

# A Computational Approach for Full Nonparametric Bayesian Inference Under Dirichlet Process Mixture Models

Alan E. GELFAND and Athanasios KOTTAS

Widely used parametric generalized linear models are, unfortunately, a somewhat limited class of specifications. Nonparametric aspects are often introduced to enrich this class, resulting in semiparametric models. Focusing on single or  $k$ -sample problems, many classical nonparametric approaches are limited to hypothesis testing. Those that allow estimation are limited to certain functionals of the underlying distributions. Moreover, the associated inference often relies upon asymptotics when nonparametric specifications are often most appealing for smaller sample sizes. Bayesian nonparametric approaches avoid asymptotics but have, to date, been limited in the range of inference. Working with Dirichlet process priors, we overcome the limitations of existing simulation-based model fitting approaches which yield inference that is confined to posterior moments of linear functionals of the population distribution. This article provides a computational approach to obtain the entire posterior distribution for more general functionals. We illustrate with three applications: investigation of extreme value distributions associated with a single population, comparison of medians in a  $k$ -sample problem, and comparison of survival times from different populations under fairly heavy censoring.

**Key Words:** Dirichlet process mixing; Extreme value distributions; Functionals; Markov chain Monte Carlo; Survival data distributions.

## 1. INTRODUCTION

Parametric modeling has long dominated both classical and Bayesian inference work. Such modeling is typically developed using generalized linear models within standard exponential families. Such families are limited, being unimodal with implicit mean-variance relationship. In looking beyond standard parametric families one is naturally led to mixture models. Continuous mixture models arising through a parametric family of mixing distributions—for example,  $t$ 's, Beta-binomials and Poisson-gammas—typically achieve

---

Alan E. Gelfand is Professor, Department of Statistics, University of Connecticut, Storrs, CT 06269-4120 (E-mail: alan@stat.uconn.edu). Athanasios Kottas is Visiting Assistant Professor, Institute of Statistics and Decision Sciences, Duke University, Durham, NC 27708-0251 (E-mail: thanos@stat.duke.edu).

©2002 American Statistical Association, Institute of Mathematical Statistics,  
and Interface Foundation of North America

*Journal of Computational and Graphical Statistics*, Volume 11, Number 2, Pages 289–305

increased heterogeneity but are still limited to unimodality. Finite mixture distributions (Titterton, Smith, and Makov 1985) are more flexible and now more feasible to implement due to advances in simulation-based model fitting. See, for example, Diebolt and Robert (1994) and Richardson and Green (1997).

Paradoxically, rather than handling the very large number of parameters resulting from finite mixture models with a large number of mixands, it may be easier to work with an infinite dimensional specification by assuming a random mixing distribution which is not restricted to a specified parametric family. The Dirichlet process (Ferguson 1973, 1974) is arguably the most widely used in this context, following Antoniak (1974), Ferguson (1983), and Lo (1984), resulting in what we refer to as Dirichlet process mixing. Since the work of Ferguson there has been a rapid growth in this area which is now generally referred to as Bayesian nonparametrics. An attractive summary and synthesis of this literature was given by Walker, Damien, Laud, and Smith (1999). Computational details for simulation-based fitting of Dirichlet process mixed models were given by Escobar (1994), Escobar and West (1995), Bush and MacEachern (1996), MacEachern and Müller (1998), Walker and Damien (1998), and Neal (2000).

A slightly different perspective leads to semiparametric modeling. As Oakes (1988) observed, "A practical motivation for consideration of semiparametric models is to avoid restrictive assumptions about secondary aspects of a problem while preserving a tight formulation for the features of primary concern." In the context of regression modeling, Gelfand (1999) noted that the objective of semiparametric modeling is "to enrich the class of standard parametric models by wandering nonparametrically near, in some sense, the standard class but retaining the linear structure."

In what follows, for simplicity, we confine our development to single and multiple sample problems. These can be handled entirely nonparametrically. However, our approach for full inference is readily applicable to semiparametric settings as in Gelfand and Kottas (2001) and Kottas and Gelfand (2001, in press).

In the Bayesian literature, for a single continuous population, fully nonparametric work includes Susarla and Van Ryzin (1976) and Lo (1984) on the estimation of the survival curve and the density, respectively. However, to estimate the population median based upon a single sample, the approach has been semiparametric, introducing the median parametrically as a location shift,  $\theta$ , relative to a random median 0 distribution,  $F$ , where  $F$  is modeled using some form of Dirichlet process prior. See, for example, Dalal (1979a, b) and Doss (1985a, b). Here, the joint posterior distribution of  $\theta$  and  $F$  is obtained and then marginalized to provide the posterior for  $\theta$ .

Our contribution is to be fully nonparametric and achieve full inference within the Dirichlet process mixing framework. Working in  $R^1$ , the support of our mixture class is *dense* within the class of all distributions on  $R^1$  so there is no need to investigate a richer class of mixture models. In Gelfand and Mukhopadhyay (1995), population inference was confined to posterior moments of linear functionals of the population distribution. Here, we provide a computational approach to obtain the entire posterior distribution for more general functionals. The approach uses the Sethuraman (1994) representation, *after* fitting

the model, to obtain posterior samples of the random mixing distribution. Then, a Monte Carlo integration is used to convert each such sample to a random draw from the posterior distribution of the functional of interest. Hence, arbitrarily accurate inference is available for the functional and for comparing it across populations. Functionals illustrated here include population distribution functions, density functions, hazard functions, quantiles and interquartile ranges as well as these features for the distribution of any order statistic associated with a sample of arbitrary size from the population.

Recently, Ishwaran and Zarepour (2000) proposed an alternate method of fitting Dirichlet process mixed models which yields approximate inference for functionals of the random mixture. For Dirichlet process models inference approaches for functionals is discussed in Muliere and Tardella (1998) and Guglielmi and Tweedie (2000).

The plan of the article is as follows. Section 2 briefly reviews Dirichlet process mixing. Section 3 presents the details of our approach along with theoretical justification. Section 4 looks at the single population problem, including the distribution of the maximum, providing a nonparametric Bayesian approach for the investigation of extreme value distributions. Section 5 considers fully nonparametric Bayesian analysis of variance. Finally, Section 6 considers a survival analysis setting in the presence of fairly heavy censoring.

## 2. DIRICHLET PROCESS MIXING

### 2.1 REVIEW

Consider a space  $\Theta$  and a  $\sigma$ -field  $\mathcal{B}$  of subsets of  $\Theta$ . Following Ferguson (1973), a random distribution function  $G$  on  $(\Theta, \mathcal{B})$  follows a Dirichlet process  $DP(\nu G_0)$ , where  $\nu > 0$  is a scalar (precision parameter) and  $G_0$  a specified distribution on  $\Theta$ , if, given an arbitrary finite measurable partition,  $B_1, \dots, B_r$  of  $\Theta$ , the joint distribution of  $(G(B_1), \dots, G(B_r))$  is Dirichlet  $(\nu G_0(B_1), \dots, \nu G_0(B_r))$ .

Let  $K(\cdot; \theta)$  be a parametric family of distributions (cdf's), indexed by  $\theta \in \Theta$ , with associated densities,  $k(\cdot; \theta)$ . If  $G$  is proper we define the mixture distribution

$$F(\cdot; G) = \int K(\cdot; \theta) G(d\theta). \tag{2.1}$$

In (2.1) it is useful to think of  $G(d\theta)$  as the conditional distribution of  $\theta$  given  $G$ . Differentiating both sides of (2.1) with respect to  $(\cdot)$  defines  $f(\cdot; G) = \int k(\cdot; \theta) G(d\theta)$ .

