Statistica Sinica **34** (2024), 1281-1297
doi:https://doi.org/10.5705/ss.202022.0062

BAYESIAN CONSISTENCY WITH THE SUPREMUM METRIC

Nhat Ho and Stephen G. Walker*

University of Texas at Austin

Abstract: We present conditions for Bayesian consistency in the supremum metric. The key to the technique is a triangle inequality that 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 that could track the data too closely. Our main result is that we demonstrate supremum consistency using conditions comparable with those currently used to secure $\mathbb{L}_1$-consistency.

Key words and phrases: Fourier integral theorem, Prokhorov metric, sinc kernel, weak convergence.

1. Introduction

Bayesian consistency remains an open topic, and has seen much progress since the seminal papers of Barron, Schervish and Wasserman (1999) and Ghosal, Ghosh and Ramamoorthi (1999). A dominating sufficient, but not necessary condition is the Kullback–Leibler support condition for the prior,

$$\Pi\big(D(f_0, f) < \varepsilon\big) > 0, \tag{1.1}$$

for all $\varepsilon > 0$. Here, $D(f_0, f) = \int f_0 \log(f_0/f)$ denotes the Kullback–Leibler divergence between f_0 and f, and f_0 represents the true density function from which the identically distributed $(X_i)_{i=1:n}$ are observed. Furthermore, we write $\Pi(df)$ to denote the prior distribution on a space of probability density functions, say, $\mathbb{P}$.

It is well known that condition (1.1) is not sufficient for strong consistency. Strong consistency holds if

$$\Pi_n(A_\varepsilon) := \Pi(A_\varepsilon \mid X_{1:n}) \to 0 \quad \text{a.s.} \quad P_0^\infty, \tag{1.2}$$

for all $\varepsilon > 0$, where $A_\varepsilon = \{f : d_H(f_0, f) > \varepsilon\}$ and d_H is the Hellinger distance between f_0 and f. Note the the Hellinger distance is equivalent to the $\mathbb{L}_1$-distance. Barron, Schervish and Wasserman (1999) provide a counterexample that shows that a posterior is not strongly consistent, given only the Kullback–

*Corresponding author.

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 $n \to \infty$, such that the size of $\mathbb{F}_n$, as measured by some suitable entropy, is bounded by $e^{n\kappa}$, 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. This automatically satisfies the entropy condition, and the $\mathbb{F}_n'$ condition is satisfied when $\sum_{j=1:\infty} \sqrt{\Pi(A_j)} < \infty$, where $(A_j)_{j=1}^{\infty}$ 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, the $\mathbb{L}_1$-metric, 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, Chae and Walker (2017) use $\bar{f}(x) = [F(x+h) - F(x-h)/(2h)$, for some smoothing parameter $h > 0$, in a 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 obtain conditions for strong consistency with respect to the supremum metric on $\mathbb{R}^d$, that is, $\mathbb{L}_\infty(\mathbb{R}^d)$. We believe that we are the first to consider this problem for density estimation. Previous works on the supremum metric have been done on $[0,1]^d$, including the work of Castillo (2014), who considers contraction rates, assuming that the true density on $(0,1)$ is bounded away from zero and satisfies $\log f_0 \in \mathcal{C}_\alpha(0,1)$; that is, it is Hölder smooth, with coefficient α. Other works on $[0,1]^d$ include that of Shen and Ghosal (2017), who consider densities of the form $f(x \mid \theta) \propto \Psi(\theta' b(x))$, for some basis function b and some fixed continuously differentiable function Ψ. Assumptions made include $\Psi^{-1}(f_0) \in \mathcal{C}_\alpha$, for a known α, and f_0 is bounded away from zero. Other works on consistency and rates using the $\mathbb{L}_r$-metrics and others include Gine and Nickl (2011), Hoffmann, Rousseau and Schmidt-Hieber (2015), Scricciolo (2014), and Li and Ghosal (2021). Related works, though with a fundamental difference, consider the supremum metric for consistency with respect to the standard nonparametric regression model $y_i = f(x_i) + \epsilon_i$; see Yoo and Ghosal (2016), Yoo, Rousseau and Rivoirard (2018), and Li and Ghosal (2020).

We focus on supremum consistency on $\mathbb{R}^d$, particularly for $d = 1$. We start with the triangle inequality

$$|f_0(x) - f(x)| \le |f_0(x) - f_{0,R}(x)| + |f_{0,R}(x) - f_R(x)| + |f(x) - f_R(x)|, \quad (1.3)$$

where f_R is an alternative kernel-smoothed version of f, specifically, using the sinc kernel. That is,

$$f_R(x) := \frac{1}{\pi^d} \int_{\mathbb{R}^d} \prod_{j=1}^{d} \frac{\sin(R(x_j - y_j))}{x_j - y_j} f(y)\, dy, \quad (1.4)$$

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

Here we focus solely on consistency, because weakening the conditions on prior distributions in order to achieve consistency remains an important topic. These weakened conditions can then be used to achieve benchmark rates of convergence, it is argued, with some technical applications. However, we derive our insights by examining how the weakening of the assumptions required for consistency arises.

The remainder of the paper is as follows. In Section 2, we outline the assumptions and initial results for the general theory. In Section 3, we establish the posterior supremum consistency for the widely used infinite normal mixture model. We conclude the paper with a discussion in Section 4. Additional proofs are provided in the Appendix.

