On Consistency of Nonparametric Normal Mixtures for Bayesian Density Estimation

Antonio LIJOI, Igor PRÜNSTER, and Stephen G. WALKER

The past decade has seen a remarkable development in the area of Bayesian nonparametric inference from both theoretical and applied perspectives. As for the latter, the celebrated Dirichlet process has been successfully exploited within Bayesian mixture models, leading to many interesting applications. As for the former, some new discrete nonparametric priors have been recently proposed in the literature that have natural use as alternatives to the Dirichlet process in a Bayesian hierarchical model for density estimation. When using such models for concrete applications, an investigation of their statistical properties is mandatory. Of these properties, a prominent role is to be assigned to consistency. Indeed, strong consistency of Bayesian nonparametric procedures for density estimation has been the focus of a considerable amount of research; in particular, much attention has been devoted to the normal mixture of Dirichlet process. In this article we improve on previous contributions by establishing strong consistency of the mixture of Dirichlet process under fairly general conditions. Besides the usual Kullback–Leibler support condition, consistency is achieved by finiteness of the mean of the base measure of the Dirichlet process and an exponential decay of the prior on the standard deviation. We show that the same conditions are also sufficient for mixtures based on priors more general than the Dirichlet process. This leads to the easy establishment of consistency for many recently proposed mixture models.

KEY WORDS: Bayesian nonparametrics; Density estimation; Mixture of Dirichlet process; Neutral to the right process; Normalized random measure; Normal mixture model; Random discrete distribution; Species sampling model; Strong consistency.

1. INTRODUCTION

Consistency of Bayesian nonparametric procedures has been the focus of a considerable amount of research in recent years. Most contributions in the literature exploit the "frequentist" approach to Bayesian consistency, also termed the "what if" method according to Diaconis and Freedman (1986). This essentially consists of verifying what would happen to the posterior distribution if the data were generated from a "true" fixed density function f_0: Does the posterior accumulate in suitably defined neighborhoods of f_0?

Early works on consistency were concerned with weak consistency. (The reader is referred to, e.g., Freedman 1963 and Diaconis and Freedman 1986 for some interesting examples of possible inconsistency.) A sufficient condition for weak consistency, which is solely a support condition, was provided by Schwartz (1965).

When considering problems of density estimation, it is natural to ask for the strong consistency of posterior distributions. An early contribution in this area was made by Barron (1988). Later developments combined techniques from the theory of empirical processes with results on uniformly consistent tests achieved by Barron (1988) and provided sufficient conditions for strong consistency relying on the construction of suitable sieves. General results have been derived by Barron, Schervish, and Wasserman (1999) and Ghosal, Ghosh, and Ramamoorthi (1999), whereas significant priors were studied by Petrone and Wasserman (2002) and Choudhuri, Ghosal, and Roy (2004), among others. The "sieve-approach" was treated in great detail in the monograph by Ghosh and Ramamoorthi (2003) (see also Wasserman 1998 for a more concise account). A review of Bayesian asymptotics from a different perspective was provided by Walker (2004a). Recently, a new approach to the study of strong consistency for Bayesian density estimation was introduced by Walker (2004b), who obtained a simple sufficient condition for strong consistency not relying on sieves.

In the framework of Bayesian density estimation, one is naturally led to think of the mixture of Dirichlet process (MDP), a cornerstone in the area. This model was introduced by Lo (1984) and later popularized by Escobar (1988) and Escobar and West (1995), who developed suitable simulation techniques (see also MacEachern 1994; MacEachern and Müller 1998). The MDP was extensively reviewed in the book edited by Dey, Müller, and Sinha (1998) and by Müller and Quintana (2004), who emphasized applications and simulation algorithms. As far as consistency is concerned, the normal MDP model was analyzed by Ghosal et al. (1999) by exploiting the sieve approach.