If  $G$  is random, say  $G \sim DP(\nu G_0)$ , then  $F(\cdot; G)$  is random. If the data  $D$  are  $Y_1, \dots, Y_n$  independent and identically distributed (iid) from  $F(\cdot; G)$  then, using the convenient bracket notation of Gelfand and Smith (1990), we write the posterior of  $F(\cdot; G)$  as  $[F(\cdot; G) | D]$ . Functionals of  $F(\cdot; G)$ , which we denote by  $H(F(\cdot; G))$ , are of interest with posteriors denoted by  $[H(F(\cdot; G)) | D]$ .

In the context of (2.1), suppose for each  $Y_i, i = 1, \dots, n$  we introduce a latent  $\theta_i$  and assume that the  $Y_i$ 's are conditionally independent given the  $\theta_i$ 's. Assume further that the  $\theta_i$ 's are conditionally independent and identically distributed given  $G$ . As a result the  $Y_i$ 's

are marginally independent, with joint density  $\prod_{i=1}^n f(y_i; G) = \prod_{i=1}^n \int k(y_i; \theta_i) G(d\theta_i)$ . Adding  $G \sim \text{DP}(\nu G_0)$  completes the Bayesian model specification, apart, perhaps, from a hyperprior on  $\nu$  (see Escobar and West 1995). Antoniak (1974) noted that this Bayesian model can be *marginalized* over  $G$  to obtain

$$\prod_{i=1}^n k(y_i; \theta_i) [\theta_1, \dots, \theta_n \mid G_0, \nu], \tag{2.2}$$

where the distribution  $[\theta_1, \dots, \theta_n \mid G_0, \nu]$  can be developed by exploiting the Polya urn characterization of the Dirichlet process. In particular,  $\theta_1 \sim G_0$  and for  $i = 2, \dots, n$ ,  $\theta_i \mid \theta_1, \dots, \theta_{i-1}$  is distributed according to the mixed distribution that places point mass  $(\nu + i - 1)^{-1}$  at  $\theta_j, j = 1, \dots, i - 1$  and continuous mass  $\nu(\nu + i - 1)^{-1}$  on  $G_0$ .

Hence, after marginalization, the  $\theta_i$  are no longer independent but a Markov chain Monte Carlo (MCMC) simulation method can be implemented (see the references in the introduction) to obtain samples essentially from the posterior,  $[\theta_1, \dots, \theta_n \mid D]$ . Gelfand and Mukhopadhyay (1995) described how to use these samples to compute posterior moments of linear functionals associated with  $F(\cdot; G)$ .

### 2.2 THE CLASS OF DISTRIBUTIONS

Using one-to-one transformations if needed, we can assume that the support of  $K(\cdot; \theta)$  is  $R^1$ . But then, as observed by Lo (1984, p. 355), the closure of the family of distributions of the form

$$F(\cdot; G) = \int K\left(\frac{\cdot - \mu}{\phi^{1/2}}\right) G(d\mu, d\phi) \tag{2.3}$$

contains all distributions on the real line. For convenience we take  $K$  to be Gaussian.

Inference about the median of (2.1), which we denote by  $\eta(F(\cdot; G))$ , requires its posterior,  $[\eta(F(\cdot; G)) \mid D]$ . This contrasts with the semiparametric approach discussed in the introduction where the median is introduced as a location parameter to a random zero-median distribution. In our setting, the semiparametric approach corresponds to creating, by Dirichlet process mixing,  $F_0(\cdot; G) = \int K(\frac{\cdot}{\sigma}) G(d\sigma^2)$  where  $K(0) = .5$  and then  $F(\cdot; G, \theta) = F_0(\cdot - \theta; G)$ , whence we would seek  $[\theta \mid D]$ . The semiparametric model is easier to work with but the class  $F_0(\cdot - \theta; G)$  is not as rich as (2.3) where we mix on both scale and location. For instance, if  $K(\cdot)$  is symmetric and/or unimodal so is  $F_0(\cdot - \theta; G)$ . Our class also differs from the mixture class discussed in Doss (1994). There, the family arises as a mixture of Dirichlet process, rather than by mixing using a mixing distribution which is drawn from a Dirichlet process.

### 3. INFERENCE FOR FUNCTIONALS ASSOCIATED WITH $F(\cdot; G)$

Section 3.1 describes our computational approach while Section 3.2 provides theoretical support.

### 3.1 THE APPROACH WITH EXAMPLES

The marginalization over  $G$  resulting in (2.2) enables one to *integrate* out the infinite dimensional part of the model and thus leaves only a finite dimensional posterior,  $[\theta_1, \dots, \theta_n \mid D]$ . Simulation based model fitting provides samples from this posterior which we denote by  $\theta_b^*, b = 1, \dots, B$ .

These samples are limited in terms of inference about  $F(\cdot; G)$ . In Escobar and West (1995), they were used to obtain the predictive distribution  $F(y \mid D)$  or the Bayesian density estimate  $f(y \mid D)$  which are the posterior mean of  $[F(y; G) \mid D]$  and  $[f(y; G) \mid D]$ , respectively. In Gelfand and Mukhopadhyay (1995) they were used to obtain posterior moments of linear functionals.

Working exclusively with expectations limits inference. For instance, how would we obtain an interval estimate for  $F(c; G)$  at a given  $c$ ? Furthermore, many features of  $F(\cdot; G)$  are not expressible as expectations. These include the median, in fact, arbitrary quantiles, and the interquartile range. Following Section 2, we denote the latter two by  $\eta_p(F(\cdot; G))$  and  $\eta_{.75}(F(\cdot; G)) - \eta_{.25}(F(\cdot; G))$ .

Also of interest are derived distributions from  $F(\cdot; G)$ , for example,  $(F(\cdot; G))^N$  and  $N(F(\cdot; G))^{N-1} f(\cdot; G)$ , the distribution function and density function, respectively, of the maximum of a sample of size  $N$  from  $F(\cdot; G)$ . In an extreme value analysis we might wish a point and interval estimate for the  $1-\alpha$ th quantile of the distribution of  $\max(Y_1, \dots, Y_N)$ . Features of the distribution of other order statistics and, perhaps, of the range might be useful as well. Other derived functions from  $F(\cdot; G)$  include the cumulative hazard,  $\Lambda(y; G) = -\log(1 - F(y; G))$  and the hazard  $\lambda(y; G) = -d \log(1 - F(y; G))/dy = f(y; G)/(1 - F(y; G))$ . Apart from inference for  $\Lambda(\cdot; G)$  or  $\lambda(\cdot; G)$ , comparison of these between populations might be desired.

In order to enrich our inferential capability we propose to approximately sample  $[H(F(\cdot; G)) \mid D]$  which in turn requires sampling  $[G \mid D]$ . First, if  $H$  is a linear functional,

$$H(F(\cdot; G)) = \int H(K(\cdot; \theta_0))G(d\theta_0). \tag{3.1}$$

Hence, given  $G$  if  $\theta_{0b}^*, b = 1, \dots, B$ , are a sample from  $G$ ,  $B^{-1} \sum_{b=1}^B H(K(\cdot; \theta_{0b}^*))$  is a Monte Carlo integration for a realization  $H(F(\cdot; G))$ . Next, since  $[\theta, G \mid D] \propto [G \mid \theta] [\theta \mid D]$ , given a posterior sample  $\theta_b^*, b = 1, \dots, B$ , if for each  $\theta_b^*$  we draw  $G_b^* \sim [G \mid \theta_b^*]$  we obtain a realization from  $[G \mid D]$ . Then, each  $G_b^*$  can be used with the Monte Carlo integration for (3.1) to obtain  $H_b^* = H(F(\cdot; G_b^*))$ , a realization from  $[H(F(\cdot; G)) \mid D]$ . The sample of  $H_b^*$  provides point and interval estimates for  $H(F(\cdot; G))$ .