2. General Theory

We start with equation (1.3), and consider the posterior $\Pi_n(d_\infty(f_0, f) > \epsilon)$, which is upper bounded by

$$\Pi_n(d_\infty(f_0, f) > \epsilon) \le \Pi_n\left(d_\infty(f_{0,R}, f_R) > \frac{\epsilon - \bar{d}}{2}\right) + \Pi_n\left(d_\infty(f, f_R) > \frac{\epsilon - \bar{d}}{2}\right),$$

where $\bar{d} = d_\infty(f_0, f_{0,R})$. Our assumption is that for any $\epsilon > 0$, there exists an $R < \infty$ for which $d_\infty(f_0, f_{0,R}) < \epsilon/2$. Equivalently, $\lim_{R\to\infty} d_\infty(f_0, f_{0,R}) = 0$. Establishing f_0 for which this holds forms the main theoretical content of this paper.

In the second subsection, we consider the term $d_\infty(f, f_R)$, and motivate the need to have R be sample-size dependent, written as R_n. We show how a prior condition allows us to achieve $\Pi_n(d_\infty(f, f_{R_n}) > \epsilon) \to 0$. In the final subsection, we examine the term $d_\infty(f_{R_n}, f_{0,R_n})$. Indeed, we need R_n to be related to the Prokhorov rate $\widetilde{\epsilon}_n$, that is, $\Pi_n(d_P(f, f_0) > \widetilde{\epsilon}_n) \to 0$, which is guaranteed by the Kullback–Leibler support condition.

2.1. The term $d_\infty(f_0, f_{0,R})$

First, we consider $d_\infty(f_0, f_{0,R})$. For supremum consistency, we require that

$$\lim_{R\to\infty} d_\infty(f_{0,R}, f_0) = 0. \tag{2.1}$$

For $f_{0,R}$ to exist and be close to f_0, we require the very mild condition that $f(x+)$ and $f_0(x-)$ exist for all x, and

$$\int_0^\delta \frac{f_0(x+t) - f_0(x-)}{t}\,dt \quad \text{and} \quad \int_0^\delta \frac{f_0(x+t) - f_0(x-)}{t}\,dt$$

both exist for some $\delta > 0$. Then,

$$\frac{1}{2}(f_0(x+) + f_0(x-)) = \pi^{-1} \lim_{R\to\infty} \int \frac{\sin(R(x-y))}{x-y} f_0(y)\,dy.$$

At points of discontinuity, we can define the value of $f_0(x)$ as $(1/2)(f_0(x+) + f_0(x-))$, though to keep things simple, we assume that all density functions are continuous; that is, $f(x+) - f(x-) = 0$, for all x.

To obtain bounds for $d_\infty(f_0, f_{0,R})$, we make certain smoothness assumptions. We define the following notions of supersmooth and ordinary smooth density functions. To simplify the presentation, $\widehat{f}$ denotes the Fourier transform of the function f. Though we use f, strictly speaking the following is only required for f_0, and we drop the subscript 0 temporarily.

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

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

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

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

The supersmooth and ordinary smooth notions have been used in deconvolution problems; see, for example, Fan (1991) and Zhang (1990). Examples of supersmooth functions include mixtures of location Gaussian distributions and mixtures of location Laplace distributions, with a similar scale parameter. In particular, when we have $f(x) = \sum_{j=1}^k \omega_i \mathrm{N}(x|\mu_j, \sigma^2 I_d)$, where $1 \leq k \leq \infty$, then f is a supersmooth density function of order 2, with scale parameter σ. When

f is a mixture of location Cauchy distributions with the same scale parameter $\sigma^2 I_d$, then f is a supersmooth density function of order 1, with scale parameter σ. Examples of ordinary smooth functions include mixtures of location Cauchy distributions with a similar scale parameter $\sigma^2 I_d$. In this case, these mixtures are ordinary smooth functions of order 2, with scale parameter σ.

Based on Definition 1, we have the following results on the difference between f_R and f. These are fundamental to our approach because they set a sup bound between f and f_R.

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*

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

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

(2) *Assume that f is an ordinary smooth density function of order $\beta > 1$, with scale parameter σ. Then, there exists a universal constant c such that*

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

The proof of Proposition 1 is provided in Appendix B. The results of Proposition 1 imply that for sufficiently large R, we have that $\sup_{x\in\mathbb{R}^d} |f_0(x) - f_{0,R}(x)|$ is arbitrarily small.

Existing works relate the smoothness of f to the tail behavior of Fourier transforms; see (Nissila (2021)). One of the results from Theorem 2.1 in Nissila (2021) is that if $f \in W_{1,1}(\mathbb{R})$ and $f \in C_\beta(\mathbb{R})$, for any $\beta > 0$, then f is ordinary smooth. That is, if $\int |f'| < \infty$ and f is Hölder smooth with coefficient β, then f is ordinary smooth.

We now discuss a direct result that avoids using tails of Fourier transforms.

