The Annals of Statistics
1999, Vol. 27, No. 1, 143–158

POSTERIOR CONSISTENCY OF DIRICHLET MIXTURES IN DENSITY ESTIMATION

BY S. GHOSAL, J. K. GHOSH AND R. V. RAMAMOORTHI

Vrije Universiteit, Amsterdam, Indian Statistical Institute, Calcutta, and Purdue University and Michigan State University

A Dirichlet mixture of normal densities is a useful choice for a prior distribution on densities in the problem of Bayesian density estimation. In the recent years, efficient Markov chain Monte Carlo method for the computation of the posterior distribution has been developed. The method has been applied to data arising from different fields of interest. The important issue of consistency was however left open. In this paper, we settle this issue in affirmative.

1. Introduction. Recent years have seen a surge of interest in nonparametric priors on densities arising out of mixtures. These priors were introduced by Lo (1984) [see also Ghorai and Rubin (1982)], who obtained expressions for the resulting posterior and predictive distributions. The mixture model uses a kernel $K(x, \theta)$ on $\mathscr{X}\times\Theta$, that is, $K(x, \theta)$ is a measurable function such that for all θ, $K(\cdot, \theta)$ is a density (with respect to some σ-finite measure) on $\mathscr{X}$. If for any probability P on Θ, $K(x, P) = \int K(x, \theta)\, dP(\theta)$, then, any prior Π on the space of probability measures on Θ gives rise to a prior on densities via the map $P \mapsto K(\cdot, P)$. The resulting model for n independent and identically distributed (i.i.d.) observations would then be $P \sim \Pi$, given P, $X_1, X_2, \dots, X_n$ are i.i.d. $K(\cdot, P)$. Another equivalent formulation, which is convenient for Bayesian computation, is to treat $\theta_1, \theta_2, \dots, \theta_k$ as i.i.d. given P and given the θ_i's, X_i are independent with X_i having density $K(\cdot, \theta_i)$.

A simple kernel when $\mathscr{X} = \mathbb{R}$ is $K(x, \theta) = 1/h$, $kh < x$, $\theta < (k+1)h$, $k = \dots, -1, 0, 1, \dots$, which leads to random histograms studied by Gasparini (1992). A general choice of the kernel might be $h^{-1}K((x-\theta)/h)$, where K is a symmetric density around 0. The scale parameter h plays a role somewhat similar to that of window length in density estimation problems, where it is chosen a priori or decided empirically from the observations. In the Bayesian context, h might be elicited a priori, and if that is not possible, then it could be treated as a hyperparameter endowed with a prior. In the later case, the predictive density given $X_1, X_2, \dots, X_n$ is a mixture of $h^{-1}K((x-\theta_i)/h)$, $i = 1, 2, \dots, k$, $k \le n$, and thus corresponds to a data driven choice of the "window lengths" h.

Of special interest is the case when K is the normal density and Π is a Dirichlet process. The base measure $\alpha = M\alpha_0$ of the Dirichlet can be elicited

Received February 1998; revised September 1998.
AMS 1991 *subject classifications*. Primary 62G07, 62G20.
Key words and phrases. Consistency, Dirichlet process, mixture, posterior consistency, posterior distribution.

up to some parameters and a hierarchical prior can be considered for these hyperparameters. Yet another possibility is to start with a Dirichlet process prior for (θ, h). Ferguson (1983) studied some of these models and more recently, West (1992) and West, Muller and Escobar (1994) have developed powerful Markov chain Monte Carlo methods to calculate Bayes estimates and other posterior quantities and have also used these priors very effectively in many applications.

This paper addresses issues related to consistency of the posterior of these mixture models and is organized as follows.

In Section 2, we give some basic definitions and state two theorems. The first theorem, due to Schwartz (1965), is used to establish weak consistency. Since the space under consideration is a set of densities, as argued in Barron, Schervish and Wasserman (1998), strong consistency, that is, consistency for L_1-neighborhoods, is more appropriate. Schwartz's theorem is not useful for establishing strong consistency. Our Theorem 2 is the key result toward the strong consistency for densities. This theorem is in the spirit of Theorem 1 in Barron, Schervish and Wasserman (1998) for consistency for L_1-neighborhoods. They use L_1-entropy with upper bracketing while we use L_1-metric entropy.

Section 3 is devoted to weak consistency. We present results for general mixture models and also for Dirichlet–normal mixture models.

In Section 4 we study strong consistency and Section 5 contains a brief discussion of some possible extensions.

2. Consistency theorems. Let $\mathscr{F}$ be the set of all densities on $\mathbb{R}$ with respect to Lebesgue measure. There are two natural topologies on $\mathscr{F}$—the weak topology and the norm topology. Thus if $f_0 \in \mathscr{F}$, a weak neighborhood of f_0 is a set containing a set of the form

$$V = \left\{ f \in \mathscr{F}\colon \left| \int \phi_i f - \int \phi_i f_0 \right| < \varepsilon,\ i = 1, 2, \dots, k \right\},$$

where ϕ_i's are bounded continuous functions on $\mathbb{R}$. A strong or L_1-neighborhood is a set containing a set of the form $V = \{f \in \mathscr{F}\colon \|f - f_0\| < \varepsilon\}$, where $\|f - f_0\| = \int |f - f_0|$.

Let Π be a prior on $\mathscr{F}$ and given f, let $X_1, X_2, \dots, X_n$ be i.i.d. with common density f. Then for any measurable subset A of $\mathscr{F}$, the posterior probability of A given $X_1, X_2, \dots, X_n$ is

$$\Pi(A | X_1, X_2, \dots, X_n) = \frac{\int_A \prod_{i=1}^n f(X_i) \Pi(df)}{\int_{\mathscr{F}} \prod_{i=1}^n f(X_i) \Pi(df)}.$$

For a density f, let P_f stand for the probability measure corresponding to f.

Definition 1. A prior Π is said to be weakly consistent at f_0, if with P_{f_0}-probability 1,

$$\Pi(U | X_1, X_2, \dots, X_n) \to 1$$

for all weak neighborhoods U of f_0.

DEFINITION 2. A prior Π is said to be strongly consistent at f_0 if with P_{f_0}-probability 1,

$$\Pi(U|X_1, X_2, \dots, X_n) \to 1$$

for all strong neighborhoods U of f_0.

If μ is a probability measure on a complete separable metric space $\mathscr{X}$, then x is said to be in the *support* of μ if every open neighborhood of x has positive μ measure. In particular, if Π is a prior on the set of all probabilities on $\mathbb{R}$, then P_0 is in the support of Π if every weak neighborhood of P_0 has positive Π measure.

