta_{\mathbf{x}}}}$ for all $\mathbf{x}$. As before, we can write the posterior probability of a set $U^c = \{\theta \in \Theta : d(\theta, \theta^*) > \epsilon\}$ as follows:

$$\Pi(U^c|Y_{1:n}) := \frac{\int_{U^c} \prod_{i=1}^n \frac{f_{\theta_{\mathbf{x}_i}}(Y_i)}{f_{\theta^*_{\mathbf{x}_i}}(Y_i)} d\Pi(\theta)}{\int_{\Theta} \prod_{i=1}^n \frac{f_{\theta_{\mathbf{x}_i}}(Y_i)}{f_{\theta^*_{\mathbf{x}_i}}(Y_i)} d\Pi(\theta)} =: \frac{R'_{1n}}{R'_{2n}}.$$

We will be interested in probabilities of complements of sup-norm neighborhoods of θ^*. The following useful lemma shows that if the functions θ and θ^* differ at a point $\mathbf{x}_0$, then they will necessarily differ on a neighborhood around $\mathbf{x}_0$ as well.

Lemma 3. *Suppose Assumption A holds. Let $\theta' \in U^c$ and $\mathbf{x}_0 \in \mathcal{X}$ be such that $|\theta'_{\mathbf{x}_0} - \theta^*_{\mathbf{x}_0}| > \epsilon$. Then $\exists\ \delta'$ such that $\forall\ \mathbf{x} : \|\mathbf{x} - \mathbf{x}_0\| < \delta'$ we have $|\theta'_{\mathbf{x}} - \theta^*_{\mathbf{x}}| \geq \frac{\epsilon}{2}$.*

Clearly, for the posterior probability of $\sup|\theta(\mathbf{x}) - \theta^*(\mathbf{x})| > \epsilon$ to go to 0, we need that $|\theta(\mathbf{x}_i) - \theta^*(\mathbf{x}_i) > \epsilon$ for infinitely many of the design values $\mathbf{x}_i$. The following assumption is a way to formalize this notion.

Assumption B. For any given $\mathbf{x}_0 \in \mathcal{X}$, $\delta' > 0$, let $A_{\mathbf{x}_0,\delta'} = \{\mathbf{x} : \|\mathbf{x} - \mathbf{x}_0\| < \delta'\}$ and $I_{A_{\mathbf{x}_0,\delta'}}(\mathbf{x})$ be the indicator function which is 1 when $\mathbf{x} \in A_{\mathbf{x}_0}$ and 0 otherwise. Then, $\kappa(\mathbf{x}_0, \delta') = \liminf_{n\geq 1} \frac{1}{n} \sum_{i=1}^n I_{A_{\mathbf{x}_0,\delta'}}(\mathbf{x}_i) > 0$.

Such a condition can be seen to hold in various situations. A simple instance would be when $\mathcal{X}$ is a finite set $\{a_1, a_2, \ldots, a_k\}$ and when the covariates X_i take each of these values a_j a fixed proportion of times. Another example would be when the design set $\{x_i,\ i \geq 1\}$ is dense in a closed interval say $[0,1]$ such that the proportion of x_i's falling in any sub-interval is proportional to the interval length or a fixed function of the interval length. More generally, it could also be designs where the proportion of x_i could vary, for example, twice as many x_i's samples on $[0, 0.5]$ versus $[0.5, 1]$, etc.

Towards deriving conditions for $\Pi(U^c|Y_{1:n}) \to 0$, we first note that the next two assumptions are equivalent to Assumption 1 of the *i.i.d.* case and help control the denominator of the posterior probability, as seen in Proposition 4 below.

Assumption C. $\exists\ \theta^* \in \Theta$ such that $\theta^*_{\mathbf{x}} = \arg\min_{t\in[-M,M]} \mathsf{E}_{\mathbf{x}} \log \frac{f_{0\mathbf{x}}}{f_t}$, $\forall\ \mathbf{x} \in \mathcal{X}$ and θ^* is in the sup-norm support of Π.

Assumption D. $\mathsf{E}_{\mathbf{x}}[\log \frac{f_t}{f_{t'}}]$ and and $\mathsf{E}_{\mathbf{x}}[(\frac{f_t}{f_{t'}})^\alpha]$ for every $\alpha \in [0,1]$ are continuous functions in $(\mathbf{x}, t, t') \in \mathcal{X} \times [-M, M]^2$ and $\mathsf{E}_{\mathbf{x}} \log^2 \frac{f_t}{f_{t'}}$ is uniformly bounded for $(\mathbf{x}, t, t') \in \mathcal{X} \times [-M, M]^2$.

This continuity assumption is in the same spirit as Assumption 4 in the *i.i.d.* case. Such an assumption will hold if $\frac{f_t(y)}{f_t'(y)}$ is continuous in (t, t') for each y and if the true density $f_{0\mathbf{x}}(y)$ is continuous in x for each y, and can be bounded by an integrable function in y. The boundedness condition on the second moment of the log-likelihood ratio is to enable the application of the strong law of large numbers (SLLN) for independent random variables and is used in the proof of Proposition 4.

Proposition 4. *Suppose Assumptions A, C and D hold, then*

$$\text{for any } \beta > 0,\ e^{n\beta} R'_{2n} \to \infty \qquad \mathsf{P}_0\text{-}a.s.$$

The next assumption helps relate the sup-norm metric with the Kullback–Leibler divergence and is analogous to Assumption 2a.

Assumption E. For any $\epsilon > 0$, $\exists\ \delta \in (0, 1)$ and $\alpha_0 \in (0, 1)$ such that

$$\left\{ t \in [-M, M] : \mathsf{E}_{\mathbf{x}} \log \frac{f_{\theta^*_{\mathbf{x}}}}{f_t} < \delta \right\} \subseteq \{t : \ |t - \theta^*_{\mathbf{x}}| < \epsilon\},\ \forall\ \mathbf{x} \in \mathcal{X}.$$

Without delving into the details, we just note here that sufficient conditions for this assumption to hold can be derived based on ideas developed in Section 4. For example, as in Corollary 1, a sufficient condition would be that $\log \frac{f_{\theta(x)}}{f_{\theta^*_x}}$ is uniformly bounded.

