

Bayesian Consistency with the Supremum Metric

Nhat Ho[◊] Stephen G. Walker^{◊,b}

Department of Statistics and Data Sciences, University of Texas at Austin[◊],

Department of Mathematics, University of Texas at Austin^b

Abstract

We present simple conditions for Bayesian consistency in the supremum metric. The key to the technique is a triangle inequality which allows us to explicitly use weak convergence, a consequence of the standard Kullback–Leibler support condition for the prior. A further condition is to ensure that smoothed versions of densities are not too far from the original density, thus dealing with densities which could track the data too closely. A key result of the paper is that we demonstrate supremum consistency using weaker conditions compared to those currently used to secure \mathbb{L}_1 consistency.

Keywords: Prokhorov metric, sin kernel, Fourier integral theorem, weak convergence.

1 Introduction

Bayesian consistency remains an open topic and has seen much progress and ideas since the seminal papers of [Barron et al. \[1999\]](#) and [Ghosal et al. \[1999\]](#). A dominating sufficient, but not necessary, condition is a Kullback–Leibler support condition for the prior;

$$\Pi(D(p_0, p) < \varepsilon) > 0 \tag{1}$$

for all $\varepsilon > 0$. Here $D(p_0, p) = \int p_0 \log(p_0/p)$ denotes the Kullback–Leibler divergence between p_0 and p and p_0 represents the true density function from which the identically distributed $(X_i)_{i=1:n}$ are observed. Further, we write $\Pi(dp)$ to denote the prior distribution on a space of probability density functions; say \mathbb{P} .

It is well known that condition (1) is not sufficient for strong consistency. Strong consistency holds if

$$\Pi_n(A_\varepsilon) := \Pi(A_\varepsilon \mid X_{1:n}) \rightarrow 0 \quad \text{a.s.} \quad P_0^\infty \tag{2}$$

for all $\varepsilon > 0$, where $A_\varepsilon = \{p : d_H(p_0, p) > \varepsilon\}$ and d_H is the Hellinger distance between p_0 and p . Note the Hellinger distance is equivalent to the \mathbb{L}_1 distance. There is a counter example in [Barron et al. \[1999\]](#) which shows that a posterior is not strongly consistent given only the Kullback–Leibler support condition.

The standard additional sufficient condition for consistency involves the existence of an increasing sequence of sieves (\mathbb{F}_n) , which become \mathbb{P} as $n \rightarrow \infty$, such that the size of \mathbb{F}_n , as measured by some suitable entropy, is bounded by e^{n^κ} , for some $\kappa > 0$, and $\Pi(\mathbb{F}'_n) < e^{-n^\xi}$ for some $\xi > 0$.

On the other hand, [Walker \[2004\]](#) found a sieve, based on Π itself, which automatically satisfies the entropy condition, and the \mathbb{F}'_n condition is satisfied when $\sum_{j=1:\infty} \Pi(A_j) < \infty$, where the $(A_j)_{j=1}^\infty$ form a partition of \mathbb{P} with respect to Hellinger neighborhoods. A recent survey of Bayesian consistency is provided in [Ghosal and van der Vaart \[2017\]](#).

A new approach to Bayesian consistency was developed by [Chae and Walker \[2017\]](#). The idea is to rely on the weak convergence of the posterior and to find a minimal extension to secure strong consistency. The triangle inequality, for some strong metric d , yields

$$d(f_0, f) \leq d(f_0, \bar{f}_0) + d(f, \bar{f}) + d(\bar{f}_0, \bar{f}),$$

where \bar{f} indicates a smoothed version of f . Specifically in [Chae and Walker \[2017\]](#)

$$\bar{f}(x) = \frac{1}{2h} [F(x+h) - F(x-h)]$$

is used for some smoothing parameter $h > 0$ under the univariate setting.

The triangle inequality is perfect for understanding the key aspects of strong consistency. The idea is that weak convergence can deal with the $d(\bar{f}, \bar{f}_0)$ term, an assumption on f_0 can deal with the $d(\bar{f}_0, f_0)$ term, and a condition on f not being too oscillating can deal with the $d(\bar{f}, f)$ term.

In this paper we use such a strategy but alter the specific details. In particular we obtain a very simple condition for strong consistency with respect to the sup metric; i.e., \mathbb{L}_∞ . Previous work on the sup metric has been done by [Castillo \[2014\]](#), who considered contraction rates, assuming the true density on $(0, 1)$ is bounded away from 0. Other papers on consistency and rates using the \mathbb{L}_r metrics include [\[Gine and Nickl, 2011\]](#), [\[Hoffmann et al., 2015\]](#) and [\[Scricciolo, 2014\]](#).