Another useful notion is that of *K–L support*. For any $f_0 \in \mathscr{F}$, we denote by $K_\varepsilon(f_0)$ the Kullback–Leibler neighborhood $\{f\colon \int f_0 \log(f_0/f) < \varepsilon\}$. Say that f_0 is in the *K–L support of* Π if $\Pi(K_\varepsilon(f_0)) > 0$ for all $\varepsilon > 0$.

An early theorem on consistency due to Schwartz (1965) implies the following.

THEOREM 1 (Schwartz). *If f_0 is in the K–L support of Π, then the posterior is weakly consistent at f_0.*

Schwartz's theorem deals with weak consistency and the following theorem is developed to handle strong consistency. This involves conditions on the size of the parameter space measured in terms of L_1-metric entropy. We first recall the definition of L_1-metric entropy.

DEFINITION 3. Let $\mathscr{G} \subset \mathscr{F}$. For $\delta > 0$, the L_1-metric entropy $J(\delta, \mathscr{G})$ is defined as the logarithm of the minimum of all k such that there exist $f_1, f_2, \dots, f_k$ in $\mathscr{F}$ with the property $\mathscr{G} \subset \bigcup_{i=1}^k \{f\colon \|f - f_i\| < \delta\}$.

THEOREM 2. *Let Π be a prior on $\mathscr{F}$. Suppose $f_0 \in \mathscr{F}$ is in the K–L support of Π and let $U = \{f\colon \|f - f_0\| < \varepsilon\}$. If there is a $\delta < \varepsilon/4$, $c_1, c_2 > 0$, $\beta < \varepsilon^2/8$ and $\mathscr{F}_n \subset \mathscr{F}$ such that, for all n large*:

(i) $\Pi(\mathscr{F}_n^c) < c_1 \exp(-nc_2)$, *and*
(ii) $J(\delta, \mathscr{F}_n) < n\beta$,

then $\Pi(U|X_1, X_2, \dots, X_n) \to 1$ *a.s.* P_{f_0}.

It is worthwhile to note that the constants δ, c_1, c_2, β and $\mathscr{F}_n$ are all allowed to depend on the fixed neighborhood, equivalently on the fixed ε. The proof of the theorem is given in the Appendix.

This last theorem is very much in the spirit of Barron, Schervish and Wasserman (1998). Their theorem is in terms of L_1-entropy with upper bracketing. If $\mathscr{G} \subset \mathscr{F}$, for $\delta > 0$, the L_1-entropy with upper bracketing $J_1(\delta, \mathscr{G})$ is defined as the logarithm of the minimum of all k such that there exist $g_1, g_2, \dots, g_k$ satisfying:

1. $\int g_i \le 1 + \delta$;
2. For every $g \in \mathscr{G}$ there exists an i such that $g \le g_i$.

Since $g_i^* = g_i / \int g_i$ is in $\mathscr{F}$, it is easy to see that $J(2\delta, \mathscr{G}) \le J_1(\delta, \mathscr{G})$. Hence Theorem 2 is somewhat more general than the result of Barron, Schervish and Wasserman (1998) and is, at least in some examples, more convenient to apply. In general [see, e.g., van der Vaart and Wellner (1996), page 84], except for the uniform norm, there is no inequality in the reverse direction.

3. Dirichlet mixtures: weak consistency. Returning to the mixture model, let ϕ and ϕ_h denote, respectively, the standard normal density and the normal density with mean 0 and standard deviation h. Let $\Theta = \mathbb{R}$ and $\mathscr{M}$ be the set of probability measures on Θ. If P is in $\mathscr{M}$, then $f_{h,P}$ will stand for the density

$$f_{h,P}(x) = \int \phi_h(x-\theta)\, dP(\theta).$$

Note that $f_{h,P}$ is just the convolution $\phi_h * P$.

Our model consists of a prior μ for h and a prior Π on $\mathscr{M}$. The prior $\mu \times \Pi$ through the map $(h, P) \mapsto f_{h,P}$ induces a prior on $\mathscr{F}$. We continue to denote this prior also by Π. Thus $(h, P) \sim \mu \times \Pi$ and given (h, P), $X_1, X_2, \dots, X_n$ are i.i.d. $f_{h,P}$. This section describes a class of densities which are in the K–L support of Π. By Schwartz's theorem the posterior will be weakly consistent at these densities.

THEOREM 3. *Let the true density f_0 be of the form $f_0(x) = f_{h_0, P_0}(x) = \int \phi_{h_0}(x-\theta)\, dP_0(\theta)$. If P_0 is compactly supported and belongs to the support of* Π, *and h_0 is in the support of μ, then $\Pi(K_\varepsilon(f_0)) > 0$ for all $\varepsilon > 0$.*

PROOF. Suppose $P_0[-k, k] = 1$. Since P_0 is in the weak support of Π, it follows that $\Pi\{P\colon P[-k, k] > 1/2\} > 0$. Also it is easy to see that f_0 has moments of all orders.

For $\eta > 0$, choose k' such that $\int_{|x|>k'} \max(1, |x|) f_0(x)\, dx < \eta$. For $h > 0$, we write $\int_{-\infty}^{\infty} f_0 \log(f_{h,P_0}/f_{h,P})$ as the sum

$$\int_{-\infty}^{-k'} f_0 \log \frac{f_{h,P_0}}{f_{h,P}} + \int_{-k'}^{k'} f_0 \log \frac{f_{h,P_0}}{f_{h,P}} + \int_{k'}^{\infty} f_0 \log \frac{f_{h,P_0}}{f_{h,P}}. \tag{1}$$

Now

$$\begin{aligned}
&\int_{-\infty}^{-k'} f_0(x) \log\left(\frac{f_{h,P_0}(x)}{f_{h,P}(x)}\right) dx \\
&\quad \le \int_{-\infty}^{-k'} f_0(x) \log\left(\frac{\int_{-k}^{k} \phi_h(x-\theta)\, dP_0(\theta)}{\int_{-k}^{k} \phi_h(x-\theta)\, dP(\theta)}\right) dx \\
&\quad \le \int_{-\infty}^{-k'} f_0(x) \log\left(\frac{\phi_h(x+k)}{\phi_h(x-k) P[-k,k]}\right) dx \\
&\quad = \int_{-\infty}^{-k'} f_0(x) \frac{2k|x|}{h^2}\, dx - \log(P[-k,k]) \int_{-\infty}^{-k'} f_0(x)\, dx \\
&\quad < \left(\frac{2k}{h^2} + \log 2\right)\eta,
\end{aligned}$$

provided $P[-k, k] > 1/2$. Similarly, we get a bound for the third term in (1).

