Sankhyā : The Indian Journal of Statistics
2006, Volume 68, Part 1, pp. 90-110

Posterior Consistency of Dirichlet Location-scale Mixture of Normals in Density Estimation and Regression

Surya T. Tokdar
Purdue University, USA

Abstract

We provide sufficient conditions under which a Dirichlet location-scale mixture of normal prior achieves weak and strong posterior consistency at a true density. Our conditions involve both the prior and the true density from which observations are obtained. We consider it to be a significant improvement over the existing results since our conditions cover the case of fat tailed densities like the Cauchy, with a standard choice for the base measure of the Dirichlet process. This provides a wider choice for using these popular mixture priors for nonparametric density estimation and semiparametric regression problems.

AMS (2000) *subject classification.* Primary 62G07, 62G08, 62G20.
Keywords and phrases. Posterior consistency, Dirichlet process, location-scale mixtures, density estimation, regression.

---

1 Introduction

Priors based on Dirichlet location-scale mixture of normals are widely used to model densities as mixtures of normal kernels. A random density f arising from such a prior can be expressed as

$$f(y) = (\phi * P)(y) = \int \frac{1}{\sigma}\phi\left(\frac{y-\theta}{\sigma}\right) dP(\theta, \sigma), \tag{1.1}$$

where $\phi(\cdot)$ is the standard normal density and the mixing distribution P follows a Dirichlet process. These priors were first introduced in Ferguson (1983) and Lo (1984) and have been extensively investigated and applied by many in the subsequent years (see, for example, West et al., 1994, MacEachern, 1994, Escobar and West, 1995, Liu, 1996, Müller et al., 1996).

Ghosal et al. (1999) initiated a theoretical study of these priors for the problem of density estimation. They showed that if a density f_0 satisfies certain conditions, then a Dirichlet location mixture of normals achieves posterior consistency at f_0. Their conditions can be best summarized as f_0 having a moment generating function on an open interval containing $[-1, 1]$. Ghosal and van der Vaart (2001) extended these results to rate calculations for the more general Dirichlet location-scale mixture prior. However, they restricted the scale parameter σ to a compact interval $[\underline{\sigma}, \overline{\sigma}] \subset (0, \infty)$. Amewou-Atisso et al. (2003) generalized the results of Ghosal et al. (1999) to the case of not identically distributed (non iid) random variables and derived similar consistency results for regression problems.

Unfortunately, none of these consistency results applies if the true density has heavy tails. For example, the family of t densities and in particular the Cauchy density are not covered. It was conjectured in Amewou-Atisso et al. (2003) that one needs a Dirichlet location-scale mixture of normal prior with full support for σ to cover the case of fat tailed densities. In this paper we prove this conjecture to be true for both density estimation and regression and derive the conditions that have to be satisfied by the underlying Dirichlet process.

Our main result states that a Dirichlet location-scale mixture, with some regularity condition on the tail of its base measure, achieves posterior consistency at f_0 if $\int |x|^\eta f_0(x)dx < \infty$ for some $\eta > 0$. The tail conditions are automatically satisfied by the popular normal-inverse gamma base measures. We also establish a more general result that under other *reasonable* priors on the mixing distribution P, posterior consistency is achieved at f_0 satisfying $\int |x|^{2+\eta} f_0(x)dx < \infty$ for some $\eta > 0$.

In Sections 2 to 4, we present the new ideas and constructions needed to make the above extensions feasible. The portions of the proofs which have major overlaps with the arguments presented in Ghosal et al. (1999) and Amewou-Atisso et al. (2003) are provided in the appendix. A brief introduction to the concepts of posterior consistency is provided in Section 2.

We end this discussion by noting that posterior consistency at a large set of parameter values serves as an objective validation of the choice of a prior (see Ghosh and Ramamoorthi, 2003, Ch. 4 for more details). Since consistency of the posterior for a nonparametric prior is hard to establish, we hope that our results make a strong case for the use of Dirichlet location-scale mixture priors for density estimation and regression problems.

2 Preliminaries

To make this paper relatively self-contained, we recall the definitions of posterior consistency in the context of density estimation and regression. These definitions formalize the concept that in order to achieve consistency, the posterior should concentrate on arbitrarily small neighborhoods of the true model when more observations are made available.

Posterior Consistency for Density Estimation. Suppose $X_1, X_2, \cdots$ are independent and identically distributed according to an unknown density f_0. We take the parameter space as $\mathcal{F}$ - a set of probability densities on the space of the observations and consider a prior distribution Π on $\mathcal{F}$. Then the posterior distribution $\Pi(\cdot|X_1, \cdots, X_n)$ given a sample $X_1, \cdots, X_n$ is obtained as,

$$\Pi(A|X_1, \cdots, X_n) = \frac{\int_A \prod_{i=1}^n f(X_i) d\Pi(f)}{\int_{\mathcal{F}} \prod_{i=1}^n f(X_i) d\Pi(f)}.$$

We say that the posterior achieves weak (or strong) posterior consistency at f_0 if for any weak (or L_1) neighborhood U of f_0, $\Pi(U|X_1, X_2, \cdots, X_n) \to 1$ almost surely as $n \to \infty$.

Posterior Consistency for Regression. Suppose one observes $Y_1, Y_2, \cdots$ from the model $Y_i = \alpha_0 + \beta_0 x_i + \epsilon_i$, where x_i's are known non-random covariate values and ϵ_i's are independent and identically distributed with an unknown symmetric density f_0. The regression coefficients α_0, β_0 are also unknown. Here, it is appropriate to consider the parameter space as $\Theta = \mathcal{F}^* \times \mathbb{R} \times \mathbb{R}$, where $\mathcal{F}^*$ is a set of symmetric probability densities on $\mathbb{R}$ with a prior Π on Θ. The posterior distribution $\Pi(\cdot|Y_!, \cdots, Y_n)$ is then computed as,

$$\Pi(A|Y_1, \cdots, Y_n) = \frac{\int_A \prod_{i=1}^n f(Y_i - \alpha - \beta x_i) d\Pi(f, \alpha, \beta)}{\int_\times \prod_{i=1}^n f(Y_i - \alpha - \beta x_i) d\Pi(f, \alpha, \beta)}.$$

We say that the posterior achieves weak consistency at (f_0, α_0, β_0) if for any weak neighborhood U of f_0 and any $\delta > 0$,

$$\Pi((f, \alpha, \beta) : f \in U, |\alpha - \alpha_0| < \delta, |\beta - \beta_0| < \delta | Y_1, Y_2, \cdots, Y_n) \quad \to 1$$

almost surely as $n \to \infty$.

In this paper, we restrict ourselves to one dimensional covariates to maintain clarity of exposition. However, the arguments can be easily extended to the case of multiple regression; see the discussion given in Section 1 of Amewou-Atisso et al. (2003).

3 Density Estimation: Weak Consistency

We start with weak posterior consistency for the problem of density estimation. Our main tool is the following theorem due to Schwartz (1965).

Theorem 3.1. *A prior* Π *achieves weak posterior consistency at a density* f_0, *if*

$$\forall \epsilon > 0,\ \Pi\left(f \in \mathcal{F} : \int f_0(x) \log \frac{f_0(x)}{f(x)} dx < \epsilon\right) > 0 \tag{3.1}$$

Remark 3.1. We use the notation $f_0 \in KL(\Pi)$ to indicate that a density f_0 satisfies (3.1).