Proposition 2. *If f is Hölder smooth on $\mathbb{R}$, for some positive coefficient, and $\int |f'| < \infty$, then*

$$\lim_{R\to\infty} \sup_x |f_R(x) - f(x)| = 0.$$

Proof. We can write

$$f_R(x) = \int_{-\infty}^{+\infty} f'(s) \int_{-\infty}^{R(x-s)} \frac{\sin z}{\pi z}\, dz\, ds.$$

Now, split the outer integral into three parts, namely, those between $(-\infty, x-\epsilon_R)$,

$(x-\epsilon_R, x+\epsilon_R)$, and $(x+\epsilon_R, \infty)$, where $\epsilon_R \to 0$ and $R\epsilon_R \to \infty$. For the first part, the inner integral has $R(x-s) > R\epsilon_R$, and so for all x, this inner integral acts as $1-\delta_R$, where $\delta_R = 1/(R\epsilon_R)$, based on the asymptotic result

$$\int_{-\infty}^{\xi} \frac{\sin z}{\pi z}\, dz = 1 - \frac{c}{\xi} + o\left(\frac{1}{\xi}\right) \quad \text{as} \quad \xi \to \infty,$$

for some $c > 0$. For the third part, $R(x-s) < -R\epsilon_R$, and so the inner integral now acts as δ_R, for all x. Hence, we can write

$$\begin{aligned} f_R(x) = &\int_{-\infty}^{x-\epsilon_R} f'(s)\, ds\, (1-\delta_R) + \int_{x+\epsilon_R}^{\infty} f'(s)\, ds\, \delta_R \\ &+ \int_{x-\epsilon_R}^{x+\epsilon_R} f'(s) \int_{-\infty}^{R(x-s)} \frac{\sin z}{\pi z}\, dz\, ds + o(\delta_R). \end{aligned}$$

Because the sinc integral is bounded and $\int |f'| < \infty$, we have

$$f_R(x) = f(x+\epsilon_R) + M\ \{f(x+\epsilon_R) - f(x-\epsilon_R)\} + O(\delta_R),$$

for some $M < \infty$. Hence,

$$\sup_x |f_R(x) - f(x)| \le M^* \sup_x |f(x+\epsilon_R) - f(x)| + O(\delta_R),$$

for some finite M^*, and this goes to zero as $R \to \infty$ under the assumption that f is Hölder smooth for some coefficient. This is precisely the result in Nissila (2021). Note that Hölder smoothness arises from the condition $||f'||_\infty < \infty$, which can be proven using the mean value theorem.

Another smoothness assumption is provided in Shen, Tokdar and Ghoshal (2013). For some envelope function $L(x)$, the density $f \in C(\beta, L, \tau)$ if

$$|D^k f(x+y) - D^k f(x)| \le L(x)\, \exp(\tau\, y^2)\, |y|^{\beta - \lfloor \beta \rfloor},$$

for some $\tau > 0$ and $\beta > 0$ and for all $k \le \lfloor \beta \rfloor$, where D^k denotes the kth derivative. To obtain the desired result for this class, we require that $L(x)$ is bounded, so that f is bounded. If not, for example, if $f(0) = \infty$, then we cannot establish $\sup_x |f_R(x) - f(x)| \to 0$ as $R \to \infty$, because $f(0) = \infty$. If L is bounded and β is not an integer, then we have

$$\sup_x |f(x+\epsilon) - f(x)| \le \exp(\tau)\, \sup_x\, L(x)\, |\epsilon|^{\beta - \lfloor \beta \rfloor} \to 0 \quad \text{as} \quad \epsilon \to 0.$$

If β is an integer, for example, $\beta = 1$, then the smoothness assumption is equivalent to $|f(x+\epsilon) - f(x)|$ and $|f'(x+\epsilon) - f'(x)|$ being bounded by some constant, which is not sufficiently smooth.

Hence, with the appropriate smoothness conditions, for any $\epsilon > 0$, we can find an R large enough such that $d_\infty(f_0, f_{0,R}) < \epsilon/2$. Thus, we need only to consider

$$\Pi_n(d_\infty(f_0, f) > \epsilon) \le \Pi_n\left(d_\infty(f_{0,R}, f_R) > \frac{\epsilon}{4}\right) + \Pi_n\left(d_\infty(f, f_R) > \frac{\epsilon}{4}\right),$$

which we focus on in the following two subsections.

To deal with these two terms, we introduce the notion that allows R to be sample-size dependent. We set $R = R_n \to \infty$ and require R_n to be connected to the Prokhorov rate of convergence, guaranteed to exist under the assumption of weak convergence, via the assumption that f_0 is in the Kullback–Leibler support of the prior. This is the same assumption made in Theorem 7 of Ghosal, Ghosh and Ramamoorthi (1999).

2.2. The term $d_\infty(f, f_R)$

For this part, R_n is sample-size dependent, and we need to ensure the posterior satisfies $\Pi_n(d_\infty(f, f_{R_n}) > \epsilon) \to 0$ a.s. To achieve this, we rely on the notion that R_n satisfies a prior condition of the type

$$\Pi\big(d_\infty(f, f_{R_n}) > \epsilon\big) < \exp(-n\, c_\epsilon), \tag{2.2}$$

for some $c_\epsilon > 0$. It is well known that equation (2.2) with the Kullback-Leibler support condition implies that the posterior satisfies $\Pi_n(d_\infty(f, f_{R_n}) > \epsilon) \to 0$ a.s.; that is, if the prior mass on a sequence of sets is exponentially small, then the posterior mass on the sequence of sets tends to zero. The setting of R_n is therefore problem specific, though we assume that $d_\infty(f, f_R) > \epsilon \Rightarrow \tau(f) < \alpha(R, \epsilon)$, for some functional τ.

2.3. The Term $d_\infty(f_R, f_{0,R})$

If the prior puts positive mass on all Kullback–Leibler neighborhoods of f_0, that is, equation (1.1), then the posterior converges on weak neighborhoods of f_0 (Schwartz (1965)). Hence, there exists $\widetilde{\epsilon}_n$ for which $\Pi_n(d_P(f_0, f) > \widetilde{\epsilon}_n) \to 0$ a.s., where d_P denotes the Prokhorov distance, given by

$$d_P(g, f) = \inf_{\epsilon > 0} \{G(A) \le F(A^\epsilon) + \epsilon \quad \text{and} \quad F(A) \le G(A^\epsilon) + \epsilon \quad \text{for all} \quad A\},$$

where $A^\epsilon = \{b \,|\, \exists a \in A, |b - a| < \epsilon\}$. We also define $A/R = \{b \,|\, \exists a \in A, b = a/R\}$ and $A - x = \{b \,|\, \exists a \in A, b = a - x\}$. Our first result follows.