Clearly,

$$c := \inf_{|x| \le k'} \inf_{|\theta| \le k} \phi_h(x - \theta) > 0.$$

The family of functions $\{\phi_h(x-\theta): x \in [-k', k']\}$, viewed as a set of functions of θ in $[-k, k]$, is uniformly equicontinuous. By the Arzela–Ascoli theorem, given $\delta > 0$, there exist finitely many points $x_1, x_2, \ldots, x_m$ such that for any $x \in [-k', k']$, there exists an i with

$$\sup_{\theta \in [-k, k]} |\phi_h(x - \theta) - \phi_h(x_i - \theta)| < c\delta. \tag{2}$$

Let

$$E = \left\{ P: \left| \int \phi_h(x_i - \theta)\, dP_0(\theta) - \int \phi_h(x_i - \theta)\, dP(\theta) \right| < c\delta;\ i = 1, 2, \ldots, m \right\}.$$

Since E is a weak neighborhood of P_0, $\Pi(E) > 0$. Let $P \in E$. Then for any $x \in [-k', k']$, choosing the appropriate x_i from (2) and using a simple triangulation argument, we get

$$\left| \frac{\int \phi_h(x-\theta)\, dP(\theta)}{\int \phi_h(x-\theta)\, dP_0(\theta)} - 1 \right| < 3\delta$$

and so

$$\left| \frac{\int \phi_h(x-\theta)\, dP_0(\theta)}{\int \phi_h(x-\theta)\, dP(\theta)} - 1 \right| < \frac{3\delta}{1 - 3\delta}$$

(provided $\delta < 1/3$).

Thus for any fixed $h > 0$, for P in a set of positive Π-probability, we have

$$\int f_0 \log(f_{h, P_0}/f_{h, P}) < 2\left(\frac{2k}{h^2} + \log 2\right)\eta + \frac{3\delta}{1 - 3\delta}. \tag{3}$$

Now for any h,

$$\int f_0 \log(f_0/f_{h, P}) = \int f_0 \log(f_0/f_{h, P_0}) + \int f_0 \log(f_{h, P_0}/f_{h, P}). \tag{4}$$

The first term on the right-hand side (RHS) of (4) converges to 0 as $h \to h_0$. To see this, observe that

$$\frac{\int \phi_{h_0}(x-\theta)\, dP_0(\theta)}{\int \phi_h(x-\theta)\, dP_0(\theta)} \le \sup_{|\theta| \le k} \frac{\phi_{h_0}(x-\theta)}{\phi_h(x-\theta)}.$$

The rest follows by an application of the dominated convergence theorem.

Now given any $\varepsilon > 0$, choose a neighborhood N of h_0 (not containing 0) such that if $h \in N$, the first term on the RHS of (4) is less than $\varepsilon/2$. Next choose η and δ so that for any $h \in N$, the RHS of (3) is less than $\varepsilon/2$. Since h_0 is in the support of μ, the result follows. □

REMARK 1. In Theorem 4, the true density is a compact location mixture of normals with a fixed scale. It is also possible to obtain consistency at true densities which are (compact) location-scale mixtures of the normal, provided we use a mixture prior for h as well. More precisely, if we modify the prior so that $(\theta, h) \sim P$ [a probability on $\mathbb{R} \times (0, \infty)$] and $P \sim \Pi$, then consistency holds at $f_0 = \int \phi_h(x-\theta) P_0(d\theta, dh)$ provided P_0 has compact support and belongs to the support of Π. The proof is similar to that of Theorem 3.

Theorem 3 covers the case when the true density is normal or a mixture of normal over a compact set of locations. This theorem, however, does not cover the case when the true density itself has compact support, like, say the uniform. The next theorem takes care of such densities.

THEOREM 4. *Let* 0 *be in the support of* μ *and* f_0 *be a density in the support of* Π. *Let* $f_{0,h} = \phi_h * f_0$. *If*:

(i) $\lim_{h\to 0} \int f_0 \log(f_0/f_{0,h}) = 0$ *and*
(ii) f_0 *has compact support, then* $\Pi(K_\varepsilon(f_0)) > 0$ *for all* $\varepsilon > 0$.

PROOF. Note that for each h,

$$\int f_0 \log(f_0/f_{h,P}) = \int f_0 \log(f_0/f_{0,h}) + \int f_0 \log(f_{0,h}/f_{h,P}).$$

Choose h_0 such that for $h < h_0$, $\int f_0 \log(f_0/f_{0,h}) < \varepsilon/2$ so all that is required is to show that for all $h > 0$,

$$\Pi\left\{P\colon \int f_0 \log(f_{0,h}/f_{h,P}) < \varepsilon/2\right\} > 0$$

if f_0 has support in $[-k, k]$. Then

$$\int f_0 \log(f_{0,h}/f_{h,P}) \le \int_{-k}^{k} f_0(x) \log\left(\frac{\int_{-k}^{k} \phi_h(x-\theta) f_0(\theta)\, d\theta}{\int_{-k}^{k} \phi_h(x-\theta)\, dP(\theta)}\right) dx.$$

The rest of the argument proceeds on the same lines as that in the last theorem. □

While the last two theorems are valid for general priors on $\mathscr{M}$, the next theorem makes strong use of the properties of the Dirichlet process. For any P in $\mathscr{M}$, set $\overline{P}(x) = P(x, \infty)$ and $\underline{P}(x) = P(-\infty, x)$.

THEOREM 5. *Let* D_α *be a Dirichlet process on* $\mathscr{M}$. *Let* l_1, l_2, u_1, u_2 *be functions such that for some* $k > 0$ *for all* P *in a set of* D_α*-probability* 1, *there exists* x_0 (*depending on* P) *such that*

$$\begin{aligned} &\overline{P}(x) \ge l_1(x),\ \bar{P}(x + k\log x) \le u_1(x) \qquad \forall x > x_0 \quad \textit{and} \\ &\underline{P}(x) \ge l_2(x),\ \underline{P}(x - k\log|x|) \le u_2(x) \qquad \forall x < -x_0. \end{aligned} \tag{5}$$

For any $h > 0$, *define*

$$L_h(x) = \begin{cases} \phi_h(k \log x)(l_1(x) - u_1(x)), & \text{if } x > 0, \\ \phi_h(k \log |x|)(l_2(x) - u_2(x)), & \text{if } x < 0, \end{cases}$$

and assume that $L_h(x)$ *is positive for sufficiently large* $|x|$. *Let* f_0 *be the "true" density and* $f_{0,h} = \phi_h * f_0$. *Assume that* 0 *is in the support of the prior on* h. *If* f_0 *is in the support of* D_α [*equivalently*, $\operatorname{supp}(f_0) \subset \operatorname{supp}(\alpha)$] *and satisfies*:

(i) $\lim_{h\downarrow 0} \int f_0 \log(f_0/f_{0,h}) = 0$;

(ii) *for all* h,

$$\lim_{a\uparrow\infty} \int_{-\infty}^{\infty} f_0(x) \log\left(\frac{f_{0,h}(x)}{\int_{-a}^{a} \phi_h(x-\theta) f_0(\theta)\, d\theta}\right) dx = 0;$$

(iii) *for all* h,

$$\lim_{M\to\infty} \int_{|x|>M} f_0(x) \log\left(\frac{f_{0,h}(x)}{L_h(x)}\right) dx = 0,$$

then $\Pi(K_\varepsilon(f_0)) > 0$ *for all* $\varepsilon > 0$.