To sample  $[G \mid \theta_b^*]$  we recall that this distribution (Ferguson 1973) is  $DP(\nu^* G_{0b}^*)$  where  $\nu^* = \nu + n$ ,  $G_{0b}^* = (\nu + n)^{-1}(\nu G_0 + \sum_{i=1}^n \delta_{\theta_{0i}^*})$  with  $\theta_b^* = (\theta_{b1}^*, \dots, \theta_{bn}^*)$  and  $\delta_a$  a degenerate distribution at  $a$ . An actual realization from  $[G \mid \theta_b^*]$  is almost surely of the form  $\sum_{j=1}^\infty \omega_j \delta_{\theta_j}$  where  $\omega_1 = z_1$ ,  $\omega_j = z_j \prod_{s=1}^{j-1} (1 - z_s)$ ,  $j = 2, 3, \dots$ ,  $\{z_s, s = 1, 2, \dots\}$  are iid from  $Be(1, \nu^*)$  and independently  $\{\theta_j, j = 1, 2, \dots\}$  are iid from  $G_{0b}^*$  (Sethuraman 1994). Following Sethuraman and Tiwari (1982), Doss (1994), and Muliere and Tardella (1998), we use a partial sum approximation,  $\sum_{j=1}^J \omega_j \delta_{\theta_j}$ . To choose  $J$ , we note that  $1 - \sum_{j=1}^J \omega_j$

$= \prod_{s=1}^J (1 - z_s)$  so  $E(\sum_{j=1}^J \omega_j) = 1 - (\nu^*/(\nu^* + 1))^J$ . For a given  $n$  with  $\nu = 1$ , we choose  $\epsilon$  sufficiently small, for example,  $\epsilon = .0001$  in our examples, and solve  $((n + 1)/(n + 2))^J = \epsilon$  for  $J$ . At  $j = J$  we replace  $\omega_J$  with  $1 - \sum_{j=1}^{J-1} \omega_j = \prod_{s=1}^{J-1} (1 - z_s)$ ; an alternative would be to rescale the partial sum approximation by  $\sum_{j=1}^J \omega_j$ . If there is concern with regard to a particular empirical selection for  $J$ , obvious sensitivity analysis could be carried out. We note that Muliere and Tardella (1998) allow  $J$  to be random and offer an algorithm to select  $J$  according to a desired accuracy. Such  $\epsilon$ -Dirichlet processes are valuable for approximate fitting of Dirichlet process models. Under modeling in the form of Dirichlet process mixing, such approximation in model fitting can be avoided; we can sample  $[\theta_1, \dots, \theta_n \mid D]$  to arbitrary accuracy. For us, such approximation is employed only after the model is fitted, as a device for obtaining realizations from the posterior of (3.1). This feature also distinguishes our approach from that of Ishwaran and Zarepour (2000) where the Dirichlet process is first approximated by easier to work with random distributions and then the resulting mixture model is fitted.

Immediately, the set of  $T_b^* = T(H_b^*)$  are a sample from the posterior of  $T(H(F(\cdot; G)))$  for an arbitrary function  $T$ . In this way, we can sample the “distribution function-at-a-point” functional or the “cumulative hazard-at-a-point” functional for any order statistic from a sample of arbitrary size  $N$ . Moreover, given the linear functionals  $H_1(F(\cdot; G))$  and  $H_2(F(\cdot; G))$ , if  $G_b^* \sim [G \mid D]$  then the pair  $H_{1b}^* = H_1(F(\cdot; G_b^*))$ ,  $H_{2b}^* = H_2(F(\cdot; G_b^*))$  are a realization from  $[H_1, H_2 \mid D]$  and  $T_b^* = T(H_{1b}^*, H_{2b}^*)$  is a realization from the posterior of  $T(H_1, H_2)$ . In this way we can sample the posterior of the log-likelihood, that is,  $[\sum_{i=1}^n \log f(y_i; G) \mid D]$  which is of interest for model determination, as in Dempster (1997). In addition, we can sample the “hazard-at-a-point” functional and the “density-at-a-point” functional for any order statistic.

Next, for a grid of  $y$  values,  $y_1 < y_2 < \dots < y_L$  and a sample  $G_b^*$ ,  $b = 1, \dots, B$  approximately from  $[G \mid D]$  we can create an  $L \times B$  matrix of realizations  $F(y_l; G_b^*)$ . The  $l$ th row of this matrix provides a sample from the posterior  $[F(y_l; G) \mid D]$ . Each column provides (with interpolation) a random realization of the cdf  $F(\cdot; G)$  from  $[F(\cdot; G) \mid D]$ .

But then, given a random  $F(\cdot; G^*)$ , we can invert to find a random  $\eta_p^* = \eta_p(F(\cdot; G^*))$  and thus a sample  $\eta_{p,b}^*$  from  $[\eta_p(F(\cdot; G)) \mid D]$ . Obviously, we can also obtain a sample from the posterior of the interquartile range of  $F(\cdot; G)$ . Finally, suppose we want quantiles for the distribution of order statistics associated with  $F(\cdot; G)$ . For instance, suppose we seek a sample from the posterior of the  $p$ th quantile of the distribution of  $\max(Y_1, \dots, Y_N)$ . The preceding discussion shows how to obtain realizations from  $[(F(y; G))^N \mid D]$ . A matrix similar to the above  $L \times B$  matrix with  $(F(y_l; G_b^*))^N$  replacing  $F(y_l; G_b^*)$  can be created; by inversion a sample of quantiles can be obtained.

### 3.2 THEORETICAL JUSTIFICATION FOR THE APPROACH

Again, the approach in Section 3.1 introduces the partial sum approximation only in the posterior sampling of  $H(F(\cdot; G))$ . The  $\theta_b^*$  are obtained using customary Gibbs sampling under the Dirichlet process mixing specification. This contrasts with the work of Muliere

and Tardella (1998) who begin with an  $\epsilon$ -Dirichlet process approximation in the fitting of the model and then provide convergence results for functionals of  $G$ . Our use of the partial sum approximation, with a fixed  $J$ , after the model fitting allows stronger convergence results, in the process providing formal justification for our approach. Following Section 3.1, for  $J = 2, 3, \dots$ , let

$$G_J = \sum_{j=1}^{J-1} \omega_j \delta_{\theta_j} + \left\{ \prod_{s=1}^{J-1} (1 - z_s) \right\} \delta_{\theta}, \tag{3.2}$$

be the  $J$ th partial sum approximation to  $G = \sum_{j=1}^{\infty} \omega_j \delta_{\theta_j}$ . We study the limiting behavior of the sequence of random variables

$$H(F(\cdot; G_J)) - H(F(\cdot; G)) = \int H(K(\cdot; \theta)) G_J(d\theta) - \int H(K(\cdot; \theta)) G(d\theta),$$

where  $H$  is a linear functional, providing convergence results for linear functionals and continuous functions of linear functionals. A result for the quantile functional is established as well.