Theorem 1. *If* $d_P(f, f_0) < \epsilon$ *and* $f_{x,R}(t) = f(t/R - x)/R$, *then* $\sup_x d_P(f_{x,R}, f_{0,x,R}) < R\epsilon$.

Proof. Given the assumption for f and f_0, we have that $F(A) \leq F_0(A^\epsilon) + \epsilon$, for all sets A. Now,

$$F_{x,R}(A) = F\left(\frac{A}{R} - x\right) \leq F_0\left(\left(\frac{A}{R} - x\right)^\epsilon\right) + \epsilon$$

and $(A/R - x)^\epsilon = A^{R\epsilon}/R - x$. For example, if $A = (a, b)$, then

$$\left(\frac{A}{R} - x\right)^\epsilon = \left(\frac{a}{R} - x - \epsilon, \frac{b}{R} - x + \epsilon\right) = \frac{a - R\epsilon, b + R\epsilon}{R} - x.$$

Thus, for all x, A, and R, we have $F_{x,R}(A) \leq F_{0,x,R}(A^{R\epsilon}) + \epsilon$, and because $R \to \infty$, we can put the last term ϵ as $R\epsilon$. This implies $\sup_x d_P(f_{x,R}, f_{0,x,R}) < R\epsilon$, completing the proof.

Theorem 1 yields

$$\Pi_n\left(\sup_x \left|\int \frac{\sin t}{t}\,(f_{x,R_n}(t) - f_{0,x,R_n}(t))\; dt\right| \gtrsim \widetilde{\epsilon}_n\, R_n\right) \to 0,$$

because $(\sin t)/t$ is a continuous and bounded function. Hence,

$$\Pi_n\left(\sup_x \left|\frac{\sin(R_n(x-y))}{x-y}\,(f(y) - f_0(y))\; dy\right| \gtrsim \widetilde{\epsilon}_n\, R_n^2\right) \to 0;$$

that is, $\Pi_n(d_\infty(f_{R_n}, f_{0,R_n}) > \epsilon) \to 0$, for all $\epsilon > 0$, under the constraint on R_n that $\widetilde{\epsilon}_n R_n^2 \to 0$.

Combining these conditions, the assumptions required are extremely mild. Other than the smoothness condition on f_0, the only substantial requirement is that of (2.2). Another problem is related to the Prokhorov rate of convergence, which we discuss briefly. An upper bound for the rate is needed. Given a prior Π with the Kullback–Leibler support condition, there exists a rate $\widetilde{\epsilon}_n$ such that $\Pi_n(d_P(f, f_0) > \widetilde{\epsilon}_n) \to 0$. Additional conditions, specifically, setting the parameters of the prior to ensure L_1-consistency, do not disturb the Prokhorov rate. The posterior $\mathbb{L}_1$-rates, say ϵ_n^*, are typically known and are well documented in the literature. Thus, an upper bound for $\widetilde{\epsilon}_n$, which is required, is provided by ϵ_n^*. This follows easily, because the Prokhorov distance is upper bounded by the $\mathbb{L}_1$-distance. Therefore, we obtain the following general consistency result.

Theorem 2. *Suppose the prior has f_0 in the Kullback–Leibler support of the prior Π, and f_0 satisfies $\lim_{R\to\infty} d_\infty(f_0, f_{0,R}) = 0$ (see Section 2.1 for details). If $\widetilde{\epsilon}_n$ is the Prokhorov posterior rate and R_n is set to satisfy $\widetilde{\epsilon}_n R_n^2 \to 0$, with the prior satisfying $\Pi(d_\infty(f, f_{R_n}) > \epsilon) < \exp(-nc_\epsilon)$, for some $c_\epsilon > 0$ and $c_0 = 0$, with c_ϵ increasing as ϵ increases, then the posterior is consistent with respect to the supremum metric.*

We now present an illustration of Theorem 2 using normal mixtures.

3. Mixture of Normal Distribution

Here, we consider normal mixtures, one of the most widely used nonparametric models. To keep things simple, we consider the normal mixture model in dimension $d = 1$, where

$$f(x) = \sum_{j=1}^{\infty} \frac{w_j\,\phi((x-\mu_j)/\sigma)}{\sigma}, \tag{3.1}$$

$(w_j)_{j=1}^{\infty}$ are a set of weights, $(\mu_j)_{j=1}^{\infty}$ are a set of locations, and σ is a variance term common to each normal component. Furthermore, ϕ represents the usual standard normal density function. In a Bayesian model, prior distributions are assigned to the weights, locations, and variance. The conditions we require for consistency for all f_0 in the Kullback–Leibler support of the prior amount to a condition on the prior for σ. For f_0 to be found in the Kullback–Leibler support, see Wu and Ghosal (2008). In the following, we have three main parts, corresponding to the three subsections in Section 2, labelled (i), (ii) and (iii), respectively.
(i) 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 by 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 that uses closed-form computations.