In this article, we face the issue of consistency of the MDP by exploiting the approach set out by Walker (2004b). This leads to quite a dramatic improvement on previous results. We essentially show that an MDP model is consistent if the base measure of the Dirichlet process has finite mean and the prior on the standard deviation has an exponentially decaying tail in a neighborhood of 0. Our results carry over to normal mixture models, where the Dirichlet process is replaced by a general discrete nonparametric prior, thus establishing consistency of many models recently proposed in the literature. In particular, one can easily establish consistency of normal mixtures based on (a) species sampling models and relevant subclasses, namely homogeneous normalized random measures with independent increments, stick-breaking priors and neutral to the right species sampling models; (b) nonhomogeneous normalized random measures with independent increments; and (c) neutral to the right processes.

Antonio Lijoi is Assistant Professor, Department of Economics and Quantitative Methods, University of Pavia, Pavia, Italy (E-mail: *lijoi@unipv.it*) and Research Associate, Institute of Applied Mathematics and Computer Science (IMATI), National Research Council (CNR), Milan, Italy. Igor Prünster is Assistant Professor, Department of Economics and Quantitative Methods, University of Pavia, Pavia, Italy (E-mail: *igor.pruenster@unipv.it*) and Research Fellow, International Center of Economic Research (ICER), Turin, Italy. Stephen G. Walker is Professor, Institute of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury, U.K. (E-mail: *S.G.Walker@kent.ac.uk*). Lijoi and Prünster were partially supported by the Italian Ministry of University and Research (MIUR), research project "Bayesian Nonparametric Methods and Their Applications." The work of Walker is financed by an EPSRC Advanced Research Fellowship. The authors are grateful to a referee for valuable comments.

Journal of the American Statistical Association
December 2005, Vol. 100, No. 472, Theory and Methods
DOI 10.1198/016214505000000358

The article is structured as follows. In Section 2, after giving a concise description of the normal mixture model, we state the main result. Then we provide illustrations describing how the result applies to a variety of nonparametric priors. Finally, in Section 3 we provide a detailed proof.

2. THE CONSISTENCY RESULT

2.1 The Bayesian Normal Mixture Model

Nowadays the most common use of Bayesian nonparametric procedures is represented by density estimation via a mixture model based on a random discrete distribution. In particular, attention has been focused on normal mixtures, that is,

$$\tilde{f}_{\sigma,\tilde{P}}(x) = \phi_\sigma * \tilde{P} = \int \phi_\sigma(x-\theta)\tilde{P}(\mathrm{d}\theta), \tag{1}$$

where for each positive σ, ϕ_σ is the density function of the normal distribution with mean 0 and variance σ^2. $\tilde{P}$ is a random probability distribution on $\mathbb{R}$ whose law, Λ, selects discrete distributions almost surely. Moreover, σ has a prior distribution, which we denote by μ. The model (1) can be equivalently expressed in hierarchical form as

$$(X_i|\theta_i,\sigma) \overset{\text{ind}}{\sim} \mathrm{N}(X_i;\theta_i,\sigma^2), \qquad i=1,\dots,n;$$
$$(\theta_i|\tilde{P}) \overset{\text{iid}}{\sim} \tilde{P}, \qquad i=1,\dots,n;$$
$$\tilde{P} \sim \Lambda;$$

and

$$\sigma \sim \mu,$$

where μ and Λ are independent and $\mathrm{N}(\cdot;\theta,\sigma^2)$ denotes for the normal distribution with mean θ and variance σ^2. Clearly, the MDP model is obtained when $\tilde{P}$ in (1) coincides with the Dirichlet process with parameter-measure α, a finite nonnull measure (see Ferguson 1973).

An important element in prior specification that we consider later is the prior guess at the shape of $\tilde{P}$, that is,

$$P_0(C) = E[\tilde{P}(C)], \tag{2}$$

for any C belonging to the Borel σ-field of $\mathbb{R}$, denoted by $\mathscr{B}(\mathbb{R})$.