In this work, we consider the triangle inequality

$$|f_0(x) - f(x)| \leq |f_0(x) - f_{0,R}(x)| + |f_{0,R}(x) - f_R(x)| + |f(x) - f_R(x)|$$

where f_R is an alternative kernel smoothed version of f ; specifically using the sin kernel. That is

$$f_R(x) := \int_{\mathbb{R}^d} \prod_{j=1}^d \frac{\sin(R(x_j - y_j))}{x_j - y_j} f(y) dy, \quad (3)$$

for any $x = (x_1, \dots, x_d) \in \mathbb{R}^d$. As R approaches infinity and $f \in \mathbb{L}_1(\mathbb{R}^d)$, $f_R(x)$ converges to $f(x)$ according to the Fourier integral theorem [\[Wiener, 1933, Bochner, 1959\]](#).

The present paper focuses solely on consistency. The idea being that weakening the conditions on prior distributions for consistency to be achieved is and remains an important topic. These weakened conditions can then be used to achieve current rates of convergence, it is argued, with some technical applications; but the insights are coming from how the weakening of assumptions required for consistency arise.

The layout of the paper is as follows. In [Section 2](#) we outline the assumptions and initial results needed to support the general theory in [Section 4](#). [Section 3](#) provides an illustration of the proof strategy for establishing posterior strong consistency under widely used infinite normal mixtures. We conclude the paper with some discussion in [Section 5](#).

2 Assumptions and Initial Results

In order to study Bayesian consistency, we define the following notion of supersmooth and ordinary smooth density functions. To simplify the presentation, \hat{f} denotes the Fourier transform of the function f .

Definition 1. (1) We say that the density function f is supersmooth of order α with scale parameter σ if there exist universal constants C, C_1 such that for almost all $x \in \mathbb{R}^d$, we obtain

$$|\hat{f}(x)| \leq C \exp \left(-C_1 \sigma^2 \left(\sum_{j=1}^d |x_j|^\alpha \right) \right).$$

(2) The density function f is ordinary smooth of order β with scale parameter σ if there exists universal constant c such that for almost all $x \in \mathbb{R}^d$, we have

$$|\hat{f}(x)| \leq c \cdot \prod_{j=1}^d \frac{1}{(1 + \sigma^2 |x_j|^\beta)}.$$

Examples of supersmooth functions include mixtures of location Gaussian distributions or mixture of location Cauchy distributions with similar scale parameter. In particular, when $f(x) = \sum_{i=1}^k \omega_i \mathcal{N}(x | \mu_i, \sigma^2 I_d)$ where $1 \leq k \leq \infty$, then f is supersmooth density function of order 2 with scale parameter σ . When f is mixture of location Cauchy distributions with scale parameter $\sigma^2 I_d$, then f is supersmooth density function of order 1 with scale parameter σ .

Examples of ordinary smooth functions include mixtures of location Laplace distribution with similar scale parameter σI_d . In this case, these mixtures are ordinary smooth functions of order 2 with scale parameter σ .

Based on Definition 1, we have the following result regarding the difference between f_R and f . The proof of Proposition 1 is similar to that of Theorem 1 in Ho and Walker [2021]; therefore, it is omitted.

Proposition 1. (1) Assume that f is a supersmooth density function of order $\alpha > 0$ with scale parameter σ . Then, there exist universal constants C and C' such that for $R \geq C'$, we have that

$$\sup_{x \in \mathbb{R}^d} |f_R(x) - f(x)| \leq C \frac{R^{\max\{1-\alpha, 0\}}}{\sigma^{2d}} \exp(-C_1 \sigma^2 R^\alpha),$$

where C_1 is a universal constant associated with the supersmooth density function f from Definition 1.

(2) Assume that f is a ordinary smooth density function of order $\beta > 0$ with scale parameter σ . Then, there exists a universal constants c such that

$$\sup_{x \in \mathbb{R}^d} |f_R(x) - f(x)| \leq \frac{c}{\sigma^{2+2(d-1)/\beta} R^{\beta-1}}.$$

Hence, for sufficiently large R , we have that $\sup_{x \in \mathbb{R}^d} |f(x) - f_R(x)|$ is sufficiently small.