The next lemma and proposition help obtain a result that is analogous to Lemma 2.

Lemma 4. *Let* $U^c = \{\theta : \ \sup_{\mathbf{x} \in \mathcal{X}} |\theta(\mathbf{x}) - \theta^*(\mathbf{x})| > \epsilon\}$*. If Assumptions A to E hold, then* $\exists\ \delta_1 \in (0, 1)$ *such that for every* $\theta' \in U^c$*, an* $\alpha' \in (0, 1)$ *can be chosen such that*

$$\mathsf{E}_0 \left(\prod_{i=1}^n \frac{f_{\theta'_{\mathbf{x}_i}}(Y_i)}{f_{\theta^*_{\mathbf{x}_i}}(Y_i)} \right)^{\alpha'} < e^{-n\delta_1} \quad \text{for all sufficiently large } n.$$

The next proposition is analogous to Lemma 2 and helps control the numerator of the posterior probability.

Proposition 5. *Suppose Assumptions A to E hold. Then for any* $\theta' \in U^c$, $\exists$ *an open set* $A_{\theta'}$ *containing* θ' *such that for some* $\alpha \in (0, 1)$, $\delta \in (0, 1)$ *and for any probability measure* $\nu(\cdot)$ *on* $A_{\theta'}$*, for all sufficiently large* n*, we have*

$$\mathsf{E}_0 \left[\left(\int_{A_{\theta'}} \prod_{i=1}^n \frac{f_{\theta_{\mathbf{x}_i}}(Y_i)}{f_{\theta^*_{\mathbf{x}_i}}(Y_i)} d\nu(\theta) \right)^{\alpha} \right] < e^{-n\alpha \frac{\delta}{2}}.$$

Finally, we obtain the result for the *i.n.i.d.* case.

Theorem 9. *Suppose that Assumptions A to E hold. Then,*

$$\Pi(U^c | Y_{1:n}) \to 0 \qquad \mathsf{P}_0\text{-}a.s.$$

Proof. Note that Proposition 5 can be applied by taking $\nu(\cdot) = \frac{\Pi(\cdot)}{\Pi(A_{\theta'})}$. So, for any $\theta' \in U^c$, $\exists\, \delta > 0$ and an open set $A_{\theta'}$ containing θ' such that, for sufficiently large n,

$$\begin{aligned}
&\mathsf{P}_0\left(\left(e^{n\frac{\delta}{4}} \int_{A_{\theta'}} \prod_{i=1}^{n} \frac{f_{\theta_{\mathbf{x}_i}}(Y_i)}{f_{\theta^*_{\mathbf{x}_i}}(Y_i)} d\Pi(\theta)\right)^{\alpha} > \epsilon^{\alpha}\right) \\
&\leq \frac{\Pi^{\alpha}(A_{\theta'})}{\epsilon^{\alpha}} \cdot \mathsf{E}_0\left(e^{n\frac{\delta}{4}} \int_{A_{\theta'}} \prod_{i=1}^{n} \frac{f_{\theta_{\mathbf{x}_i}}(Y_i)}{f_{\theta^*_{\mathbf{x}_i}}(Y_i)} \frac{d\Pi(\theta)}{\Pi(A_{\theta'})}\right)^{\alpha} \leq \frac{e^{-n\alpha\frac{\delta}{4}}}{\epsilon^{\alpha}}.
\end{aligned}$$

Therefore, by Borel–Cantelli lemma, we can conclude that

$$e^{n\frac{\delta}{4}} \int_{A_{\theta'}} \prod_{i=1}^{n} \frac{f_{\theta_{\mathbf{x}_i}}(Y_i)}{f_{\theta^*_{\mathbf{x}_i}}(Y_i)} d\Pi(\theta) \to\ 0 \qquad \mathsf{P}_0\text{-a.s.}$$

By Proposition 4, it follows in particular that

$$e^{n\frac{\delta}{4}} \int_{\Theta} \prod_{i=1}^{n} \frac{f_{\theta_{\mathbf{x}_i}}(Y_i)}{f_{\theta^*_{\mathbf{x}_i}}(Y_i)} d\Pi(\theta) \to\ \infty \qquad \mathsf{P}_0\text{-a.s.}$$

Considering the ratio of the above two quantities immediately gives $\Pi(A_{\theta'}|Y_{1:n}) \to 0$ P_0-a.s. By compactness, U^c can be covered by finitely many sets of the form $A_{\theta'}$. Hence the result follows. ☐

The proofs of lemmas and propositions discussed in this section are provided in Appendix C. As an application, we briefly discuss an example based on Bayesian quantile regression with *i.n.i.d.* response.

6.1 Example: Bayesian nonlinear quantile regression

Similar to Section 5.3, consider a family of densities $\mathbb{F}_0 = \{f_t : t \in [-M, M]\}$, where $f_t(y) = \varphi(y - \theta(x))g(x)$ where $\varphi(\cdot)$ is the asymmetric Laplace density given by $\varphi(z) = \tau(1-\tau)e^{-z(\tau - I_{(z\leq 0)})}$, $z \in (-\infty, \infty)$ with $I_{(\cdot)}$ being the indicator function, $0 < \tau < 1$. Let the "true" quantile function of Y given covariate $\mathbf{X}$ be $\theta_0(\mathbf{X})$.