2.2 A Sufficient Condition for Strong Consistency

The relevance, both theoretical and applied, of normal mixture models motivates a study of their asymptotic properties. Among these properties, consistency plays a prominent role. Because the aim is density estimation, the appropriate notion to deal with is strong consistency. Consider a sequence of observations $(X_n)_{n\ge1}$ each taking values in $\mathbb{R}$, and let $\mathscr{F}$ be the space of probability density functions with respect to the Lebesgue measure on $\mathbb{R}$. Let Π be the prior distribution of the random density function $\tilde{f}_{\sigma,\tilde{P}}$ in (1). Then the posterior distribution, given the observations $(X_1,\dots,X_n)$, coincides with

$$\Pi_n(B) = \frac{\int_B \prod_{i=1}^n f(X_i)\Pi(\mathrm{d}f)}{\int_{\mathscr{F}} \prod_{i=1}^n f(X_i)\Pi(\mathrm{d}f)}$$

for all measurable subsets B of $\mathscr{F}$. Let us assume that there exists a "true" density function f_0 such that the observations X_n's are iid from f_0, and let F_0 denote the probability distribution corresponding to f_0. Hence Π is said to be *strongly consistent* at f_0 in $\mathscr{F}$ if, for any $\varepsilon > 0$,

$$\Pi_n(A_\varepsilon) \to 1 \quad \text{a.s. } [F_0^\infty]$$

as $n \to +\infty$, where A_ε is an L_1-neighborhood of f_0 with radius ε and F_0^∞ denotes the infinite product measure on $\mathbb{R}^\infty$.

Hereinafter we assume that the density f_0 is in the Kullback–Leibler support of the prior Π. This means that Π assigns positive masses to any Kullback–Leibler neighborhood of f_0. It is known that such an assumption is sufficient to ensure *weak consistency* of Π at f_0 (see Schwartz 1965). Conditions for f_0 to be in the Kullback–Leibler support of the normal mixture model prior Π defined in (1) have been given by Ghosal et al. (1999). However, because we aim to establish the stronger property of L_1 consistency, the Kullback–Leibler support condition is not enough.

Note that Π is determined both by Λ and by the prior distribution for σ, which we have denoted by μ. As for the latter, from the standpoint of consistency the most important values of σ are those included in a right-neighborhood of 0. Thus, with no loss of generality, we can choose μ such that its support coincides with $(0, M]$ for some positive and finite M.

The main result on strong consistency of normal mixture models can now be stated. In the sequel $g(x) \sim h(x)$, as x tends to $+\infty$, means that $g(x)/h(x)$ tends to 1 as x tends to $+\infty$. Recall the definition of P_0 as the prior guess at $\tilde{P}$ given in (2). The proof to the following result can be found in Section 3.

Theorem 1. Let f_0 be a density in the Kullback–Leibler support of Π. Suppose that the following conditions hold:

(a) $\int_{\mathbb{R}} |\theta| P_0(\mathrm{d}\theta) < +\infty$

(b) $\mu\{\sigma < \sigma_k\} \le \exp\{-\gamma k\}$ for some sufficiently large γ, where $(\sigma_k)_{k\ge1}$ is any sequence such that $\sigma_k \sim k^{-1}$ as $k \to \infty$.

Then Π is consistent at f_0.

By the foregoing result, strong consistency follows from a simple condition on the prior guess P_0 combined with a condition on the probabilities assigned by μ on shrinking neighborhoods of the origin. Note that the value for which γ can be considered sufficiently large is determined in the proof; see (8).

Theorem 1 can be compared with the results obtained in Ghosal et al. (1999) for the MDP. Their theorem has three conditions (i)–(iii). Indeed, our condition (a) improves on their condition (i), which essentially requires that α have exponential tails. Moreover, our condition (b) and their condition (ii) coincide. Finally, we have no need for their condition (iii).