If the prior puts positive mass on all Kullback–Leibler neighborhoods of f_0 ; i.e., equation (1), then the posterior converges on weak neighborhoods of f_0 . That is:

$$\Pi_n \left(\left| \int_{\mathbb{R}^d} g(y) (f(y) - f_0(y)) dy \right| > \varepsilon \right) \rightarrow 0 \quad \text{a.s.} \quad P_0^\infty$$

for all continuous and bounded functions g . For our purpose, we will be using the product of sinc functions, which is given by:

$$g_{x,R}(y) = \prod_{j=1}^d \frac{\sin R(y_j - x_j)}{R(y_j - x_j)},$$

which is continuous and bounded for every $x = (x_1, \dots, x_d) \in \mathbb{R}^d$, $y = (y_1, \dots, y_d) \in \mathbb{R}^d$ and $R > 0$. Hence, with equation (1), we have

$$\Pi_n \left(\left| \int g_{x,R}(y) (f(y) - f_0(y)) dy \right| > \varepsilon \right) \rightarrow 0 \quad \text{a.s.} \quad P_0^\infty$$

for all $x \in \mathbb{R}^d$ and $R > 0$. However, we have a stronger result to this, which is that

$$\Pi_n \left(\sup_{x \in \mathbb{R}^d} \left| \int g_{x,R}(y) (f(y) - f_0(y)) dy \right| > \varepsilon \right) \rightarrow 0 \quad \text{a.s.} \quad P_0^\infty \quad (4)$$

for any fixed $R > 0$. This is direct to show since we can write equation (4) as

$$\Pi_n \left(\sup_{x \in \mathbb{R}^d} \left| \int \prod_{j=1}^d \frac{\sin(Rt_j)}{Rt_j} (f(t-x) - f_0(t-x)) dt \right| > \epsilon \right) \rightarrow 0 \quad \text{a.s.} \quad P_0^\infty.$$

This holds since if $d_P(f, f_0) \rightarrow 0$ then $\sup_{x \in \mathbb{R}^d} d_P(f(\cdot - x), f_0(\cdot - x)) \rightarrow 0$, which follows from the definition of the Prokhorov metric d_P and is tantamount to demonstrating that $(A - x)^\epsilon = A^\epsilon - x$, where $A^\epsilon = \{b : \|b - a\| < \epsilon, a \in A\}$ and $A - x = \{a - x : a \in A\}$, which is straightforward to do. It is worth writing this out as the following lemma:

Lemma 1. *If $d_P(f, f_0) < \epsilon$ then $\sup_{x \in \mathbb{R}^d} d_P(f(\cdot - x), f_0(\cdot - x)) < \epsilon$.*

Assumption on f_0 : Throughout this paper, we utilize the following mild assumption on the true density function f_0 :

$$\varepsilon_R := \sup_{x \in \mathbb{R}^d} |f_{0,R}(x) - f_0(x)| \leq \bar{C}/R, \quad (5)$$

for all $R \geq R_0$, where \bar{C} and R_0 are some positive universal constants. This assumption is satisfied when f_0 is ordinary smooth function of order 2 with any fixed scale parameter. When $d = 1$, this condition is reduced to f_0 being almost surely second order differentiable function.

3 Illustration

Before stating our general posterior consistency results in Section 4, we consider an illustration of our proof strategy for the posterior consistency under normal mixtures, one of the most widely used nonparametric models. To keep thing simple, we consider the normal mixture models in dimension $d = 1$; whereby

$$f(x) = \sum_{j=1}^{\infty} w_j \phi((x - \mu_j)/\sigma)/\sigma,$$

the $(w_j)_{j=1}^{\infty}$ are a set of weights, the $(\mu_j)_{j=1}^{\infty}$ are a set of locations and the σ is a common variance term to each normal component. Further, ϕ represents the usual standard normal density function. In a Bayesian model, prior distributions are assigned to the weights, locations and the variance. Our proof for the posterior consistency of normal mixtures consists of two main steps.

Step 1: First, we find an appropriate upper bound for $\sup_x |f_R(x) - f(x)|$. Note that, the bound for $\sup_x |f_R(x) - f(x)|$ falls within the supersmooth setting in Proposition 1 and can be proved via bounding the tail of the Fourier transform of normal mixtures; nevertheless, in this section we show a different approach for deriving this bound for the normal mixture models via some closed-form computations.

To this end, we first show that

$$I(R) = \int_{-\infty}^{\infty} \cos(Rx) \phi(x) dx = e^{-\frac{1}{2}R^2} \quad (6)$$

for all $R \geq 0$. Now, $I'(R) = -\int_{-\infty}^{\infty} \sin(Rx) x \phi(x) dx$, and using integration by parts, with $x \phi(x) = -\phi'(x)$, we have $I'(R) = -R I(R)$ and hence equation (6) holds since $I(0) = 1$.