Theorem 3. *If f is a mixture model as in* (3.1)*, then*

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

Proof. We first show that

$$I(R) = \int_{-\infty}^{\infty} \cos(Rx)\,\phi(x)\,dx = e^{-(1/2)R^2}, \tag{3.2}$$

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 (3.2) holds because $I(0) = 1$. Now we consider

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

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

$$\int_{-\infty}^{\infty} \cos(R(y-x))\, \frac{\phi((x-\mu)/\sigma)}{\sigma} dx = \cos(R(y-\mu))\, e^{-(1/2)\sigma^2 R^2}, \tag{3.3}$$

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 (3.2). Thus, $J(R) = \int_0^R e^{-(1/2)s^2}\, ds$, because $J(0) = 0$. Hence, we have that

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

We want to examine $f_R(x) - f(x) = (1/\pi)J(x; \mu, \sigma, R) - \phi((x-\mu)/\sigma)/\sigma$. From equation (3.2), we have that

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

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

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

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

$$\sup_{x \in \mathbb{R}} |f_R(x) - f(x)| < \frac{1}{\pi \sigma^2 R} e^{-(1/2)\sigma^2 R^2}. \tag{3.4}$$

(ii) To set R_n, as a result of (3.4), we require

$$\Pi \left(\frac{e^{-(1/2)\sigma^2 R_n^2}}{\pi \sigma^2 R_n^2} > \frac{\epsilon}{R_n} \right) < \exp(-n c_\epsilon),$$

for all large n and for some $c_\epsilon > 0$. Ignoring the π term, we require Π on σ^2 such that

$$\Pi\left(\frac{\exp(-0.5\sigma^2 R_n^2)}{\sigma^2 R_n^2} \gtrsim \frac{\epsilon}{R_n} \right) < \Pi \left(\sigma^2 \lesssim \frac{1}{\epsilon R_n^{3/2}} \right).$$

We then take $\Pi(\sigma^2 < \xi) = \exp(-(1/\xi)^b)$, an Inverse Weibull distribution, for some $b > 0$, and thus can take $R_n = n^{2/(3\widetilde{b})}$, for any $\widetilde{b} \leq b$.

(iii) To investigate the $\mathbb{L}_1$-rate, which gives an upper bound for the Prokhorov rate and a possible range of b-values, we follow Ghosal and van der Vaart (2007), who consider a mixture of normal distributions under assumed smooth conditions for the true f_0. Specifically, they assume f_0 is twice continuously differentiable,

$$\int \left(\frac{f_0''}{f_0}\right)^2 f_0 < \infty \quad \text{and} \quad \int \left(\frac{f_0'}{f_0}\right)^4 f_0 < \infty,$$

and $F_0[-a,a]^c < \exp(-ca^\gamma)$, for some $c,\gamma > 0$. These imply one of our conditions, namely, $\int |f_0'| < \infty$. Our conditions are covered by those in Ghosal and van der Vaart (2007). The rate of convergence is given in Theorem 2 of Ghosal and van der Vaart (2007) and is of the form $(\log n)^\kappa n^{-(1/2-\delta)}$, for some $\delta > 0$. Hence, we require $R_n < n^{1/4-\delta/2}/(\log n)^\kappa$. Thus, overall, we have $n^{1/(2\widetilde{b})} < R_n < n^{1/4-\delta/2}/(\log n)^\kappa$, implying that we must take $b > 1/((1/2)-\delta)$.

To summarize, we assume $\int |f_0'| < \infty$ and f_0 is Hölder smooth on $\mathbb{R}$, and with $\Pi(\sigma^2 < \xi) = \exp(-1/\xi^b)$, with $1/b < (1/2) - \delta$, where δ determines the $\mathbb{L}_1$-rate of convergence, yields consistency with respect to the supremum metric. We argue that is comparable to the conditions under which $\mathbb{L}_1$-consistency is guaranteed.

4. Discussion

At the heart of our study is the inequality

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

which is valid for all $R > 0$. The first term in the inequality enforces smoothness on f, the second term provides smoothness conditions on f_0, and the final term is handled by weak convergence.

For the one-dimensional setting, another inequality based on the triangle inequality involves using $f_h(x) = [F(x+h) - F(x-h)]/(2h)$, as in Chae and Walker (2017). We can now determine that $\sup_x |f_h(x) - f(x)| \le \sup_{|x-y|<h} |f(x) - f(y)|$, and so if f and f_0 belong to a Hölder class with radius L and smoothness parameter β, then

$$\sup_x |f(x) - f_0(x)| \le \frac{d_K(f,f_0)}{h} + 2h^\beta,$$

for any $h > 0$. Here, we define $d_K(f,f_0) := \sup_x |F(x) - F_0(x)|$ as the Kolmogorov distance, where F and F_0 are the cumulative distribution functions of f and f_0, respectively. This can be upper bounded by the Prokhorov metric, $d_K(f,f_0) \le d_P(f,f_0)\,(1+\min\{||f||_\infty, ||f_0||_\infty\})$; see, for example, Gibbs and Su (2002).

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

It is possible to use an alternative kernel to the sinc, for example, the Gaussian kernel. We would then consider $d_h = \sup_x |f_{0,h}(x) - f_0(x)|$, where

$$f_{0,h}(x) = h^{-1} \int \phi\left(\frac{x-y}{h}\right) f_0(y)\, dy.$$

It is straightforward to show that if f_0 is Hölder smooth, then $d_h < C\, h^\beta$, where $\beta > 0$ is the coefficient of smoothness. However, for f a mixture of normal distributions, we obtain $\sup_x |f_h(x) - f(x)| \le c\, h^2/\sigma^3$, which does not compare well with the bound from the sinc kernel; see equation (3.4).

Acknowledgments

The authors are grateful for the detailed comments and suggestions of the reviewers.