By the properties of ALD, it can be observed (see Proposition 1, Lemmas 1 and 2 of Sriram et al. 2013) that (a) $\theta^* = \theta_0$, (b) $|\log \frac{f_{t'}}{f_t}| \leq |t - t'|$, and (c) that if $|t - \theta^*_{\mathbf{x}_i}| > \epsilon$ then

$$\mathsf{E}_{\mathbf{x}_i} \log \frac{f_{\theta^*_{\mathbf{x}_i}}}{f_{\theta}} > \delta_{\mathbf{x}_i} = \frac{\epsilon}{2} \cdot \min\left\{P_{0\mathbf{x}_i}\left(0 < Y_i - \theta^*_{\mathbf{x}_i} < \frac{\epsilon}{2}\right), P_{0\mathbf{x}_i}\left(-\frac{\epsilon}{2} < Y_i - \theta^*_{\mathbf{x}_i} < 0\right)\right\}.$$

As discussed above, there are various examples where Assumptions A and B would hold. One may consider any of those possibilities for the current example as well.

If we consider any prior that puts positive mass on sup-norm neighborhoods of θ_0, that along with observation (a) ensures that Assumption C holds.

It follows by Assumption A and observation (b) above that $\frac{f_t}{f_{t'}}$ is uniformly bounded and by property of ALD that it is continuous in (t, t'). If we assume that the true density function $f_{0\mathbf{x}}(y)$ is continuous in x for each y and can be dominated by an integrable function in y, then an application of dominated convergence theorem (DCT) would ensure that Assumption D holds.

Finally, using observation (c), if $\mathsf{P}_{0\mathbf{x}}(0 < Y - \theta^*_{\mathbf{x}} < \frac{\epsilon}{2})$ and $\mathsf{P}_{0\mathbf{x}}(-\frac{\epsilon}{2} < Y - \theta^*_{\mathbf{x}} < 0)$ (where $Y \sim \mathsf{P}_{0\mathbf{x}}$) are continuous and positive functions of $\mathbf{x}$, then $\{\delta_{\mathbf{x}_i},\ i \geq 1\}$ can be uniformly bounded below by a positive number. Hence, Assumption E would be satisfied.

Appendix A: Supporting results and proofs

We now state some technical results used in the paper. The first useful result given below is same as Lemma 6.3 of Kleijn and van der Vaart (2006).

Lemma 5. *As $\alpha \downarrow 0$, $\frac{1-h^\star_\alpha(f_0,f)}{\alpha} \uparrow K^\star(f_0, f)$.*

Proposition 2 in Section 2.2 shows that the three notions of divergence we consider in this paper are equivalent for convex sets. The proof of this result proceeds via the minimax theorem. We state below the minimax theorem due to Sion (1958) Corollary 3.3. and relevant lemmas leading up to the proof of Proposition 2.

Let M, N be convex sets. A function g on $M \times N$ is *quasi-concave* in M if $\{\nu : g(\mu, \nu) \geq c\}$ is a convex set for any $\mu \in M$ and real c. Likewise, g is *quasi-convex* in N if $\{\mu : g(\mu, \nu) \leq c\}$ is a convex set for any $\nu \in N$ and real c. The function g is *quasi-concave–convex* if it is quasi-concave in M and quasi-convex in N. Similarly, if $g(\cdot, \nu)$ is upper semi-continuous (usc) for any $\nu \in N$ and if $g(\mu, \cdot)$ is lower semi-continuous (lsc) for any $\mu \in M$, then it is said to be *usc–lsc*. Then Sion (1958) proved the following.

Theorem 10. *Let M and N be convex sets one of which is compact, and g a quasi-concave–convex and usc–lsc function on $M \times N$. Then*

$$\sup_{\mu \in M} \inf_{\nu \in N} g(\mu, \nu) = \inf_{\nu \in N} \sup_{\mu \in M} g(\mu, \nu).$$

The next lemma investigates the relevant properties needed on the function $h^\star_\alpha(f_0, f)$ so as to apply the minimax theorem. For clarity, we recall the definition of $h^\star_\alpha$ and note that

$$h^\star_\alpha(f_0, f) = \begin{cases} \mathsf{E}_0 \left(\frac{f}{f^\star}\right)^\alpha & \text{if } 0 < \alpha < 1, \\ 1 & \text{if } \alpha = 0, \\ \mathsf{E}_0 \left(\frac{f}{f^\star}\right) & \text{if } \alpha = 1. \end{cases}$$

Lemma 6. *The function $h^\star_\alpha(f_0, f)$ is concave in f and convex in α. Further, for fixed α, it is continuous in f in the $L_1(\mu_0)$-topology, where $d\mu_0 = (f_0/f^\star)\, d\mu$. Also, for fixed f, $h^\star_\alpha(f_0, f)$ is continuous in α.*

Proof. Concavity and convexity are easy to check. Continuity in f follows by noting that, for $\varphi \in \langle \mathbb{F}_0 \rangle$,

$$
\begin{aligned}
|h^\star_\alpha(f_0,\varphi) - h^\star_\alpha(f_0,f)| &= \Big| \int \Big(\frac{\varphi}{f^\star}\Big)^\alpha f_0 d\mu - \int \Big(\frac{f}{f^\star}\Big)^\alpha f_0 d\mu \Big| \\
&\le \int \Big| \frac{\varphi}{f^\star} - \frac{f}{f^\star} \Big|^\alpha f_0 d\mu \\
&\le \Big(\int |\varphi - f| d\mu_0 \Big)^\alpha .
\end{aligned}
$$

Continuity in α follows from the dominated convergence theorem since

$$(f/f^\star)^\alpha \le I_{\{f \le f^\star\}} + (f/f^\star) I_{\{f > f^\star\}},$$

where $I_{\{\cdot\}}$ is the indicator function. □

The following theorem is an immediate consequence of the minimax theorem and Lemma 6.

Proposition 6. *For any convex set $A \subset \langle \mathbb{F}_0 \rangle$,*

$$\inf_{0\le\alpha\le 1} \sup_{f\in A} h^\star_\alpha(f_0,f) = \sup_{f\in A} \inf_{0\le\alpha\le 1} h^\star_\alpha(f_0,f)$$

Another useful application of the minimax theorem is the following.