Now consider

$$I(R) = \int_{-\infty}^{\infty} \cos(Rx) \phi(x - \mu) dx = \int_{-\infty}^{\infty} \cos(R(x + \mu)) \phi(x) dx$$

and recall $\cos(R(x + \mu)) = \cos(Rx) \cos(R\mu) - \sin(Rx) \sin(R\mu)$, so, $I(R) = \cos(R\mu) e^{-\frac{1}{2}R^2}$ since $\sin(Rx)$ is an odd function. Further, it is straightforward to show that

$$\int_{-\infty}^{\infty} \cos(R(y - x)) \phi((x - \mu)/\sigma)/\sigma dx = \cos(R(y - \mu)) e^{-\frac{1}{2}\sigma^2 R^2}, \quad (7)$$

using suitable transforms. If we denote

$$J(R) = \int_{-\infty}^{\infty} \frac{\sin(Rx)}{x} \phi(x) dx,$$

then $J'(R)$ is given by equation (6), so $J(R) = \int_0^R e^{-\frac{1}{2}s^2} ds$ since $J(0) = 0$. Hence, we find that

$$J(y; \mu, \sigma, R) = \int_{-\infty}^{\infty} \frac{\sin(R(y - x))}{y - x} \phi((x - \mu)/\sigma)/\sigma dx = \int_0^R e^{-\frac{1}{2}\sigma^2 s^2} \cos(s(y - \mu)) ds.$$

We want to look at $f_R(x) - f(x) = \frac{1}{\pi} J(x; \mu, \sigma, R) - \phi((x - \mu)/\sigma)/\sigma$, and from equation (6), we have that

$$\int_0^{\infty} e^{-\frac{1}{2}\sigma^2 s^2} \cos(s(x - \mu)) ds = \pi \phi((x - \mu)/\sigma)/\sigma.$$

Therefore, for all $x \in \mathbb{R}$ we have

$$\pi |f_R(x) - f(x)| = \left| \int_R^{\infty} e^{-\frac{1}{2}\sigma^2 s^2} \cos(s(x - \mu)) ds \right| \leq \int_R^{\infty} e^{-\frac{1}{2}\sigma^2 s^2} ds < \frac{1}{\sigma^2 R} e^{-\frac{1}{2}\sigma^2 R^2}.$$

As a consequence, for any $R > 0$ we obtain that

$$\sup_{x \in \mathbb{R}} |f_R(x) - f(x)| < \frac{1}{\pi \sigma^2 R} e^{-\frac{1}{2}\sigma^2 R^2}. \quad (8)$$

Step 2: Now, for any $R > 0$, an application of triangle inequality leads to

$$\sup_{x \in \mathbb{R}} |f(x) - f_0(x)| \leq \sup_{x \in \mathbb{R}} |f(x) - f_R(x)| + \sup_{x \in \mathbb{R}} |f_R(x) - f_{0,R}(x)| + \sup_{x \in \mathbb{R}} |f_{0,R}(x) - f_0(x)|.$$

From the assumption with f_0 in equation (5), we can rewrite the above bound as follows:

$$\sup_{x \in \mathbb{R}} |f(x) - f_0(x)| \leq \sup_{x \in \mathbb{R}} |f(x) - f_R(x)| + \sup_{x \in \mathbb{R}} |f_R(x) - f_{0,R}(x)| + \bar{C}/R,$$

as long as $R \geq R_0$. If we choose R such that $R \geq \max\{2\bar{C}/\varepsilon, R_0\}$ then we have $\bar{C}/R < \varepsilon/2$. Therefore, a direct application of union bounds shows that

$$\begin{aligned} \Pi_n \left(\sup_{x \in \mathbb{R}} |f(x) - f_0(x)| > \varepsilon \right) &\leq \Pi_n \left(\sup_{x \in \mathbb{R}} |f_R(x) - f(x)| + \sup_{x \in \mathbb{R}} |f_R(x) - f_{0,R}(x)| > \varepsilon/2 \right) \\ &\leq \Pi_n \left(\sup_{x \in \mathbb{R}} |f_R(x) - f(x)| > \varepsilon/4 \right) \\ &\quad + \Pi_n \left(\sup_{x \in \mathbb{R}} |f_R(x) - f_{0,R}(x)| > \varepsilon/4 \right). \end{aligned}$$

Step 2.1: For the second term $\Pi_n(\sup_{x \in \mathbb{R}} |f_R(x) - f_{0,R}(x)| > \varepsilon/4)$, equation (4) indicates that

$$\Pi_n \left(\sup_{x \in \mathbb{R}} \frac{|f_R(x) - f_{0,R}(x)|}{R} > \varepsilon'/4 \right) \rightarrow 0 \quad \text{a.s.} \quad P_0^\infty$$

when $\varepsilon' = \varepsilon/R$. It is equivalent to

$$\Pi_n \left(\sup_{x \in \mathbb{R}} |f_R(x) - f_{0,R}(x)| > \varepsilon/4 \right) \rightarrow 0 \quad \text{a.s.} \quad P_0^\infty.$$