3.1. General mixture priors. First consider the case when the mixing distribution P in (1.1) follows some general distribution $\tilde{\Pi}$, not necessarily a Dirichlet process. It is *reasonable* to assume that the weak support of $\tilde{\Pi}$ contains all probability measures on $\mathbb{R} \times \mathbb{R}^+$ that are compactly supported. The next lemma reveals the implication of this property.

Lemma 3.1. *Consider an* $f_0 \in \mathcal{F}$ *such that* $\int x^2 f_0(x)dx < \infty$. *Suppose* $\tilde{f} = \phi * \tilde{P}$ *is such that* $\tilde{P}((-a, a) \times (\underline{\sigma}, \overline{\sigma})) = 1$ *for some* $a > 0, 0 < \underline{\sigma} < \overline{\sigma}$. *Then for any* $\epsilon > 0$, *there exists a weak neighborhood* W *of* $\tilde{P}$ *such that for any* $f = \phi * P$ *with* $P \in W$,

$$\int f_0(x) \log \frac{\tilde{f}(x)}{f(x)} dx < \epsilon$$

The proof of this lemma is similar to the proof of Theorem 3 of Ghosal et al. (1999) and we present it in the appendix. Here we state and prove the main result.

Theorem 3.2. *Let* $f_0(x)$ *be a continuous density on* $\mathbb{R}$ *satisfying:*

1. f_0 *is nowhere zero and bounded above by* $M < \infty$.
2. $\left|\int_{\mathbb{R}} f_0(x) \log f_0(x) dx\right| < \infty$.
3. $\int_{\mathbb{R}} f_0(x) \log \frac{f_0(x)}{\psi_1(x)} dx < \infty$ *where* $\psi_1(x) = \inf_{t \in [x-1, x+1]} f_0(t)$.
4. $\exists \eta > 0$ *such that* $\int_{\mathbb{R}} |x|^{2(1+\eta)} f_0(x) dx < \infty$.

Then, $f_0 \in KL(\Pi)$.

Remark 3.2. Assumption *4* provides the important moment condition on f_0. Assumption *2* is satisfied by most of the common densities and assumption *3* can be viewed as a regularity conditions. The interval $[x-1, x+1]$ that appears in assumption *3* can be replaced by $[x-a, x+a]$ for any $a > 0$.

Proof of Theorem 3.2. Note that,

$$\int f_0(x) \log \frac{f_0(x)}{f(x)} dx = \int f_0(x) \log \frac{f_0(x)}{\tilde{f}(x)} dx + \int f_0(x) \log \frac{\tilde{f}(x)}{f(x)} dx.$$

Therefore, the result would follow if for any $\epsilon > 0$, we can find an $\tilde{f}$ which makes $\int f_0 \log \frac{f_0}{\tilde{f}} dx < \epsilon/2$ and also satisfies the condition of Lemma 3.1. Next we show how to construct such an $\tilde{f}$.

Consider the densities $f_n = \phi * P_n$, $n \geq 1$, with P_n's constructed as,

$$dP_n(\theta, \sigma) = t_n I_{(\theta \in [-n,n])} f_0(\theta) \delta_{\sigma_n}(\sigma)$$

where $\sigma_n = n^{-\eta}$, $t_n = (\int_{-n}^{n} f_0(y)dy)^{-1}$, I_A is the indicator function of a set A and δ_x is the point mass at a point x. Note that f_n can be simply written as,

$$f_n(x) = t_n \int_{-n}^{n} \frac{1}{\sigma_n} \phi\left(\frac{x-\theta}{\sigma_n}\right) f_0(\theta) d\theta.$$

Find a positive constant ξ such that $\int_{-\xi}^{\xi} \phi(t)dt > 1-\epsilon$. Now fix an $x \in \mathbb{R}$. For sufficiently large n such that $[x - \xi\sigma_n, x + \xi\sigma_n] \subset [-n, n]$, one obtains,

$$\inf_{y \in (x-\xi\sigma_n, x+\xi\sigma_n)} f_0(y)(1-\epsilon) < \frac{f_n(x)}{t_n} < \sup_{y \in (x-\xi\sigma_n, x+\xi\sigma_n)} f_0(y) + M\epsilon \qquad (3.2)$$

Since $t_n \to 1$ and $\sigma_n \to 0$, (3.2) would imply that $f_n(x) \to f_0(x)$ as $n \to \infty$ by continuity of f_0. Therefore one can conclude,

$$\log \frac{f_0(x)}{f_n(x)} \to 0 \text{ for all } x \in \mathbb{R} \qquad (3.3)$$

Since t_n is a decreasing sequence and $f_0(\theta) < M$ for all $\theta \in \mathbb{R}$, one can readily see that for all $n \geq 1$ and all $x \in \mathbb{R}$,

$$f_n(x) = t_n \int_{-n}^{n} \frac{1}{\sigma_n} \phi\left(\frac{x-\theta}{\sigma_n}\right) f_0(\theta) d\theta \leq M t_n \leq M t_1. \qquad (3.4)$$

Now, fix an $x \in \mathbb{R}$. Since, $|x-\theta| \le |x| + n$ for all $\theta \in [-n, n]$ and $t_n \ge 1$, it follows that for all $n \le |x|$,

$$f_n(x) \ge \frac{1}{\sigma_n}\phi\left(\frac{|x|+n}{\sigma_n}\right) = n^\eta \phi(n^\eta(|x|+n)) \ge |x|^\eta \phi(2|x|^{1+\eta}).$$

The last inequality follows from the fact that $\tau^\eta\phi(\tau^\eta(|x|+\tau))$ is decreasing in τ for $\tau \ge 1$.

Let $\psi_n(x) = \inf_{t\in[x-\sigma_n, x+\sigma_n]} f_0(t)$. It may be noted that the function $\psi_1(x)$ of assumption *3* is consistent with this definition. Let $A_n = [-n, n] \cap [x-\sigma_n, x+\sigma_n]$ and $c = \int_0^1 \phi(t)dt < 1$. Observe that for all $n > |x|$,

$$f_n(x) \ge t_n \int_{A_n} \frac{1}{\sigma_n}\phi\left(\frac{x-\theta}{\sigma_n}\right) f_0(\theta)d\theta \ge t_n\psi_n(x)\int_{A_n}\frac{1}{\sigma_n}\phi\left(\frac{x-\theta}{\sigma_n}\right)d\theta \quad (3.5)$$

Since $t_n \ge 1$, $\psi_n(x) \ge \psi_1(x)$ and $\int_{A_n}\frac{1}{\sigma_n}\phi(\frac{x-\theta}{\sigma_n})d\theta \ge \int_0^1\phi(t)dt = c$ for all $n \ge 1$ and all $x \in \mathbb{R}$ it follows from (3.5) that $f_n(x) \ge c\psi_1(x)$ for all $n > |x|$.. Therefore,

$$f_n(x) \ge \begin{cases} c\psi_1(x) & |x| < 1 \\ \min(|x|^\eta\phi(2|x|^{1+\eta}), c\psi_1(x)) & |x| \ge 1 \end{cases} \quad (3.6)$$