Proposition 7. *For any convex set $A \subset \mathbb{F}$ and $f \in A$, define:*

$$g(\alpha,f) = \begin{cases} K^\star(f_0,f) & \text{if } \alpha = 0, \\ (1 - h^\star_\alpha(f_0,f))/\alpha & \text{if } \alpha \in (0,1), \\ 1 - \int (f/f^\star) f_0 \, d\mu & \text{if } \alpha = 1. \end{cases}$$

Then

$$\sup_{0\le\alpha\le 1} \inf_{f\in A} g(\alpha,f) = \inf_{f\in A} \sup_{0\le\alpha\le 1} g(\alpha,f).$$

Proof. On A, we give the $L_1(\mu_0)$-topology. From Lemma 6 it follows that for each $\alpha \in (0,1)$, $g(\alpha,f)$ is continuous in f. Next, we argue that $g(\alpha,f)$ is lsc in f when $\alpha = 0$, i.e., we need to show that, if $\int |f_k - f| d\mu_0 \to 0$, for $f \in A$, then $\liminf K^\star(f_0,f_k) \ge K^\star(f_0,f)$. Suppose, on the contrary, that $\liminf K^\star(f_0,f_k) = \delta < K^\star(f_0,f)$. Then, there exists a subsequence $\{f_{k'}\}$ such that, $K^\star(f_0,f_{k'})$ is increasing and $\lim K^\star(f_0,f_{k'}) = \delta$. Let α_n decrease to 0, and set $A_n = \{f \in A : g(\alpha_n,f) \le \delta\}$. By Lemma 5, $g(\alpha_n,f) \le K^\star(f_0,f)$ and hence for each n, $\{f_{k'}\} \subset A_n$. Further, the continuity of $g(\alpha_n,f)$ with respect to f implies that f itself is in A_n. Thus $f \in \bigcap_n A_n$, which implies $K^\star(f_0,f) \le \delta$, a contradiction. Continuity at $\alpha = 1$ is trivial. Similarly, continuity of $g(\alpha,f)$ in α for $\alpha \in (0,1)$ follows from Lemma 6 and at $\alpha = 0$ from Lemma 5. Finally, $g(\alpha,f)$ is convex in f and by monotonicity in α, quasi-concave in α. Applying the minimax theorem gives the result. □

Proof of Proposition 2. (iii) $\implies$ (ii) is immediate. Now suppose (ii) holds, i.e.,

$$\sup_{f \in A} \inf_{0 \le \alpha \le 1} h^\star_\alpha(f_0, f) < e^{-\delta} \text{ for some } \delta > 0.$$

This means for a given, $f \in A$, $\exists\, 0 < \alpha \le 1$ such that $h^\star_\alpha(f_0, f) < e^{-\delta}$. Now, for any $\alpha' < \alpha$, setting $p = \frac{\alpha}{\alpha'}, q = \frac{p}{p-1}$ and applying Holder's inequality to $(\frac{f}{f^\star})^{\alpha'} 1$ with respect to the measure $f_0 d\mu$,

$$\int \left(\frac{f}{f^\star}\right)^{\alpha'} f_0 d\mu \le \left[\int \left(\frac{f}{f^\star}\right)^{\alpha} f_0 d\mu\right]^{\frac{\alpha'}{\alpha}},$$

or $h^\star_{\alpha'}(f_0, f) \le [h^\star_\alpha(f_0, f)]^{\frac{\alpha'}{\alpha}}$.

Consequently, for any $\alpha' \le \alpha, h^\star_{\alpha'}(f_0, f) \le e^{-\alpha'\delta}$. Equivalently, for all $\alpha' < \alpha$,

$$\frac{1 - h^\star_{\alpha'}(f_0, f)}{\alpha'} \ge \frac{1 - e^{-\alpha'\delta}}{\alpha'}.$$

As $\alpha' \downarrow 0$, from Lemma 5 the left-hand side of the last expression converges to $K^\star(f_0, f)$ and the right-hand side converges to $-\frac{d}{d\alpha'} e^{-\alpha'\delta}|_{\alpha'=0} = \delta$. This holds for each $f \in A$, and hence (i) holds. This completes the proof of $(iii) \implies (ii) \implies (i)$.

We will now show that $(i) \implies (iii)$ when A is a convex set, thus concluding that in this case, the three conditions are equivalent. Suppose (i) holds, then from Lemma 7,

$$\sup_{0 \le \alpha \le 1} \inf_{f \in A} g(\alpha, f) = \inf_{f \in A} \sup_{0 \le \alpha \le 1} g(\alpha, f) > \delta.$$

Since $\inf_{f \in A} g(\alpha, f)$ is increasing as $\alpha \downarrow 0$, given any $\delta' < \delta$, there is a $\alpha_0 > 0$, such that for all $f \in A$,

$$g(\alpha_0, f) = \frac{1 - h^\star_{\alpha_0}(f_0, f)}{\alpha_0} > \delta'.$$

So that $h^\star_{\alpha_0}(f_0, f) < 1 - \alpha_0 \delta' \le e^{-\eta}$ where $\eta = \alpha_0 \delta'$. Therefore, condition (iii) holds. □

Appendix B: Illustrating need for assumptions on topology

Example 1. This example shows that while $f^\star$ in the L_1 support of Π is necessary for consistency, this may not follow from Assumption 1.

Let f_0 be the $\mathsf{Unif}(0, 1)$ density and $f^\star$ be the $\mathsf{Unif}(0, 2)$ density. Let $\mathbb{F}_0 = \{f_k : k \ge 1\}$, where

$$f_k(y) = \begin{cases} b_k & \text{if } y \in (0, 1), \\ 2(1 - b_k), & \text{if } y \in (1, 3/2), \end{cases}$$

with $b_k \uparrow 1/2$. Take any prior such that $\Pi(f_k) > 0$ for all k. Then Assumption 1 is seen to be satisfied because $K^\star(f_0, f_k) \downarrow K^\star(f_0, f^\star) = \log 2$. However, the L_1-distance does not vanish, i.e., $\|f_k - f^\star\| > 1/4$ for all k.