REMARK 2. It follows from Doss and Sellke (1982) that if $\alpha = M\alpha_0$, where α_0 is a probability measure, then

$$l_1(x) = \exp[-2 \log |\log \overline{\alpha_0}(x)| / \overline{\alpha_0}(x)],$$
$$l_2(x) = \exp[-2 \log |\log \underline{\alpha_0}(x)| / \underline{\alpha_0}(x)],$$
$$u_1(x) = \exp\left[-\frac{1}{\overline{\alpha_0}(x + k\log x) |\log \overline{\alpha_0}(x - k \log x)|^2}\right],$$
$$u_2(x) = \exp\left[-\frac{1}{\underline{\alpha_0}(x - k\log |x|) |\log \underline{\alpha_0}(x - k \log |x|)|^2}\right]$$

satisfy the requirements of (5). For example, when α_0 is double exponential, we may choose any $k > 2$ and the requirements of the theorem are satisfied if f_0 has finite moment generating function in an open interval containing $[-1, 1]$.

REMARK 3. The following argument provides a method for the verification of Condition 1 of Theorem 4 and Theorem 5 for many densities. Suppose that f_0 is continuous a.e., $\int f_0 \log f_0 < \infty$ and further assume that, as for unimodal densities, there exists an interval $[a, b]$ such that, $\inf\{f(x)\colon x \in [a, b]\} = c > 0$ and f_0 is increasing in $(-\infty, a)$ and is decreasing in (b, ∞). Note that $\{x\colon f_0(x) \geq c\}$ is an interval containing $[a, b]$. Replacing the original $[a, b]$ by this new interval, we may assume that $f_0(x) \leq c$ outside $[a, b]$. Choose h_0 such that $N(0, h_0)$ gives probability 1/3 to $(0, b-a)$. Let $h < h_0$. Let Φ denote the cumulative distribution function of $N(0, 1)$. If $x \in [a, b]$ then

$$f_{0,h}(\theta) \geq \int_a^b f_0(\theta)\phi_h(x-\theta)\, d\theta \geq c(\Phi((b-x)/h) + \Phi((x-a)/h) \geq c/3.$$

If $x > b$, then

$$f_{0,h}(\theta) \ge \int_a^x f_0(\theta)\phi_h(x-\theta)\,d\theta \ge f_0(x)\big(\tfrac{1}{2} + \Phi((b-a)/h) - 1\big) \ge f_0(x)/3.$$

Using a similar argument when $x < a$, we have that the function

$$g(x) = \begin{cases} \log(3f_0(x)/c), & \text{if } x \in [a,b], \\ \log 3, & \text{otherwise}, \end{cases}$$

dominates $\log(f_0/f_{0,h})$ for $h < h_0$ and is P_{f_0}-integrable. Since $f_0(x)/f_{0,h}(x) \to 1$ as $h \to 0$ whenever x is a continuity point of f_0 and $\int f_0 \log(f_0/f_{0,h}) \ge 0$, an application of (a version of) Fatou's lemma shows that $\int f_0 \log(f_0/f_{0,h}) \to 0$ as $h \to 0$.

PROOF OF THEOREM 6. Let $\varepsilon > 0$ be given and $\delta > 0$, to be chosen later. First find h_0 so that $\int f_0 \log(f_0/f_{0,h}) < \varepsilon/2$ for all $h < h_0$. Fix $h < h_0$. Choose k_1 such that

$$\int_{-\infty}^{\infty} f_0(x) \log\left(\frac{f_{0,h}(x)}{\int_{-k_1}^{k_1} \phi_h(x-\theta) f_0(\theta)\,d\theta}\right) dx < \delta.$$

Let $p = P[-k_1, k_1]$ and let p_0 denote the corresponding value under P_0. We may assume that $p_0 > 0$. Let P^* denote the conditional probability under P given $[-k_1, k_1]$, that is, $P^*(A) = P(A \cap [-k_1, k_1])/p$ (if $p > 0$) and P_0^* denoting the corresponding objects for P_0. Let E be the event $\{P\colon |p/p_0 - 1| < \delta\}$. Since P_0 is in the support of D_α, $D_\alpha(E) > 0$. Now choose $x_0 > k_1$ such that:

(i) $\displaystyle\int_{|x|>x_0} f_0(x) \log\left(\frac{f_{0,h}(x)}{L_h(x)}\right) dx < \delta$;

(ii) $D_\alpha(E \cap F) > 0$, where

$$F = \left\{ P\colon \begin{matrix} \overline{P}(x) \ge l_1(x),\ \overline{P}(x + k\log x) \le u_1(x)\ \forall\ x > x_0 \\ \text{and} \\ \underline{P}(x) \ge l_2(x),\ \underline{P}(x - k\log|x|) \le u_2(x)\ \forall\ x < -x_0 \end{matrix} \right\}.$$

By Egoroff's theorem, it is indeed possible to meet (ii).

Consider the event

$$G = \left\{ P\colon \sup_{-x_0 < x < x_0} \log\left(\frac{\int_{-k_1}^{k_1} \phi_h(x-\theta)\,dP_0^*(\theta)}{\int_{-k_1}^{k_1} \phi_h(x-\theta)\,dP^*(\theta)}\right) < 2\delta \right\}.$$

We shall argue that $D_\alpha(E \cap F \cap G) > 0$ and if $P \in (E \cap F \cap G)$ then $\int f_0 \log(f_0/f_{h,P}) < \varepsilon$ for a suitable choice of δ.

The events $E \cap F$ and G are independent under D_α, and hence, to prove the first statement, it is enough to show that $D_\alpha(G) > 0$. By intersecting G with E and using the fact that $\{\phi_h(x-\theta)\colon -x_0 \le x \le x_0\}$ is uniformly equicontinuous when $\theta \in [-k_1, k_1]$, we can conclude that $D_\alpha(G) \ge D_\alpha(G \cap E) > 0$ (see the proof of Theorem 3).