A little algebraic manipulation with (3.4) and (3.6) produces, $\forall n \ge 1$,

$$\left|\log\frac{f_0(x)}{f_n(x)}\right| \le \log\frac{Mt_1}{f_0(x)} + \log\frac{f_0(x)}{c\psi_1(x)} + I_{\{|x|>1\}}\log\frac{f_0(x)}{|x|^\eta\phi(2|x|^{1+\eta})}$$

From the assumptions of Theorem 3.3, it can be easily verified that the function on the right hand side of the above display is f_0 integrable. Therefore an application of DCT on (3.3) implies that,

$$\lim_{n\to\infty}\int f_0(x)\log\frac{f_0(x)}{f_n(x)}dx = 0.$$

Therefore we can simply choose $\tilde{f} = f_{n_0}$ for some large enough n_0. □

3.2. Dirichlet mixture of normals. Next we consider $\tilde{\Pi} = Dir(\alpha G_0)$, a Dirichlet process with parameter αG_0. Here α is a positive constant and G_0 is a probability measure on $\mathbb{R} \times \mathbb{R}^+$.

Lemma 3.2. *Suppose* $f_0 \in \mathcal{F}$ *satisfies the following property: For any* $0 < \tau < 1, \epsilon > 0$, *there exist a set* $\mathcal{A}$ *and a positive number* x_0 *such that* $\tilde{\Pi}(\mathcal{A}) > 1 - \tau$ *and for any* $f = \phi * P$ *with* $P \in \mathcal{A}$,

$$\int_{|x|>x_0} f_0(x) \log \frac{f_0(x)}{f(x)} dx < \epsilon.$$

Then, $f_0 \in KL(\Pi)$.

The proof of this lemma uses some tools developed by Ghosal et al. (1999) and is given later in the appendix. Here we move on to our main result. Note that the moment condition of Theorem 3.2 is substantially reduced.

Theorem 3.3. *Let* f_0 *be a density on* $\mathbb{R}$ *satisfying*

1. $\int f_0(x) \log f_0(x) dx < \infty$.
2. $\exists\, \eta \in (0,1)$ *such that* $\int |x|^\eta f_0(x) dx < \infty$.

Further assume that there exist $\sigma_0 > 0$, $0 < \beta < \eta$, $\gamma > \beta$ *and* $b_1, b_2 > 0$ *such that for large* $x > 0$

3. $$\max\left(G_0\left(\left[x - \sigma_0 x^{\frac{\eta}{2}}, \infty\right) \times [\sigma_0, \infty)\right),\right. \\ \left. G_0\left([0,\infty) \times (x^{1-\frac{\eta}{2}}, \infty)\right)\right) \geq b_1 x^{-\beta}$$
4. $$G_0\left((-\infty, x) \times (0, e^{|x|^\eta - \frac{1}{2}})\right) > 1 - b_2 |x|^{-\gamma}.$$

and for large $x < 0$,

3′. $$\max\left(G_0\left(\left(-\infty, x + \sigma_0 |x|^{\frac{\eta}{2}}\right] \times [\sigma_0, \infty)\right),\right. \\ \left. G_0\left((-\infty, 0] \times (|x|^{1-\frac{\eta}{2}}, \infty)\right)\right) \geq b_1 |x|^{-\beta}$$

4′. $$G_0\left((x, \infty) \times (0, e^{|x|^\eta - \frac{1}{2}})\right) > 1 - b_2 |x|^{-\gamma}.$$

then $f_0 \in KL(\Pi)$.

Remark 3.3. Other than the important moment condition on f_0 this theorem also requires some regularity in the tail of the base measure G_0. For example, assumption *3, 3′* requires the tail of G_0 not to decay faster than a polynomial rate for the scale parameter σ. This condition seems very reasonable since the Cauchy density itself can be written as a scale mixture of normals with the mixing density having a polynomial decay towards infinity.

REMARK 3.4. A standard choice for G_0 is the conjugate normal-inverse gamma distribution (see Escobar and West, 1995), under which, $\theta|\sigma \sim N(0, \xi\sigma^2)$ and $\sigma^{-2} \sim Gamma(r, \lambda)$, for some $\xi, r, \lambda > 0$. For such a G_0 with $r \in (1/2, 1)$, one can show that the conditions of Theorem 3.3 hold true with $\eta \in (2r/(1+r), 1)$, $\beta = r(2-\eta)$ and $\gamma = 2r$. For example, the conditions in *Assumptions 3, 3′* are satisfied since,

$$\begin{aligned} G_0\left([0,\infty)\times(x^{1-\frac{\eta}{2}},\infty)\right) &= \frac{1}{2}\Pr(\sigma^{-2} \leq x^{-(2-\eta)}) \\ &= c\int_0^{x^{-(2-\eta)}} v^{r-1}e^{-\lambda v}dv \leq c'x^{-r(2-\eta)}, \end{aligned}$$

for some positive constants c, c'. To see that the conditions of *Assumptions 4, 4′* also hold, note that,

$$1 - G_0\left((-\infty, x) \times (0, e^{|x|^\eta - \frac{1}{2}})\right) \leq \Pr(\theta > x) + \Pr(\sigma^{-2} < e^{-2|x|^\eta + 1}).$$

An argument similar to the one provided above shows that the second term, namely, $\Pr(\sigma^{-2} < e^{-2|x|^\eta+1})$ is bounded by a constant times $e^{-2r|x|^\eta + r}$. Therefore, this term can be made smaller than $c|x|^{-\gamma}$ for a suitable constant c. Now, using the inequality $1 - \Phi(x) \leq (1/x)\phi(x)$, where $\Phi(\cdot)$ and $\phi(\cdot)$ are the standard normal distribution and density functions, we obtain

$$\Pr(\theta > x) \leq \frac{c}{x}\int_0^\infty v^{r-1/2-1}e^{-(\frac{x^2}{2\xi}+\lambda)v}dv = \frac{c'}{x(\frac{x^2}{2\xi}+\lambda)^{r-1/2}} \leq \frac{c''}{x^{2r}}$$

for some positive constants c, c', c''. The desired inequality follows from these two bounds. Therefore, such a choice of G_0 would lead to posterior consistency, for example, when f_0 is a Cauchy density.