Step 2.2: For the first term $\Pi_n(\sup_{x \in \mathbb{R}} |f_R(x) - f(x)| > \varepsilon/4)$, the bound in equation (8) indicates that

$$\Pi_n \left(\sup_{x \in \mathbb{R}} |f_R(x) - f(x)| > \varepsilon/4 \right) \leq \Pi_n \left(\frac{1}{\pi\sigma^2 R} e^{-\frac{1}{2}\sigma^2 R^2} > \varepsilon/4 \right) \leq \Pi_n \left(\frac{1}{\pi\sigma^2 R} e^{-\frac{1}{2}\sigma^2 R^2} > \frac{\bar{C}}{2R} \right),$$

where the second inequality is due to $\varepsilon > 2\bar{C}/R$. Putting the above results together, for strong consistency with respect to L_∞ we need to ensure that

$$\Pi_n \left(\frac{1}{\pi\sigma^2 R^2} e^{-\frac{1}{2}\sigma^2 R^2} > \frac{\bar{C}}{2R^2} \right) \rightarrow 0$$

as $n \rightarrow \infty$. It can be achieved by taking the prior on σ^2 to be sample size dependent and

$$\Pi \left(\sigma^2 < \tau(\tilde{C})/R^2 \right) < e^{-n \cdot g(R)} \tag{9}$$

where $\tau(\cdot)$ is the inverse function of $\exp(-z/2)/z$, \tilde{C} is some universal constant, and $g(R)$ is an arbitrary choice of increasing function with $g(0) = 0$ and $g(\infty) = \infty$.

The condition (9) on the prior σ^2 is an instance of the general theory for strong consistency of posterior distribution for a family of supersmooth and ordinary smooth density functions. The proof strategy for this specific class of normal mixtures provides a general recipe for obtaining such general theory that we will establish in the next section. We should add that it is a common feature in the literature to consider a sequence of sample size dependent prior distributions; see for example [Ghosal et al., 2000].

4 General Theory

Based on the insight from the specific class of normal mixtures in Section 3 and the insight on the convergence of f_R to $f_{0,R}$ uniformly as long as f converges weakly to f_0 , we are now ready to state our conditions for the strong consistency with respect to the \mathbb{L}_∞ norm.

To simplify the presentation, we denote by $\mathcal{S}_{\alpha,\sigma}$ the set of supersmooth density function f of order $\alpha > 0$ with scale parameter $\sigma > 0$. Furthermore, we define $\mathcal{O}_{\beta,\sigma}$ to be the set of ordinary smooth density function f of order $\beta > 0$ with scale parameter $\sigma > 0$. We have the following result for the strong consistency with respect to \mathbb{L}_∞ norm.

Theorem 1. (1) (*Supersmooth setting*) Assume that $\alpha > 0$ and the sample size dependent prior Π on σ satisfies

$$\Pi\left(\sigma^2 \leq \tilde{C}_1/R^{\alpha/2}\right) \leq \exp(-\tilde{C}_2 n \cdot g_1(R))$$

for any $R \geq \tilde{R}_1$ where $\tilde{C}_1, \tilde{C}_2, \tilde{R}_1$ are some universal constants and g_1 is an increasing function in R with $g_1(0) = 0$ and $g_1(\infty) = \infty$. Then, we obtain that

$$\Pi_n\left(f \in \mathcal{S}_{\alpha,\sigma} : \sup_{x \in \mathbb{R}^d} |f(x) - f_0(x)| > \varepsilon\right) \rightarrow 0 \quad a.s. \quad P_0^\infty$$

for all $\varepsilon > 0$.

(2) (*Ordinary smooth setting*) Assume that $\beta > 1$ and the sample size dependent prior Π on σ satisfies that

$$\Pi\left(\sigma^{2+2(d-1)/\beta} < \bar{c}_1/R^{(\beta-1)/2}\right) < \exp(-\bar{c}_2 n \cdot g_2(R)),$$

for any $R \geq \bar{R}_1$ where $\bar{c}_1, \bar{c}_2, \bar{R}_1$ are some universal constants and g_2 is an increasing function in R with $g_2(0) = 0$ and $g_2(\infty) = \infty$. Then, we obtain that

$$\Pi_n\left(f \in \mathcal{O}_{\beta,\sigma} : \sup_{x \in \mathbb{R}^d} |f(x) - f_0(x)| > \varepsilon\right) \rightarrow 0 \quad a.s. \quad P_0^\infty$$

for all $\varepsilon > 0$.