Appendix

A. Continuation of discussion in Section 4

In this appendix, we continue our discussion in Section 4. Since Prokhorov consistency follows directly from equation (1.1), our first result is moving from Prokhorov consistency to Kolmogorov consistency. Here we use the inequality,

$$d_K(f, f_0) \le d_P(f, f_0)\,(1 + \max\{||f||_\infty, ||f_0||_\infty\});$$

see, for example, (Gibbs and Su (2002)), and we assume $||f_0||_\infty < \infty$. Define the increasing sequence (M_n) to be such that $\Pi_n(d_P(f, f_0) > \epsilon/(1 + M_n)) \to 0$ a.s. for all $\epsilon > 0$. For example, since the Prokhorov rate of convergence will not be slower than $1/\log n$ we can take $M_n = \log n$. In fact, any sequence converging to ∞ slow enough works.

Theorem 4. *If we take the sample size dependent prior to be* $\Pi(||f||_\infty > M_n) < \exp(-n\tau)$ *for all large* n *and for some* $\tau > 0$*, then* $\Pi_n(d_K(f, f_0) > \epsilon) \to 0$ *a.s. for all* $\epsilon > 0$.

Proof. Now $\Pi_n(d_K(f, f_0) > \epsilon) < \Pi_n(d_P(f, f_0)\,(1 + \max\{||f||_\infty, ||f_0||_\infty\}) > \epsilon)$, which can be written as:

$$\begin{aligned}\Pi_n(d_P(f, f_0) > &\frac{\epsilon}{1 + \max\{||f||_\infty, ||f_0||_\infty\}) \cap ||f||_\infty < M_n} + \Pi_n(d_P(f, f_0)\\ &> \frac{\epsilon}{1 + \max\{||f||_\infty, ||f_0||_\infty\}) \cap ||f||_\infty > M_n}.\end{aligned}$$

The second term on the right is bounded above by $\Pi_n(||f||_\infty > M_n)$, which converges to 0. The first term on the right is, for all large n, upper bounded by

$\Pi_n\left(d_P(f,f_0) > \epsilon/(1+M_n)\right)$. This converges to 0 by virtue of weak consistency and that for large n the $\epsilon/(1+M_n)$ is greater than the Prokhorov rate.

The second result is concerned with the supremum norm; i.e., $d_\infty(f,f_0) = \sup_x |f(x) - f_0(x)|$, assuming the posterior is consistent with respect to the Kolmogorov metric, which we have just established in Theorem 4. Recall that we define $f_h(x) = (F(x+h) - F(x-h))/(2h)$, which is also a density function on $\mathbb{R}$, as used by Chae and Walker (2017). We then exploit the following triangle inequality:

$$|f(x) - f_0(x)| \le |f_h(x) - f(x)| + |f_h(x) - f_{h,0}(x)| + |f_0(x) - f_{h,0}(x)|. \tag{A.1}$$

Looking at the terms on the right side, the first can be made small with a suitable condition on f, the second term can be made small using the notion that f and f_0 are close with respect to a weak metric, and the final term will be small with a continuity condition on f_0.

Now $|f_h(x) - f(x)| \le \sup_{y:|x-y|<h} |f(x) - f(y)|$, which follows using $F(x+h) = F(x) + h\, f(x_h)$, for some x_h lying between x and $x+h$. Further, we have

$$|f_h(x) - f_{h,0}(x)| = \frac{1}{2h}\left|F(x+h) - F_0(x+h) - F(x-h) + F_0(x-h))\right|.$$

We can bound this using the Kolmogorov metric; i.e., $|f_h(x) - f_{h,0}(x)| \le d_K(f,f_0)$ $/h$. Hence, for all $h > 0$, $d_\infty(f,f_0) \le d_K(f,f_0) + \phi_h(f) + \phi_h(f_0)$, where $\phi_h(f) := \sup_{|x-y|<h} |f(x) - f(y)|$ and $\phi_h(f_0)$ is assumed to converge to 0 as $h \to 0$. In the following we let h depend on n, written as h_n, and choose the sequence to ensure that $\Pi_n(d_K(f,f_0) > ch_n) \to 0$ a.s. for any $c > 0$. We can assume that h_n is any slow enough sequence going to 0 and take it formally as $h_n = 1/\log n$.

Theorem 5. *If we take the sample size dependent prior* $\Pi(\phi_{h_n}(f) > \epsilon) < \exp(-n\epsilon)$ *for all* $\epsilon > 0$ *then* $\Pi_n(d_\infty(f,f_0) > \epsilon) \to 0$ *a.s.*

Proof. Using the triangle inequality we get

$$\Pi_n(d_\infty(f,f_0) > \epsilon) \le \Pi_n\bigg(\frac{d_K(f,f_0)}{h_n} + \phi_{h_n}(f) > \epsilon - \phi_{h_n}(f_0)\bigg).$$

The right term is easily seen to be bounded by $\Pi_n(d_K(f,f_0)$ $> h_n\epsilon_n) + \Pi_n(\phi_{h_n}(f) > \epsilon_n)$ where $\epsilon_n = (1/2)(\epsilon - \phi_{h_n}(p_0))$. Both terms can be easily shown to converge to 0.

B. Proof of Proposition 1

The proof of Proposition 1 follows the proof argument of Theorem 1 in Ho and Walker (2021). Here, we provide the proof for the completeness.