Some comments on our condition (a) are in order. Notice, for instance, that it is satisfied by even heavy-tailed distributions, that is, by those P_0's for which $P_0([-\theta,\theta]^c) \sim \theta^{-\gamma}$, for some $\gamma > 1$, as $\theta \to +\infty$. Weakening the tail condition for P_0 from an exponential to a power law decay seems to be a quite remarkable achievement.

2.3 Illustrations

In this section we show how condition (a) translates for a variety of normal mixture models, thus giving a simple criterion for establishing their strong consistency. It is worth stressing that strong consistency for the more general mixtures that we are going to consider has not yet been considered in the literature. Note, moreover, that Theorem 1 also applies to mixture models directed by random probability measures $\tilde{P}$, whose support contains continuous distributions. But such cases seem to be not of particular interest, because, commonly in applications one wishes to exploit the clustering behavior arising from a discrete random probability measure $\tilde{P}$.

First, recall that the celebrated MDP is recovered by setting $\tilde{P}$ to be the Dirichlet process with parameter-measure α. In this case $P_0 = \alpha/\alpha(\mathbb{R})$, and condition (a) reduces to

$$\int_{\mathbb{R}} |\theta| \alpha(\mathrm{d}\theta) < \infty. \tag{3}$$

Let us now consider more general nonparametric mixture models. Kingman (1975) proposed modeling random probabilities in the context of storage problems by normalizing a subordinator, that is, an increasing and purely discontinuous Lévy process with stationary increments. Recently, Regazzini, Lijoi, and Prünster (2003) extended this class of random probability measures to the so-called "nonhomogeneous" case and developed it from a Bayesian perspective. Such random probabilities include, as a special case, the Dirichlet process and here are referred to, according to the terminology set of Regazzini et al. (2003), as *normalized random measures with independent increments* (NRMI). A wide class of mixtures can then be achieved by setting $\tilde{P}$ to be an NRMI, as was done by Nieto-Barajas, Prünster, and Walker (2004). It can be shown that the prior guess is given by

$$P_0(C) = \int_C \int_0^{+\infty} \mathrm{e}^{-\psi(u)} \times \left\{ \int_0^{+\infty} \mathrm{e}^{-uv} v \rho(\mathrm{d}v|\theta)\, \mathrm{d}v \right\} \mathrm{d}u\, \alpha(\mathrm{d}\theta) \tag{4}$$

for any C in $\mathscr{B}(\mathbb{R})$, where $\nu_\alpha(\mathrm{d}v, \mathrm{d}\theta) = \rho(\mathrm{d}v|\theta)\alpha(\mathrm{d}\theta)$ is the Poisson intensity measure on $(0, +\infty) \times \mathbb{R}$ associated with the increasing additive process ξ that generates $\tilde{P}$. Moreover, ψ denotes the Laplace exponent of ξ, which can be determined via the well-known Lévy–Khintchine representation theorem (for details, see Regazzini et al. 2003; James 2002). When $\rho(\mathrm{d}v|\theta) = \rho(\mathrm{d}v)$, for each $\theta \in \mathbb{R}$, $\tilde{P}$ is said to be homogeneous, the prior guess in (4) reduces to

$$P_0(C) = \frac{\alpha(C)}{\alpha(\mathbb{R})} \quad \text{for any } C \in \mathscr{B}(\mathbb{R}),$$

and condition (a) coincides with (3). Apart from the Dirichlet process, the most notable prior within this class, which leads to explicit forms for quantities of statistical interest, is the so-called *normalized inverse Gaussian process* studied by Lijoi, Mena, and Prünster (2005).

Another interesting class of mixture models, first considered by Ishwaran and James (2001) and further developed by Ishwaran and James (2003a), arises when $\tilde{P}$ is chosen to be a *stick-breaking prior*. Such a prior depends on the specification of a stick-breaking procedure and of a measure α that is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}$. The prior guess P_0 coincides with $\alpha/\alpha(\mathbb{R})$, and again (a) becomes (3). Among these priors, it is worth mentioning the two-parameter Poisson–Dirichlet process (see Pitman 1996).

