

Bayesian nonparametric inference on stochastic ordering

DAVID B. DUNSON and SHYAMAL PEDDADA

Biostatistics Branch, MD A3-03

National Institute of Environmental Health Sciences

Research Triangle Park, NC 27709

dunson1@niehs.nih.gov

SUMMARY

This article considers Bayesian inference on collections of unknown distributions subject to a partial stochastic ordering. To address problems in testing of equalities between groups and estimation of group-specific distributions, we propose classes of restricted dependent Dirichlet process (rDDP) priors. These rDDP priors have full support in the space of stochastically ordered distributions, and can be used for collections of unknown mixture distributions to obtain a flexible class of rDDP mixture models. Theoretical properties are discussed, efficient methods are developed for posterior computation using MCMC, and the methods are illustrated using data from a study of DNA damage and repair.

Some key words: Dependent Dirichlet process; Hypothesis testing; Mixture model; Nonparametric Bayes; Order restriction.

1. INTRODUCTION

Our focus is inference on K group-specific distributions. For example, in toxicology studies, the groups may correspond to different doses of a potentially adverse chemical exposure. In such settings, it is of interest to assess whether the response distribution changes across groups, while also estimating the group-specific distributions. Although parametric assumptions may be difficult to justify, it is common to have prior knowledge that the magnitude of a response for a particular experimental unit would not decrease if that unit had been exposed to a higher dose. This implies stochastic ordering in the response distributions.

Focusing on the two sample case, Arjas and Gasbarra (1996) proposed nonparametric Bayes methods for ordered hazard functions, while Gelfand and Kottas (2001) induced priors on stochastic ordered distributions through products of independent Dirichlet process (DP) (Ferguson, 1973; 1974) components. More recently, Hoff (2003b) developed general methods for estimating probability measures constrained to belong to a convex set, considering applications to mean, mode, quantile and stochastic ordering constraints. The problem of Bayesian estimation of probability measures subject to a partial stochastic order was further considered by Hoff (2003a), relying on the theory in Hoff (2003b) to develop latent variable and rejection sampling methods for posterior computation.

These methods can be difficult to implement routinely, particularly for moderate to large K . In addition, interest often focuses not only on estimation of the group specific distributions but also on testing hypotheses of equalities between groups against stochastically ordered alternatives. To our knowledge, this hypothesis testing problem has not been addressed in the literature from a Bayesian perspective. In fact, there has been surprisingly little work on nonparametric Bayes hypothesis testing and model selection. Gutierrez-Pena and Walker (2005) imbed parametric model selection problems within a Bayesian nonparametric framework. Berger and Guglielmi (2001) test the fit of a parametric model through

comparison to a nonparametric alternative. Dass and Lee (2004) study consistency of Bayes factors for testing point nulls versus nonparametric alternatives. Basu and Chib (2003) develop methods for calculating Bayes factors for comparing Dirichlet process mixture models.

There has been recent interest in Bayesian methods for unconstrained collections of probability measures. Much of this work relies on extending the Ferguson (1973, 1974) Dirichlet process (DP). Suppose P is an unknown probability measure on $(\mathcal{X}, \mathcal{B})$, with $\mathcal{X} \subset \mathfrak{R}$ a Borel subset of Euclidean space and \mathcal{B} the Borel σ -algebra of subsets of \mathcal{X} . Then, $P \sim DP(\alpha P_0)$ denotes that P is assigned a DP prior with precision α and base measure P_0 . Sethuraman (1994) showed that this is equivalent to the stick-breaking representation:

$$P(\cdot) = \sum_{h=1}^{\infty} \pi_h \delta_{\Theta_h}(\cdot), \quad \pi_h = V_h \prod_{l < h} (1 - V_l), \quad (1)$$

where $V_h \stackrel{iid}{\sim} \text{beta}(1, \alpha)$, $\Theta_h \stackrel{iid}{\sim} P_0$, for $h = 1, \dots, \infty$. Here, δ_{Θ} is a probability measure concentrated at the atom Θ , $\mathbf{V} = \{V_h, h = 1, \dots, \infty\}$ is an infinite sequence of stick-breaking weights, and $\Theta = \{\Theta_h, h = 1, \dots, \infty\}$ is an infinite sequence of atoms, with \mathbf{V} and Θ mutually independent.

In order to generalize the DP to place a prior on a collection of probability measures (P_1, \dots, P_K) , MacEachern (1999, 2001) proposed a dependent DP (DDP). The DDP represents each P_k as in (1), but incorporates dependence through dependent stick-breaking weights and/or dependent atoms. De Iorio et al. (2004) used the DDP to create an ANOVA-type dependence structure for random measures, assuming fixed stick-breaking weights but dependent atoms. Gelfand et al. (2005) developed a related approach for spatially-dependent probability measures. More flexible methods that also allow the stick-breaking weights to vary, with some computational expense, have been proposed by Griffin and Steel (2006) and Duan et al. (2005). Alternative strategies for modeling of dependent collections of random measures through convex combinations of independent DPs have been proposed by Müller, Quintana and Rosner (2004), Pennell and Dunson (2006) and Dunson, Pillai and

Park (2006).

These methods are appealing in allowing flexible borrowing of information across groups in estimating unknown distributions. However, to our knowledge, there has been no consideration of the incorporation of stochastic ordering constraints in the DDP literature. One of our contributions is to demonstrate that the prior proposed by Hoff (2003a) for collections of stochastically ordered random measures can be equivalently specified as a DDP, leading to improvements in computational efficiency.

Another contribution of this article is the development of a framework for testing of equalities in distributions between groups against stochastically ordered alternatives. The current Bayes literature on nonparametric modeling of collections of distributions focuses almost entirely on estimation, and it is not clear how to use such methods to assess whether two groups have an equivalent (or essentially equivalent) response distribution. Recent work by Pennell and Dunson (2007) considers nonparametric Bayes testing of equalities in groups, but without incorporation of stochastic ordering constraints.

Section 2 proposes our formulation, relating our specification to the DDP and Hoff (2003a,b) methods. Section 3 considers properties and methods for hypothesis testing. Section 4 describes a simple and efficient Gibbs sampler for posterior computation. Section 5 applies the approach to simulated data, Section 6 considers a genotoxicology application, and Section 7 discusses the results.

2. STOCHASTICALLY ORDERED RANDOM PROBABILITY MEASURES

2.1. *Formulation and Background*

Let $(P_1, \dots, P_K) \in \mathcal{P}^K$, with \mathcal{P}^K the set of $K \times 1$ collections of probability measures on $(\mathcal{X}, \mathcal{B})$. In addition, define the following convex subset of \mathcal{P}^K :

$$C_E = \{(P_1, \dots, P_K) \in \mathcal{P}^K : P_i \preceq P_j \forall (i, j) \in E\},$$

where $E \subset (1, \dots, K)^2$ is a partial ordering. Here, $P_i \preceq P_j$ if $P_i(x, \infty) \leq P_j(x, \infty)$ for all x , so that P_j is stochastically larger than P_i . The collections of probability measures belonging to C_E satisfy the partial ordering defined by E .

As shown by Hoff (2003a,b), C_E is a weakly closed convex set with extreme points $\{(\delta_{s_1}, \dots, \delta_{s_K}) : \mathbf{s} \in \mathcal{S}_E\}$, where $\mathcal{S}_E = \{(s_1, \dots, s_K) \in \mathcal{X}^K : s_i \leq s_j \forall (i, j) \in E\}$. Using a corollary to Choquet's theorem, Hoff (2003a,b) shows that every $(P_1, \dots, P_K) \in C_E$ can be represented as a mixture over the extreme points. Hence, by placing an unconstrained prior on the mixing measure, one can obtain a prior for collections of stochastically ordered random measures. Let $\mathbf{x} = (x_1, \dots, x_n)'$ denote the observed data, with $x_i \stackrel{ind}{\sim} P_{a_i}$, where $a_i \in \{1, \dots, K\}$ is a group index. Then, Hoff (2003a) induces a prior on (P_1, \dots, P_K) with weak support on C_E by letting $x_i = s_{i, a_i}$, with $\mathbf{s}_i \sim Q$, where $Q \sim DP(\alpha Q_0)$ and Q_0 is a Borel probability measure on \mathcal{S}_E .