The following theorem, whose proof is given in the Appendix, yields a strong result for linear functionals with  $H(K(\cdot; \theta))$  bounded, as a function of  $\theta$ .

**Theorem 1.** *Suppose  $H$  is a bounded linear functional. Then  $H(F(\cdot; G_J)) - H(F(\cdot; G))$  converges almost surely to 0 as  $J \rightarrow \infty$ .*

In particular, for the “distribution function-at-a-point” functional we obtain Corollary 1.

**Corollary 1.** *For any fixed  $c$  on the support of  $K(\cdot; \theta)$ ,  $F(c; G_J) - F(c; G)$  converges to 0 almost surely as  $J \rightarrow \infty$ .*

Note that the “density function-at-a-point” functional is not bounded for  $\frac{1}{\phi^{1/2}} k(\frac{\cdot - \mu}{\phi^{1/2}})$  over  $\theta = (\mu, \phi)$ . Theorem 1 applies if we bound  $\phi$  away from 0. Alternatively, Theorem 2 provides a quadratic mean convergence result, whose proof is given in the Appendix.

**Theorem 2.** *Consider a linear functional  $H$  for which  $E\{(H(K(\cdot; \theta)))^2 \mid \theta \sim G_0\} < \infty$ . Then  $E\{H(F(\cdot; G_J)) - H(F(\cdot; G))\}^2 \rightarrow 0$  as  $J \rightarrow \infty$ .*

Applied to the “density-at-a-point” functional we have Corollary 2.

**Corollary 2.** *For any fixed  $c$  on the support of  $K(\cdot; \theta)$ ,  $E\{f(c; G_J) - f(c; G)\}^2 \rightarrow 0$  as  $J \rightarrow \infty$ , provided  $E\{(k(c; \theta))^2 \mid \theta \sim G_0\} < \infty$ .*

The condition of Corollary 2 depends on the specific choice of  $k(\cdot; \theta)$  and  $G_0$ . For the class  $F(\cdot; G)$  defined in (2.3), where  $k(c; \theta) \equiv k(c; \mu, \phi)$  is the density of a  $N(\mu, \phi)$  distribution, it holds if we take  $G_0(\mu, \phi) = N(\mu \mid m, \tau\phi)IG(\phi \mid a, b)$ , with  $m, \tau, a$  and  $b$  specified hyperparameters.

Combining Corollaries 1 and 2 and using standard results, Theorem 3 provides convergence results for all the functionals discussed in Section 3.1 other than the quantile functional.

**Theorem 3.** *Consider fixed points  $c$  and  $d$  on the support of  $K(\cdot; \theta)$ . Let  $g_1, g_2$  and  $g_3$  be continuous  $R$ -valued functions on  $[0, 1]$ ,  $[0, 1] \times R^+$  and  $R^+ \times R^+$ , respectively. Then*

Table 1. A Sample of Velocities for 82 Galaxies of the Corona Borealis Region, from Roeder (1990). A “+” indicates the observation is in the sample of size 10 and a “\*” that it is in the sample of size 40.

Velocity (in km/sec $\times 10^{-3}$ )				
9.172*	19.349*	20.196	22.209	23.706*
9.350*	19.440*	20.215	22.242	23.711
9.483	19.473*	20.221	22.249	24.129*
9.558 <sup>+</sup> *	19.529*	20.415*	22.314*	24.285*
9.775	19.541*	20.629 <sup>+</sup>	22.374*	24.289*
10.227*	19.547*	20.795*	22.495*	24.366*
10.406	19.663*	20.821	22.746*	24.717*
16.084 <sup>+</sup>	19.846*	20.846*	22.747	24.990
16.170 <sup>+</sup>	19.856	20.875 <sup>+</sup> *	22.888	25.633 <sup>+</sup>
18.419	19.863	20.986	22.914	26.960*
18.552*	19.914*	21.137	23.206	26.995
18.600 <sup>+</sup> *	19.918	21.492	23.241 <sup>+</sup>	32.065*
18.927*	19.973 <sup>+</sup> *	21.701	23.263	32.789 <sup>+</sup>
19.052	19.989*	21.814	23.484	34.279
19.070*	20.166	21.921	23.538*	
19.330*	20.175	21.960	23.542*	
19.343*	20.179*	22.185	23.666	

- (a)  $g_1(F(c; G_J))$  converges almost surely to  $g_1(F(c; G))$  as  $J \rightarrow \infty$ .
  - (b)  $g_2(F(c; G_J), f(c; G_J))$  converges in probability to  $g_2(F(c; G), f(c; G))$  as  $J \rightarrow \infty$ .
  - (c)  $g_3(f(c; G_J), f(d; G_J))$  converges in probability to  $g_3(f(c; G), f(d; G))$  as  $J \rightarrow \infty$ .
- Finally, Theorem 4, whose proof is in the Appendix, handles the quantile functional.

**Theorem 4.** For any fixed  $p$  in  $(0, 1)$ , outside a set of Lebesgue measure 0,  $\eta_p(F(\cdot; G_J)) - \eta_p(F(\cdot; G))$  converges in probability to 0 as  $J \rightarrow \infty$ .

The proof of Theorem 4 makes use only of the convergence in distribution of  $F(c; G_J)$  to  $F(c; G)$ , for any fixed  $c$ . Hence, implementing Theorem 3, similar convergence results for the quantiles of the distributions of order statistics associated with  $F(\cdot; G)$  can be obtained.

### 4. EXTREME VALUE ANALYSIS IN THE ONE SAMPLE PROBLEM

Suppose  $Y_1, \dots, Y_n$  is a sample from  $F(\cdot; G)$  with  $F(\cdot; G)$  defined in (2.3). We illustrate the approach of Section 3 in terms of extreme value distributions. That is, we assume  $[(F(\cdot; G))^N \mid D]$  is of interest for an arbitrary  $N$ . Note that we are inferring about the distribution of  $Y_{(N)}$  without observing samples of  $Y_{(N)}$ 's, as is assumed in customary parametric extreme value theory. See, for example, Johnson, Kotz, and Balakrishnan (1995, chap. 22). We are not aware of any other Bayesian nonparametric approach in this context.

We consider an astronomical dataset (Table 1) from Roeder (1990) which records velocities in  $\text{km/sec} \times 10^{-3}$  for 82 galaxies of the Corona Borealis region. Roeder focused on estimating the velocity density, as did Escobar and West (1995) and Richardson and Green (1997). Using random samples of size 10 and 40, from the 82, we seek inference regarding the distribution of  $Y_{(82)}$ , which is, in fact, observed to be 34.279. A Normal-Inverse Gamma form  $G_0(\mu, \phi) = N(\mu \mid m, \tau\phi)IG(\phi \mid a, b)$ , facilitates sampling from the