PROOF OF THEOREM 3.3. We simply need to show that such an f_0 satisfies the condition of Lemma 3.2. Let $w(x) = \exp(-x^\eta)$, $x \geq 0$. Define a class of subsets of $\mathbb{R} \times \mathbb{R}^+$ indexed by $x \in \mathbb{R}$, as follows:

$$K_x = \left\{(\theta, \sigma) \in \mathbb{R} \times \mathbb{R}^+ : \frac{1}{\sigma}\phi\left(\frac{x-\theta}{\sigma}\right) \geq \frac{1}{(2\pi)^{1/2}}w(|x|)\right\}$$

These sets are of particular interest, since for $f = \phi * P$,

$$\begin{aligned}
&\int_{|x|>x_0} f_0(x) \log \frac{f_0(x)}{f(x)} dx \\
&\quad \le \int_{|x|>x_0} f_0(x) \log \frac{f_0(x)}{\int_{K_x} \frac{1}{\sigma}\phi\left(\frac{x-\theta}{\sigma}\right) dP(\theta,\sigma)} dx \\
&\quad \le \int_{|x|>x_0} f_0(x) \log \frac{f_0(x)}{\frac{1}{(2\pi)^{1/2}} w(|x|) P(K_x)} dx \\
&\quad \le \int_{|x|>x_0} f_0(x) \left\{ \log f_0(x) + |x|^\eta + \log \frac{(2\pi)^{1/2}}{P(K_x)} \right\} dx. \qquad (3.7)
\end{aligned}$$

By the assumptions of the Theorem, this quantity can be made arbitrarily small for a suitably large x_0 if we can show that $P(K_x) > c_1 \exp(-c_2|x|^\eta)$ for all $|x| > x_0$ for some fixed constants $c_1, c_2 > 0$. Therefore it suffices to prove that,

Lemma 3.3. *For any $\tau > 0$ there exists an $x_0 > 0$ and a set $\mathcal{A}$ with $\tilde{\Pi}(\mathcal{A}) > 1 - \tau$ such that $P \in \mathcal{A} \Rightarrow P(K_x) \ge (1/2)\exp(-2|x|^\eta/b_1)$ for all $|x| > x_0$.*

The proof of this Lemma is fairly technical. It makes an extensive use of the tail behavior of a random probability P arising from a Dirichlet process. For clarity of reading, we present details of the proof in the appendix. □

4 Density Estimation: Strong Consistency

We establish L_1-consistency of a Dirichlet location-scale mixture of normal prior Π by verifying the conditions of Theorem 8 of Ghosal et al. (1999) (similar conditions are derived in Barron et al., 1999, see also Walker, 2004). This theorem is reproduced below.

Theorem 4.1. *Let Π be a prior on $\mathcal{F}$ such that $f_0 \in KL(\Pi)$. If there is a $\delta < \epsilon/4$, c_1, $c_2 > 0$, $\beta < \epsilon^2/8$ and $\mathcal{F}_n \subseteq \mathcal{F}$ such that for all n large,*

1. $\Pi(\mathcal{F}_n^c) < c_1 e^{-nc_2}$,
2. $J(\delta, \mathcal{F}_n) < n\beta$,

then Π achieves strong posterior consistency at f_0.

Remark 4.1. Here $J(\delta, \mathcal{G})$ denotes logarithm of the covering number of $\mathcal{G}$ by L_1 balls of radii δ.

We first show how to calculate $J(\delta, \mathcal{G})$ for certain type of sets $\mathcal{G}$.

LEMMA 4.1. *For some* $a > 0,\ u > l > 0$ *define*

$$\mathcal{F}_{a,l,u} = \{f = \phi * P : P\left((-a,a] \times (l,u]\right) = 1\}$$

Then,

$$J(2\kappa, \mathcal{F}_{a,l,u}) \leq b_0 \left(b_1 \frac{a}{l} + b_2 \log \frac{u}{l} + 1\right).$$

where b_0, b_1 *and* b_2 *depend upon* κ *but not on* a, l *or* u.

PROOF. Let $\phi_{\theta,\sigma}$ denote the normal density with mean θ and standard deviation σ. For $\sigma_2 > \sigma_1 > \sigma_2/2$, it can be shown that,

$$\begin{aligned} \|\phi_{\theta_1,\sigma_1} - \phi_{\theta_2,\sigma_2}\| &\leq \|\phi_{\theta_1,\sigma_2} - \phi_{\theta_2,\sigma_2}\| + \|\phi_{\theta_1,\sigma_1} - \phi_{\theta_1,\sigma_2}\| \\ &\leq \left(\frac{2}{\pi}\right)^{1/2} \frac{|\theta_2 - \theta_1|}{\sigma_2} + 3\frac{\sigma_2 - \sigma_1}{\sigma_1}. \end{aligned} \tag{4.1}$$

Let $\zeta = \min(\kappa/6, 1)$. Define $\sigma_m = l(1+\zeta)^m,\ m \geq 0$. Let M be the smallest integer such that $\sigma_M = l(1+\zeta)^M \geq u$. This implies $M \leq (1+\zeta)^{-1}\log(u/l) + 1$. For $1 \leq j \leq M$, let $N_j = \left\lceil \left(\frac{32}{\pi}\right)^{1/2} a/(\kappa\sigma_{j-1})\right\rceil$. For $1 \leq i \leq N_j;\ 1 \leq j \leq M$, define

$$E_{ij} = \left(-a + \frac{2a(i-1)}{N_j}\ ,\ -a + \frac{2ai}{N_j}\right] \times (\sigma_{j-1}, \sigma_j].$$

Then, $(\theta,\sigma),\ (\theta',\sigma') \in E_{ij} \Rightarrow \left\|\phi_{\theta,\sigma} - \phi_{\theta',\sigma'}\right\| < \kappa$. Take $N = \sum_{j=1}^{M} N_j$ and let

$$\mathcal{P}_N = \left\{(P_{11}, \cdots, P_{N_1 1}, \cdots, P_{1M}, \cdots, P_{N_M M}) : P_{ij} \geq 0, \sum_{ij} P_{ij} = 1\right\}$$

be the N dimensional probability simplex and $\mathcal{P}_N^*$ be a κ-net in $\mathcal{P}_N$. Let τ_j's be as before and $\theta_{ij} = -a + 2a(i-1/2)/N_j,\ 1 \leq i \leq N_j,\ 1 \leq j \leq M$. So $(\theta_{ij}, \sigma_j) \in E_{ij}\ \forall i, j$. It can be shown by following an argument similar to the one presented in the proof of Lemma 1 of Ghosal et al. (1999) that ,

$$\mathcal{F}\left\{\sum_{j=1}^{M}\sum_{i=1}^{N_j} P_{ij}^* \phi_{\theta_{ij},\sigma_j} : P^* \in \mathcal{P}_N^*\right\}$$

is a 2κ-net in $\mathcal{F}_{a,l,u}$ and consequently,

$$J(2\kappa, \mathcal{F}_{a,l,u}) \leq J(\kappa, \mathcal{P}_N) \leq N\left(1 + \log\frac{1+\kappa}{\kappa}\right).$$

But,

$$\begin{aligned} N &\leq \sum_{j=1}^{M}\left(\left(\frac{32}{\pi}\right)^{1/2}\frac{a}{\sigma_{j-1}\kappa} + 1\right) = \left(\frac{32}{\pi}\right)^{1/2}\frac{a}{l\kappa}\sum_{j=0}^{M-1}(1+\zeta)^{-j} + M \\ &\leq \left(\frac{32}{\pi}\right)^{1/2}\frac{a}{l}\frac{1+\zeta}{\kappa\zeta} + \frac{1}{1+\zeta}\log\frac{u}{l} + 1 \\ &= b_1\frac{a}{l} + b_2\log\frac{u}{l} + 1 \qquad (4.2) \end{aligned}$$

From this inequality, the result follows with $b_0 = 1 + \log\frac{1+\kappa}{\kappa}$. □

LEMMA 4.2. *Let* $\mathcal{F}^{\kappa}_{a,l,u} = \{f = \phi * P : P((-a,a] \times (l,u]) \geq 1-\kappa\}$. *Then* $J(3\kappa, \mathcal{F}^{\kappa}_{a,l,u}) \leq J(\kappa, \mathcal{F}_{a,l,u})$.