2.2 Restricted Dependent Dirichlet Process

Instead of inducing a prior for $(P_1, \dots, P_K) \in C_E$ indirectly through a prior for the mixing measure Q , we propose a direct constructive specification relying on a restricted dependent Dirichlet process (rDDP). In particular, we let

$$P_k(\cdot) = \sum_{h=1}^{\infty} \pi_h \delta_{\Theta_{hk}}(\cdot), \quad \pi_h = V_h \prod_{l < h} (1 - V_l), \quad k = 1, \dots, K, \quad (2)$$

where $V_h \stackrel{iid}{\sim} \text{beta}(1, \alpha)$, $h = 1, \dots, \infty$, are stick-breaking weights (as in expression 1), and $\Theta_h = (\Theta_{h1}, \dots, \Theta_{hK})' \stackrel{iid}{\sim} P_0$ are random atoms, with P_0 a Borel probability measure on \mathcal{S}_E . Here, $\mathbf{V} = \{V_h, h = 1, \dots, \infty\}$ and $\Theta = \{\Theta_h, h = 1, \dots, \infty\}$ are mutually independent sequences. Prior (2) is a type of DDP prior for the collection (P_1, \dots, P_K) . To highlight the incorporation of restrictions, we use the rDDP terminology.

Lemma 1. The prior for (P_1, \dots, P_K) in (2) has weak support on C_E .

The rDDP formulation has some advantages over the Hoff (2003a) specification. First, it is immediately apparent that P_k has a marginal DP prior: $P_k \sim DP(\alpha P_{0k})$, where P_{0k} is the k th marginal of P_0 . Hence, $E\{P_k(B)\} = P_{0k}(B)$, for any $B \in \mathcal{B}$, so that P_{0k} can be chosen as the prior expectation of the probability measure for group k . In order to incorporate a prior guess P_{0k} for P_k , Hoff (2003a) instead recommends an iterative procedure for finding a measure Q_0 with marginals P_{01}, \dots, P_{0K} . Also, by using the constructive representation, it is more straightforward to obtain insight into properties and to develop methods for hypothesis testing and posterior computation, as will be clear in §3 and §4.

THEOREM 1. *Let $x_i = s_{a_i}$, with $\mathbf{s}_i \sim Q = \sum_h \pi_h \delta_{\Theta_h}$ and $\{\pi_h, \Theta_h\}_{h=1}^\infty$ defined as in (2). Letting $\Pi_{kl} = Pr(\Theta_k = \Theta_l)$ denote the probability that atoms in groups k and l are equivalent under $\Theta = (\Theta_1, \dots, \Theta_K)' \sim P_0$,*

$$Pr(s_{ik} = s_{il}) \sim \text{beta}(\alpha \Pi_{kl}, \alpha(1 - \Pi_{kl})),$$

with $Pr(s_{ik} = s_{il}) = 0$ iff $\Pi_{kl} = 0$.

Theorem 1 implies that $Pr(s_{ik} = s_{il})$ is stochastically increasing with Π_{kl} for fixed α . The elements of \mathbf{s}_i can be viewed as potential outcomes had subject i been assigned to each of the K groups, with $x_i = s_{a_i}$ the only outcome that is observed directly.

As shown in Theorem 1, the prior forces strict inequalities: $s_{ik} < s_{il}$, for all $(k, l) \in E$, unless P_0 is chosen to assign probability mass to boundaries of S_E . As illustration, suppose $K = 2$, with $k = 1$ a control group, $k = 2$ an exposed group, and $P_1 \preceq P_2$. Then, s_{i1} is the potential response given randomization to the control group, s_{i2} is the potential response given randomization to the exposure group, $s_{i1} = s_{i2}$ if treatment had no effect, and $s_{i1} < s_{i2}$ if the treatment had an effect. Often, inferences on whether the treatment had an effect are of primary interest, so it is necessary to choose a prior which allows uncertainty in whether $s_{i1} = s_{i2}$ or $s_{i1} < s_{i2}$. For example, one could choose P_0 as the probability measure

corresponding to the following joint prior density:

$$f(\Theta_1, \Theta_2) = f_1(\Theta_1)\{\pi_0\delta_0(\Theta_2 - \Theta_1) + (1 - \pi_0)f_2(\Theta_2 - \Theta_1)\}, \quad (3)$$

where $\mathcal{X} = \mathfrak{R}$, $f_1(\cdot)$ is a density on \mathfrak{R} (e.g, Gaussian), $0 \leq \pi_0 \leq 1$ is the prior probability of $\Theta_1 = \Theta_2$, and $f_2(\cdot)$ is a density with support on \mathfrak{R}^+ (e.g., truncated Gaussian).

Previous Bayesian nonparametric analyses of stochastically ordered distributions have made strict constraints except when \mathcal{X} is discrete. In addition to not allowing uncertainty in group differences, strict constraints can lead to a tendency to over estimate group differences, particularly when the true difference is small, sample sizes are small to moderate, and the number of groups is moderate to large. By incorporating prior mass at the boundary, one obtains a shrinkage estimator of the density, which borrows information across groups.

2.3 Restricted DDP Mixtures

When \mathcal{X} is continuous it is not appropriate to assume $x_i \sim P_{a_i}$, as expression (2) implicitly assumes almost sure discreteness of each P_k . To avoid this assumption, we propose a class of rDDP mixtures. Following a related approach to Lo (1984), let $g_k(x) = \int_{\mathfrak{R}} \mathcal{K}(x, s)P_k(ds)$ represent the density function for group k , where $\mathcal{K}(x, s)$ is a non-negative bounded kernel defined on $(\mathcal{X} \times \mathfrak{R}, \mathcal{B} \otimes \mathcal{A})$ such that $\int_{\mathcal{X}} \mathcal{K}(x, s)dx = 1$ for all $s \in \mathfrak{R}$, with \mathcal{A} the Borel sets of \mathfrak{R} . For any $(P_1, \dots, P_K) \in C_E$, we obtain a corresponding collection of density functions $(g_1, \dots, g_K) \in \mathcal{L}_E$, with the choice of kernel an important determinate of the properties of the mapping from $C_E \rightarrow \mathcal{L}_E$ and the richness of \mathcal{L}_E .

Letting $G_k(B) = \int_B g_k(z)dz$, for all $B \in \mathcal{B}$ and $k = 1, \dots, K$, $(G_1, \dots, G_K) \in \mathcal{P}^K$ denotes the collection of probability measures on $(\mathcal{X}, \mathcal{B})$ corresponding to the density functions (g_1, \dots, g_K) . From proposition 2.17 of Hoff (thesis), the collection (G_1, \dots, G_K) satisfies the stochastic ordering constraint given by E if and only if there exists a K -variate measure Q^* with k th marginal equal to G_k , for $k = 1, \dots, K$, and $Q^*(s_k \leq s_l) = 1$ for all $(k, l) \in E$. This

motivates the following hierarchical model:

$$\begin{aligned}
x_i &= s_{i,a_i}, \quad i = 1, \dots, n \\
\mathbf{s}_i &= \mathbf{u}_i + \mathbf{1}_K \epsilon_i, \quad \epsilon_i \stackrel{iid}{\sim} \mathcal{K}(\cdot, 0), \quad i = 1, \dots, n \\
\mathbf{u}_i &\sim Q = \sum_{h=1}^{\infty} \pi_h \delta_{\Theta_h}, \quad i = 1, \dots, n,
\end{aligned} \tag{4}$$

where $\mathcal{K}(\cdot, \mu)$ is a non-degenerate density on \mathcal{X} centered at μ , $\mathbf{1}_K$ denotes the $K \times 1$ vector of ones, $\{\pi_h, \Theta_h\}_{h=1}^{\infty}$ are as defined previously, $Q \sim DP(\alpha P_0)$, and $\mathbf{s}_i \sim Q^*$, for $i = 1, \dots, n$, with Q^* the induced K -variate probability measure defining the joint distribution of \mathbf{s}_i .

As $Q(u_k \leq u_l) = 1$ and $s_k = u_k + e, s_l = u_l + e$, it is clear that $Q^*(s_k \leq s_l) = 1$ for all $(k, l) \in E$. In addition, $x_i \sim G_{a_i}, i = 1, \dots, n$, with G_k equal to the k th marginal of Q^* , for $k = 1, \dots, K$. Hence, $(G_1, \dots, G_K) \in C_E$. Note that Q has weak support in C_E and the mapping from C_E to \mathcal{L}_E is weakly continuous. In addition, \mathcal{L}_E is a convex set with extreme points:

$$\text{ex}\mathcal{L}_E = \left[\{K(\cdot, s_1), \dots, K(\cdot, s_K)\} : \mathbf{s} \in \mathcal{S}_E \right].$$

From Lo (1984), it is apparent that \mathcal{L}_E contains all $K \times 1$ collections of densities satisfying the partial ordering E in its closure.

Note that we can induce an equivalent prior to (4) on (G_1, \dots, G_K) as follows:

$$\begin{aligned}
x_i &\sim \mathcal{K}(\cdot, \mu_i), \\
\mu_i &\sim P_{a_i}, \quad i = 1, \dots, n,
\end{aligned} \tag{5}$$

where P_k is as defined in expression (2). This specification has considerable computational advantages over (4), as will be apparent in §4.