A further possibility is represented by the choice of the background driving $\tilde{P}$ as a *neutral to the right* (NTR) process, a class of priors introduced by Doksum (1974) (see also Ferguson 1974; Ferguson and Phadia 1979). In such a case the probability distribution function corresponding to P_0 can be expressed as

$$P_0\big((-\infty, \theta]\big) = 1 - \mathrm{E}(\mathrm{e}^{-\xi_\theta}) \quad \text{for any } \theta \in \mathbb{R},$$

where ξ is a suitable increasing additive process characterizing $\tilde{P}$. For instance, if $\tilde{P}$ is the beta–Stacy process, due to Walker and Muliere (1997), that $\tilde{P}$ can be easily centered on any desired choice of P_0, leading to a straightforward verification of condition (a). James (2003) considered mixtures of NTR processes, and extended NTR priors to a spatial setting. By deriving the corresponding exchangeable probability partition function and combining it with an absolutely continuous finite measure α, James obtained a new family of random probability measures termed *spatial NTR species sampling* models. Again the prior guess coincides with $\alpha/\alpha(\mathbb{R})$, and condition (a) reduces to (3).

It must to be remarked that spatial NTR species sampling models, stick-breaking priors, and homogeneous (but not nonhomogeneous) NRMI essentially belong to the class of *species sampling models* due to Pitman (1996), for which

$$P_0(C) = \frac{\alpha(C)}{\alpha(\mathbb{R})} \quad \text{for any } C \in \mathscr{B}(\mathbb{R}).$$

Hence condition (a) is again equivalent to (3). The class of species sampling models is quite rich and allows investigation of the structural properties of its members. However, unless it is possible to effectively assess the weights in the species sampling representation, no explicit expression for quantities of statistical interest is achievable. Indeed, the three subclasses mentioned earlier are to our current knowledge the only species sampling models that are sufficiently tractable to be useful in applications. Species sampling mixture models, with emphasis on the two-parameter Poisson–Dirichlet process, were dealt with by Ishwaran and James (2003b).

These illustrations stress the usefulness of Theorem 1 in checking the consistency of normal mixture models based on a number of alternatives to the Dirichlet process as a mixing distribution.

3. THE PROOF

3.1 Preliminary Result

Recall that A_ε^c is the complement of the L_1-neighborhood of f_0 with radius ε. By separability of $\mathscr{F}$, such a set can be covered by a countable union of disjoint sets B_j, where $B_j \subseteq B_j^* := \{f : \|f - f_j\| < \eta\}$, f_j are densities in A_ε^c, $\|\cdot\|$ is the L_1-norm and η is any number in $(0, \varepsilon)$. An extension of a result of Walker (2004b) can be stated as follows: If for some $\beta \in (0, 1)$ and for some covering $(B_j)_{j \ge 1}$ as before,

$$\sum_{j \ge 1} (\Pi(B_j))^\beta < +\infty, \tag{5}$$

then Π is consistent at f_0, with the proviso that f_0 is in the Kullback–Leibler support of Π. This result is a key ingredient in the following proof.

3.2 The Proof

Let us first set some useful notation. For any $a > 0$ and $\sigma > 0$, let

$$\mathcal{F}^U_{\sigma,a,\delta} = \{\phi_\sigma * P : P([-a,a]) \ge 1-\delta\},$$
$$\mathcal{F}^L_{\sigma,a,\delta} = \{\phi_\sigma * P : P([-a,a]) < 1-\delta\},$$

and $\overline{\mathcal{F}}^M_{\sigma,a,\delta} = \bigcup_{\sigma<\sigma'<M} \mathcal{F}^U_{\sigma',a,\delta}$. For $\mathcal{G} \subset \mathcal{F}$ and $\eta > 0$, define $J(\eta,\mathcal{G})$ to be the L_1-metric entropy of the set $\mathcal{G}$. This means that $J(\eta,\mathcal{G})$ is the logarithm of the minimum number of L_1-balls of radius η that cover $\mathcal{G}$. From Ghosal et al. (1999), we have

$$J(\delta, \overline{\mathcal{F}}^M_{\sigma,a,\delta}) \le \frac{Ca}{\sigma},$$

where C depends only on M and δ.