Now,

$$\begin{aligned}\int f_0 \log(f_0/f_{h,P}) &\le \int_{-\infty}^{\infty} f_0(x)\log(f_0(x)/f_{0,h}(x))\,dx\\ &\quad+\int_{|x|\le x_0} f_0(x)\log\left(\frac{f_{0,h}(x)}{\int_{-k_1}^{k_1}\phi_h(x-\theta)f_0(\theta)\,d\theta}\right)dx\\ &\quad+\int_{|x|\le x_0} f_0(x)\log\left(\frac{\int_{-k_1}^{k_1}\phi_h(x-\theta)f_0(\theta)\,d\theta}{\int_{-k_1}^{k_1}\phi_h(x-\theta)\,dP(\theta)}\right)dx\\ &\quad+\int_{|x|>x_0} f_0(x)\log\left(\frac{f_{0,h}(x)}{\int\phi_h(x-\theta)\,dP(\theta)}\right)dx.\end{aligned}$$

If $P \in E\cap F\cap G$, then for $x > x_0$,

$$\begin{aligned}\int_{-\infty}^{\infty}\phi_h(x-\theta)\,dP(\theta) &\ge \int_x^{x+k\log x}\phi_h(x-\theta)\,dP(\theta)\\ &\ge \phi_h(k\log x)[\overline{P}(x)-\overline{P}(x+k\log x)]\end{aligned}$$

and since $P\in F$, the expression above is greater than or equal to

$$\phi_h(k\log x)[l_1(x)-u_1(x)] = L_h(x).$$

Using a similar argument for $x < -x_0$, we get

$$\int_{|x|>x_0} f_0(x)\log\left(\frac{f_{0,h}(x)}{f_{h,P}(x)}\right)dx \le \int_{|x|>x_0} f_0(x)\log\left(\frac{f_{0,h}(x)}{L_h(x)}\right)dx < \delta.$$

Since $P\in E\cap G$, for each x in $[-x_0, x_0]$,

$$\log\left(\frac{\int_{-k_1}^{k_1}\phi_h(x-\theta)f_0(\theta)\,d\theta}{\int_{-k_1}^{k_1}\phi_h(x-\theta)\,dP(\theta)}\right) = \log\left(\frac{p_0}{p}\frac{\int_{-k_1}^{k_1}\phi_h(x-\theta)\,dP_0^*(\theta)}{\int_{-k_1}^{k_1}\phi_h(x-\theta)\,dP^*(\theta)}\right) < 3\delta.$$

All these imply that if δ is sufficiently small, then $P\in E\cap F\cap G$ implies that $\int f_0\log(f_{0,h}/f_{h,P}) < \varepsilon$. □

A few remarks about the case when h is fixed a priori to be, say h_0, are in order. In this case, the induced prior is supported by $\mathscr{F}_{h_0} = \{f_{h_0,P}\colon P\in\mathscr{M}\}$, and the following facts are easy to establish from Scheffe's theorem.

1. The map $P\mapsto f_{h_0,P}$ is one-to-one, onto $\mathscr{F}_{h_0}$. Further $P_n\to P_0$ weakly if and only if $\|f_{h_0,P_n}-f_{h_0,P}\|\to 0$.
2. $\mathscr{F}_{h_0}$ is a closed subset of $\mathscr{F}$.

Fact (2) shows that $\mathscr{F}_{h_0}$ is the support of Π and hence consistency is to be sought only for densities of the form $f_{h_0,P}$. Theorem 3 implies consistency for such densities. Fact (1) shows that if the interest is in the posterior distribution of P, then weak consistency at P_0 is equivalent to strong consistency of the posterior of the density at $f_{h_0, P}$.

4. Dirichlet mixtures: strong consistency. As before, we consider the prior which picks a random density $\phi_h * P$, where h is distributed according to μ and P is chosen independently of h according to D_α. Since we view h as corresponding to window length, it is only the small values of h that are relevant, and hence we assume that the support of μ is $[0, M]$ for some finite M.

In this model the prior is concentrated on

$$\mathscr{F} = \bigcup_{0<h<M} \mathscr{F}_h,$$

where $\mathscr{F}_h = \{\phi_h * P\colon P \in \mathrm{M}\}$.

In order to apply Theorem 2, given $U = \{f\colon \|f - f_0\| < \varepsilon\}$, for some $\delta < \varepsilon/4$, we need to construct sieves $\{\mathscr{F}_n\colon n \geq 1\}$ such that $J(\delta, \mathscr{F}_n) \leq n\beta$ and $\mathscr{F}_n^c$ has exponentially small prior probability. Since, as $a_n \to \infty$, $D_\alpha\{P\colon P[-a_n, a_n] > 1 - \delta\} \to 1$, a natural candidate for $\mathscr{F}_n$ is

$$\mathscr{F}_n = \bigcup_{h_n<h<M} \mathscr{F}_h^{a_n}$$

where $h_n \downarrow 0$, a_n increases, and $\mathscr{F}_h^{a_n} = \{\phi_h * P\colon P[-a_n, a_n] > 1 - \delta\}$. What is then needed is an estimate of $J(\delta, \mathscr{F}_n)$. The next theorem, whose proof is deferred to the Appendix, provides such an estimate.

THEOREM 6. *Let* $\mathscr{F}_{h,a,\delta}^M = \bigcup_{h<h'<M}\{f_{h,P}\colon P[-a, a] \geq 1 - \delta\}$. *Then*

$$J(\delta, \mathscr{F}_{h,a,\delta}^M) \leq K\frac{a}{h},$$

where K *is a constant that depends on* δ *and* M, *but not on* a *or* h.

The next theorem formulates the above discussion in terms of strong consistency for Dirichlet-normal mixtures.

THEOREM 7. *Suppose that the prior* μ *has support in* $[0, M]$. *If for each* $\delta > 0$, $\beta > 0$, *there exists sequences* a_n, $h_n \downarrow 0$ *and constants* β_0, β_1 (*all depending on* δ, β *and* M) *such that*:

(i) *for some* β_0, $D_\alpha\{P\colon P[-a_n, a_n] < 1 - \delta\} < \exp(-n\beta_0)$;
(ii) $\mu\{h < h_n\} \leq \exp(-n\beta_1)$;
(iii) $a_n/h_n < n\beta$,

then f_0 *is in the K–L support of the prior implies that the posterior is strongly consistent at* f_0.

REMARK 4. What is involved above is a balance between a_n and h_n. Since δ and M are fixed, the constant K obtained in Theorem 6 does not play any role. If α has compact support, say $[-a, a]$, then we may trivially choose $a_n = a$ and so h_n may be allowed to take values of the order of n^{-1} or larger. If α is chosen as a normal distribution and h^2 is given a (right truncated) inverse gamma prior, then the conditions of the theorem are satisfied if a_n is of the order $\sqrt{n}$ and $h_n = C/\sqrt{n}$ for a suitable (large) C (depending on δ and β).