Since the function f is supersmooth or ordinary smooth, its Fourier transform $\widehat{f}$ is integrable. Therefore, the Fourier inversion transform and integral theorem hold. The Fourier integral theorem (Wiener (1933); Bochner (1959)) indicates that

$$\begin{aligned}
&|f_R(x) - f(x)| \\
&= \left| \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d \setminus [-R,R]^d} \int_{\mathbb{R}^d} \cos(s^\top (x-t)) f(t) ds dt \right| \\
&= \left| \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d \setminus [-R,R]^d} \left[\cos(s^\top x) \mathrm{Re}(\widehat{f}(s)) - \sin(s^\top x) \mathrm{Im}(\widehat{f}(s)) \right] ds \right| \\
&\leq \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d \setminus [-R,R]^d} \left[\left| \cos(s^\top x) \right| \left| \mathrm{Re}(\widehat{f}(s)) \right| + \left| \sin(s^\top x) \right| \left| \mathrm{Im}(\widehat{f}(s)) \right| \right] ds \\
&\leq \frac{\sqrt{2}}{(2\pi)^d} \int_{\mathbb{R}^d \setminus [-R,R]^d} |\widehat{f}(s)| ds \leq \frac{\sqrt{2}}{(2\pi)^d} \sum_{i=1}^d \int_{A_i} |\widehat{f}(s)| ds, \qquad \text{(B.1)}
\end{aligned}$$

where the second inequality in equation (B.1) is based on Cauchy-Schwarz inequality. Here, we respectively denote $\mathrm{Re}(\widehat{f})$, $\mathrm{Im}(\widehat{f})$ the real and imaginary parts of the Fourier transform $\widehat{f}$. Furthermore, we define $A_i = \{x \in \mathbb{R}^d : |x_i| \geq R\}$ for all $i \in [d]$.

(a) Since f is supersmooth density function of order α with scale parameter σ, we have

$$\begin{aligned}
\int_{A_i} |\widehat{p}_0(s)| ds &\leq C \int_{A_i} \exp\left(-C_1 \sigma^2 \left(\sum_{i=1}^d |s_i|^\alpha \right) \right) ds \\
&= \frac{C \alpha^{d-1}}{(2C_1 \sigma^2 \Gamma(1/\alpha))^{d-1}} \cdot \int_{|t| \geq R} \exp(-C_1 \sigma^2 |t|^\alpha) dt,
\end{aligned}$$

where C and C_1 are universal constants from Definition 1.

When $\alpha \geq 1$, we obtain that

$$\int_R^\infty \exp\left(-C_1 \sigma^2 t^\alpha\right) dt \leq \int_R^\infty t^{\alpha-1} \exp\left(-C_1 \sigma^2 t^\alpha\right) dt = \frac{\exp(-C_1 \sigma^2 R^\alpha)}{C_1 \sigma^2 \alpha}.$$

When $\alpha \in (0,1)$, we find that

$$\begin{aligned}
&\int_R^\infty \exp(-C_1 \sigma^2 t^\alpha) dt \\
&= \int_R^\infty t^{1-\alpha} t^{\alpha-1} \exp(-C_1 \sigma^2 t^\alpha) dt \\
&= \frac{R^{1-\alpha} \exp\left(-C_1 \sigma^2 R^\alpha\right)}{C_1 \sigma^2 \alpha} + \frac{1-\alpha}{C_1 \sigma^2 \alpha} \int_R^\infty t^{-\alpha} \exp(-C_1 \sigma^2 t^\alpha) dt \\
&\leq \frac{R^{1-\alpha} \exp\left(-C_1 \sigma^2 R^\alpha\right)}{C_1 \sigma^2 \alpha} + \frac{1-\alpha}{C_1 \sigma^2 \alpha R^\alpha} \int_R^\infty \exp(-C_1 \sigma^2 t^\alpha) dt.
\end{aligned}$$

We choose R such that $R^\alpha \geq \frac{2(1-\alpha)}{C_1\sigma^2\alpha}$. Then, the inequality in the above display becomes

$$\int_R^\infty \exp(-C_1\sigma^2 t^\alpha)dt \leq \frac{2R^{1-\alpha}\exp\left(-C_1\sigma^2 R^\alpha\right)}{C_1\sigma^2\alpha}.$$

Collecting the above results, we obtain

$$\int_{|t|\geq R} \exp(-C_1\sigma^2|t|^\alpha)dt \leq \frac{4R^{\max\{1-\alpha,0\}}}{C_1\sigma^2\alpha}\exp(-C_1\sigma^2R^\alpha).$$

Hence, for each $i \in [d]$, we have the following upper bound:

$$\int_{A_i} |\widehat{f}(s)|ds \leq \frac{C\alpha^{d-2}R^{\max\{1-\alpha,0\}}}{2^{d-3}C_1^d\sigma^{2d}(\Gamma(1/\alpha))^{d-1}}\exp(-C_1\sigma^2R^\alpha). \tag{B.2}$$

The results from equations (B.1) and (B.2) lead to

$$|f_R(x) - f(x)| \leq \frac{\sqrt{2}Cd\cdot\alpha^{d-2}R^{\max\{1-\alpha,0\}}}{\pi^d 2^{2d-3}C_1^d\sigma^{2d}(\Gamma(1/\alpha))^{d-1}}\exp(-C_1\sigma^2R^\alpha).$$

As a consequence, we reach the conclusion of part (a).