Take $(a_n)_{n\ge1}$ to be any increasing sequence of positive numbers such that $\lim_n a_n = +\infty$, and let $(\sigma_n)_{n\ge1}$ be a decreasing sequence of positive numbers such that $\lim_n \sigma_n = 0$. For our purposes, it is useful to consider sets of the type

$$\mathcal{G}_{\sigma,a_j,\delta} := \{\phi_\sigma * P : P([-a_j,a_j]) \ge 1-\delta,$$
$$P([-a_{j-1},a_{j-1}]) < 1-\delta\}.$$

These sets are pairwise disjoint, and $\lim_j \mathcal{G}_{\sigma,a_j,\delta} = \emptyset$ for any positive σ and δ. This definition entails the following inclusions: $\mathcal{G}_{\sigma,a_j,\delta} \subset \mathcal{F}^U_{\sigma,a_j,\delta}$ and $\mathcal{G}_{\sigma,a_j,\delta} \subset \mathcal{F}^L_{\sigma,a_{j-1},\delta}$. Moreover, $\mathcal{F}^L_{\sigma,a_j,\delta} \downarrow \emptyset$ as j tends to $+\infty$. Thus for any $\eta > 0$, there exists an integer N such that for any $j \ge N$,

$$J(\eta, \mathcal{F}^L_{\sigma,a_j,\delta}) \le J(\eta, \mathcal{F}^U_{\sigma,a_N,\delta}).$$

Set

$$\mathcal{G}^M_{\sigma_k,a_j,\delta} = \bigcup_{\sigma_k<\sigma<M} \mathcal{G}_{\sigma,a_j,\delta},$$

and note that

$$\bigcup_{j,k\ge1} \mathcal{G}^M_{\sigma_k,a_j,\delta} = \mathcal{F}.$$

Because $\mathcal{G}^M_{\sigma_k,a_j,\delta}$ is included in $\bigcup_{\sigma_k<\sigma<M} \mathcal{F}^L_{\sigma,a_j,\delta}$, we have

$$J(\eta, \mathcal{G}^M_{\sigma_k,a_j,\delta}) \le \frac{Ca_N}{\sigma_k} \qquad (6)$$

for any $j \ge N$. But the inclusion $\mathcal{G}^M_{\sigma_k,a_j,\delta} \subset \overline{\mathcal{F}}^M_{\sigma_k,a_j,\delta}$ entails that (6) holds true also for any $j < N$. These findings can be summarized by saying that $\mathcal{G}^{\sigma_{k-1}}_{\sigma_k,a_j,\delta}$ has a finite η-covering $\{C_{j,k,l} : l = 1,2,\ldots,N_{j,k}\}$, where $N_{j,k} \le [\exp(Ca_N/\sigma_k)] + 1$. Here we let $[x]$ denote the integer part of a real number x. Now define the sets

$$B_{j,\delta} = \{P : P([-a_j,a_j]) \ge 1-\delta,\ P([-a_{j-1},a_{j-1}]) < 1-\delta\}$$

for each $j \ge 1$. The condition for convergence (5) would be implied by

$$\sum_{j,k\ge1} \sum_{l=1}^{N_{j,k}} (\Pi(C_{j,k,l}))^\beta$$
$$\le \sum_{j,k\ge1} N_{j,k} \{\Pi(\mathcal{G}^{\sigma_{k-1}}_{\sigma_k,a_j,\delta})\}^\beta$$
$$\le \sum_{k\ge1} e^{Ca_N/\sigma_k} \{\mu(\sigma_k < \sigma \le \sigma_{k-1})\}^\beta \sum_{j\ge1} \{\Lambda(B_{j,\delta})\}^\beta$$
$$< +\infty, \qquad (7)$$