5. Extensions. The methods developed in this paper towards the simple mixture models can be used to study many of the variations used in practice. Some of these are discussed in this section.

1. It is often sensible to let the prior depend on the sample size; see, for instance, Roeder and Wasserman (1995). A case in point in our context would be when the precision parameter $M = \alpha(\mathbb{R})$ is allowed to depend on the sample size.

   If Π_n is the prior at stage n, then Theorem 2 goes through if the assumption $\Pi(K_\varepsilon(f_0)) > 0$ is replaced by $\liminf_{n\to\infty} \Pi_n(K_\varepsilon(f_0)) > 0$. This follows from the fact that Barron's theorem (see Appendix) goes through with a similar change. The only stage that needs some care is an argument which involves Fubini, but it can be handled easily.
2. Another way the Dirichlet mixtures can be extended is by including a further mixing. Formally, let $X_1, X_2, \dots$ be observations from a density f where $f = \phi_h * P$, $P \sim D_{\alpha_\tau}$, $h \sim \pi$, τ is a finite-dimensional mixing parameter which is also endowed with some prior ρ. Let f_0 be the true density. We are interested in verifying the Schwartz condition at f_0 and conditions for strong consistency.

   By Fubini's theorem, Schwartz's condition is satisfied for the mixture if

$$\rho\{\tau\colon \text{the Schwartz condition is satisfied with } \alpha_\tau\} > 0. \tag{6}$$

   (a) In particular, if f_0 has compact support, then (6) reduces to

$$\rho\{\tau\colon \operatorname{supp}(f_0) \subset \operatorname{supp}(\alpha_\tau)\} > 0. \tag{7}$$

   (b) Suppose f_0 is not of compact support and $\tau = (\mu, \sigma)$ gives a location-scale mixture. So we have to seek for the condition so that the Schwartz condition holds with the base measure $\alpha((\cdot - \mu)/\sigma)$. We report results only for $\alpha_0 = \alpha/\alpha(\mathbb{R})$ double exponential or normal.

   When α_0 is double exponential, a sufficient condition is that $f_0(\mu + \sigma x)$ has finite moment generating function on an open interval containing $[-1, 1]$. When α is normal, we need the integrability of $x \log|x| \exp[x^2/2]$ with respect to the density $f_0(\mu + \sigma x)$. For example, if the true density is $N(\mu_0, \sigma_0)$, then the required condition will be $\sigma < \sigma_0$, so we need

$$\rho\{(\mu, \sigma)\colon \sigma < \sigma_0\} > 0.$$

   We omit the proof of these statements.

   (c) For strong consistency, we further assume that the support of the prior ρ [for (μ, σ)] is compact. For each (μ, τ), find the corresponding $a_n(\mu, \tau)$ of Theorem 7, that is, satisfying

$$D_{\alpha(\mu,\tau)}\{P\colon P[-a_n(\mu, \tau), a_n(\mu, \tau)] < 1 - \delta\} < \exp(-n\beta_0)$$

for some $\beta_0 > 0$. Now choose $a_n = \sup_{\mu, \sigma} a_n(\mu, \sigma)$. The order of a_n will then be the same as that of the individual $a_n(\mu, \sigma)$'s.

(d) In some special cases, it is also possible to allow unbounded location mixtures. For example, when the base measure is normal, a normal prior for the location parameter is both natural and convenient. Strong consistency continues to hold in this case as long as σ has a compactly supported prior. To see this, observe that $\rho\{|\mu| > \sqrt{n}\}$ is exponentially small and $\sup_{|\mu| \le \sqrt{n}, \sigma} a_n(\mu, \sigma)$ is again of the order of $\sqrt{n}$.

APPENDIX

Schwartz (1965) showed that for a set U, $\Pi(U|X_1, X_2, \dots, X_n) \to 1$ a.s. P_{f_0} if:

1. $\Pi(K_\varepsilon(f_0)) > 0$ for all $\varepsilon > 0$;
2. there exists a uniformly consistent sequence of tests for testing H_0: $f = f_0$ versus H_1: $f \in U^c$, that is, there exist tests $\phi_n(X_1, X_2, \dots, X_n)$ such that as $n \to \infty$,

$$E_{f_0}\phi_n(X_1, X_2, \dots, X_n) \to 0 \quad \text{and} \quad \inf_{f \in U^c} E_f \phi_n(X_1, X_2, \dots, X_n) \to 1.$$

When U is a weak neighborhood of f_0, it is not hard to see that condition (ii) of Theorem 1 holds. This immediately leads to the statement in Section 2.

Barron (1989) and LeCam (1973) show that, in general, when U is a strong neighborhood of f_0, there does not exist a uniformly consistent sequence of tests for testing H_0: $f = f_0$ versus H_1: $f \in U^c$. This fact renders that the Schwartz theorem inapplicable in establishing strong consistency. Our approach to strong consistency is based on the following result of Barron (1988), which is also discussed in Barron, Schervish and Wasserman (1998).

THEOREM 8 (Barron). *Let Π be a prior on $\mathscr{F}$, $f_0 \in \mathscr{F}$ and U be a neighborhood of f_0. Assume that $\Pi(K_\varepsilon(f_0)) > 0$ for all $\varepsilon > 0$. Then the following are equivalent*:

(i) *There exists a β_0 such that*

$$P_{f_0}\big\{\Pi(U^c|X_1, X_2, \dots, X_n) > \exp(-n\beta_0) \textit{ infinitely often}\big\} = 0.$$

(ii) *There exist subsets V_n, W_n of $\mathscr{F}$, positive numbers $c_1, c_2, \beta_1, \beta_2$ and a sequence of tests $\{\phi_n(X_1, X_2, \dots, X_n)\}$ such that*:

(a) $U^c \subset V_n \cup W_n$,

(b) $\Pi(W_n) \le c_1 \exp(-n\beta_1)$,

(c) $P_{f_0}\{\phi_n(X_1, X_2, \dots, X_n) > 0$ *infinitely often*$\} = 0$ *and*

$$\inf_{f \in V_n} E_f \phi_n \ge 1 - c_2 \exp(-n\beta_2).$$

A proof may be found in Barron (1998).

PROOF OF THEOREM 2. Let $U = \{f\colon \|f - f_0\| < \varepsilon\}$, $V_n = \mathscr{F}_n \cap U^c$ and $W_n = \mathscr{F}_n^c$. We will argue that the pair (V_n, W_n) satisfies (ii) of Theorem 2. Clearly $U^c \subset V_n \cup W_n$ and $\Pi(W_n) < c_1 \exp(-nc_2)$.