Before putting the proof we again emphasize the weakness of the conditions to secure the supremum consistency, weaker even than those currently used for the \mathbb{L}_1 consistency.

Proof. (1) We first consider the supersmooth setting. The proof argument is a generalization of that in Section 3 for normal mixtures. Here, we provide the proof for the completeness. For any $R > 0$ and $x \in \mathbb{R}^d$, an application of triangle inequality leads to

$$\sup_{x \in \mathbb{R}^d} |f(x) - f_0(x)| \leq \sup_{x \in \mathbb{R}^d} |f(x) - f_R(x)| + \sup_{x \in \mathbb{R}^d} |f_R(x) - f_{0,R}(x)| + \sup_{x \in \mathbb{R}^d} |f_{0,R}(x) - f_0(x)|.$$

From equation (5), by choosing $R \geq \max\{2\bar{C}/\varepsilon, R_0\}$, we have $\varepsilon_R < \varepsilon/2$. A direct application

of union bounds leads to

$$\begin{aligned}\Pi_n \left(\sup_{x \in \mathbb{R}^d} |f(x) - f_0(x)| > \varepsilon \right) &\leq \Pi_n \left(\sup_{x \in \mathbb{R}^d} |f_R(x) - f(x)| + \sup_{x \in \mathbb{R}^d} |f_R(x) - f_{0,R}(x)| > \varepsilon/2 \right) \\ &\leq \Pi_n \left(\sup_{x \in \mathbb{R}^d} |f_R(x) - f(x)| > \varepsilon/4 \right) \\ &\quad + \Pi_n \left(\sup_{x \in \mathbb{R}^d} |f_R(x) - f_{0,R}(x)| > \varepsilon/4 \right).\end{aligned}$$

For the second term $\Pi_n(\sup_{x \in \mathbb{R}^d} |f_R(x) - f_{0,R}(x)| > \varepsilon/4)$, equation (4) indicates that

$$\Pi_n \left(\sup_{x \in \mathbb{R}^d} \frac{|f_R(x) - f_{0,R}(x)|}{R^d} > \varepsilon'/4 \right) \rightarrow 0 \quad \text{a.s.} \quad P_0^\infty$$

when $\varepsilon' = \varepsilon/R^d$. It is equivalent to

$$\Pi_n \left(\sup_{x \in \mathbb{R}^d} |f_R(x) - f_{0,R}(x)| > \varepsilon/4 \right) \rightarrow 0 \quad \text{a.s.} \quad P_0^\infty.$$

To consider the first term $\Pi_n(\sup_{x \in \mathbb{R}^d} |f_R(x) - f(x)| > \varepsilon/4)$; we look at the smoothness conditions. Since f is supersmooth density function, from part (1) of Proposition 1, we obtain that

$$\begin{aligned}\Pi_n \left(\sup_{x \in \mathbb{R}^d} |f_R(x) - f(x)| > \varepsilon/4 \right) &\leq \Pi \left(C \frac{R^{\max\{1-\alpha, 0\}}}{\sigma^{2d}} \exp(-C_1 \sigma^2 R^\alpha) > \varepsilon/4 \right) \\ &\leq \Pi \left(C \frac{R^{\max\{1-\alpha, 0\}}}{\sigma^{2d}} \exp(-C_1 \sigma^2 R^\alpha) > \bar{C}/(2R) \right),\end{aligned}$$

where the second inequality is due to the fact that $R \geq \max\{2\bar{C}/\varepsilon, R_0\}$; since the inequality

$$C R^{\max\{1-\alpha, 0\}} \exp(-C_1 \sigma^2 R^\alpha) / \sigma^{2d} > \bar{C}/(2R)$$

implies $\lim_{R \rightarrow \infty} \sigma^2 R^\alpha / \log R < \infty$. Therefore, there exist R_1 and \bar{C}_1 such that $\sigma^2 \leq \bar{C}_1 \log R / R^\alpha$ as long as $R \geq R_1$. Putting the above results together, we find that

$$\Pi \left(C \frac{R^{\max\{1-\alpha, 0\}}}{\sigma^{2d}} \exp(-C_1 \sigma^2 R^\alpha) > \bar{C}/(2R) \right) \leq \Pi(\sigma^2 \leq \bar{C}_1 \log R / R^\alpha)$$

as long as $R \geq \max\{2\bar{C}/\varepsilon, R_0, R_1\}$. Since $\log(R)/R^{\alpha/2} \rightarrow 0$ as $R \rightarrow \infty$, we can find R_2 such that $\log R / R^{\alpha/2} \leq \tilde{C}_1 / \bar{C}_1$ as long as $R \geq R_2$. Collecting these results, when $R \geq \max\{2\bar{C}/\varepsilon, R_0, R_1, R_2\}$, we have