PROOF. Let $f = \phi * P \in \mathcal{F}^{\kappa}_{a,l,u}$. Consider the probability measure defined by $P^*(A) = P(A \cap (-a,a] \times (l,u])/P((-a,a] \times (l,u])$. Then the density $f^* = \phi * P^*$ clearly belongs to $\mathcal{F}_{a,l,u}$ and further satisfies $\|f - f^*\| < 2\kappa$. This proves the lemma. □

THEOREM 4.2. *Suppose for each* $\kappa > 0, \beta > 0$, *there exist sequences of positive numbers* $a_n, u_n \uparrow \infty, l_n \downarrow 0$ *with* $l_n < u_n$ *and constant* β_0, *all depending on* κ *and* β *such that*

1. $\tilde{\Pi}(\{P : P((-a_n, a_n] \times (l_n, u_n]) < 1-\kappa\}) < e^{-n\beta_0}$,
2. $a_n/l_n < n\beta$, $\log(u_n/l_n) < n\beta$.

then $f_0 \in KL(\Pi)$ *implies that* Π *achieves strong posterior consistency at* f_0.

PROOF. Take $\mathcal{F}_n = \mathcal{F}^{\kappa}_{a_n,l_n,u_n}$. Then the conditions of Theorem 4.1 are easily verified using Lemma 4.2 for a suitable choice of $\kappa > 0$. □

REMARK 4.2. If $\tilde{\Pi} = Dir(\alpha G_0)$, verification of conditions 1 and 2 becomes particularly simple. For example, if G_0 is a product of a normal on θ and an inverse gamma on σ^2, then the conditions of theorem 4.2 are satisfied if $a_n = O(\sqrt{n}), l_n = O(1/\sqrt{n})$ and $u_n = O(e^n)$.

5 Posterior Consistency: Regression

For regression we consider the simple linear regression model outlined in Section 2 with nonrandom covariates x_i's. A location-scale mixture prior $\Pi = \Pi^* \times \mu$ is defined on the parameter space $\Theta = \mathcal{F}^* \times \mathbb{R} \times \mathbb{R}$ by taking a (symmetrized) location-scale mixture prior Π^* on $\mathcal{F}^*$ and any prior μ on (α, β). The symmetrization of Π^* can be obtained by defining the random densities as $f = \phi * P^s$ where $dP^s(\theta,\sigma) = .5dP(\theta,\sigma) + .5dP(-\theta,\sigma)$, $P \sim Dir(\alpha G_0)$.

Following Amewou-Atisso et al. (2003), we start with two necessary assumptions on the covariates and some useful notations.

Assumption A. There exists $\epsilon_0 > 0$ such that

$$\liminf_{n\to 1} \frac{1}{n}\sum_{i=1}^{n} I\{x_i < -\epsilon_0\} > 0 \quad , \quad \liminf_{n\to 1} \sum_{i=1}^{n} I\{x_i > \epsilon_0\} > 0.$$

Assumption B. For some L, $|x_i| < L$ for all i.

Notations. Define,

$$f_{\alpha,\beta,i}(y) = f_{\alpha+\beta x_i}(y) = f(y - \alpha - \beta x_i)$$

with $f_{0i} = f_{0,\alpha_0,\beta_0,i}$. For any two densities f and g, let

$$K(f,g) = \int f \log \frac{f}{g}, \quad V(f,g) = \int f \left(\log_+ \frac{f}{g}\right)^2.$$

and put

$$K_i(f,\alpha,\beta) = K(f_{0i}, f_{\alpha,\beta,i}), \quad V_i(f,\alpha,\beta) = V(f_{0i}, f_{\alpha,\beta,i}).$$

The main tool in assessing weak posterior consistency for this non iid case is the following variant of Schwartz theorem presented in Amewou-Atisso et al. (2003).

Theorem 5.1. *If*

1. *Assumptions* A *and* B *hold*

2. *For all* $\delta > 0$,

$$\Pi\left\{(f,\alpha,\beta) : K_i(f,\alpha,\beta) < \delta\ \forall i, \sum_{i=1}^{n} \frac{V_i(f,\alpha,\beta)}{i^2} < \infty\right\} > 0,$$

then Π *has weak posterior consistency at* (f_0, α_0, β_0).

Our main result that verifies the conditions of the above theorem for f_0 with flat tails is given in Theorem 5.2. But first we present an important lemma which is similar in nature to lemma 6.1 in Amewou-Atisso et al. (2002), but the higher moment condition on the true density has been relaxed.

LEMMA 5.1. *Fix a* P, *and take* $f(x) = \phi * P$. *If*

1. $\int f_0(x)(\log f_0(x))^2 dx < \infty$.
2. $\exists\ \eta > 0$ *such that* $\int |x|^{2\eta} f_0(x) dx < \infty$.
3. $\sigma_P = \int ((2\pi)^{1/2}\sigma)^{-1} dP(\theta,\sigma) < \infty$.
4. $\exists M, b_1, b_2, \eta$ *(all positive) such that* $\forall |x| > M$,

$$f(x) > b_1 \exp(-b_2|x|^{\eta}),$$

then,

(a) $\lim_{t\to 0} \int f_0(y) \log \frac{f_0(y)}{f_t(y)} dy = \int f_0(y) \log \frac{f_0(y)}{f(y)} dy$,

(b) $\lim_{t\to 0} \int f_0(y) \left(\log \frac{f_0(y)}{f_t(y)}\right)^2 dy = \int f_0(y) \left(\log \frac{f_0(y)}{f(y)}\right)^2 dy$.

The proof of this lemma is presented in the appendix. Here we state our main result which uses assumptions similar to the assumptions of Theorem 3.3.

THEOREM 5.2. *Suppose* $\Pi = \Pi^* \times \mu$ *be a location-scale mixture prior induced by a symmetrized Dirichlet location scale mixture on the densities. Let* G_0 *be the base measure of the underlying Dirichlet process prior. Assume that,*

1. *Assumptions 3,* $3'$, *4 and* $4'$ *of Theorem 3.3 hold for* G_0 *for some* $0 < \beta < \eta, \gamma > \beta$.

2. $\int f_0(x)(\log f_0(x))^2 dx < \infty$.

3. $\int |x|^{2\eta} f_0(x) dx < \infty$.

4. $\sigma_P = \int (1/\sigma) dP(\theta, \sigma) < \infty$ *almost surely.*

Then, Π *achieves weak posterior consistency at* (f_0, α_0, β_0) *provided* (α_0, β_0) *is in the support of* μ.