Let $g_1, g_2, \dots, g_k$ in $\mathscr{F}$ be such that $V_n \subset \bigcup_{i=1}^k G_i$ where $G_i = \{f\colon \|f - g_i\| < \delta\}$. Let $f_i \in V_n \cap G_i$. Then for each $i = 1, 2, \dots, k$, $\|f_0 - f_i\| > \varepsilon$ and if $f \in G_i$, then $\|f_i - f\| < 2\delta$. Consequently, for each $i = 1, 2, \dots, k$, if $A_i = \{x\colon f_0(x) < f_i(x)\}$ then

$$P_{f_0}(A_i) = \alpha_i \quad \text{and} \quad P_{f_i}(A_i) = \gamma_i > \alpha_i + \varepsilon/2.$$

Hence if $f \in G_i$, then $P_f(A_i) > \gamma_i - \delta > \alpha_i + \varepsilon/2 - \delta$.

Let

$$B_i = \left\{(x_1, x_2, \dots, x_n)\colon \frac{1}{n}\sum_{j=1}^n I_{A_i}(x_j) \geq \frac{(\gamma_i + \alpha_i)}{2}\right\}.$$

A straightforward application of Hoeffding's inequality shows that

$$P_{f_0}(B_i) \leq \exp[-n\varepsilon^2/8].$$

On the other hand, if $f \in G_i$,

$$\text{(8)} \qquad \begin{aligned} P_f(B_i) &\geq P_f\left\{\frac{1}{n}\sum_{j=1}^n I_{A_i}(x_j) - P_f(A_i) \geq \frac{(\alpha_i - \gamma_i)}{2} + \delta\right\} \\ &\geq P_f\left\{n^{-1}\sum_{j=1}^n I_{A_i}(x_j) - P_f(A_i) \geq -\frac{\varepsilon}{4} + \delta\right\}. \end{aligned}$$

Applying Hoeffding's inequality to the negative of the indicator variables, the above probability is greater than or equal to

$$1 - \exp[-2n(\varepsilon/4 - \delta)^2].$$

If we set

$$\phi_n(X_1, X_2, \dots, X_n) = \max_{1 \leq i \leq k} I_{B_i}(X_1, X_2, \dots, X_n),$$

then

$$E_{f_0}\phi_n \leq k \exp[-n\varepsilon^2/8]$$

and

$$\inf_{f \in V_n} E_f \phi_n \geq 1 - \exp[-2n(\varepsilon/4 - \delta)^2].$$

By choosing $\log k = J(\delta, \mathscr{F}_n) < n\beta$, we have $E_{f_0}\phi_n \leq \exp[-n(\varepsilon^2/8 - \beta)]$. Since $\beta < \varepsilon^2/8$, all that is left to show is

$$P_{f_0}\{\phi_n > 0 \text{ infinitely often}\} = 0.$$

This follows easily from an application of Borel–Cantelli and from the fact that ϕ_n takes only values 0 or 1. □

PROOF OF THEOREM 6. We prove Theorem 6 through a sequence of lemmas. Let $\mathscr{F}_{h,a} = \{f_{h,P}\colon P(-a, a] = 1\}$. Without loss of generality, we shall assume that $a \geq 1$.

LEMMA 1. $J(2\delta, \mathscr{F}_{h,a}) \leq (\sqrt{\frac{8}{\pi}}\frac{a}{h\delta} + 1)(1 + \log(\frac{1+\delta}{\delta}))$.

PROOF. For any $\theta_1 < \theta_2$,

$$\begin{aligned}
\|\phi_{\theta_1,h} - \phi_{\theta_2,h}\| &= \frac{1}{\sqrt{2\pi}h}\int_{x>(\theta_1+\theta_2)/2} \exp[-(x-\theta_2)^2/(2h^2)]\,dx \\
&\quad - \frac{1}{\sqrt{2\pi}h}\int_{x>(\theta_1+\theta_2)/2} \exp[-(x-\theta_1)^2)/(2h^2)]\,dx \\
&\quad + \frac{1}{\sqrt{2\pi}h}\int_{x<(\theta_1+\theta_2)/2} \exp[-(x-\theta_1)^2/(2h^2)]\,dx \\
&\quad - \frac{1}{\sqrt{2\pi}h}\int_{x<(\theta_1+\theta_2)/2} \exp[-(x-\theta_2)^2/(2h^2)]\,dx \\
&= 4\frac{1}{\sqrt{2\pi}}\int_0^{(\theta_2-\theta_1)/(2h)} \exp[-x^2/2]\,dx \\
&\leq \sqrt{\frac{2}{\pi}}\frac{(\theta_2-\theta_1)}{h}.
\end{aligned}$$

Given δ, let N be the smallest integer greater than $\sqrt{8}a/(\sqrt{\pi}h\delta)$. Divide $(-a, a]$ into N intervals. Let

$$E_i = \left(-a + \frac{2a(i-1)}{N}, -a + \frac{2ai}{N}\right]\colon i = 1, 2, \ldots, N,$$

and let θ_i be the midpoint of E_i. Note that if $\theta, \theta' \in E_i$, then $|\theta - \theta'| < 2a/N$, and consequently $\|\phi_{\theta,h} - \phi_{\theta',h}\| < \delta$.

Let $\mathscr{P}_N = \{(P_1, P_2, \ldots, P_N)\colon P_i \geq 0, \sum_{i=1}^N P_i = 1\}$ be the N-dimensional probability simplex and let $\mathscr{P}_N^*$ be a δ-net in $\mathscr{P}_N$, that is, given $P \in \mathscr{P}_N$, there is $P^* = (P_1^*, P_2^*, \ldots, P_N^*) \in \mathscr{P}_N^*$ such that $\sum_{i=1}^N |P_i - P_i^*| < \delta$.