3. PROPERTIES AND HYPOTHESIS TESTING

3.1 Two Group Case

In further considering properties and addressing the hypothesis testing problem, we focus initially on the two group problem in which $K = 2$ and $E = \{(1, 2)\}$, so that $P_1 \preceq P_2$.

Suppose that P_0 is specified as in expression (3) and $x_i \stackrel{iid}{\sim} P_{a_i}$, with $a_i = 1$ for $i = 1, \dots, n_1$ and $a_i = 2$ for $i = n_1 + 1, \dots, n$. Updating prior (2) with the data from group one, $\mathbf{x}_1 = (x_1, \dots, x_{n_1})$, we obtain

$$(P_1 | \mathbf{x}_1) \sim DP\left(\alpha P_{01} + \sum_{i=1}^{n_1} \delta_{x_i}\right),$$

where $P_{01}(B) = \int_B f_1(\Theta_1) d\Theta_1 \forall B \in \mathcal{B}$. This updated DP can be expressed as

$$P_1(\cdot) = \sum_{h=1}^{\infty} V_h \prod_{l < h} (1 - V_l) \delta_{\Theta_{h1}}(\cdot),$$

with $V_h \stackrel{iid}{\sim} \text{beta}(1, \alpha + n_1)$, $\Theta_{h1} \stackrel{iid}{\sim} P_{01}^{(1)}$, where

$$P_{01}^{(1)}(\cdot) = \left(\frac{\alpha}{\alpha + n_1}\right) P_{01}(\cdot) + \left(\frac{1}{\alpha + n_1}\right) \sum_{i=1}^{n_1} \delta_{x_i}(\cdot).$$

Conditionally on P_1 and \mathbf{x}_1 , the prior for P_2 is as follows:

$$P_2(\cdot) = \sum_{h=1}^{\infty} V_h \prod_{l < h} (1 - V_l) \delta_{\Theta_{h1} + \beta_h}(\cdot),$$

with $\mathbf{V} = (V_h, h = 1, \dots, \infty)'$, $\Theta_1 = (\Theta_{h1}, h = 1, \dots, \infty)'$ fixed in conditioning on P_1 , and $\beta_h = \Theta_{h2} - \Theta_{h1} \sim \pi_0 \delta_0 + (1 - \pi_0) f_2$. Hence, in the limit as $\pi_0 \rightarrow 1$, $P_2 \equiv P_1$.

To formally assess how *close* P_2 is to P_1 , we consider the distance metric:

$$d_{12} = \max_{x \in \mathcal{X}} |P_2(x, \infty) - P_1(x, \infty)|. \quad (6)$$

Under the rDDP prior, we obtain the simplified expression:

$$d_{12} = \max_{x \in \mathcal{X}} \left\{ \sum_{h=1}^{\infty} \pi_h 1(\Theta_{h1} \leq x, \Theta_{h2} > x) \right\} = \sum_{h=1}^{\infty} \pi_h 1(\beta_h > 0).$$

Letting $\gamma_h = 1(\beta_h > 0)$ and relying on Lemma 2, we have

$$d_{12} = \sum_{h=1}^{\infty} \pi_h \gamma_h \sim \text{beta}(\alpha(1 - \pi_0), \alpha\pi_0), \quad (7)$$

so that the distance between P_1 and P_2 stochastically increases with $1 - \pi_0$. In addition, $E(d_{12}) = 1 - \pi_0$, with α controlling uncertainty in this prior expectation.

In many applications, the primary focus is on comparing the null hypothesis that P_1 and P_2 are *close* in some sense to an alternative of stochastic ordering. Using the distance metric in (6), these null and alternative hypotheses can be formalized as follows:

$$H_0 : d_{12} \leq \epsilon \quad \text{and} \quad H_1 : d_{12} > \epsilon, \quad (8)$$

where $\epsilon > 0$ is a small positive constant. In the limit as $\pi_0 \rightarrow 0$, $d_{12} \xrightarrow{\mathcal{D}} \delta_1$, so that $\Pr(d_{12} < \epsilon) \rightarrow 0$, for any $\epsilon < 1$. In the general case, the prior probability of H_0 is

$$\Pi_0 = \Pr(H_0) = \Pr(d_{12} < \epsilon) = \frac{1}{B(\alpha(1 - \pi_0), \alpha\pi_0)} \int_0^\epsilon t^{\alpha(1-\pi_0)} (1-t)^{\alpha\pi_0-1} dt, \quad (9)$$

where $B(a, b)$ is the beta function.

Let $\mathcal{I}_1 = \{1, \dots, n_1\}$, $\mathcal{I}_2 = \{n_1 + 1, \dots, n\}$, and $\mathcal{I}_{2(-i)} = \mathcal{I}_2 \setminus \{i\}$. Then, using a generalization of the Blackwell and MacQueen (1973) Pólya urn scheme, we can obtain the conditional distribution of x_i , $i \in \mathcal{I}_2$, given $\mathbf{x}_{2(-i)}$ and \mathbf{x}_1 as follows:

$$(x_i | \mathbf{x}_{2(-i)}, \mathbf{x}_1) \sim \left(\frac{\alpha}{\alpha + n - 1} \right) f_2^* + \sum_{h=1}^{n_1} \left\{ \left(\frac{\pi_0}{\alpha + n - 1} \right) \delta_{x_h} + \left(\frac{1 - \pi_0}{\alpha + n - 1} \right) f_2(x_h) \right\} + \sum_{h \in \mathcal{I}_{2(-i)}} \left(\frac{1}{\alpha + n - 1} \right) \delta_{x_h}, \quad (10)$$

where $f_2^* = f_1 \oplus f_2$ is the convolution of f_1 and f_2 , and $f_2(x_h) = \delta_{x_h} \oplus f_2$. Here, $\Pr(x_i = x_h) = \pi_0/(\alpha + 1)$, for $h \in \mathcal{I}_1$, is the probability that randomly selected subjects in groups 1 and 2 have equivalent values. Hence, π_0 controls the probability of ties in the two groups.

From (7), it is clear that dependence in the random measures, P_1 and P_2 , is controlled by π_0 , though one can also induce dependence by favoring values close to zero in f_2 . Dependence arises through the incorporation of identical atoms in P_1 and P_2 , which further leads to ties in the data for groups 1 and 2. Hence, if the data actually do not contain ties (i.e., $\mathbf{x}_1 \cap \mathbf{x}_2 = \emptyset$), the posterior will tend to concentrate on values of d_{12} in a neighborhood of 1 as the sample size increases. This behavior leads to posterior inconsistency, with $\lim_{n_1, n_2 \rightarrow \infty} \Pr(H_0 | \mathbf{x}_1, \mathbf{x}_2) \rightarrow 0$ when the true distributions in groups 1 and 2 are identical but continuous.

When the sampling distributions are continuous, it is more appropriate to use an rDDP mixture as in expression (5). Then, hypotheses of near equivalence in the two groups against alternatives of stochastic ordering can be formalized as in expression (8), with P_1 and P_2 as the random mixing measures. Theorem 2 provides justification for using hypotheses on the mixing measures as a basis for inferences on G_1 and G_2 .

Theorem 2. Specify H_0 and H_1 as in (8) and let $G_k(B) = \int_B g_k(x)dx$, $k = 1, 2$, with $g_k(x) = \int \mathcal{K}(x, s)dP_k(s)$. Then, H_0 implies

$$\max_{x \in \mathcal{X}} \left| G_2(x, \infty) - G_1(x, \infty) \right| < \epsilon.$$

3.2 Multiple Group Case

The approach in §3.1 is straightforward to generalize to the multiple group setting. When there are K groups subject to partial stochastic ordering E , it is first necessary to specify P_0 . For most orderings of interest in applications, one can express the h th atom in group k as $\Theta_{hk} = \mathbf{w}'_k \boldsymbol{\beta}_h^*$, where \mathbf{w}_k is a fixed $K \times 1$ vector of 0s and 1s, and $\boldsymbol{\beta}_h^* = (\beta_{h1}^*, \dots, \beta_{hK}^*)'$ is a parameter vector having a subset of elements constrained to be non-negative or non-positive.