(b) Since the density function is ordinary smooth of order β with scale parameter σ, for each $i \in [d]$ we obtain

$$\begin{aligned}
\int_{A_i} |\widehat{f}(s)|ds &\leq c\int_{A_i}\prod_{j=1}^d \frac{1}{(1+\sigma^2|s_j|^\beta)}ds \\
&= c\left(\int_{-\infty}^\infty \frac{1}{1+\sigma^2|t|^\beta}dt\right)^{d-1}\cdot\int_{|t|\geq R}\frac{1}{1+\sigma^2|t|^\beta}dt \\
&= \frac{c}{\sigma^{2(d-1)/\beta}}\left(\int_{-\infty}^\infty \frac{1}{1+|t|^\beta}dt\right)^{d-1}\cdot\int_{|t|\geq R}\frac{1}{1+\sigma^2|t|^\beta}dt.
\end{aligned}$$

Since $\beta > 1$, $I_\beta = \int_{-\infty}^\infty (1/1+|t|^\beta)dt < \infty$. Furthermore, we have

$$\int_{|t|\geq R}\frac{1}{1+\sigma^2|t|^\beta}dt \leq 2\int_R^\infty \frac{1}{\sigma^2 t^\beta}ds = \frac{2}{(\beta-1)\sigma^2}R^{-\beta+1}.$$

Combining the above results, we find that

$$\int_{A_i} |\widehat{f}(s)|ds \leq \frac{2cI_\beta^{d-1}}{(\beta-1)\sigma^{2+2(d-1)/\beta}}R^{1-\beta}. \tag{B.3}$$

The results from equations (B.1) and (B.3) lead to the conclusion of part (b).

References

Barron, A., Schervish, M. and Wasserman, L. (1999). The consistency of posterior distributions in nonparametric problems. *The Annals of Statistics* **27**, 536–561.

Bochner, S. (1959). *Lectures on Fourier Integrals*. Princeton University Press.

Castillo, I. (2014). On Bayesian supremum norm contraction rates. *The Annals of Statistics* **42**, 2058–2091.

Chae, M. and Walker, S. (2017). A novel approach to Bayesian consistency. *Electronic Journal of Statistics* **11**, 4723–4745.

Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *The Annals of Statistics* **19**, 1257–1272.

Ghosal, S., Ghosh, J. and Ramamoorthi, R. (1999). Posterior consistency of Dirichlet mixtures in density estimation. *The Annals of Statistics* **27**, 143–158.

Ghosal, S. and van der Vaart, A. (2007). Posterior convergence rates of dirichlet mixtures at smooth densities. *The Annals of Statistics* **35**, 697–723.

Ghosal, S. and van der Vaart, A. (2017). *Fundamentals of Nonparametric Bayesian Inference*. Cambridge University Press.

Gibbs, A. L. and Su, F. E. (2002). On choosing and bounding probability metrics. *International Statistical Review* **70**, 419–435.

Gine, E. and Nickl, R. (2011). Rates on contraction for posterior distributions in $\mathbb{L}_r$ metrics, $1 \leq r \leq \infty$. *The Annals of Statistics* **39**, 2883–2911.

Ho, N. and Walker, S. (2021). Multivariate smoothing via the Fourier integral theorem and Fourier kernel. *arXiv:2012.14482*.

Hoffmann, M., Rousseau, J. and Schmidt-Hieber, J. (2015). On adaptive posterior concentration rates. *The Annals of Statistics* **43**, 2259–2295.

Li, W. and Ghosal, S. (2020). Posterior contraction and credible sets for filaments of regression functions. *Electronic Journal of Statistics* **14**, 1707–1743.

Li, W. and Ghosal, S. (2021). Posterior contraction and credible regions for level sets. *Electronic Journal of Statistics* **15**, 2647–7524.

Nissila, J. (2021). Fourier decay of absolutely and Hölder continuous functions with infinitely or finitely many oscillations. *arXiv:1805.0244v2*.

Schwartz, L. (1965). On Bayes procedures. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **4**, 10–26.

Scricciolo, C. (2014). Adaptive Bayesian density estimation in $\mathbb{L}_p$ metrics with Pitman–Yor or normalized inverse–Gaussian process kernel mixtures. *Bayesian Analysis* **9**, 475–520.

Shen, W. and Ghosal, S. (2017). Posterior contraction rates of density derivative estimation. *Sankhyā: The Indian Journal of Statistics, Series A* **79**, 336–354.

Shen, W., Tokdar, S. and Ghoshal, S. (2013). Adaptive Bayesian multivariate density estimation with dirichlet mixtures. *Biometrika* **100**, 623–640.

Walker, S. (2004). New approaches to Bayesian consistency. *The Annals of Statistics* **32**, 2028–2043.

Wiener, N. (1933). *The Fourier Integral and Certain of its Applications*. Cambridge University Press.

Wu, Y. and Ghosal, S. (2008). Kullback Leibler property of kernel mixture priors in Bayesian density estimation. *Electronic Journal of Statistics* **2**, 298–331.

Yoo, W. and Ghosal, S. (2016). Supremum norm posterior contraction and credible sets for nonparametric multivariate regression. *The Annals of Statistics* **44**, 1069–1102.

Yoo, W., Rousseau, J. and Rivoirard, V. (2018). Adaptive supremum norm posterior contraction:

Wavelet spike-and-slab and anisotropic Besov spaces. Web: `https://arxiv.org/abs/1708.01909.pdf`.

Zhang, C. (1990). Fourier methods for estimating mixing densities and distributions. *The Annals of Statistics* **18**, 806–831.

Nhat Ho

Department of Mathematics, The University of Texas at Austin, TX 78712, USA.

E-mail: minhnhat@math.utexas.edu

Stephen G. Walker

Department of Mathematics, The University of Texas at Austin, TX 78712, USA.

E-mail: s.g.walker@math.utexas.edu

(Received February 2022; accepted October 2022)