$$\Pi(\sigma^2 \leq \bar{C}_1 \log R / R^\alpha) \leq \Pi(\sigma^2 \leq \tilde{C}_1 / R^{\alpha/2}) \leq \exp(-\tilde{C}_2 n \cdot g_1(R)) \leq \exp(-\tilde{C}_2 n \cdot g_1(R_0)),$$

where the final inequality is due to the increasing property of g_1 . As a consequence, we have

$$\Pi_n \left(\sup_{x \in \mathbb{R}^d} |f(x) - f_0(x)| > \varepsilon \right) \rightarrow 0 \quad \text{a.s.} \quad P_0^\infty$$

for all $\varepsilon > 0$.

(2) We now move to the ordinary smooth setting. We denote $\gamma = \min\{1, (\beta - 1)/4\}$. From equation (5), as long as $R \geq \max\{R_0, 1\}$, we have

$$\varepsilon_R = \sup_{x \in \mathbb{R}^d} |f_{0,R}(x) - f_0(x)| \leq \bar{C}/R^\gamma.$$

We follow the similar argument as the supersmooth case by choosing $R \geq \max\{(2\bar{C}/\varepsilon)^{1/\gamma}, R_0, 1\}$. This choice of R is to guarantee that $\varepsilon_R < \varepsilon/2$. Therefore, we also obtain that

$$\begin{aligned} \Pi_n \left(\sup_{x \in \mathbb{R}^d} |f(x) - f_0(x)| > \varepsilon \right) &\leq \Pi_n \left(\sup_{x \in \mathbb{R}^d} |f_R(x) - f(x)| > \varepsilon/4 \right) \\ &\quad + \Pi_n \left(\sup_{x \in \mathbb{R}^d} |f_R(x) - f_{0,R}(x)| > \varepsilon/4 \right). \end{aligned}$$

The second term also approaches 0 via similar argument as the supersmooth case. For the first term, the result of part (2) of Proposition 1 for ordinary smooth function indicates that

$$\begin{aligned} \Pi_n \left(\sup_{x \in \mathbb{R}^d} |f_R(x) - f(x)| > \varepsilon/4 \right) &\leq \Pi \left(\frac{c}{\sigma^{2+2(d-1)/\beta} R^{\beta-1}} > \bar{C}/(2R^\gamma) \right) \\ &= \Pi \left(\sigma^{2+2(d-1)/\beta} < 2c/(\bar{C}R^{\beta-(1+\gamma)}) \right). \end{aligned}$$

Since $2c/(\bar{C}R^{\beta-(1+\gamma)}) < \bar{c}_1/R^{(\beta-1)/2}$ where \bar{c}_1 is a constant in Theorem 1 as long as $R^{\frac{\beta-1}{2}-\gamma} > c/(\bar{c}_1\bar{C})$, it indicates that as long as $R \geq \max\{(2\bar{C}/\varepsilon)^{1/\gamma}, (c/(\bar{c}_1\bar{C}))^{2/(\beta-1-2\gamma)}, R_0, 1\}$, we have

$$\Pi \left(\sigma^{2+2(d-1)/\beta} < 2c/(\bar{C}R^{\beta-(1+\gamma)}) \right) \leq \Pi \left(\sigma^{2+2(d-1)/\beta} < \bar{c}_1/R^{(\beta-1)/2} \right) < \exp(-\bar{c}_2 n \cdot g_2(1)) \rightarrow 0$$

as $n \rightarrow \infty$ where \bar{c}_2 is the universal constant in Theorem 1 and the final inequality is due to the increasing property of g_2 . Putting the above result together, we obtain that

$$\Pi_n \left(\sup_{x \in \mathbb{R}^d} |f(x) - f_0(x)| > \varepsilon \right) \rightarrow 0 \quad \text{a.s.} \quad P_0^\infty$$

for all $\varepsilon > 0$. □

5 Discussion

At the heart of the paper is the inequality

$$\sup_x |f(x) - f_0(x)| \leq \sup_x |f_R(x) - f(x)| + \sup_x |f_{0,R}(x) - f_0(x)| + \sup_x |f_{0,R}(x) - f_R(x)|,$$

valid for all $R > 0$, where the final term can be expressed as

$$\left| \int g_R(y) (f(y-x) - f_0(y-x)) dy \right|,$$

with $g_R(y) = \prod_{j=1}^d \sin(Ry_j)/y_j$ for any $y = (y_1, \dots, y_d)$. The first term is about enforcing some notion of smoothness on f and the final term is handled by weak convergence.