For example, when there is a simple stochastic ordering: $P_1 \preceq P_2 \preceq \dots \preceq P_K$, we let $\mathbf{w}_k = (\mathbf{1}'_k, \mathbf{0}'_{K-k})'$, which implies that $\beta_{hk}^* = \Theta_{hk} - \Theta_{h,k-1}$, for $h = 2, \dots, K - 1$. Hence, constraining $\beta_{hk}^* \geq 0$, for $h = 2, \dots, K$, induces a P_0 with appropriate support. Generalizing prior (3), we let

$$f(\boldsymbol{\beta}_h^*) = f_1(\beta_{h1}^*) \left\{ \prod_{k=2}^K \pi_{0k} \delta_0(\beta_{hk}^*) + (1 - \pi_{0k}) f_k(\beta_{hk}^*) \right\}, \quad (11)$$

where $\pi_{0k} = \Pr(\Theta_{hk} = \Theta_{h,k-1})$ and f_k is a density with support on \mathfrak{R}^+ , for $k = 2, \dots, K$. To modify the prior to accommodate tree stochastic order: $P_1 \preceq P_k$, $k = 2, \dots, K$, we can use the same P_0 but with $\mathbf{w}_k = (1, \mathbf{0}_{k-2}, 1, \mathbf{0}_{K-k})'$. Umbrella orderings and cases involving multiple factors (refer to §5) can also be accommodated by changing $\mathbf{W} = (\mathbf{w}_1, \dots, \mathbf{w}_K)'$.

For ease in exposition, we focus our discussion of multiple group hypothesis testing on the simple ordering case, though modifications to other cases are automatic. Generalizing the hypotheses in (8), we let

$$H_{0k} : d_{k,k+1} \leq \epsilon \quad \text{and} \quad H_{1k} : d_{k,k+1} > \epsilon, \quad (12)$$

for $k = 1, \dots, K - 1$, with $d_{k,k+1} = \max_{x \in \mathcal{X}} |P_{k+1}(x, \infty) - P_k(x, \infty)|$. We refer to H_{0k} as the *local* null hypothesis of near equivalence in groups k and $k + 1$. Under prior (11), expression (9) and Theorem 2 generalize automatically to the multiple group case. One can also consider the *global* null hypothesis: $H_0 : \bigcap_{k=1}^{K-1} H_{0k}$ of near equivalence in all K groups. Under the assumption of mutual independence of the local hypotheses, the prior probability of H_0 is

$$\begin{aligned} \Pr(H_0) &= \Pr(d_{k,k+1} < \epsilon, k = 1, \dots, K - 1) \\ &= \prod_{k=1}^{K-1} \Pr(H_{0k}) = \prod_{k=2}^K \frac{1}{B(\alpha(1 - \pi_{0k}), \alpha\pi_{0k})} \int_0^\epsilon t^{\alpha(1 - \pi_{0k})} (1 - t)^{\alpha\pi_{0k} - 1} dt. \end{aligned} \quad (13)$$

Using the Gibbs sampler of §4, we can calculate local and global hypothesis probabilities from a single chain, while also obtaining model-averaged group-specific density estimates.

4. POSTERIOR COMPUTATION

4.1 Model and Background

In describing algorithms for posterior computation, we focus on the following case:

$$\begin{aligned} x_i &\sim N(\mathbf{w}'_{a_i} \boldsymbol{\beta}_i, \tau^{-1}) \\ \boldsymbol{\beta}_i &\sim P = \sum_{h=1}^{\infty} V_h \prod_{l < h} (1 - V_l) \delta_{\boldsymbol{\beta}_h^*}, \quad V_h \stackrel{iid}{\sim} \text{beta}(1, \alpha), \quad \boldsymbol{\beta}_h^* \stackrel{iid}{\sim} P_0, \end{aligned} \quad (14)$$

where $a_i = k$ for $i \in \{n_{k-1} + 1, \dots, n_{k-1} + n_k\}$, $k = 1, \dots, K$, with $n_0 = 0$ and n_k the number of subjects in group k , \mathbf{w}_k chosen as discussed in §3.2, and $\boldsymbol{\beta}_i = (\beta_{i1}, \dots, \beta_{iK})'$. Note that this expression is a special case of (5) with \mathcal{K} corresponding to a normal kernel and with $\mu_i = \mathbf{w}'_{a_i} \boldsymbol{\beta}_i$. Assuming prior (11), we let $f_1(\boldsymbol{\beta}_{h1}^*) = N(\boldsymbol{\beta}_{h1}^*; \mu_0, \sigma_0^2)$ and $f_k(\boldsymbol{\beta}_{hk}^*) = N_+(\boldsymbol{\beta}_{hk}^*; 0, \kappa^{-1})$, for $k = 2, \dots, K$, with $N_+(\cdot)$ denoting the normal distribution truncated to be positive.

Due to the reparameterization used in expression (14), we effectively have a typical Dirichlet process mixture of normal linear regression models, with a constrained mixture structure used in the base measure, P_0 . Hence, we can potentially use standard MCMC algorithms for posterior computation in DP mixture models. Because the structure of the base measure creates some difficulties in implementing Pólya urn-based algorithms, we focus on a blocked Gibbs sampler (Ishwaran and James, 2001). This algorithm is based on updating the random weights and atoms in a truncation approximation to the infinite stick-breaking representation.

In particular, let $P = \sum_{h=1}^T V_h \prod_{l < h} (1 - V_l) \delta_{\beta_k^*}$, with the components defined as in (14) but with $V_N = 1$ so that the $N + 1, \dots, \infty$ terms can be excluded. Following the approach of Ishwaran and James (2001), one can show that this approximation tends to be accurate for moderate N (e.g., $N = 20$), particularly if $\alpha \leq 1$. As such values of α favor a small number of mixture components and it is well known that a modest number of normals can approximate any density accurately, it seems justified to use $N = 20$ as a default in most applications.

4.2 *Prior Specification and Posterior Inference*

Hyperparameter elicitation is an important component of the analysis. Here, we recommend a default specification. To avoid sensitivity to the measurement scale, standardize the data by subtracting the group one mean and dividing by the group one standard deviation. Then, $\mu_0 = 0$ and $\sigma_0^2 = 1$ are chosen for the mean and variance of the group one atoms. To assign high probability to a wide range of mild to moderate shifts in the response density, let $\kappa \sim \text{gamma}(1/2, 1/2)$ to induce a Cauchy prior on β_{hk}^* , for $k > 1$. The Cauchy is often used as a robust default in parametric model selection. A diffuse prior can be chosen for the error precision, τ^{-1} , which is common to all the models.

There is substantial information in the data about α , so a diffuse prior can potentially

be used, but we recommend letting $a_\alpha = 1, b_\alpha = K - 1$ to favor a small to moderate number of clusters. Letting $\epsilon = 0.05$ and fixing $\alpha = 1$, one can assign $\Pr(H_{0k}) = \Pr(H_{1k}) = 1/2$ by choosing $\pi_{0k} = 0.792$, the solution to the equation: $0.5 = \int_0^{0.05} \text{beta}(\pi; 1 - \pi_{0k}, \pi_{0k}) d\pi$. Letting $a_\pi + b_\pi = 1$ to correspond to a unit information prior, we obtain a hyperprior for π_{0k} with mean 0.792 by letting $a_\pi = 0.792, b_\pi = 0.208$.

With the prior specification complete, the Gibbs sampler described in the Appendix can be run to obtain draws from the posterior distribution. These draws can be used to obtain group-specific density estimates and posterior hypothesis probabilities. The posterior distribution of the distance between any two groups can also be obtained.

5. SIMULATION EXAMPLES

We considered two simulation cases. In both cases, $K = 2, n_1 = n_2 = 100$ and data in group 1 were simulated from the following mixture of three normals:

$$f(y) = 0.2N(y; -2.5, \tau^{-1}) + 0.7N(y; 0, \tau^{-1}) + 0.1N(y; 1.5, \tau^{-1}),$$

with $\tau = 3$. In case 1 data in group 2 were simulated from this same density, while in case 2 we increased the component-specific means slightly from $(-2.5, 0, 1.5)$ to $(-2.4, 0.4, 2.2)$.

For each case, we simulated 100 data sets, analyzing each using the Gibbs sampler of §4, with 2000 iterations collected after a 500 iteration burn-in. Apparent convergence was rapid and mixing was good. The entire set of simulations was run using Matlab on a Mac PowerBook G4 laptop in approximately 10 hours.

Figure 1 plots the group-specific Bayesian density estimates for each of the 100 data sets simulated in case 1. The estimates are distributed about the true value for each group, with no evidence of bias. In addition, 95% pointwise credible intervals obtained in the individual simulations provided a good measure of uncertainty in the estimates. We also obtained posterior distributions for the d_{12} distance measure in each of the simulations. The average

value across the simulations was 0.044, and the empirical 95% interval was [0.003, 0.215]. In addition, the posterior probability of H_0 (using $\epsilon = 0.05$) had an average value of 0.849 and a 95% interval of [0.300, 0.990], with only 3% of the simulations showing even weak evidence against H_0 .