PROOF. Following the arguments presented in Lemma 3.3 it can be shown that for any $\delta > 0$ there exist some positive constants $b_1, b_2, x_0 > 0$ such that

$$\Pi^* \left(f : f(x) \geq b_1 e^{-b_2 |x|^\eta} \quad \forall |x| > x_0 \right) > 0$$

This combined with assumptions *2* and *4* gives

$$\Pi^* \{ f : K(f_0, f) < \delta, \;\; V(f_0, f) < \infty \} > 0. \tag{5.1}$$

As can be seen from the proof of Lemma 3.3, with probability 1, condition *4* of lemma 5.1 would be satisfied (these constants would depend on P). This property and assumption *4* of this theorem ensures that the results a) and b) of lemma 5.1 hold almost surely. Put $\theta_i = \alpha - \alpha_0 + (\beta - \beta_0) x_i$. Hence, applying lemma 5.1 we get that there exists an $\delta_f > 0$ such that for $|\theta_i| < \delta_f$,

$$\begin{aligned} K_i(f, \alpha, \beta) &= \int f_{0i} \log \frac{f_{0i}}{f_{\alpha,\beta,i}} = \int f_0 \log \frac{f_0}{f_{\theta_i}} = \int f_0 \log \frac{f_0}{f} + \int f_0 \log \frac{f}{f_{\theta_i}} \\ &< K(f_0, f) + \delta \end{aligned} \tag{5.2}$$

and

$$\begin{aligned} V_i(f, \alpha, \beta) &= \int f_{0i} \left(\log \frac{f_{0i}}{f_{\alpha,\beta,i}} \right)^2 = \int f_0 \left(\log \frac{f_0}{f_{\theta_i}} \right)^2 \\ &\leq 2 \int f_0 \left[\left(\log \frac{f_0}{f} \right)^2 + \left(\log \frac{f}{f_{\theta_i}} \right)^2 \right] \\ &< 2V(f_0, f) + 2\delta \end{aligned} \tag{5.3}$$

Thus, combining (5.1) with (5.2) and (5.3) above we see that condition *2* of theorem 5.1 is satisfied if the prior puts positive probabilities on $|\theta_i| < \delta$ for all small $\delta > 0$. But this is ensured by assumption B and the assumption that μ puts positive probability on neighborhoods of (α_0, β_0). This proves the theorem. □

Appendix: Proofs

Proof of Lemma 3.1. Denote by G the set $(-a, a) \times (\underline{\sigma}, \overline{\sigma})$. Choose $k > a + \overline{\sigma}$ such that $\int_{|x|>k} (|x| + a)^2/(2\underline{\sigma}^2) f_0(x) dx < \epsilon/2$ (possible since $\int |x|^2 f_0(x) dx < \infty$). Take $V = \{P : P(G) > \underline{\sigma}/\overline{\sigma}\}$. Then V is weak neighborhood of $\tilde{P}$ and for any $f = \phi * P$ with $P \in V$,

$$\begin{aligned}\int_{|x|>k} f_0(x) \log \frac{\tilde{f}(x)}{f(x)} dx &\leq \int_{|x|>k} f_0(x) \log \frac{\int_G \frac{1}{\sigma}\phi\left(\frac{x-\theta}{\sigma}\right) d\tilde{P}(\theta, \sigma)}{\int \frac{1}{\sigma}\phi(\frac{x-\theta}{\sigma}) dP(\theta, \sigma)} dx \\ &\leq \int_{|x|>k} f_0(x) \log \frac{\frac{1}{\overline{\sigma}}\phi\left(\frac{|x|-a}{\overline{\sigma}}\right)}{\frac{1}{\underline{\sigma}}\phi(\frac{|x|+a}{\underline{\sigma}}) P(G)} dx \\ &\leq \int_{|x|>k} f_0(x) \left\{\frac{(|x| + a)^2}{2\underline{\sigma}^2}\right\} dx < \frac{\epsilon}{2} \qquad \text{(A.1)}\end{aligned}$$

An almost exact repetition of the arguments given in the proof of Theorem 3 of Ghosal et al. (1999) shows that there exists a weak open neighborhood U of $\tilde{P}$ such that for any $f = \phi * P$ with $P \in U$,

$$\int_{|x|\leq k} f_0(x) \log \frac{\tilde{f}(x)}{f(x)} dx < \epsilon/2.$$

This proves the lemma with $W = V \cap U$. □

Proof of Lemma 3.2. Fix $0 < \tau < 1$ and $\epsilon > 0$ and find x_0 and $\mathcal{A}$ using the property of f_0, such that $\tilde{\Pi}(\mathcal{A}) > 1 - \tau$ and

$$\int_{|x|>x_0} f_0(x) \log \frac{f_0(x)}{f(x)} dx < \epsilon/2 \qquad \text{(A.2)}$$

Let g_0 be the density obtained by truncating f_0 to the interval $[-x_0, x_0]$. It follows from the arguments presented in Remark 3 of Ghosal et al. (1999), that $\exists \sigma_1 > 0$ such that

$$\begin{aligned}&\int_{-x_0}^{x_0} f_0(x) \log \frac{f_0(x)}{\int_{-x_0}^{x_0} \frac{1}{\sigma_1}\phi(\frac{x-\theta}{\sigma_1}) f_0(\theta) d\theta} dx \\ &\qquad = \frac{1}{t_0} \int g_0(x) \log \frac{g_0(x)}{\int \frac{1}{\sigma_1}\phi(\frac{x-\theta}{\sigma_1}) g_0(\theta) d\theta} dx < \frac{\epsilon}{4} \qquad \text{(A.3)}\end{aligned}$$

Let P_0 be such that such that $dP_0 = f_0 \times \delta_{\sigma_1}$. Fix a $0 < \kappa$ and find $\lambda > 0$ such that $1 - \lambda/(\kappa^2(1 - \lambda)^2) > \tau$. Choose a big compact set K

such that $[-x_0, x_0] \times \{\sigma_1\} \subset K$ and $G_0(K) > 1 - \lambda, P_0(K) > 1 - \lambda$. Let $\mathcal{B} = \{P : |P(K)/P_0(K) - 1| < \kappa\}$. Since $P(K) \sim Beta(\alpha G_0(K), \alpha G_0(K^c))$, an application of Chebyshev's inequality obtains,

$$\tilde{\Pi}(\mathcal{B}) \geq 1 - \frac{E(P(K) - P_0(K))^2}{\kappa^2 P_0(K)^2} \geq 1 - \frac{\lambda}{\kappa^2(1-\lambda)^2} > \tau$$

Thus, $\Pi(\mathcal{A} \cap \mathcal{B}) > 0$.

Following Ghosal et al. (1999), it can be shown that there exists a set $\mathcal{C}$ such that $\tilde{\Pi}(\mathcal{A} \cap \mathcal{B} \cap \mathcal{C}) > 0$. and $P \in \mathcal{B} \cap \mathcal{C} \Rightarrow$

$$\int_{-x_0}^{x_0} f_0(x) \log \frac{\int_K \frac{1}{\sigma}\phi(\frac{x-\theta}{\sigma}) dP_0}{\int_K \frac{1}{\sigma}\phi(\frac{x-\theta}{\sigma}) dP} dx < \frac{\kappa}{1-\kappa} + 2\kappa < \epsilon/4 \tag{A.4}$$

by a suitable choice of κ. But, for $f = \phi * P$ with $P \in \mathcal{A} \cap \mathcal{B} \cap \mathcal{C}$,