For one dimensional setting, another inequality based on the triangle inequality involves using

$$f_h(x) = \frac{1}{2h} [F(x+h) - F(x-h)],$$

as used by [Chae and Walker \[2017\]](#). We can now determine that

$$\sup_x |f_h(x) - f(x)| \leq \sup_{|x-y|<h} |f(x) - f(y)|$$

and so if f and f_0 belong to a Hölder class with radius L and smoothness parameter β , then

$$\sup_x |f(x) - f_0(x)| \leq d_K(f, f_0)/h + 2h^\beta,$$

for any $h > 0$. Here

$$d_K(f, f_0) = \sup_x |F(x) - F_0(x)|$$

is the Kolmogorov distance where F and F_0 are cumulative distribution functions of f and f_0 . This can be upper bounded by the Prokhorov metric,

$$d_K(f, f_0) \leq d_P(f, f_0) (1 + \min\{\|f\|_\infty, \|f_0\|_\infty\}).$$

See for example [\[Gibbs and Su, 2002\]](#).

Hence, we should also be able to demonstrate sup norm consistency for a β Hölder class of density once we have established weak consistency. The only condition for which we might need to construct a specific suitable prior for is the required boundedness of $\|f\|_\infty$.

A succinct summary of the general technique using

$$d(f, f_0) \leq d(\bar{f}, f) + d(\bar{f}_0, f_0) + d(\bar{f}_0, \bar{f}),$$

where \bar{f} is a smooth version of f , is that the last term is handled using weak convergence, the middle term is assumed to be small, and so consistency with metric d follows with some condition on $d(\bar{f}, f)$. For example, a samples size dependent prior would consider $\Pi(d(\bar{f}, f) > \epsilon) < e^{-n\epsilon}$ for all large n and $\epsilon > 0$.

Finally, we would like to mention posterior rates of convergence. It is not a difficult task to demonstrate rates of convergence equivalent to those currently appearing in the literature, and we would anticipate with weaker conditions on the types of prior considered. However, we believe this is best reported in a future paper. The aim of the present paper is merely to point out the new technique and how elegant and useful it has turned out to be.

References

- A. Barron, M.J. Schervish, and L. Wasserman. The consistency of posterior distributions in nonparametric problems. *Annals of Statistics*, 27, 1999. (Cited on page [1](#).)
- S. Bochner. *Lectures on Fourier Integrals*. Princeton University Press, 1959. (Cited on page [2](#).)
- I Castillo. On bayesian supremum norm contraction rates. *Annals of Statistics*, 42:2058–2091, 2014. (Cited on page [2](#).)
- M Chae and S.G. Walker. A novel approach to bayesian consistency. *Electronic Journal of Statistics*, 11:4723–4745, 2017. (Cited on pages [2](#) and [10](#).)

- S. Ghosal and A.W. van der Vaart. *Fundamentals of Nonparametric Bayesian Inference*. Cambridge University Press, 2017. (Cited on page 1.)
- S. Ghosal, J.K. Ghosh, and R.V. Ramamoorthi. Posterior consistency of Dirichlet mixtures in density estimation. *Annals of Statistics*, 27:143–158, 1999. (Cited on page 1.)
- S. Ghosal, J. K. Ghosh, and A. van der Vaart. Convergence rates of posterior distributions. *Annals of Statistics*, 28:500–531, 2000. (Cited on page 6.)
- A L Gibbs and F E Su. On choosing and bounding probability metrics. *International Statistical Review*, 70:419–435, 2002. (Cited on page 10.)
- E Gine and R Nickl. Rates on contraction for posterior distributions in l_r metrics, $1 \leq r \leq \infty$. *Annals of Statistics*, 39:2883–2911, 2011. (Cited on page 2.)
- N. Ho and S.G. Walker. Multivariate smoothing via the Fourier integral theorem and Fourier kernel. *Arxiv preprint*, 2021. (Cited on page 3.)
- M Hoffmann, J Rousseau, and J Schmidt-Hieber. On adaptive posterior concentration rates. *Annals of Statistics*, 43:2259–2295, 2015. (Cited on page 2.)
- C Scricciolo. Adaptive bayesian density estimation in l_p metrics with pitman–yor or normalized inverse–gaussian process kernel mixtures. *Bayesian Analysis*, 9:475–520, 2014. (Cited on page 2.)
- S.G. Walker. New approaches to Bayesian consistency. *Annals of Statistics*, 32, 2004. (Cited on page 1.)
- N. Wiener. *The Fourier Integral and Certain of its Applications*. Cambridge University Press, 1933. (Cited on page 2.)