Example 2. This example demonstrates that even if $f^\star$ is in the L_1 support of Π and Assumption 1 is satisfied, consistency still may not hold.

Let μ be the measure obtained as a sum of the Lebesgue measure on $[0, 2]$ and point masses on integers $k \geq 3$. For $k \geq 3$ and $a \in (0, 1]$, let f_k and g_a be densities with respect to the measure μ, defined as follows:

$$f_k(y) = \begin{cases} \frac{1}{2} - \frac{1}{k} & \text{if } y \in [0, 1], \\ \frac{1}{2} + \frac{1}{k} & \text{if } y = k, \end{cases}$$
$$g_a(y) = \begin{cases} \frac{1}{2} & \text{if } y \in [0, a], \\ 1 - \frac{a}{2} & \text{if } y \in [1, 2]. \end{cases}$$

Let $\mathbb{F} = \{g_a, f_k : a \in (0, 1), k \geq 3\}$ endowed with the $L_1(\mu)$ norm $\|\cdot\|$. Take the following priors on f_k and g_a:

$$\pi(f_k) = \frac{1}{2^{k-1}}, \quad k \geq 3,$$
$$\pi(g_a) = \frac{1}{4} a^{-\frac{1}{2}}, \quad a \in (0, 1].$$

Note that $\sum_{k \geq 3} \pi(f_k) = \frac{1}{2}$ and $\int_0^1 \pi(g_a) da = \frac{1}{2}$.

Let f_0 be the density on $(0, 1)$ whose distribution function F_0 is

$$F_0(y) = \begin{cases} 2y\Big[1 - \Big(-\log(1 - 1/2^{1/2})\Big)^{-1/2}\Big] & \text{if } y \in (0, 1/2], \\ 1 - \big(-\log(1 - y^{1/2})\big)^{-1/2} & \text{if } y \in (1/2, 1). \end{cases}$$

We will see later that, for this f_0, we get $1 - M_n^{1/2} < e^{-n}$ P_0-almost surely for all large n, where $M_n = \max\{Y_1, \ldots, Y_n\}$. Let $f^\star$ be the $\mathsf{Unif}(0, 2)$, density, i.e., $g_a(\cdot)$ for $a = 1$. Then it is easy to see the following:

$$\int \log\Big(\frac{f^\star}{f_k}\Big) f_0 \, d\mu = \log\left(\frac{1/2}{1/2 - 1/k}\right),$$
$$\int \log\Big(\frac{f^\star}{g_a}\Big) f_0 \, d\mu = \begin{cases} \infty & \text{if } a \in (0, 1), \\ 0 & \text{if } a = 1. \end{cases}$$

These show that, indeed, $f^\star$ minimizes the Kullback–Leibler divergence from f_0. Also, $\{f : K^\star(f_0, f) < \varepsilon\}$ contains $\{f_k : \log(\frac{1/2}{1/2 - 1/k}) < \varepsilon\} = \{f_k : k \geq 2(1 - e^{-\varepsilon})^{-1}\}$. Clearly, the assumed prior puts positive mass on the latter set, so Assumption 1 is satisfied. Further, note that, for $k \geq 3$ and $a \in (0, 1)$,

$$\|f_k - f^\star\| > 1/2 \quad \text{and} \quad \|g_a - f^\star\| = 1 - a.$$

Therefore, for any $\varepsilon \in (0, 1/2)$, we have

$$\Pi_n(\|f - f^\star\| > \varepsilon) = \frac{\int_{M_n}^{1-\varepsilon} 2^{-n}\, \frac{1}{4} a^{-1/2}\, da + \sum_{k\geq 3}(\frac{1}{2} - \frac{1}{k})^n\, \frac{1}{2^{k-1}}}{\int_{M_n}^{1} 2^{-n}\, \frac{1}{4} a^{-1/2}\, da + \sum_{k\geq 3}(\frac{1}{2} - \frac{1}{k})^n\, \frac{1}{2^{k-1}}} \geq \frac{0 + A_n(M_n)}{1 + A_n(M_n)},$$

where

$$A_n(m) = \frac{\sum_{k\geq 3}(\frac{1}{2} - \frac{1}{k})^n\, \frac{1}{2^{k-1}}}{\int_m^1 2^{-n}\, \frac{1}{4} a^{-1/2}\, da},$$

and $M_n = \max\{Y_1, \ldots, Y_n\}$ as before. We claim that $1 - M_n^{1/2} < e^{-n}$ for all large n with P_0-probability 1. To see this, write

$$\mathsf{P}_0(1 - M_n^{1/2} > e^{-n}) = \mathsf{P}_0\{M_n < (1 - e^{-n})^2\} = \Big(1 - \frac{1}{\sqrt{n}}\Big)^n \leq e^{-\sqrt{n}}.$$

Since this upper bound is summable over $n \geq 1$, the claim follows from the Borel–Cantelli lemma. Therefore, when n is large,

$$A_n(M_n) \geq \sum_{k\geq 3} e^n \left(\frac{1}{2} - \frac{1}{k}\right)^n \frac{1}{2^{k-1}}.$$

Since for large enough k, $e\,(\frac{1}{2} - \frac{1}{k}) > \frac{e}{2.5}$, this implies that $\liminf A_n(M_n) = \infty$ P_0-almost surely. Consequently, $\Pi_n(\|f - f^\star\| > \varepsilon) \not\to 0$, i.e., Assumption 1, together with $f^\star$ in the L_1 support of Π, is not enough to guarantee L_1-consistency.

Appendix C: Proofs of results for the *i.n.i.d.* case

Here, we provide details of proofs of results for the *i.n.i.d.* case discussed in Section 6.