$$\begin{aligned} \int f_0(x) \log \frac{f_0(x)}{f(x)} dx \quad &\leq \quad \int_{-x_0}^{x_0} f_0(x) \log \frac{f_0(x)}{\int_{-x_0}^{x_0} \frac{1}{\sigma_1}\phi(\frac{x-\theta}{\sigma_1}) f_0(\theta) d\theta} dx \\ & \quad + \int_{-x_0}^{x_0} f_0(x) \log \frac{\int_K \frac{1}{\sigma}\phi(\frac{x-\theta}{\sigma}) dP_0}{\int_K \frac{1}{\sigma}\phi(\frac{x-\theta}{\sigma}) dP} dx \\ & \quad + \int_{|x|>x_0} f_0(x) \log \frac{f_0(x)}{f(x)} dx \\ &< \quad \epsilon \end{aligned} \tag{A.5}$$

by (A.3), (A.4) and (A.2). This concludes the proof of theorem 3.3. □

PROOF OF LEMMA 3.3. We will show how to do this for $x > 0$, the same argument can be repeated for $x < 0$. Note that, one can also represent K_x as:

$$K_x \left\{ (\theta, \sigma) : x - \tau_x(\sigma) \leq \theta \leq x + \tau_x(\sigma),\ 0 < \sigma \leq \frac{1}{w(x)} \right\}$$

where

$$\tau_x(\sigma) = (-2\sigma^2 \log(\sigma w(x)))^{1/2}, \ \ 0 < \sigma \leq \ 1/w(x),$$

is the varying width of K_x. It follows from simple calculations that for a fixed x, $\tau_x(\sigma) \to 0$ when $\sigma \to 0$ or $\sigma \to 1/w(x)$ and the maximum value of $\tau_x(\sigma)$ is attained at $\sigma = \exp(-1/2)/w(x)$. This gives K_x a balloon shape and makes it difficult to work with. To tackle this problem we further define for $x \geq 0$:

$$A_x = \begin{cases} \mathbb{R} \times \mathbb{R}^+, & x \leq t_0 \\ \{[x - \tau_x(\sigma_0), \infty) \times [\sigma_0, \infty)\} \bigcup \left\{[0, \infty) \times (x^{1-\eta/2}, \infty)\right\}, & x > t_0 \end{cases}$$

and

$$B_x = \begin{cases} \mathbb{R} \times \mathbb{R}^+, & x = 0 \\ \{(x, \infty) \times (\sigma_0, \infty)\} \bigcup \left\{[0, \infty) \times (e^{|x|^\eta - 1/2}, \infty)\right\}, & x > 0 \end{cases}$$

where t_0 is such that the function $\psi(x) = x - \tau_x(\sigma_0)$, defined on $\{x : w(x) < \sigma_0^{-1}\}$, is monotonically increasing for $x > t_0$. Existence of such a t_0 can be directly verified by differentiating $\psi(x)$. We summarize the connections between A_x, B_x and K_x in the following lemma which we would prove later.

LEMMA A.1. *For a suitable choice of* t_0,

$$1) \quad B_x \subseteq A_x, \forall x \geq 0. \tag{A.6}$$

$$2) \quad 0 < x < y \Rightarrow A_x^c \subset A_y^c \text{ and } B_x^c \subset B_y^c. \tag{A.7}$$

$$3) \quad A_0^c = B_0^c = \phi \text{ and } A_x^c, B_x^c \uparrow \mathbb{R} \times \mathbb{R}^+ \text{ as } x \uparrow \infty. \tag{A.8}$$

$$4) \quad A_x \setminus B_x \subset K_x. \tag{A.9}$$

Lemma A.1 ensures that $F^A(x) = P(A_x^c)$ and $F^B(x) = P(B_x^c)$ are random distribution functions on $[0, \infty)$, following $Dir(\alpha G_0^A)$ and $Dir(\alpha G_0^B)$ respectively, where these parameters are given by $G_0^A([0, x]) = G_0(A_x^c)$ and $G_0^B([0, x]) = G_0(B_x^c)$. It follows from Doss and Sellke (1982) (see Ghosal et al., 1999), that

$$\Pr\left(\limsup_{x\to\infty} \frac{P(B_x)}{h_1(G_0(B_x))} = 0\right) = 1$$

and

$$\Pr\left(\liminf_{x\to\infty} \frac{P(A_x)}{h_2(G_0(A_x))} = \infty\right) = 1$$

where $h_1(t) = \exp(-1/(t|\log t|^2))$ and $h_2(t) = \exp(-2\log|\log t|/t)$ for small positive values of t. An application of Egoroff's theorem to the above yields that there exist an $x_1 > 0$ and a set $\mathcal{A}$ with $\Pi(\mathcal{A}) > 1 - \tau/2$, such that

$$P \in \mathcal{A} \Rightarrow \begin{cases} P(B_x) < h_1(G_0(B_x)) \\ P(A_x) > h_2(G_0(A_x)) \end{cases} \quad \forall x > x_1.$$

Therefore $P \in \mathcal{A}$ implies that for all $x > x_1$,

$$\begin{aligned} P(K_x) &\geq P(A_x \setminus B_x) \geq h_2(G_0(A_x)) - h_1(G_0(B_x)) \\ &= \exp\left(-\frac{2\log|\log G_0(A_x)|}{G_0(A_x)}\right) - \exp\left(-\frac{1}{G_0(B_x)(\log G_0(B_x))^2}\right) \\ &= \exp(-l(x)) - \exp(-u(x)) \qquad \text{(A.10)} \end{aligned}$$

where

$$l(x) = (2\log|\log G_0(A_x)|)/(G_0(A_x))$$

and

$$u(x) = 1/(G_0(B_x)(\log G_0(B_x))^2).$$

By assumption 3 and the inequality $x - \sigma_0 x^{\eta/2} > x - \tau_x(\sigma_0)$ for large $x > 0$, that we would establish later during the proof of Lemma A.1 (see relation (A.11)), we conclude $G_0(A_x) \geq b_1 x^{-\beta}$ for large $x > 0$. On the other hand, by assumption 4, $G_0(B_x) \leq G_0\left(\left\{(-\infty, x) \times (0, e^{x^\eta - 1/2})\right\}^c\right) \leq b_2 x^{-\gamma}$ for large $x > 0$. Hence, for large $x > 0$,

$$\frac{l(x)}{u(x)} \leq \frac{2b_2}{b_1} x^{\beta-\gamma} \log|\beta\log x - \log b_1| \left(\gamma\log x - \log b_2\right)^2$$

and so, $l(x)/u(x) \to 0$ as $x \to \infty$ because $\beta < \gamma$. Therefore, $\exists x_0 > x_1 > 0$ such that $P \in \mathcal{A}$ and $x > x_0$ implies

$$P(K_x) \geq \frac{1}{2}\exp(-l(x)) \geq \frac{1}{2}\exp(-\frac{2}{b_1}x^\beta \log|\beta\log x - \log 2|) \geq \frac{1}{2}\exp(-\frac{2}{b_1}x^\eta)$$

since $\beta < \eta$. This completes the proof. □

PROOF OF LEMMA A.1.

1. Note that, $x > x - \tau_x(\sigma_0)$ for all x such that the second term is defined. Also, $\exp(x^\eta - 1/2) \geq x^{1-\eta/2}$ for large enough $x > 0$. Hence (A.6) follows.