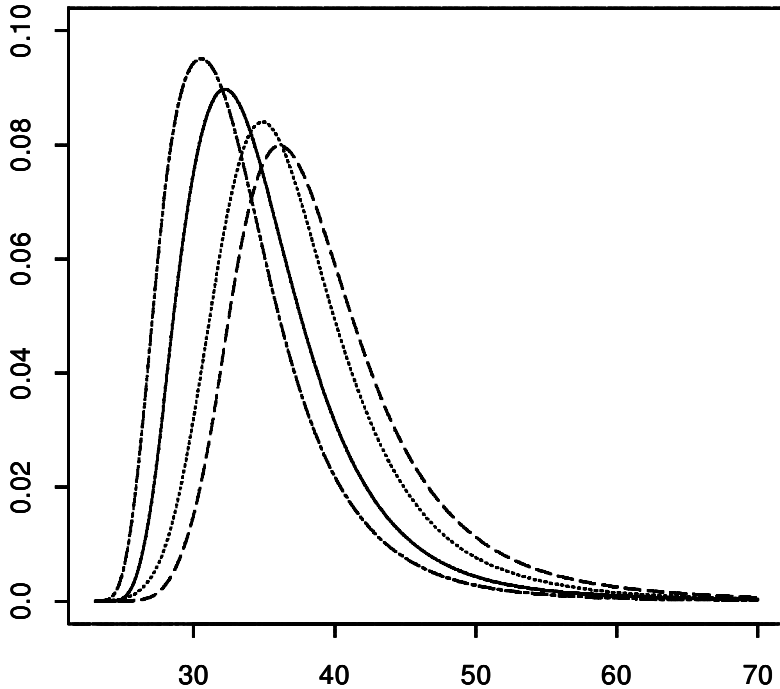


Figure 1. For the galaxy data, Bayesian point estimates of the density. For  $N = 82$ , the dotted line denotes the density corresponding to the sample of size 10 and the dashed-dotted line to that of size 40. For  $N = 150$ , the dashed line corresponds to the sample of size 10 and the solid line to that of size 40.

full conditionals for the  $\theta_i$ 's and makes Corollary 2 and Theorem 3 applicable. Based upon a range which is conservatively between 5 and 40, we set the prior mean for  $\mu$  at  $m = 22.5$  with  $\tau = 1.0$ . The choice  $a = 2.0$  and  $b = 0.03$  provides an infinite variance for  $\phi$  with expectation roughly  $(35/6)^2$ . Finally we set  $\nu = 1.0$ .

The posterior mean of the density is shown for each of the two sample sizes in Figure 1. Additionally, the posterior median and 95% equal tail interval estimates for the median and .95 quantile of the distribution of  $Y_{(82)}$ , obtained from the sample of size 10, are 36.347 (30.600,49.551) and 43.168 (35.405,64.395), respectively. The corresponding estimates obtained from the sample of size 40 are 32.263 (27.965,41.827) and 39.294 (32.591,57.932). Were we to sample say 150 galaxies what might we expect to see for the distribution of  $Y_{(150)}$ ? The density estimates again using the two samples are also shown in Figure 1. Some shift to the right relative to that of  $Y_{(82)}$  is seen but not a lot, reflecting the lack of skewness in our samples. In fact, there is essentially no skewness in the entire sample of 82 velocities (Roeder 1990, fig. 1). Again we can obtain point and interval estimates for the median and .95 quantile of the distribution of  $Y_{(150)}$ . For the sample of size 10, they are 37.859 (31.710,53.839) and 44.561 (36.566,68.064), respectively, and for the sample of size 40, 33.821 (29.149,46.897) and 40.694 (33.595,62.789), respectively.

Regarding computational details, convergence of the Gibbs sampler was assessed through multiple chains. Then for the first chain, after discarding the first 10,000 itera-

Table 2. Per Capita Public School Expenditures for 1977 (in \$1,000) for the Contiguous United States Broken into Five Regions, from Snedecor and Cochran (1989)

	NE	SE	SC	NC	MP
	1.33	1.66	1.16	1.74	1.76
	1.26	1.37	1.07	1.78	1.75
	2.33	1.21	1.25	1.39	1.60
	2.10	1.21	1.11	1.28	1.69
	1.44	1.19	1.15	1.88	1.42
	1.55	1.48	1.15	1.27	1.60
	1.89	1.19	1.16	1.67	1.56
	1.88		1.26	1.40	1.24
	1.86		1.30	1.51	1.45
	1.99			1.74	1.35
				1.53	1.16
$n$	10	7	9	11	11
$\bar{y}$	1.763	1.330	1.179	1.563	1.507
$S^2$	0.1240	0.0335	0.0057	0.0448	0.0404

tions, we thinned the output keeping every 50th and 150th iteration for the samples of size 10 and 40, respectively. In both cases, we finally obtain  $B = 1,000$  draws from the posterior. For the algorithm providing inference for functionals we work with a grid of 275 equidistant points over the interval (15,70). Following the suggestion of Section 3.1 for the number of terms in the partial sum approximation, we set  $J = 140$ , for the sample of size 10, and  $J = 450$  for that of size 40. Finally, the sample size for the Monte Carlo integration was taken to be 1,000. Similar choices were used for the examples in the following sections, adjusting  $J$  appropriately to the data sample sizes. All the computations were implemented using FORTRAN 90, with IMSL subroutines for random number generations, on a PC equipped with a Pentium III processor. The programs (not the most efficient), providing all the results for the samples of size 10 and 40, executed at 21 and 29 minutes, respectively. Computing times were essentially identical for  $N = 82$  and  $N = 150$ .

## 5. THE $K$ -SAMPLE PROBLEM

Consider data  $Y_{ij}$ ,  $i = 1, \dots, k$ ,  $j = 1, \dots, n_i$  with  $Y_{ij}$  drawn from  $F_i$ . Standard analysis of variance modeling sets  $F_i = N(\mu_i, \sigma^2)$  and seeks comparison of means. Inference relies heavily upon the normality and homogeneity of variance assumptions. We assume the  $F_i$  arise under Dirichlet process mixing and investigate comparison of medians.

We consider a small dataset taken from Snedecor and Cochran (1989, p. 231) consisting of public school expenditure data per capita in 1977 for the 48 contiguous states of the U.S. broken into five regions, Northeast (NE), Southeast (SE), South Central (SC), North Central (NC), and Mountain Pacific (MP). The data are supplied in Table 2 along with the usual summary statistics. It will be hard to *confirm* normality with the small sample sizes and the homogeneity of variance assumption clearly seems untenable. A Kruskal–Wallis test finds significant differences between the regions ( $p < .0001$ ). Pairwise comparisons at the .05 level, using average ranks, finds NE different from SE, NE different from SC, NC different

Table 3. For the Public School Expenditure Data, Bayesian and Classical Point and Interval Estimates for the Population Median

Region	Bayesian estimates (median)		Classical estimates (median)	
	Point estimate	Interval estimate	Point estimate	Interval estimate
NE	1.7934	(1.4674,2.0258)	1.87	(1.402,2.028)
SE	1.3494	(1.1869,1.5414)	1.21	(1.190,1.528)
SC	1.1970	(1.1095,1.3023)	1.16	(1.119,1.258)
NC	1.5707	(1.4318,1.7154)	1.53	(1.381,1.743)
MP	1.5255	(1.3750,1.6569)	1.56	(1.341,1.695)

from SC and MP different from SC.

Instead, we adopt the “analysis of densities” framework of Tomlinson and Escobar (1999), assuming  $F_i = F(\cdot; G_i)$  as in (2.3). The  $G_i$  are assumed to be independent realizations from  $DP(\nu G_0)$ , where  $\nu = 1.0$  and  $G_0(\mu, \phi) = N(\mu | m, \tau\phi)IG(\phi | a, b)$  with  $m = 1.75$ ,  $\tau = 10.0$ ,  $a = 2.0$ , and  $b = 16.0$ . To clarify these choices, using a range for  $Y$  which is conservatively between 1.0 and 2.5 suggest a prior mean for  $\mu$  centered at 1.75.  $\tau = 10.0$  implies our prior knowledge about  $\mu$  is worth 1/10 of an observation. If  $2.5 - 1.0 = 1.5$  is roughly a 6 standard deviation range, we can set  $E(\phi) = (1.5/6)^2$ . The choices for  $a$  and  $b$  provide this mean with infinite variance.