where $\sigma_0 = M$. Now consider the part concerning the mixing measure Λ. Let $A_j = (-\infty, -a_{j-1}) \cup (a_{j-1}, +\infty)$ and note that

$$B_{j,\delta} \subset \{P : P(A_j) > \delta'\},$$

with $\delta' > \delta$. Hence, by Markov's inequality,

$$\Lambda(B_{j,\delta}) \le \Lambda(\{P : P(A_j) > \delta'\}) \le \frac{1}{\delta'} P_0(A_j),$$

and thus (7) is implied by

$$\sum_{k\ge1} e^{C'/\sigma_k} \{\mu(\sigma_k < \sigma \le \sigma_{k-1})\}^\beta \sum_{j\ge1} \{P_0(A_j)\}^\beta < +\infty.$$

At this stage, we can fix $a_j \sim j$ as $j \to +\infty$. Condition (a) is then equivalent to $P_0(A_j) = O(j^{-(1+r)})$, which in turn ensures the convergence of $\sum_{j\ge1} \{P_0(A_j)\}^\beta$ for any β such that $(1+r)^{-1} < \beta < 1$. Moreover, take

$$\gamma > C'/\beta, \qquad (8)$$

so that condition (b) implies the prior μ to be such that

$$\sum_{k\ge1} e^{C'/\sigma_k} \{\mu(\sigma_k < \sigma \le \sigma_{k-1})\}^\beta$$

converges. The proof of Theorem 1 is now complete.

[Received November 2004. Revised February 2005.]

REFERENCES

Barron, A. (1988), "The Exponential Convergence of Posterior Probabilities With Implications for Bayes Estimators of Density Functions," Technical Report 7, University of Illinois, Dept. of Statistics.

Barron, A., Schervish, M. J., and Wasserman, L. (1999), "The Consistency of Distributions in Nonparametric Problems," *The Annals of Statistics*, 27, 536–561.

Choudhuri, N., Ghosal, S., and Roy, A. (2004), "Bayesian Estimation of the Spectral Density of a Time Series," *Journal of the American Statistical Association*, 99, 1050–1059.

Dey, D., Müller, P., and Sinha, D. (eds.) (1998), *Practical Nonparametric and Semiparametric Bayesian Statistics*, New York: Springer-Verlag.

Diaconis, P., and Freedman, D. (1986), "On the Consistency of Bayes Estimates," *The Annals of Statistics*, 14, 1–26.

Doksum, K. (1974), "Tailfree and Neutral Random Probabilities and Their Posterior Distributions," *The Annals of Probability*, 2, 183–201.

Escobar, M. D. (1988), "Estimating the Means of Several Normal Populations by Nonparametric Estimation of the Distribution of the Means," unpublished doctoral dissertation, Yale University, Dept. of Statistics.

Escobar, M. D., and West, M. (1995), "Bayesian Density Estimation and Inference Using Mixtures," *Journal of the American Statistical Association*, 90, 577–588.

Ferguson, T. S. (1973), "A Bayesian Analysis of Some Nonparametric Problems," *The Annals of Statistics*, 1, 209–230.

——— (1974), "Prior Distributions on Spaces of Probability Measures," *The Annals of Statistics*, 2, 615–629.

Ferguson, T. S., and Phadia, E. G. (1979), "Bayesian Nonparametric Estimation Based on Censored Data," *The Annals of Statistics*, 7, 163–186.

Freedman, D. A. (1963), "On the Asymptotic Behavior of Bayes's Estimates in the Discrete Case," *The Annals of Mathematical Statistics*, 34, 1386–1403.