Figure 2 plots the group-specific Bayesian density estimates for the simulated data sets in case 2. Again, the estimates are distributed about the true value. The average posterior mean of the d_{12} measure was 0.560 and the 95% empirical interval was [0.091, 0.884]. Figure 3 shows histograms of the posterior mean distances in each of the simulations, illustrating that the distances tend to be close to zero in the null simulations but not under the alternative. In case 2, the posterior probability of H_1 had an average value of 0.853 and there was evidence against the null in 89% of the simulations, with 63% having a posterior probability of H_1 greater than 90% and 51% greater than 95%.

We purposely chose a simulation case in which both the sample size and the magnitude of the difference between the groups was modest. The results are very encouraging, suggesting minimal bias even in moderate samples and good performance in testing, though these results are preliminary.

6. GENOTOXICOLOGY APPLICATION

6.1 *Data Structure and Scientific Problem*

We applied the approach to data from a study of DNA damage and repair. Batches of cells were exposed to 0, 5, 20, 50 or 100 micromoles H_2O_2 and DNA damage was then measured in individual cells after allowing a repair time of 0, 60 or 90 minutes. Letting $i = 1, \dots, n$ index the cells under study, the measured response for cell i , x_i , was the Olive tail moment, which is a surrogate of the frequency of DNA strand breaks obtained using the comet assay.

The goal of the study is to assess the sensitivity of the comet assay to detecting damage induced by the known genotoxic agent H_2O_2 , while also investigating how rapidly damage is

repaired. Let $a_i \in \{1, \dots, K\}$ be a group index denoting the level of H_2O_2 and repair time for cell i . The following table shows the value of a_i for each dose \times repair time value:

Repair time	Dose of H_2O_2				
	0	5	20	50	100
0	1	2	3	4	5
60	6	7	8	9	10
90	11	12	13	14	15

The total sample size is 1400, with 100 cells per group except for groups 9 and 13 which had 50.

DNA damage should be non-decreasing with dose of H_2O_2 and non-increasing with repair time. Hence, we make the ordering assumption: $G_1 \preceq G_2 \preceq G_3 \preceq G_4 \preceq G_5$, $G_{11} \preceq G_6 \preceq G_1$, $G_{12} \preceq G_7 \preceq G_2$, $G_{13} \preceq G_8 \preceq G_3$, $G_{14} \preceq G_9 \preceq G_4$, $G_{15} \preceq G_{10} \preceq G_5$. However, as we wish to assess the sensitivity of the comet assay and to investigate whether damage is significantly reduced across each increment of the repair time, it is important to avoid strict constraints.

6.2 Analysis and Results

Previous authors analyzed the data in groups 1-5 using a dynamic mixture of Dirichlet processes (DMDP) (Dunson, 2006; Pennell and Dunson, 2006), which allows for dependence in the distributions within adjacent dose groups but does not enforce stochastic ordering restrictions. In addition, the DMDP does not allow the incorporation of both dose of H_2O_2 and repair time, though extensions are possible.

We implemented the approach described in §4. The Gibbs sampler was run for 20,000 iterations after a 1,000 iteration burn-in. As in the simulation study, the chain appeared to converge rapidly and mix efficiently based on standard diagnostics.

Figure 4 plots group-specific Bayesian density estimates and 95% pointwise credible intervals, along with frequentist kernel density estimates obtained using only the data in a given group. Although the Bayes estimates have been constrained to follow a stochastic ordering, there is no evidence of systematic deviations from the frequentist estimates. Table

1 provides posterior probabilities of local null hypotheses for different group comparisons. The results suggest highly significant increases in DNA damage between the 0, 5 and 20 micromole H_2O_2 dose groups given a repair time of 0 minutes, with the evidence of further increases at higher doses less clear. As expected, there is no evidence of a change in the distribution between groups 1, 6 and 11, since there was no induced damage to repair. However, there were highly significant decreases in DNA damage in each of the exposed groups after a repair time of 60 minutes. Allowing an additional 30 minutes of repair did not significantly alter the distribution.

These results are consistent with the raw data and plots shown in Figure 4, and are both scientifically reasonable and interesting. It is known that the type of damage induced by hydrogen peroxide can be repaired quickly by base excision and repair mechanisms. The result that there is no further improvement after 60 minutes suggests that it may be unnecessary to collect data for repair times exceeding an hour in future molecular epidemiology studies using the comet assay to identify genotypes predictive of DNA repair rates.

To assess sensitivity, we repeated analyses using a variety of alternative hyperprior settings. In particular, we tried (i) $a_\pi + b_\pi = 5$ instead of 1 to correspond to a more informative prior; (ii) $a_\alpha = 1, b_\alpha = 1$ to favor more clusters; (iii) $\kappa = 1$; and (iv) $\kappa = 2$. The estimated densities did not change noticeably across these analyses and the conclusions were robust.

7. DISCUSSION

This article has proposed a general class of Bayesian nonparametric methods for inference on distributions subject to a partial stochastic ordering. Although inspired by pioneering work by Peter Hoff, this article make several important contributions. First, we develop general methods for stochastically ordered mixtures, which can be applied to nonparametrically estimate densities for multiple groups subject to a stochastic ordering constraint. In addition to density estimation, we also develop methods for hypothesis testing of near equalities

between groups against stochastically ordered alternatives. Using a simple and efficient Gibbs sampling algorithm, the methods can be implemented easily, estimating posterior hypothesis probabilities and group-specific densities from a single chain.

Although we have focused on a relatively simple setting, it is straightforward to imbed our prior for stochastically ordered mixture distributions within a larger hierarchical model. For example, one can incorporate a parametric adjustment for covariates within the kernel. In addition, one can allow for stochastically ordered latent variable distributions. Ongoing work focuses on the extension to incorporate continuous predictors, which can conceptually be accomplished by replacing the atoms with non-decreasing stochastic processes.

ACKNOWLEDGEMENTS

APPENDIX

Proof of lemma 1

First note that expression (2) implies the following sampling model:

$$\begin{aligned} x_i &= s_{i,a_i}, \quad i = 1, \dots, n, \\ \mathbf{s}_i &\stackrel{iid}{\sim} Q = \sum_{h=1}^{\infty} \pi_h \delta_{\Theta_h}, \quad i = 1, \dots, n. \end{aligned} \tag{15}$$

which implies that $Q \sim DP(\alpha P_0)$. Hence, expression (2) induces the same prior on (P_1, \dots, P_K) as the prior of Hoff (2003a), so that Lemma 1 follows directly from the result that his prior has full support on C_E with respect to the weak convergence topology.

Proof of theorem 1

We rely on the following lemma:

Lemma 2. Suppose $\{V_h, h = 1, \dots, \infty\}$ and $\{\gamma_h, h = 1, \dots, \infty\}$ are mutually independent random sequences, with $V_h \stackrel{iid}{\sim} \text{beta}(1, \alpha)$ and $\gamma_h \stackrel{iid}{\sim} \text{Bernoulli}(\pi_\gamma)$,

$h = 1, \dots, \infty$, then:

$$\sum_{h=1}^{\infty} V_h \prod_{l < h} (1 - V_l) \gamma_h \sim \text{beta}(\alpha \pi_\gamma, \alpha(1 - \pi_\gamma)).$$

To prove Lemma 2, first note that $G \sim DP(\alpha G_0)$ implies that

$$G(B) = \sum_{h=1}^{\infty} V_h \prod_{l < h} (1 - V_l) 1(\Theta_h \in B) \sim \text{beta}(\alpha G_0(B), \alpha \{1 - G_0(B)\}), \quad \forall B \in \mathcal{B},$$

where $V_h \stackrel{iid}{\sim} \text{beta}(1, \alpha)$, $\Theta_h \stackrel{iid}{\sim} G_0$, $h = 1, \dots, \infty$. In addition, $1(\Theta_h \in B) \stackrel{iid}{\sim} \text{Bernoulli}(G_0(B))$.

Theorem 1 follows from Lemma 2 after noting that, under expression (15), we have $s_{ik} = s_{il}$ with probability

$$\Pr(s_{ik} = s_{il} \mid \mathbf{V}, \Theta) = \sum_{h=1}^{\infty} \pi_h 1(\Theta_{hk} = \Theta_{hl}),$$

which is the probability that individual i is drawn from a component having identical atoms for groups k and l .

Proof of theorem 2

The conditions in Theorem 2 imply that

$$G_1(x, \infty) = \sum_{h=1}^{\infty} \pi_h \mathcal{K}^*(x, \Theta_{h1}) \quad \text{and} \quad G_2(x, \infty) = \sum_{h=1}^{\infty} \pi_h \mathcal{K}^*(x, \Theta_{h1} + \beta_h), \quad \forall x \in \mathcal{X},$$

where $\mathcal{K}^*(x, \Theta) = \int_x^\infty \mathcal{K}(z, \Theta) dz$. Letting $\mathcal{H}_0 = \{h : \beta_h = 0, h = 1, 2, \dots, \infty\}$ and $\bar{\mathcal{H}}_0 = \{1, 2, \dots, \infty\} \setminus \mathcal{H}_0$, we have

$$G_2(x, \infty) - G_1(x, \infty) = \sum_{h \in \bar{\mathcal{H}}_0} \pi_h \{\mathcal{K}^*(x, \Theta_{h1} + \beta_h) - \mathcal{K}^*(x, \Theta_{h1})\}.$$

Noting that $\sum_{h \in \bar{\mathcal{H}}_0} \pi_h < \epsilon$ under H_0 and, for any fixed x , $0 \leq \mathcal{K}^*(x, \Theta) \leq 1$ is a monotone increasing function of Θ , Theorem 2 follows directly.