In Table 3, for each group, we provide the posterior median and a 95% equal tail interval estimate for the population median using our approach. Shown, for comparison, are the sample median and an interval estimate based upon the sign test. In Figure 2 we plot the posterior mean,  $E(f(y; G_i) | D)$  over a grid of  $y$  values for  $i = 1, \dots, 5$ . Differences in center and in dispersion are evident. Also noteworthy is the bimodal form for the Northeast region, not surprising given that four observations in this sample lie between 1.33 and 1.55 while the other six lie between 1.86 and 2.33. Finally, in Table 4 we contrast the population medians in pairs, summarizing the  $\binom{5}{2}$  posteriors,  $[\eta(F(\cdot; G_i)) - \eta(F(\cdot; G_j)) | D]$  with posterior median and interval estimates. We agree with the Kruskal–Wallis results though the latter does not provide interval estimates. For such comparisons, assuming a shift model, we present the customary point and interval estimates, based upon the Mann–Whitney test. The inference is similar though the shift estimates find two more significances than we do.

## 6. A COMPARISON OF POPULATIONS OF SURVIVAL DATA

The literature on nonparametric modeling and inference for survival data with censoring is very large. No review is attempted here. After the early fully nonparametric Bayesian work of Susarla and Van Ryzin (1976) and Cornfield and Detre (1977), the Bayesian effort moved in the direction of semiparametric modeling as summarized in the recent article by Sinha and Dey (1997). In current work such models are fitted by MCMC methods. Censoring is handled through the introduction of latent survival time variables.

We use Dirichlet process mixing to provide nonparametric modeling of the survival distribution. We also handle censoring using latent variables, as in, for example, Kuo and

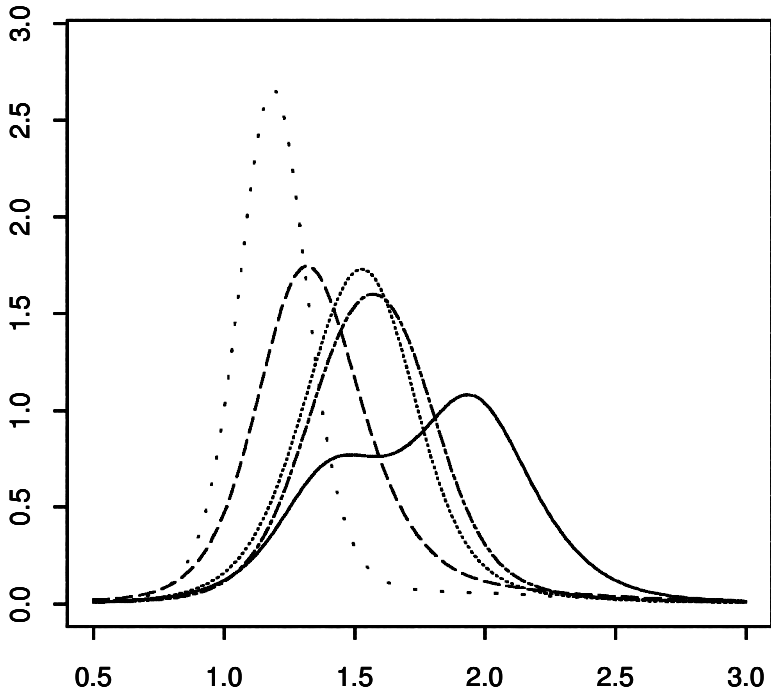


Figure 2. For the public school expenditure data, Bayesian point estimates of the five densities. The solid line denotes NE, the dashed line SE, the more sparse dotted line SC, the dashed-dotted line NC, and the dotted line MP.

Mallick (1997). We work on the log-scale in order to more comfortably employ (2.3). Semiparametric accelerated failure time models also use  $y_i = \log t_i$  in order to model errors on  $R^1$ . We make no claim that our analysis is *better* than the other nonparametric ones. However, we are drawn to the richness of (2.3) and the full inference with regard to individual distributions and comparison of distributions.

We work with a dataset involving survival time after bone marrow transplantation for treatment of leukemia, taken from Klein and Moeschberger (1997, pp. 464–467). The data samples three populations: an ALL group with sample size 38; an AML low-risk group with sample size 54; and an AML high-risk group with sample size 45. See Klein and Moeschberger for further details. Censoring is fairly heavy, 14/38, 31/54, and 11/45, respectively. In the AML high-risk group, the censored times are larger than all but one of the observed survival times, suggesting the possibility of a second mode or at the least a very heavy tail for this population. We adopt priors following the same logic as in the previous two sections. We take  $\nu = 1.0$ ,  $\tau = 20.0$  and, using a conservative range of  $(0, 12)$  for  $\log t$ ,  $m = 6.0$ ,  $a = 2.0$  and  $b = 0.25$ .

In Figure 3 we plot the Kaplan–Meier estimates for each of the groups and overlay the posterior mean of the survival function. Note the implicit data-determined smoothing which our model provides. In Table 5 the median survival time for each population is summarized by posterior median and interval estimate. In the classical theory, to obtain interval

Table 4. For the Public School Expenditure Data, Bayesian and Classical Point and Interval Estimates for Contrasts of Population Medians in Pairs

Contrast	Bayesian estimates		Classical estimates	
	Point estimate	Interval estimate	Point estimate	Interval estimate
NE-SE	0.4318	(0.0441,0.7121)	0.4250	(0.0700,0.7301)
NE-SC	0.5934	(0.2436,0.8327)	0.6650	(0.2501,0.8401)
NE-NC	0.2106	(-0.1483,0.4911)	0.1800	(-0.0900,0.5000)
NE-MP	0.2674	(-0.0798,0.5724)	0.2800	(-0.0501,0.5398)
SE-SC	0.1512	(-0.0454,0.3719)	0.1000	(0.0300,0.3300)
SE-NC	-0.2215	(-0.4626,0.0231)	-0.2100	(-0.5100,-0.0499)
SE-MP	-0.1745	(-0.3893,0.0818)	-0.2100	(-0.3899,0.0300)
SC-NC	-0.3740	(-0.5437,-0.1996)	-0.3800	(-0.5801,-0.2101)
SC-MP	-0.3223	(-0.4924,-0.1473)	-0.3400	(-0.4900,-0.1700)
NC-MP	0.0482	(-0.1537,0.2573)	0.0500	(-0.1600,0.2499)

estimates, either a density estimate at the median is required or, alternatively, the interval estimate is only available implicitly through an approximate probability statement regarding the estimated survival function. Also in Table 5 we contrast the median survival times, obtaining the posterior medians and interval estimates for the differences. The median survival time for group 2 is significantly larger than for groups 1 and 3. Also the median for