Ghosal, S., Ghosh, J. K., and Ramamoorthi, R. V. (1999), "Posterior Consistency of Dirichlet Mixtures in Density Estimation," *The Annals of Statistics*, 27, 143–158.

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

Ishwaran, H., and James, L. F. (2001), "Gibbs Sampling Methods for Stick-Breaking Priors," *Journal of the American Statistical Association*, 96, 161–173.

——— (2003a), "Some Further Developments for Stick-Breaking Priors: Finite and Infinite Clustering and Classification," *Sankhyā*, 65, 577–592.

——— (2003b), "Generalized Weighted Chinese Restaurant Processes for Species Sampling Mixture Models," *Statistica Sinica*, 13, 1211–1235.

James, L. F. (2002), "Poisson Process Partition Calculus With Applications to Exchangeable Models and Bayesian Nonparametrics," *Mathematics ArXiv*, math.PR/0205093, available at *http://arxiv.org/abs/math.PR/0205093*.

——— (2003), "Poisson Calculus for Spatial Neutral to the Right Processes," *Mathematics ArXiv*, math.ST/0305053, available at *http://arxiv.org/abs/math.ST/0305053*.

Kingman, J. F. C. (1975), "Random Discrete Distributions," *Journal of the Royal Statistical Society*, Ser. B, 37, 1–22.

Lijoi, A., Mena, R. H., and Prünster, I. (2005), "Hierarchical Mixture Modelling With Normalized Inverse Gaussian Priors," *Journal of the American Statistical Association*, 100, 1278–1291.

Lo, A. Y. (1984), "On a Class of Bayesian Nonparametric Estimates: I. Density Estimates," *The Annals of Statistics*, 12, 351–357.

MacEachern, S. N. (1994), "Estimating Normal Means With a Conjugate Style Dirichlet Process Prior," *Communications in Statistics, Part B—Simulation and Computation*, 23, 727–741.

MacEachern, S. N., and Müller, P. (1998), "Estimating Mixture of Dirichlet Process Models," *Journal of Computational and Graphical Statistics*, 7, 223–239.

Müller, P., and Quintana, F. A. (2004), "Nonparametric Bayesian Data Analysis," *Statistical Science*, 19, 95–110.

Nieto-Barajas, L. E., Prünster, I., and Walker, S. G. (2004), "Normalized Random Measures Driven by Increasing Additive Processes," *The Annals of Statistics*, 32, 2343–2360.

Petrone, S., and Wasserman, L. (2002), "Consistency of Bernstein Polynomial Posteriors," *Journal of the Royal Statistical Society*, Ser. B, 64, 79–100.

Pitman, J. (1996), "Some Developments of the Blackwell–MacQueen Urn Scheme," in *Statistics, Probability and Game Theory. Papers in Honor of David Blackwell*, eds. T. S. Ferguson, L. S. Shapley, and J. B. MacQueen, Hayward, CA: Institute of Mathematical Statistics, pp. 245–267.

Regazzini, E., Lijoi, A., and Prünster, I. (2003), "Distributional Results for Means of Random Measures With Independent Increments," *The Annals of Statistics*, 31, 560–585.

Schwartz, L. (1965), "On Bayes Procedures," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 4, 10–26.

Walker, S. G. (2004a), "Modern Bayesian Asymptotics," *Statistical Science*, 19, 111–117.

——— (2004b), "New Approaches to Bayesian Consistency," *The Annals of Statistics*, 32, 2028–2043.

Walker, S., and Muliere, P. (1997), "Beta–Stacy Processes and a Generalization of the Pólya Urn Scheme," *The Annals of Statistics*, 25, 1762–1780.

Wasserman, L. (1998), "Asymptotic Properties of Nonparametric Bayesian Procedures," in *Practical Nonparametric and Semiparametric Bayesian Statistics*, eds. D. Dey, P. Müller, and D. Sinha, New York: Springer-Verlag, pp. 293–304.