Proof of Lemma 3. Since Assumption A holds, by Arzelà–Ascoli theorem, we have the following:

(i) Θ is uniformly bounded, i.e., $\exists\, M$ such that $|\theta(\mathbf{x})| \leq M \;\; \forall\, \theta \in \Theta$ and $\mathbf{x} \in \mathcal{X}$.

(ii) Θ is equi-uniformly-continuous, i.e., for $\mathbf{x}_0 \in \mathcal{X}$, given $\epsilon > 0$, $\exists\, \delta > 0$ such that $\forall\, \mathbf{x} :\; \|\mathbf{x} - \mathbf{x}_0\| < \delta$, $|\theta_{\mathbf{x}} - \theta_{\mathbf{x}_0}| < \epsilon$, $\;\forall\, \theta \in \Theta$.

Without loss of generality, for $\theta' \in U^c$, we have $\theta'_{\mathbf{x}_0} - \theta^*_{\mathbf{x}_0} > \epsilon$. By (ii) above, i.e., equicontinuity, $\exists\, \delta'$ such that $\forall\, \|\mathbf{x} - \mathbf{x}_0\| < \delta'$, we have $|\theta_{\mathbf{x}} - \theta_{\mathbf{x}_0}| < \frac{\epsilon}{4}$, $\forall\, \theta \in \Theta$. In particular, for such $\mathbf{x}$, $|\theta^*_{\mathbf{x}} - \theta^*_{\mathbf{x}_0}| < \frac{\epsilon}{4}$. Therefore,

$$\theta'_{\mathbf{x}} - \theta^*_{\mathbf{x}} = \theta'_{\mathbf{x}} - \theta'_{\mathbf{x}_0} + \theta'_{\mathbf{x}_0} - \theta^*_{\mathbf{x}_0} + \theta^*_{\mathbf{x}_0} - \theta^*_{\mathbf{x}} \geq -\frac{\epsilon}{4} + \epsilon - \frac{\epsilon}{4} = \frac{\epsilon}{2}.$$

□

Proof of Proposition 4. Due to the compactness of $\Theta \times [-M, M]^2$ (Assumption A), and continuity (Assumption D), it follows that $E_{\mathbf{x}} \log \frac{f_t}{f_{t'}}$ is uniformly continuous in (x, t, t'). Hence, the collection $\{\mathsf{E}_{\mathbf{x}} \log \frac{f_{\theta^*(\mathbf{x}_i)}}{f_{\theta(\mathbf{x}_i)}},\ i \geq 1\}$ is equicontinuous w.r.t. $\theta \in \Theta$. Further, Assumption D implies that $\{\mathsf{E}_{\mathbf{x}} \log^2 \frac{f_{\theta^*(\mathbf{x}_i)}}{f_{\theta(\mathbf{x}_i)}},\ i \geq 1\}$ is uniformly bounded. Hence, $\exists\ \delta \in (0, 1)$ such that

$$\left\{\sup_{\mathbf{x}\in\mathcal{X}} |\theta(\mathbf{x}) - \theta_1(\mathbf{x})| < \delta\right\}$$
$$\subseteq V_\epsilon = \left\{\theta : \sup_{i\geq 1} \mathsf{E}_{\mathbf{x}_i} \log \frac{f_{\theta^*_{\mathbf{x}_i}}(Y_i)}{f_{\theta_{\mathbf{x}_i}}(Y_i)} < \epsilon,\ \sum_{i=1}^{\infty} \frac{1}{i^2} \mathsf{E}_{\mathbf{x}_i} \left(\log \frac{f_{\theta^*_{\mathbf{x}_i}}(Y_i)}{f_{\theta_{\mathbf{x}_i}}(Y_i)}\right)^2 < \infty\right\}.$$

Assumption C will therefore ensure that the prior gives positive mass for the set V_ϵ. Now, observing that $R'_{2n} \geq \int_{V_\epsilon} e^{\sum_{i=1}^n \log(\frac{f_{\theta_{\mathbf{x}_i}}(Y_i)}{f_{\theta^*_{\mathbf{x}_i}}(Y_i)})} d\Pi(\theta)$ and an application of the strong law of large numbers for independent random variables leads to

$$\sum_{i=1}^{n} \log\left(\frac{f_{\theta_{\mathbf{x}_i}}(Y_i)}{f_{\theta^*_{\mathbf{x}_i}}(Y_i)}\right) > -2n\epsilon \quad a.s.$$

Rest of the proof is along the lines of Lemma 4.4.1 of Ghosh and Ramamoorthi (2003). ☐

Proof of Lemma 4. For $\theta' \in U^c$, let $\mathbf{x}_0$ be such that $|\theta'(\mathbf{x}_0) - \theta^*(\mathbf{x}_0)| > \epsilon$. Then by Lemma 3, $\exists\ \delta'$ such that $\forall\ \mathbf{x} \in A_{\mathbf{x}_0,\delta'} := \{\mathbf{x}\ :\ \|\mathbf{x} - \mathbf{x}_0\| < \delta'\}$, we have $|\theta'_{\mathbf{x}} - \theta^*_{\mathbf{x}}| \geq \frac{\epsilon}{2}$. Therefore, by Assumption E, $\exists\ \delta \in (0, 1)$ such that $\mathsf{E}_{\mathbf{x}} \log \frac{f_{\theta^*_{\mathbf{x}}}}{f_{\theta'_{\mathbf{x}}}} \geq \delta$ for all $\mathbf{x} \in A_{\mathbf{x}_0,\delta'}$.