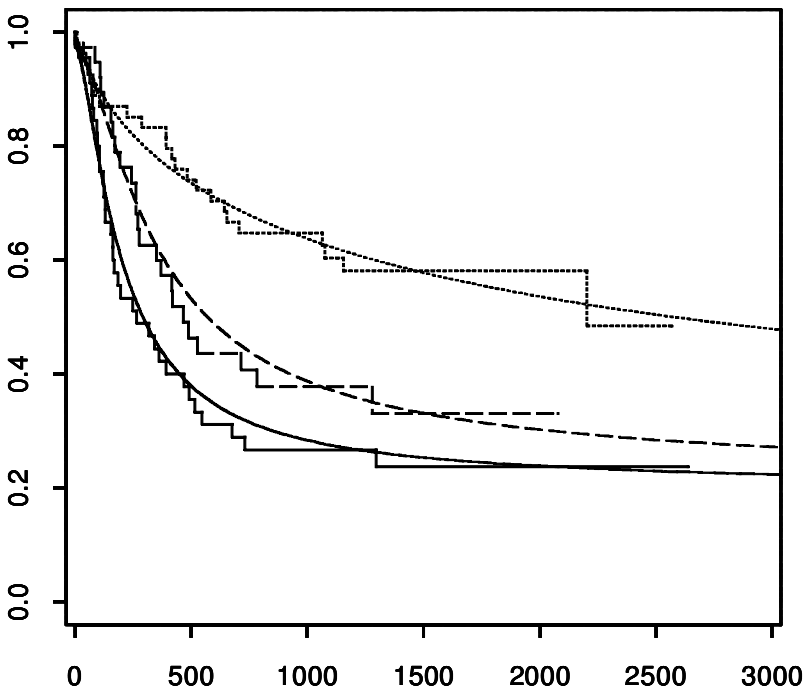


Figure 3. For the bone marrow transplantation data, Kaplan–Meier and Bayesian point estimates of the three survival functions. For the ALL group both are denoted by the dashed line, for the AML low-risk group by the dotted line and for the AML high-risk group by the solid line.

Table 5. Inference Summary for the Bone Marrow Transplantation Data

Group	Median survival time	
	Point estimate	Interval estimate
ALL	573.49	(333.85,1418.58)
AML low-risk	2545.53	(1162.84,24771.04)
AML high-risk	292.92	(187.71,524.88)
Contrast	Contrasts of median survival times	
	Point estimate	Interval estimate
AML low-risk-ALL	2009.75	(382.69,24370.11)
ALL-AML high-risk	271.67	(-52.76,1123.29)
AML low-risk-AML high risk	2271.15	(829.49,24489.15)

group 1 is nearly significantly larger than that for group 3. We are unaware of any classical nonparametric approach to obtain such interval estimates.

### APPENDIX: PROOFS

**Proof of Theorem 1:** Since  $\sum_{j=1}^{\infty} \omega_j \delta_{\theta_j}$  is almost surely a realization of  $G$ , realizations of  $H(F(\cdot; G))$  are almost surely of the form  $\sum_{j=1}^{\infty} \omega_j H(K(\cdot; \theta_j))$ . For any  $J$ , let  $X_J^H = H(F(\cdot; G)) - H(F(\cdot; G_J))$ . By (3.2) we have, almost surely, that  $X_J^H = \sum_{j=J}^{\infty} A_j$ , where  $A_j = \omega_j(H(K(\cdot; \theta_j)) - H(K(\cdot; \theta_J)))$ . Note that by assumption, there exists  $M$  such that  $|A_j| \leq 2M\omega_j$ , almost surely, for all  $j$ . Hence  $P(|X_J^H| < \infty) = 1$ , since  $\sum_{j=1}^{\infty} \omega_j = 1$ . By the same argument  $|X_J^H| \leq 2MY_J$ , almost surely, for all  $J$ , where  $Y_J = \prod_{s=1}^{J-1} (1 - z_s)$ . Since the  $z_s$  are drawn independently from  $\text{Be}(1, \nu + n)$ , it is straightforward to show that  $E(|Y_J|^r) = \{(\nu + n)/(\nu + n + r)\}^{J-1}$ , for any integer  $r > 0$ . Consequently, for each  $r$ ,  $Y_J$  converges to 0 in  $r$ th mean as  $J \rightarrow \infty$  and therefore  $Y_J$  converges to 0 in probability as  $J \rightarrow \infty$ . Moreover  $P(0 \leq Y_{J+1} \leq Y_J) = 1$  yielding almost sure convergence of  $Y_J$  to 0 as  $J \rightarrow \infty$ . Thus,  $|X_J^H|$  converges almost surely to 0 as  $J \rightarrow \infty$ , establishing the result.  $\square$

**Proof of Theorem 2:** Let  $X_J^{*H} = \sum_{j=J}^{\infty} \omega_j |H(K(\cdot; \theta_j)) - H(K(\cdot; \theta_J))|$ , for  $J = 1, 2, \dots$ . Using the Cauchy-Schwarz inequality and  $\sum_{j=1}^{\infty} \omega_j = 1$ , we obtain  $(X_J^{*H})^2 \leq S_J$ , almost surely, where  $S_J = \sum_{j=J}^{\infty} \omega_j (H(K(\cdot; \theta_j)) - H(K(\cdot; \theta_J)))^2$ . Now  $E(S_J) = \kappa_n \sum_{j=J+1}^{\infty} E(\omega_j)$ , where  $\kappa_n = 2\text{var}(H(K(\cdot; \theta)) | \theta \sim G_{0b}^*)$  with  $G_{0b}^*$  defined in Section 3.1. In this calculation we have used, in turn, the monotone convergence theorem, the independence of the sequences  $\{z_s, s = 1, 2, \dots\}$  and  $\{\theta_j, j = 1, 2, \dots\}$  and the fact that the  $\theta_j$  are a sample from  $G_{0b}^*$ . Since  $G_{0b}^*$  is a mixed distribution with continuous mass on  $G_0$  and a finite number of point masses, the assumption of the theorem yields  $\kappa_n < \infty$ . Moreover,  $E(\omega_j) = (\nu + n)^{-1} \{(\nu + n)/(\nu + n + 1)\}^j$ , for each  $j$  and therefore finally  $E(S_J) = \kappa_n \{(\nu + n)/(\nu + n + 1)\}^J < \infty$ . This proves that  $P(X_J^{*H} < \infty) = 1$  and hence  $P(X_J^H < \infty) = 1$ , with  $X_J^H$  defined in the proof of Theorem 1. Furthermore,  $|X_J^H|^2 \leq (X_J^{*H})^2 \leq S_J$ , almost surely, with  $\lim_{J \rightarrow \infty} E(S_J) = 0$ , yielding  $\lim_{J \rightarrow \infty} E(|X_J^H|^2) = 0$  and thus the result.  $\square$

**Proof of Theorem 4:** Working with a continuous distribution with distribution function  $F(\cdot; G)$ , we have for any  $p \in (0,1)$  that  $F(\eta_p(F(\cdot; G)); G) = p$ . Fix some  $\varepsilon > 0$ . Then

$$\begin{aligned} P(|\eta_p(F(\cdot; G_J)) - \eta_p(F(\cdot; G))| \leq \varepsilon) &= P(\eta_p(F(\cdot; G_J)) \leq y_1) \\ &\quad - P(\eta_p(F(\cdot; G_J)) < y_2) \\ &= P(F(y_1; G_J) \geq p) - P(F(y_2; G_J) > p), \end{aligned} \tag{A.1}$$

where  $y_1 = \eta_p(F(\cdot; G)) + \varepsilon$  and  $y_2 = \eta_p(F(\cdot; G)) - \varepsilon$ . If  $p$  is a point of continuity of the distribution functions of the random variables  $F(y_i; G)$ ,  $i = 1,2$ , then the right hand side in (A.1) converges to  $P(F(y_1; G) \geq p) - P(F(y_2; G) > p) = 1$ , as  $J \rightarrow \infty$ , as a consequence of Corollary 1. Since the sets of points of discontinuity of these distribution functions are at most countable, the result is established.  $\square$

## ACKNOWLEDGMENTS

The work of the first author was supported in part by NSF grant DMS 96-25383. The authors thank two reviewers whose suggestions have improved the article.