Let $\mathscr{F}^* = \{\sum_{i=1}^N P_i^*\phi_{\theta_i,h}\colon P^* \in \mathscr{P}_N^*\}$. We shall show that $\mathscr{F}^*$ is a 2δ net in $\mathscr{F}_{h,a}$. If $f_{h,P} = \phi_h * P \in \mathscr{F}_{h,a}$, set $P_i = P(E_i)$ and let $P^* \in \mathscr{P}_N^*$ be such that $\sum_{i=1}^N |P_i - P_i^*| < \delta$. Then

$$\begin{aligned}
&\left\|\int \phi_{\theta,h}\,dP(\theta) - \sum_{i=1}^N P_i^*\phi_{\theta_i,h}\right\| \\
&\qquad \leq \left\|\int \phi_{\theta,h}\,dP(\theta) - \sum_{i=1}^N \int I_{E_i}(\theta)\phi_{\theta_i,h}\,dP(\theta)\right\| + \left\|\sum_{i=1}^N P_i\phi_{\theta_i,h} - \sum_{i=1}^N P_i^*\phi_{\theta_i,h}\right\|
\end{aligned}$$

$$\le \int \sum_{i=1}^{N} I_{E_i}(\theta)\|\phi_{\theta,h} - \phi_{\theta_i,h}\|\, dP(\theta) + \sum_{i=1}^{N} |P_i - P_i^*|$$
$$\le 2\delta.$$

This shows that $J(2\delta, \mathscr{F}_{h,a}) \le J(\delta, \mathscr{P}_N)$, and we calculate $J(\delta, \mathscr{P}_N)$ along the lines of Barron, Schervish and Wasserman (1998) as follows.

Since $|P_i - P_i^*| < \delta/N$ for all i implies that $\sum_{i=1}^{N} |P_i - P_i^*| < \delta$, an upper bound for the cardinality of the minimal δ-net of $\mathscr{P}_N$ is given by

$$\begin{aligned} &\#\text{ cubes of length } \delta/N \text{ covering } [0,1]^N \\ &\quad \times \text{volume of } \left\{(P_1, P_2, \dots, P_N)\colon P_i \ge 0,\ \sum_{i=1}^{N} P_i \le 1+\delta\right\} \\ &\quad = (N/\delta)^N (1+\delta)^N \frac{1}{N!}. \end{aligned}$$

So,

$$\begin{aligned} J(\delta, \mathscr{P}_N) &\le N\log N - N\log\delta + N\log(1+\delta) - \log N! \\ &\le N\log N - N\log\delta + N\log(1+\delta) - N\log N + N \\ &= N\left(1 + \log\tfrac{1+\delta}{\delta}\right) \\ &\le \left(\sqrt{\frac{8}{\pi}}\frac{a}{h\delta} + 1\right)\left(1 + \log\frac{1+\delta}{\delta}\right). \end{aligned}$$ □

LEMMA 2. *Let* $\mathscr{F}_{h,a,\delta} = \{f_{h,P}\colon P(-a,a] \ge 1-\delta\}$. *Then* $J(3\delta, \mathscr{F}_{h,a,\delta}) \le J(\delta, \mathscr{F}_{h,a})$.

PROOF. Let $f = \phi_h * P \in \mathscr{F}_{h,a,\delta}$. Consider the probability measure P^* defined by $P^*(A) = P(A \cap (-a,a])/P(-a,a]$. Then the density $f^* = \phi_h * P^*$ clearly belongs to $\mathscr{F}_{h,a}$ and further satisfies $\|f - f^*\| < 2\delta$. □

LEMMA 3. *Let* $M > 0$ *and let* $\mathscr{F}^M_{h,a,\delta} = \bigcup_{h<h'<M} \mathscr{F}_{h',a,\delta}$. *If* $a > M/\sqrt{\delta}$, *then* $\mathscr{F}^M_{h,a,\delta} \subset \mathscr{F}_{h,2a,2\delta}$.

PROOF. By Chebyshev's inequality, if $h' < M$ then the probability of $(-a,a]$ under $N(0,h')$ is greater than $1-\delta$. If $f = \phi_{h'} * P$, then since $\phi_{h'} = \phi_h * \phi_{h^*}$, where $h^* < M$, $f = \phi_h * \phi_{h^*} * P$ and $(\phi_{h^*} * P)(-a,a] > 1 - 2\delta$. □

Putting Lemma 1, Lemma 2 and Lemma 3 together, we have Theorem 6. □

Acknowledgments. We thank A. Barron, M. Schervish and L. Wasserman for providing us with a preprint of their paper. We are grateful to L. Wasserman for his careful reading and comments on some of the calculations carried out in this paper.

REFERENCES

BARRON, A. R. (1988). The exponential convergence of posterior probabilities with implications for Bayes estimators of density functions. Unpublished manuscript.

BARRON, A. R. (1989). Uniformly powerful goodness of fit tests. *Ann. Statist.* **17** 107–124.

BARRON, A. R. (1998). Information theoretic analysis of Bayes performance and the choice of priors in parametric and nonparametric problems. In *Bayesian Statistics 6* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.) Oxford Univ. Press. To appear.

BARRON, A. R., SCHERVISH, M. and WASSERMAN, L. (1998). The consistency of posterior distributions in nonparametric problems. *Ann. Statist.* To appear.

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

FERGUSON, T. S. (1983). Bayesian density estimation by mixtures of Normal distributions. In *Recent Advances in Statistics* (M. Rizvi, J. Rustagi, and D. Siegmund, eds.) 287–302. Academic Press, New York.

GASPARINI, M. (1992). Bayes nonparametrics for biased sampling and density estimation. Ph.D. dissertation, Univ. Michigan.

GHORAI, J. K. and RUBIN, H. (1982). Bayes risk consistency of nonparametric Bayes density estimates. *Austral. J. Statist.* **24** 51–66.

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

LECAM, L. (1973). Convergence of estimates under dimensionality restrictions. *Ann. Statist.* **1** 38–53.

ROEDER K. and WASSERMAN, L. (1995). Practical Bayesian density estimation using mixtures of normal. *J. Amer. Statist. Assoc.* **92** 894–902.

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

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York.

WEST, M. (1992). Modeling with mixtures. In *Bayesian Statistics 4* (J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, eds.) 503–524. Oxford Univ. Press.

WEST, M., MULLER, P. and ESCOBAR, M. D. (1994). Hierarchical priors and mixture models, with applications in regression and density estimation. In *Aspects of Uncertainty: A Tribute to D. V. Lindley* (P. R. Freeman and A. F. M. Smith, eds.) 363–386. Wiley, Chichester.

S. GHOSAL
FACULTY OF MATHEMATICS
AND COMPUTER SCIENCE
VRIJE UNIVERSITEIT
DE BOELELAAN 1081A
1081 HV AMSTERDAM
THE NETHERLANDS

J. K. GHOSH
DIVISION OF THEORETICAL
STATISTICS AND MATHEMATICS
INDIAN STATISTICAL INSTITUTE
203 BARRACKPORE TRUNK ROAD
CALCUTTA 700035
INDIA

R. V. RAMAMOORTHI
DEPARTMENT OF STATISTICS
AND PROBABILITY
MICHIGAN STATE UNIVERSITY
WELLS HALL
EAST LANSING, MICHIGAN 48824
E-MAIL: ramamoorthi@stt.msu.edu