For $(\mathbf{x}, t, t') \in \mathcal{X} \times [-M, M]^2$, let $g_\alpha(\mathbf{x}, t, t') := \frac{1-\mathsf{E}_{\mathbf{x}}(\frac{f_t}{f_{t'}})^\alpha}{\alpha}$. By Lemma 5, we have that, $g_\alpha(\mathbf{x}, t, t')$ increases to $\mathsf{E}_{\mathbf{x}} \log \frac{f_{t'}}{f_t}$ as $\alpha \downarrow 0$. By Assumption D, both $g_\alpha(\cdot, \cdot, \cdot)$ and the limiting function are continuous in $(\mathbf{x}, t, t')$, which is in the compact set $\mathcal{X} \times [-M, M]^2$. Hence, it follows by Dini's theorem that this convergence is uniform, i.e.,

$$\lim_{\alpha\downarrow 0} \frac{1 - \mathsf{E}_{\mathbf{x}}\left(\frac{f_t}{f_{t'}}\right)^\alpha}{\alpha} \uparrow \mathsf{E}_{\mathbf{x}} \log \frac{f_{t'}}{f_t} \text{ uniformly on } \mathcal{X} \times [-M, M]^2.$$

Let $\kappa := \kappa(\mathbf{x}_0, \delta')$ as in Assumption B. Then, $\exists\ 0 < \alpha' < 1$ such that $g_{\alpha'}(\mathbf{x}, t, t') > \mathsf{E}_{\mathbf{x}} \log \frac{f_{t'}}{f_t} - \kappa\frac{\delta}{2}, \forall\ (\mathbf{x}, t, t') \in\ \mathcal{X} \times [-M, M]^2$. In particular, $g_{\alpha'}(\mathbf{x}_i, \theta_{\mathbf{x}_i}, \theta^*_{\mathbf{x}_i}) \geq \mathsf{E}_{\mathbf{x}_i} \log \frac{f_{\theta^*_{\mathbf{x}_i}}}{f_{\theta_{\mathbf{x}_i}}} - \kappa\frac{\delta}{2}\ \forall\ i \geq 1, \theta \in \Theta$. Also, in general $\mathsf{E}_{\mathbf{x}_i} \log \frac{f_{\theta^*_{\mathbf{x}_i}}}{f_{\theta_{\mathbf{x}_i}}} \geq 0$. Combining this with the observation we made at the beginning of the proof that $\mathsf{E}_{\mathbf{x}} \log \frac{f_{\theta^*_{\mathbf{x}}}}{f_{\theta'_{\mathbf{x}}}} \geq \delta\ \forall\ x \in A_{x_0,\delta'}$, we get

$$g_{\alpha'}(\mathbf{x}_i, \theta'_{\mathbf{x}_i}, \theta^*_{\mathbf{x}_i}) \geq \delta \cdot I_{A_{\mathbf{x}_0,\delta'}}(\mathbf{x}_i) - \kappa\frac{\delta}{2}, \tag{10}$$

where $I_{A_{\mathbf{x}_0,\delta'}}(\mathbf{x})$ is the indicator function which is 1 when $\mathbf{x} \in A_{\mathbf{x}_0,\delta'}$ and 0 otherwise. Note that, by Assumption B, for sufficiently large n, $\frac{1}{n}\sum_{i=1}^n I_{A_{\mathbf{x}_0,\delta'}}(\mathbf{x}_i) > \frac{3\kappa}{4}$. Using this, along with a bit of algebra on (10), we can conclude that the following inequality holds for sufficiently large n:

$$\mathsf{E}_0\left(\prod_{i=1}^n \frac{f_{\theta'_{\mathbf{x}_i}}(Y_i)}{f_{\theta^*_{\mathbf{x}_i}}(Y_i)}\right)^{\alpha'} \leq e^{-\delta\sum_{i=1}^n \cdot I_{A_{\mathbf{x}_0,\delta'}}(\mathbf{x}_i)+n\kappa\frac{\delta}{2}\cdot} \leq e^{-n\kappa\frac{\delta}{4}}.$$

The result follows by assigning $\delta_1 := \kappa\frac{\delta}{4}$. □

Proof of Proposition 5. First, we claim by Assumption D that the collection of functions $\{\mathsf{E}_{\mathbf{x}}\frac{f_{\theta(\mathbf{x}_i)}}{f_{\theta'(\mathbf{x}_i)}},\ i \geq 1\}$ is equicontinuous w.r.t. the sup-norm metric on Θ. Note that $\mathsf{E}_{\mathbf{x}}\frac{f_t}{f_{t'}}$ is a continuous function on a compact set $\mathcal{X} \times [-M, M]^2$. Hence, it is uniformly continuous. So, given $\epsilon > 0$, $\exists\ \delta$ such that if $\|\mathbf{x} - \mathbf{x}_1\| < \delta$, $|t - t_1| < \delta$ and $|t' - t'_1| < \delta$ then $|\mathsf{E}_{\mathbf{x}_1}\frac{f_{t_1}}{f_{t'_1}} - \mathsf{E}_{\mathbf{x}}\frac{f_t}{f_{t'}}| < \epsilon$. In particular, let $\theta, \theta_1 \in \Theta$ be such that $\sup_{\mathbf{x}\in\mathcal{X}} |\theta(\mathbf{x}) - \theta_1(\mathbf{x})| < \delta$. Then for any $\mathbf{x} \in \mathcal{X}$, taking $\mathbf{x}_1 = \mathbf{x}$, $t' = t'_1 = \theta(\mathbf{x})$ and $t = \theta(\mathbf{x})$, $t_1 = \theta_1(\mathbf{x})$, we get $|\mathsf{E}_{\mathbf{x}}\frac{f_{\theta(\mathbf{x})}}{f_{\theta'(\mathbf{x})}} - \mathsf{E}_{\mathbf{x}}\frac{f_{\theta_1(\mathbf{x})}}{f_{\theta'(\mathbf{x})}}| < \epsilon$. Hence the collection of functions $\{\mathsf{E}_{\mathbf{x}}\frac{f_{\theta(\mathbf{x}_i)}}{f_{\theta'(\mathbf{x}_i)}},\ i \geq 1\}$ is equicontinuous in θ w.r.t. sup-norm metric.