2. Proof of (A.7) follows directly by monotonicity of $exp(x^\eta - 1/2)$, $x^{1-\eta/2}$ and $x - \tau_x(\sigma_0)$ for $x > t_0$.

3. It is straightforward to see that $A_0^c = B_0^c = \phi$ and $B_x^c \to \mathbb{R} \times \mathbb{R}^+$. A direct calculation also shows that,

$$x - \sigma_0 x^{\eta/2} > x - \tau_x(\sigma_0) > x - 2\sigma_0 x^{\eta/2}. \tag{A.11}$$

   for large $x > 0$. From this it follows that $A_x^c \to \mathbb{R} \times \mathbb{R}^+$ since $x - 2\sigma_0 x^{\eta/2} \to \infty$ as $x \to \infty$. This completes the proof of (A.8).

4. Note that for large x, $x^{1-\eta/2} < \exp(x^\eta - 1/2) = \exp(-1/2)/w(x)$ which means $\tau_x(x^{1-\eta/2})$ is well defined and $x - \tau_x(x^{1-\eta/2}) = x - x(2 - (2-\eta)(\log x/x^\eta))^{1/2} < 0$. Therefore,

$$\begin{aligned} A_x \setminus B_x \;=\; & \left\{(\theta,\sigma):\ x - \tau_x(\sigma_0) \le \theta \le x, \sigma_0 < \sigma \le \frac{e^{-1/2}}{w(x)}\right\} \bigcup \\ & \left\{(\theta,\sigma):\ 0 \le \theta \le x, x^{1-\eta/2} < \sigma \le \frac{e^{-1/2}}{w(x)}\right\} \\ & \subset \left\{(\theta,\sigma):\ x - \tau_x(\sigma_0) \le \theta \le x, \sigma_0 < \sigma \le \frac{e^{-1/2}}{w(x)}\right\} \bigcup \\ & \left\{(\theta,\sigma):\ x - \tau_x(x^{1-\eta/2}) \le \theta \le x, x^{1-\eta/2} < \sigma \le \frac{e^{-1/2}}{w(x)}\right\} \\ & \subset K_x. \end{aligned} \tag{A.12}$$

   The last inclusion follows because $\tau_x(\sigma)$ takes its maximum value at $h = e^{-1/2}/w(x)$ and this ensures that for a fixed $h' < e^{-1/2}/w(x)$,

$$x - \tau_x(\sigma) \le x - \tau_x(\sigma') \qquad \forall\, \sigma \in \left[h', \frac{e^{-1/2}}{w(x)}\right].$$

□

Proof of Lemma 5.1. Under assumption 3, $f(x)$ is a continuous function. This follows from the fact that

$$(1/\sigma)\phi\left((y-\theta-t)/\sigma\right) \to (1/\sigma)\phi\left((y-\theta)/\sigma\right) \text{ as } t \to 0,$$

and if we consider the collection of functions $(1/\sigma)\phi\left((y-\theta-t)/\sigma\right)$ indexed by t, it has a P-integrable upper bound in the function $1/((2\pi)^{1/2}\sigma)$. So

Dominated Convergence Theorem applies in establishing that $f(y-t) \to f(y)$ as $t \to 0$. This means that for each fixed y, as $t \to 0$,

$$\log \frac{f_0(y)}{f_t(y)} \to \log \frac{f_0(y)}{f(y)} \text{ and } \left(\log \frac{f_0(y)}{f_t(y)}\right)^2 \to \left(\log \frac{f_0(y)}{f(y)}\right)^2. \tag{A.13}$$

Fix a $t_0 > 0$. By assumption 4, $\forall |y| > M + t_0, \forall |t| < t_0$,

$$f_t(y) = f(y-t) > b_1 \exp(-b_2|y-t|^\eta) \geq b_1 \exp(-b_2(|y| + |t_0|)^\eta).$$

It is clear from the definition of f that $c_P := \inf\{f(x) : |x| \leq M + 2t_0\} > 0$. Then $\forall |y| \leq M + t_0, \forall |t| < t_0$,

$$f_t(y) = f(y-t) \geq c_P.$$

We also have the obvious bound that $\forall y, \forall |t| < t_0$,

$$f_t(y) \leq \int (1/\sigma) dP(\theta, \sigma) = \sigma_P < \infty.$$

Hence $\forall y$,

$$|\log f_t(y)| \leq \max\left(|\log \sigma_P|, |\log c_P|, |\log b_1| + b_2(|y| + |t_0|)^\eta\right). \tag{A.14}$$

Denote the RHS of (A.14) by $\xi(y)$. Then, $\int \xi(y)^2 f_0(y) dy < \infty$, by assumption 2 and $\forall y, \forall |t| < t_0$,

$$\left|\log \frac{f_0(y)}{f_t(y)}\right| < |\log f_0(y)| + \xi(y)$$

and

$$\left(\log \frac{f_0(y)}{f_t(y)}\right)^2 < 2\left((\log f_0(y))^2 + (\xi(y))^2\right).$$

Hence by assumption 1 and 2, DCT applies to (A.13) proving the lemma. □

References

Amewou-Atisso, M., Ghosal, S., Ghosh, J.K. and Ramamoorthi, R.V. (2003). Posterior consistency for semiparametric regression problems. *Bernoulli* **9**, 291–312.

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

Doss, H. and Sellke, T. (1982). The tails of probabilities chosen from a Dirichlet prior. *Ann. Statist.* **10**, 1302–1305.

Escobar, M. and West, M. (1995). Bayesian density estimation and inference using mixtures. *J. Amer. Statist. Assoc.* **90**, 577–588.

Ferguson, T.S. (1983). Bayesian density estimation by mixtures of Normal distributions. In *Recent Advances in Statistics*, (Rizvi M., Rustagi, J. and Siegmund, D., eds.). 287-302.

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

Ghosal, S. and van der Vaart, A. W. (2001). Entropies and rates of convergence for Bayes and maximum likelihood estimation for mixture of normal densities. *Ann. Statist.* **29**, 1233–1263.

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

Liu, J.S. (1996). Nonparametric hierarchical Bayes via sequential imputations. *Ann. Statist.* **24**, 911–930.

Lo, A.Y. (1984). On a class of Bayesian nonparametric estimates I: Density estimates. *Ann. Statist.* **12**, 351–357.

MacEachern, S. (1994). Estimating normal means with a conjugate style Dirichlet process prior. *Comm. Statist. Simulation Comput.* **23**, 727–741.

Müller, P., Erkanli, A. and West, M. (1996). Bayesian curve fitting using multivariate normal mixtures. *Biometrika* **83**, 67–79.

Schwartz, L. (1965). On Bayes procedures. *Z. Wahr. Verw. Gebiete* **4**, 10–26.

Walker, S.G. (2004). New approaches to Bayesian consistency. *Ann. Statist.* **32**, 2028–2043.

West, M., Müller, P. and Escobar, M. (1994). Hierarchical priors and mixture models with applications in regression and density estimation. *Aspects of Uncertainty: A Tribute to D. V. Lindley*, (Smith, A. and Freeman, P., eds.). Wiley, New York. 363–386.

Surya T. Tokdar
Department of statistics
Purdue University
150 N University St
W Lafayette, IN 47906, USA
E-mail: stokder@stat.purdue.edu

Paper received June 2005; revised November 2005.