Gibbs Sampling Steps

Let $\zeta_i \in \{1, \dots, n\}$ denote that individual i is sampled from component h , so that $\beta_i = \beta_h^*$.

Then, the blocked Gibbs sampler proceeds through the following steps:

1. Update ζ_i , for $i = 1, \dots, n$, by sampling from the multinomial conditional with

$$\Pr(\zeta_i = h) = \frac{\pi_h \mathbf{N}(x_i; \mathbf{w}'_{a_i} \boldsymbol{\beta}_h^*, \tau^{-1})}{\sum_{l=1}^T \pi_l \mathbf{N}(x_i; \mathbf{w}'_{a_i} \boldsymbol{\beta}_l^*, \tau^{-1})}, \quad h = 1, \dots, T.$$

2. Assuming $\tau \sim \text{gamma}(a_\tau, b_\tau)$, update τ by sampling from the full conditional:

$$\text{gamma}\left(a_\tau + \frac{n}{2}, b_\tau + \frac{1}{2} \sum_{i=1}^n (x_i - \mathbf{w}'_{a_i} \boldsymbol{\beta}_i)^2\right).$$

with $\text{gamma}(a, b)$ parameterized to have mean a/b and variance a/b^2 .

3. Update V_h , for $h = 1, \dots, T-1$, by sampling from the full conditional distribution:

$$\text{beta}\left(1 + \sum_{i=1}^n 1(\zeta_i = h), \alpha + \sum_{i=1}^n 1(\zeta_i > h)\right).$$

4. Update $\boldsymbol{\beta}_h^*$, for $h = 1, \dots, T$, via the following Gibbs sub-steps:

- (a) Update $\boldsymbol{\beta}_{h1}^*$ from $\mathbf{N}(E_{h1}, V_{h1})$, where

$$E_{h1} = V_{h1} \left(\sigma_0^{-2} \mu_0 + \tau \sum_{i:\zeta_i=h} \left(x_i - \sum_{k=2}^K w_{a_i,k} \beta_{ik} \right) \right), \quad V_{h1} = \left(\sigma_0^{-2} + \tau \sum_{i=1}^n 1(\zeta_i = h) \right)^{-1}$$

- (b) Update $\boldsymbol{\beta}_{hk}^*$, for $h = 2, \dots, K$, from the full conditional: $\hat{\pi}_{hk} \delta_0 + (1 - \hat{\pi}_{hk}) \mathbf{N}_+(E_{hk}, V_{hk})$,

where the conditional probability of $\beta_{hk}^* = 0$ is

$$\hat{\pi}_{hk} = \frac{\pi_{0k}}{\pi_{0k} + (1 - \pi_{0k}) \frac{\sqrt{2\kappa} \int_0^\infty \mathbf{N}(z; E_{hk}, V_{hk}) dz}{\sqrt{\pi} \mathbf{N}(0; E_{hk}, V_{hk})}},$$

and the mean and variance in the normal component are, respectively:

$$E_{hk} = V_{hk} \left(\tau \sum_{i:\zeta_i=h} \left(x_i - \sum_{l \neq k} w_{a_i,l} \beta_{il} \right) \right), \quad V_{hk} = \left(\kappa + \tau \sum_{i:\zeta_i=h} w_{a_i,k}^2 \right)^{-1}.$$

5. Assuming $\alpha \sim \text{gamma}(a_\alpha, b_\alpha)$, update α from its full conditional distribution:

$$\text{gamma}\left(a_\alpha + T - 1, b_\alpha - \sum_{h=1}^{T-1} \log(1 - V_h)\right).$$

6. Assuming $\pi_{0k} \sim \text{beta}(a_\pi, b_\pi)$, for $k = 2, \dots, K$, update π_{0k} from its full conditional distribution:

$$\text{beta}\left(a_\pi + \sum_{h=1}^T 1(\beta_{hk}^* = 0), b_\pi + \sum_{h=1}^T 1(\beta_{hk}^* > 0)\right).$$

7. Assuming $\kappa \sim \text{gamma}(a_\kappa, b_\kappa)$, update κ from its full conditional distribution:

$$\text{gamma}\left(a_\kappa + \frac{1}{2} \sum_{h=1}^T \sum_{k=2}^K 1(\beta_{hk}^* > 0), b_\kappa + \sum_{h=1}^T \sum_{k=2}^K (\beta_{hk}^*)^2\right).$$

REFERENCES

- ARJAS, E. & GASBARRA, D. (1996). Bayesian inference of survival probabilities, under stochastic ordering constraints. *J. Am. Statist. Assoc.* **91**, 1101-9.
- BASU, S. & CHIB, S. (2003). Marginal likelihood and Bayes factors for Dirichlet process mixture models. *J. Am. Statist. Assoc.* **98**, 224-35.
- BERGER, J.O. & GUGLIELMI, A. (2001). Bayesian and conditional frequentist testing of a parametric model versus nonparametric alternatives. *J. Am. Statist. Assoc.* **96**, 174-81.
- DASS, S.C. & LEE, J. (2004). A note on the consistency of Bayes factors for testing point null versus non-parametric alternatives. *J. Statist. Plan. Inf.* **119**, 143-52.
- DE IORIO, M., MÜLLER, P., ROSNER, G.L. & MACEACHERN, S.N. (2004). An ANOVA model for dependent random measures. *J. Am. Statist. Assoc.* **99**, 205-15.
- DUAN, J.A., GUIDANI, M. & GELFAND, A.E. (2005). Generalized spatial Dirichlet process models. *ISDS Discussion Paper 2005-23*, Duke University, Durham, NC.
- DUNSON, D.B., PILLAI, N. & PARK, J-H. (2006). Bayesian density regression. *J. Royal Statist. Soc. B*, in press.

- GELFAND, A.E. & KOTTAS, A. (2001). Nonparametric Bayesian modeling for stochastic order. *Ann. Inst. Statist. Math.* **53**, 865-76.
- GELFAND, A.E., KOTTAS, A., & MACEACHERN, S.N. (2005). Bayesian nonparametric spatial modeling with Dirichlet process mixing. *J. Am. Statist. Assoc.* **100**, 1021-35.
- GRIFFIN, J.E. & STEEL, M.F.J. (2006). Order-based dependent Dirichlet processes. *J. Am. Statist. Assoc.* **101**, 179-94.
- GUTIERREZ-PENA, E. & WALKER, S.G. (2005). Statistical decision problems and Bayesian nonparametric methods. *Intern. Statist. Rev.* **73**, 309-30.
- HANSEN, M.B. & LAURITZEN, S.L. (2002). Nonparametric Bayes inference for concave distribution functions. *Statist. Neerland.* **56**, 110-27.
- HOFF, P.D. (2000). Constrained maximum likelihood estimation via mixtures. *J. Comp. & Graph. Statist.* **9**, 633-41.
- HOFF, P.D. (2003a). Bayesian methods for partial stochastic orderings. *Biometrika*, **90**, 303-17.
- HOFF, P.D. (2003b). Nonparametric estimation of convex models via mixtures. *Ann. Statist.* **31**, 174-200.
- ISHWARAN, H. & JAMES, L.F. (2001). Gibbs sampling methods for stick-breaking priors. *J. Am. Statist. Assoc.* **96**, 161-73.
- LO, A.Y. (1984). On a class of Bayesian nonparametric estimates: I. Density estimates. *Ann. Statist.* **12**, 351-7.
- MACEACHERN, S.N. (1999). Dependent nonparametric processes. In *ASA Proceedings of the Section of Bayesian Statistical Science*, Alexandria, VA: American Statistical Association.

- MACEACHERN, S.N. (2001). Decision theoretic aspects of dependent nonparametric processes. In *Bayesian Methods with Applications to Science, Policy and Official Statistics*, ed. E. George,. Creta: ISBA, 551-560.
- MÜLLER, P., QUINTANA, F. & ROSNER, G. (2004). A method for combining inference across related nonparametric Bayesian models. *J. Royal Statist. Soc. B.* **66**, 735-49.
- PENNELL, M.L. & DUNSON, D.B. (2006). Bayesian semiparametric dynamic frailty models for multiple event time data. *Biometrics* **62**, 1044-52.
- PENNELL, M.L. & DUNSON, D.B. (2007). Nonparametric Bayes testing of changes in a response distribution with an ordinal predictor. *Biometrics*, revision submitted.
- SETHURAMAN, J. (1994). A constructive definition of Dirichlet priors. *Statist. Sinica* **4**, 639-650.