Define $A_{\theta'} := \{\theta \in \Theta :\ \mathsf{E}_{\mathbf{x}_i}[\frac{f_{\theta_{\mathbf{x}_i}}}{f_{\theta'_{\mathbf{x}_i}}}] < e^{\frac{\delta}{2}}, \forall i \geq 1\}$. This set clearly contains θ' and it is an open set due to equicontinuity. By Lemma 4, $\exists\ \alpha' \in (0, 1)$ such that

$$\mathsf{E}_0\left(\prod_{i=1}^n \frac{f_{\theta'_{\mathbf{x}_i}}(Y_i)}{f_{\theta^*_{\mathbf{x}_i}}(Y_i)}\right)^{\alpha'} < e^{-n\alpha'\delta} \quad \text{for all sufficiently large } n.$$

Let $\alpha = \alpha'/2$. Then, for sufficiently large n,

$$\begin{aligned}
&\mathsf{E}_0\left[\left(\int_{A_{\theta'}}\prod_{i=1}^n \frac{f_{\theta_{\mathbf{x}_i}}(Y_i)}{f_{\theta^*_{\mathbf{x}_i}}(Y_i)}d\nu(\theta)\right)^{\alpha}\right]\\
&= \mathsf{E}_0\left[\left(\frac{f_{\theta'_{\mathbf{x}_i}}(Y_i)}{f_{\theta^*_{\mathbf{x}_i}}(Y_i)}\right)^{\alpha}\left(\int_{A_{\theta'}}\prod_{i=1}^n \frac{f_{\theta_{\mathbf{x}_i}}(Y_i)}{f_{\theta'_{\mathbf{x}_i}}(Y_i)}d\nu(\theta)\right)^{\alpha}\right]\\
&\text{(By Cauchy–Schwartz inequality)}\\
&\leq \left(\mathsf{E}_0\left[\left(\frac{f_{\theta'_{\mathbf{x}_i}}(Y_i)}{f_{\theta^*_{\mathbf{x}_i}}(Y_i)}\right)^{2\alpha}\right]\right)^{\frac{1}{2}}\cdot\left(\mathsf{E}_0\left[\left(\int_{A_{\theta'}}\prod_{i=1}^n \frac{f_{\theta_{\mathbf{x}_i}}(Y_i)}{f_{\theta'_{\mathbf{x}_i}}(Y_i)}d\nu(\theta)\right)^{2\alpha}\right]\right)^{\frac{1}{2}}\\
&\text{(By Jensen's inequality)}\\
&\leq \left(\mathsf{E}_0\left[\left(\frac{f_{\theta'_{\mathbf{x}_i}}(Y_i)}{f_{\theta^*_{\mathbf{x}_i}}(Y_i)}\right)^{\alpha'}\right]\right)^{\frac{1}{2}}\cdot\left(\int_{A_{\theta'}}\mathsf{E}_0\left[\prod_{i=1}^n \frac{f_{\theta_{\mathbf{x}_i}}(Y_i)}{f_{\theta'_{\mathbf{x}_i}}(Y_i)}\right]d\nu(\theta)\right)^{\frac{\alpha'}{2}}\\
&< e^{-n\alpha'\frac{\delta}{2}}\cdot e^{n\alpha'\frac{\delta}{4}} = e^{-n\alpha\frac{\delta}{2}}.
\end{aligned}$$

□

References

Berk, R. H. (1966). "Limiting behavior of posterior distribution when the model is incorrect." *The Annals of Mathematical Statistics*, 37(1): 51–58. MR0189176. 759

Bunke, O. and Milhaud, X. (1998). "Asymptotic behavior of Bayes estimates under possibly incorrect models." *The Annals of Statistics*, 26(2): 617–644. MR1626075. doi: http://dx.doi.org/10.1214/aos/1028144851. 759

De Blasi, P. and Walker, S. G. (2013). "Bayesian asymptotics with misspecified models." *Statistica Sinica*, 23: 169–187. MR3076163. 759, 760, 761, 770

Ghosh, J. K. and Ramamoorthi, R. V. (2003). *Bayesian Nonparametrics*. New York: Springer-Verlag. MR1992245. 762, 767, 787

Kleijn, B. J. K. and van der Vaart, A. W. (2006). "Misspecification in infinite-dimensional Bayesian statistics." *The Annals of Statistics*, 34(2): 837–877. MR2283395. doi: http://dx.doi.org/10.1214/009053606000000029. 759, 760, 761, 763, 765, 766, 767, 770, 772, 774, 782

Lee, J. and MacEachern, S. N. (2011). "Consistency of Bayes estimators without the assumption that the model is correct." *Journal of Statistical Planning and Inference*, 141(2): 748–757. MR2732945. doi: http://dx.doi.org/10.1016/j.jspi.2010.07.022. 759

Shalizi, C. R. (2009). "Dynamics of Bayesian updating with dependent data and misspecified models." *Electronic Journal of Statistics*, 3: 1039–1074. MR2557128. doi: http://dx.doi.org/10.1214/09-EJS485. 759

Sion, M. (1958). "On general minimax theorems." *Pacific Journal of Mathematics*, 8(1): 171–176. MR0097026. 782

Sriram, K., Ramamoorthi, R. V., and Ghosh, P. (2013). "Posterior consistency of Bayesian quantile regression based on the misspecified asymmetric Laplace density." *Bayesian Analysis*, 8(2): 479–504. MR3066950. doi: http://dx.doi.org/10.1214/13-BA817. 777, 781

Yu, K. and Moyeed, R. A. (2001). "Bayesian quantile regression." *Statistics and Probability Letters*, 54(2): 437–447. MR1861390. doi: http://dx.doi.org/10.1016/S0167-7152(01)00124-9. 777

Acknowledgments

We thank the Editor, Associate Editor and the Referees for their insightful comments and helping us significantly improve the paper. The work of Karthik Sriram was supported by a research grant provided by Indian Institute of Management Ahmedabad, India.