*[Received April 2000. Revised January 2001.]*

## REFERENCES

- Antoniak, C. E. (1974), "Mixtures of Dirichlet Processes With Applications to Bayesian Nonparametric Problems," *The Annals of Statistics*, 2, 1152–1174.
- Bush, C. A., and MacEachern, S. N. (1996), "A Semiparametric Bayesian Model for Randomised Block Designs," *Biometrika*, 83, 275–285.
- Cornfield, J., and Detre, K. (1977), "Bayesian Life Table Analysis," *Journal of the Royal Statistical Society*, Ser. B, 39, 86–94.
- Dalal, S. R. (1979a), "Dirichlet Invariant Processes and Applications to Nonparametric Estimation of Symmetric Distribution Functions," *Stochastic Processes and their Applications*, 9, 99–107.
- (1979b), "Nonparametric and Robust Bayes Estimation of Location," in *Optimizing Methods in Statistics*, ed. J.S. Rustagi, New York: Academic Press, pp. 141–166.
- Dempster, A. P. (1997), "The Direct Use of Likelihood for Significance Testing," *Statistics and Computing*, 7, 247–252.
- Diebolt, J., and Robert, C. P. (1994), "Estimation of Finite Mixture Distributions through Bayesian Sampling," *Journal of the Royal Statistical Society*, Ser. B, 56, 363–375.
- Doss, H. (1985a), "Bayesian Nonparametric Estimation of the Median; part I: Computation of the Estimates," *The Annals of Statistics*, 13, 1432–1444.
- (1985b), "Bayesian Nonparametric Estimation of the Median; part II: Asymptotic Properties of the Estimates," *The Annals of Statistics*, 13, 1445–1464.
- (1994), "Bayesian Nonparametric Estimation for Incomplete Data via Successive Substitution Sampling," *The Annals of Statistics*, 22, 1763–1786.
- Escobar, M. D. (1994), "Estimating Normal Means With a Dirichlet Process Prior," *Journal of the American Statistical Association*, 89, 268–277.

- 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.
- (1983), "Bayesian Density Estimation by Mixtures of Normal Distributions," in *Recent Advances in Statistics*, eds. M. H. Rizvi, J. S. Rustagi, and D. Siegmund, New York: Academic Press, pp. 287–302.
- Gelfand, A. E. (1999), "Approaches for Semiparametric Bayesian Regression," in *Asymptotics, Nonparametrics and Time Series*, ed. Subir Ghosh, New York: Marcel Dekker, Inc., pp. 615–638.
- Gelfand, A. E., and Kottas, A. (2001), "Nonparametric Bayesian Modeling for Stochastic Order," *Annals of the Institute of Statistical Mathematics*, 53, 865–876.
- Gelfand, A. E., and Mukhopadhyay, S. (1995), "On Nonparametric Bayesian Inference for the Distribution of a Random Sample," *The Canadian Journal of Statistics*, 23, 411–420.
- Gelfand, A. E., and Smith, A. F. M. (1990), "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85, 398–409.
- Guglielmi, A., and Tweedie, R. (2000), "MCMC Estimation of the Law of the Mean of a Dirichlet Process," Research Report 2000-019, Division of Biostatistics, University of Minnesota.
- Ishwaran, H., and Zarepour, M. (2000), "Markov Chain Monte Carlo in Approximate Dirichlet and Beta Two-parameter Process Hierarchical Models," *Biometrika*, 87, 371–390.
- Johnson, N. L., Kotz, S., and Balakrishnan, N. (1995), *Continuous Univariate Distributions*, Vol. 2 (Second Edition), New York: John Wiley.
- Klein, J. P., and Moeschberger, M. L. (1997), *Survival Analysis. Techniques for Censored and Truncated Data*, New York: Springer-Verlag.
- Kottas, A., and Gelfand, A. E. (2001), "Bayesian Semiparametric Median Regression Modeling," *Journal of the American Statistical Association*, 96, 1458–1468.
- (in press), "Modeling Variability Order: A Semiparametric Bayesian Approach," *Methodology and Computing in Applied Probability*.
- Kuo, L., and Mallick, B. (1997), "Bayesian Semiparametric Inference for the Accelerated Failure-Time Model," *The Canadian Journal of Statistics*, 25, 457–472.
- Lo, A. Y. (1984), "On a Class of Bayesian Nonparametric Estimates: I. Density Estimates," *The Annals of Statistics*, 12, 351–357.
- MacEachern, S. N., and Müller, P. (1998), "Estimating Mixture of Dirichlet Process Models," *Journal of Computational and Graphical Statistics*, 7, 223–238.
- Muliere, P., and Tardella, L. (1998), "Approximating Distributions of Random Functionals of Ferguson-Dirichlet Priors," *The Canadian Journal of Statistics*, 26, 283–297.
- Neal, R. M. (2000), "Markov Chain Sampling Methods for Dirichlet Process Mixture Models," *Journal of Computational and Graphical Statistics*, 9, 249–265.
- Oakes, D. (1988), "Semi-parametric Models," in *Encyclopedia of Statistical Sciences* (Vol. 8), eds. S. Kotz and N. L. Johnson, New York: John Wiley, pp. 367–369.
- Richardson, S., and Green, P. J. (1997), "On Bayesian Analysis of Mixtures With an Unknown Number of Components" (with discussion), *Journal of the Royal Statistical Society, Ser. B*, 59, 731–792.
- Roeder, K. (1990), "Density Estimation With Confidence Sets Exemplified by Superclusters and Voids in the Galaxies," *Journal of the American Statistical Association*, 85, 617–624.
- Sethuraman, J. (1994), "A Constructive Definition of Dirichlet Priors," *Statistica Sinica*, 4, 639–650.
- Sethuraman, J., and Tiwari, R. C. (1982), "Convergence of Dirichlet Measures and the Interpretation of their Parameter," in *Statistical Decision Theory and Related Topics III*, eds. S. Gupta and J. O. Berger, New York: Springer-Verlag, pp. 305–315.
- Sinha, D., and Dey, D. K. (1997), "Semiparametric Bayesian Analysis of Survival Data," *Journal of the American*



- Statistical Association*, 92, 1195–1212.
- Snedecor, G. W., and Cochran, W. G. (1989), *Statistical Methods* (Eighth Edition), Ames, IA: Iowa State Press.
- Susarla, V., and Van Ryzin, J. (1976), “Nonparametric Bayesian Estimation of Survival Curves from Incomplete Observations,” *Journal of the American Statistical Association*, 71, 897–902.
- Titterton, D. M., Smith, A. F. M., and Makov, U. E. (1985), *Statistical Analysis of Finite Mixture Distributions*, Chichester: Wiley.
- Tomlinson, G., and Escobar, M. (1999), “Analysis of Densities,” Technical Report, University of Toronto.
- Walker, S. G., and Damien, P. (1998), “Sampling Methods for Bayesian Nonparametric Inference Involving Stochastic Processes,” in *Practical Nonparametric and Semiparametric Bayesian Statistics*, eds. D. Dey, P. Müller, and D. Sinha, New York: Springer-Verlag, pp. 243–254.
- Walker, S. G., Damien, P., Laud, P. W., and Smith, A. F. M. (1999), “Bayesian Nonparametric Inference for Random Distributions and Related Functions” (with discussion), *Journal of the Royal Statistical Society, Ser. B*, 61, 485–527.