Table 1. Posterior probability of H_{0k} for different group comparisons in the DNA damage and repair study.

Groups	$\Pr(d_{k,k+1} < \epsilon \mid \text{data})$		
	$\epsilon = 0.01$	$\epsilon = 0.05$	$\epsilon = 0.10$
1 vs 2	0.000	0.000	0.000
2 vs 3	0.000	0.000	0.000
3 vs 4	0.539	0.626	0.736
4 vs 5	0.009	0.123	0.376
1 vs 6	0.806	0.830	0.838
2 vs 7	0.000	0.000	0.000
3 vs 8	0.000	0.000	0.000
4 vs 9	0.000	0.000	0.000
5 vs 10	0.000	0.000	0.000
6 vs 11	0.820	0.903	0.919
7 vs 12	0.889	0.938	0.951
8 vs 13	0.911	0.983	0.991
9 vs 14	0.887	0.948	0.973
10 vs 15	0.045	0.142	0.313

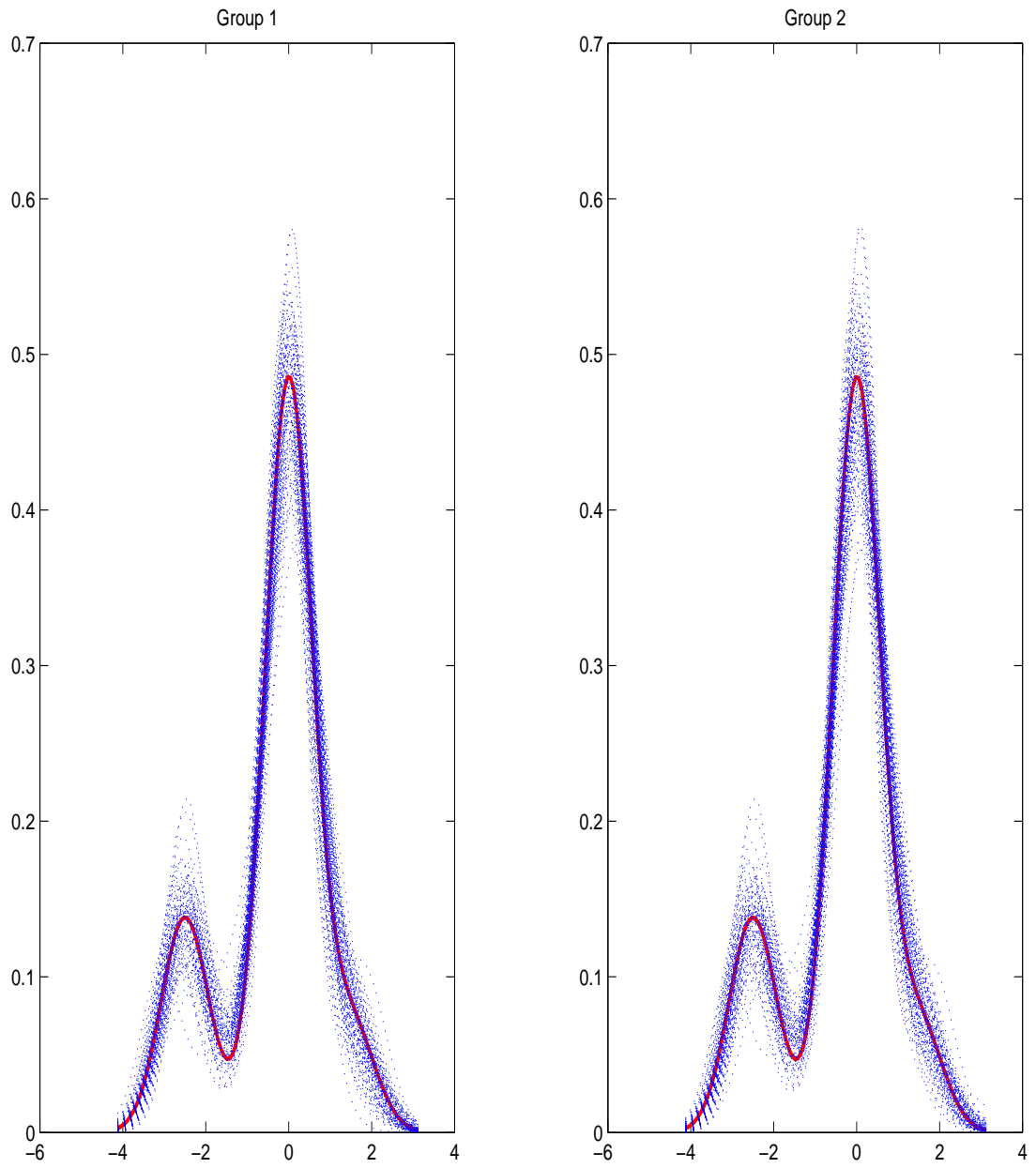


Figure 1: Group-specific density estimates for each of the 100 simulated data sets in case 1 (H_0 is true). The true density is a solid red line, while the estimates are dotted blue lines.

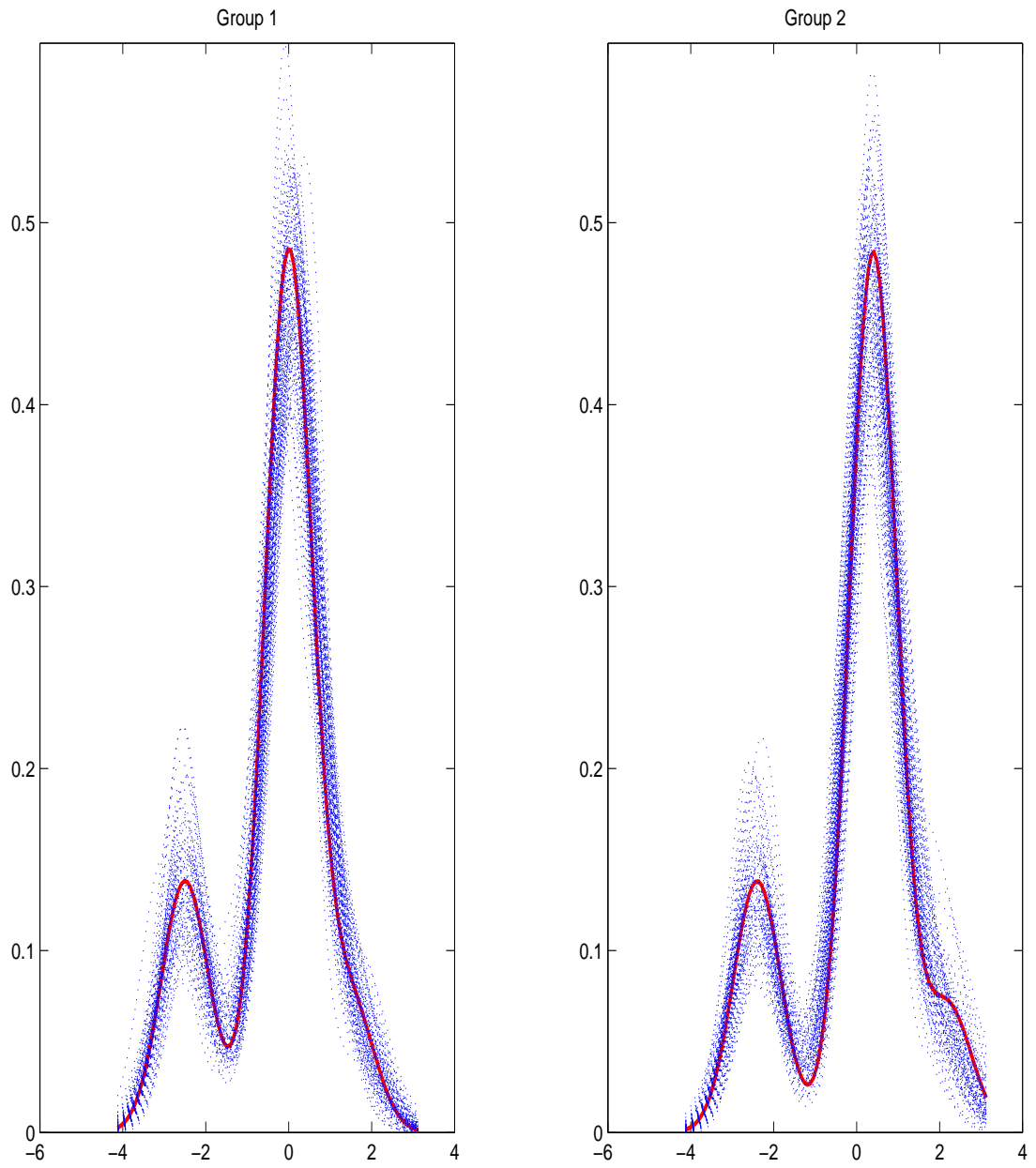


Figure 2: Group-specific density estimates for each of the 100 simulated data sets in case 2 (H_1 is true). The true density is a solid red line, while the estimates are dotted blue lines.

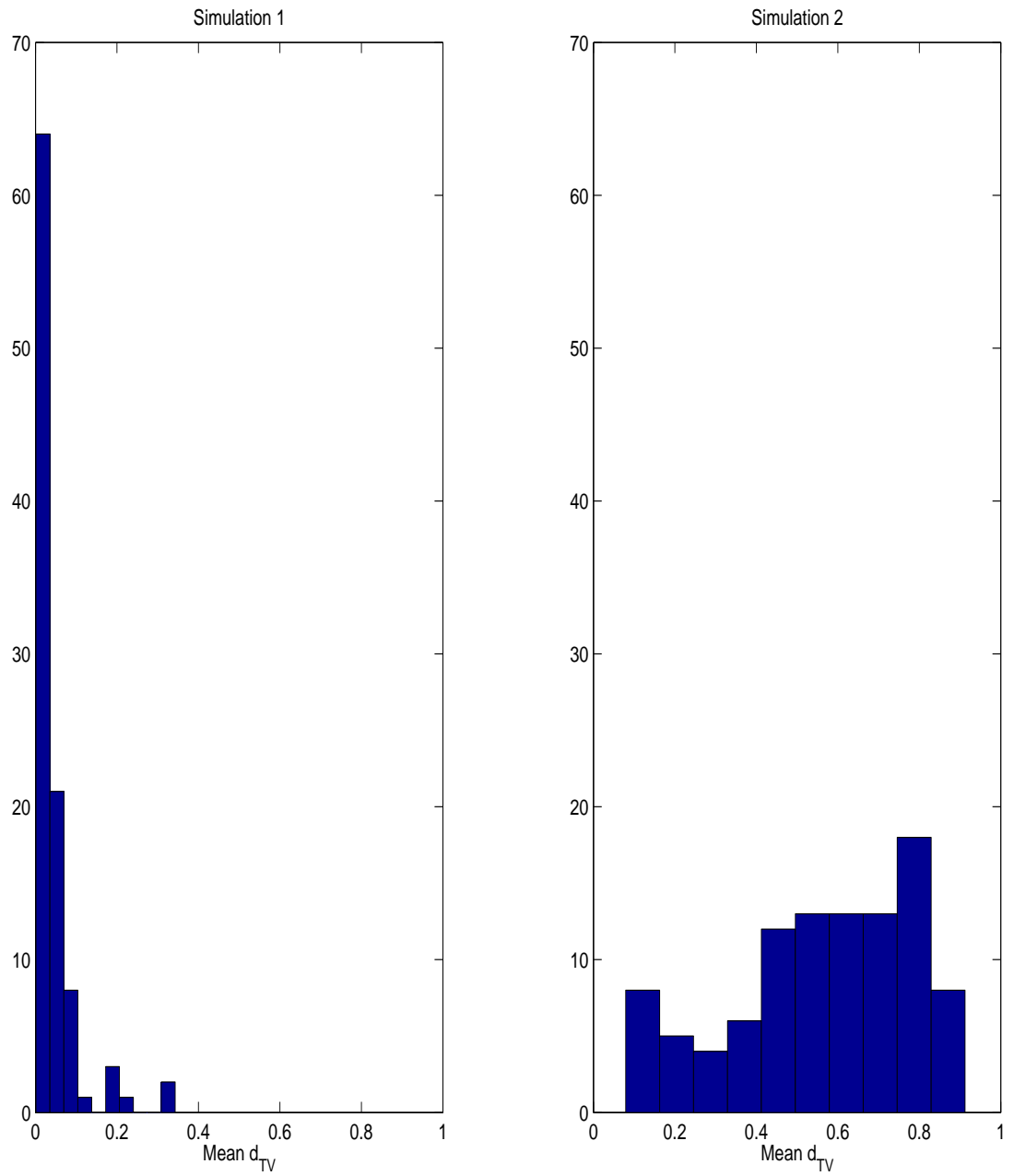


Figure 3: Histograms of the posterior means of the distance d_{12} between groups for each of the 100 simulated data sets in case 1 (H_0 is true) and case 2 (H_1 is true)

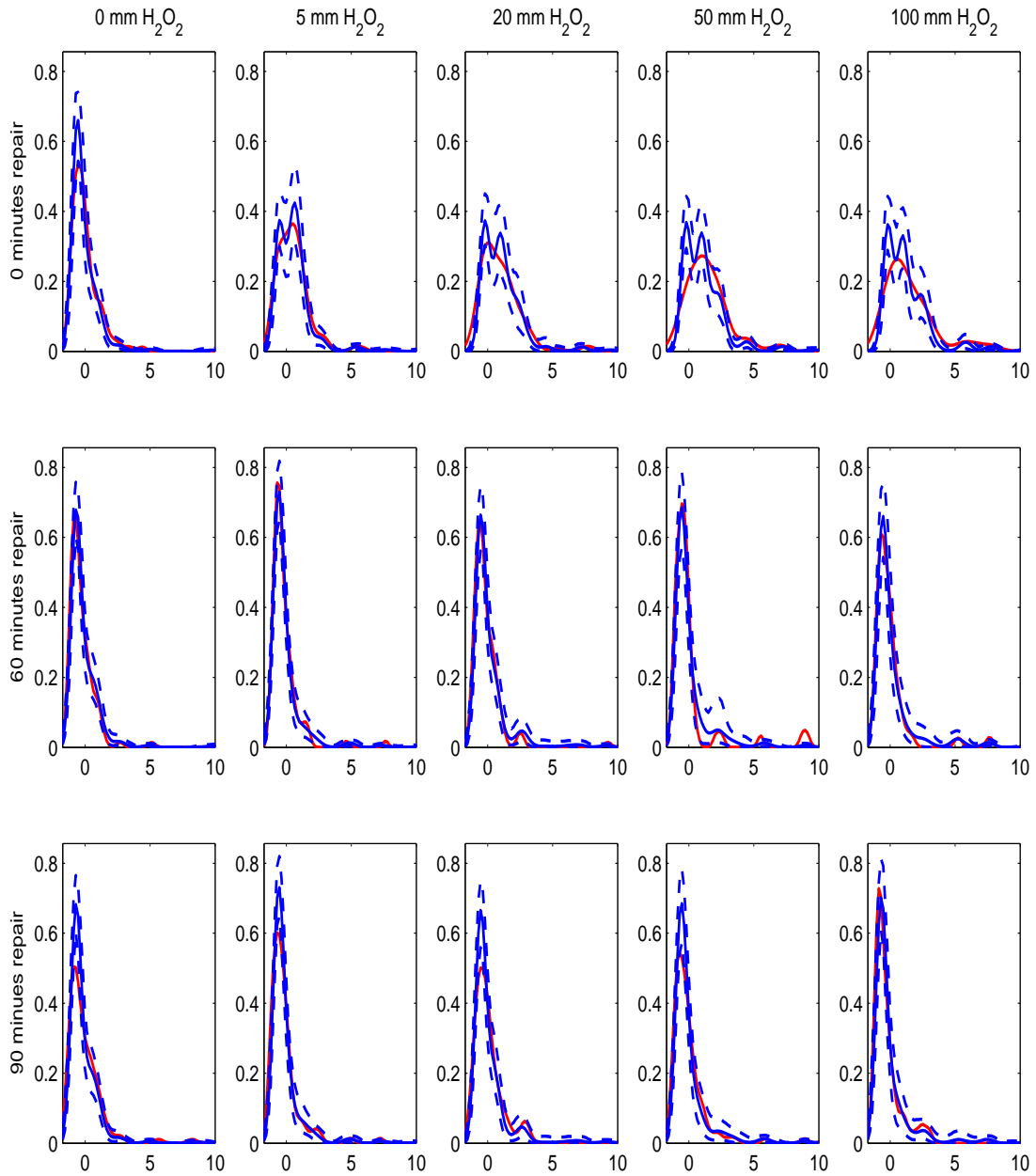


Figure 4: Estimated densities of the Olive tail moment in each of the H₂O₂ dose × repair time groups. Red lines are frequentist kernel estimates, solid blue lines are Bayes estimates, and dashed lines are 95% pointwise credible intervals.