Published in final edited form as:

Biometrics. 2008 June ; 64(2): 413–423.

Nonparametric Bayes testing of changes in a response distribution with an ordinal predictor

Michael L. Pennell^{1,*} and David B. Dunson²

¹*Division of Biostatistics, College of Public Health, The Ohio State University, B-115 Starling-Loving Hall, 320 West 10th Avenue, Columbus, OH 43210*

²*Biostatistics Branch, MD A3–03, National Institute of Environmental Health Sciences, P.O. Box 12233, Research Triangle Park, NC 27709*

Summary

In certain biomedical studies, one may anticipate changes in the shape of a response distribution across the levels of an ordinal predictor. For instance, in toxicology studies, skewness and modality might change as dose increases. To address this issue, we propose a Bayesian nonparametric method for testing for distribution changes across an ordinal predictor. Using a dynamic mixture of Dirichlet processes, we allow the response distribution to change flexibly at each level of the predictor. In addition, by assigning mixture priors to the hyperparameters, we can obtain posterior probabilities of no effect of the predictor and identify the lowest dose level for which there is an appreciable change in distribution. The method also provides a natural framework for performing tests across multiple outcomes. We apply our method to data from a genotoxicity experiment.

Keywords

Dirichlet process; Dose-response; Nonparametric Bayes; Toxicology; Trend test

1. Introduction

In many biomedical studies, there is interest in evaluating changes in a response distribution across an ordinal predictor. For example, toxicologists are often interested in assessing whether several biological responses vary in distribution with dose. In such settings, normality assumptions are typically not justified and there are biological reasons to expect changes in not only the location but also the variance and shape with dose. In particular, when there is substantial heterogeneity amongst subjects in their biological response to treatment due to variation in gene expression, timing of the cell cycle, and other factors, changes in variance, skewness, and even modality are natural.

Several frequentist methods have been proposed for testing for trends in non-Gaussian responses. Univariate approaches, such as Jonckheere's test (Terpstra, 1952; Jonckheere, 1954), are used most frequently, though other authors have proposed related methods for multivariate responses (e.g., Dietz and Killeen, 1981; Dietz, 1989; O'Brien, 1984; Huang et al., 2005). Unfortunately, these methods are most sensitive to changes in the location of a distribution and may ignore important trends in shape. In contrast, k -sample tests based on empirical distribution functions (e.g., Kiefer, 1959; Ahmad, 1976) are sensitive to changes in distribution shape, but not changes in location and scale.

* email: mpennell@cph.osu.edu

Bayesian nonparametric approaches have several advantages over frequentist alternatives, including the ability to incorporate prior information (e.g., from historical controls) and exact inferences on unknown distributions. Although there is an extensive literature on estimation (for a review see Müller and Quintana, 2004) and a growing literature on goodness of fit testing (e.g., Berger and Guglielmi, 2001; Carota and Parmigiani, 1996; Verdinelli and Wasserman, 1998), there is a paucity of methods for nonparametric testing across multiple groups. Gopalan and Berry (1998) use a nonparametric Dirichlet process prior (DPP; Ferguson, 1973, 1974) to perform multiple comparisons on distribution means. Recently, Basu and Chib (2003) proposed a less restrictive method which compares semiparametric models using Bayes factors. This method may be promising when there are only 2–3 levels of an ordinal predictor. However, in toxicology studies with several dose groups, this approach would be unwieldy when one considers each possible contrast between dose groups.

In analyzing data from multiple groups, one would typically be interested not only in testing but also in estimation of group-specific distributions. Such estimation can potentially be accomplished by fitting a DP mixture (DPM; Antoniak, 1974) model separately to the different dose groups. For some references on applications of DPMs, refer to West et al. (1994), Escobar and West (1995), Bush and MacEachern (1996), Kleinman and Ibrahim (1998), and Mukhopadhyay and Gelfand (1997). The disadvantage of such an approach is its inability to model trends and borrow information across the different dose groups, which is particularly important in applications having a modest number of subjects per group. Note that applications of the Basu and Chib (2003) procedure to test differences between groups would also have this disadvantage.

An alternative to using separate DPMs is to consider the dose group-specific distributions as a collection of dependent unknown distributions, which are assigned a dependent nonparametric prior. Dependent nonparametric priors have been the focus of a growing body of literature. In the early work of Cifarelli and Regazzini (1978), dependence was induced through a regression model within the parametric base measure of the DP (for related work see Mira and Petrone, 1996; Tomlinson and Escobar, 1999; Carota and Parmigiani, 2002; Giudici et al., 2003). Although the method is straightforward, it has limited flexibility. Other authors proposed inducing nonparametric dependence by parameterizing random measures as products of DP distributed factors, though within a different framework than ours (see, e.g., Gelfand and Kottas, 2001; Pennell and Dunson, 2006). MacEachern (1999; 2000) proposed a dependent Dirichlet process which characterizes dependency by defining a stochastic process for the atoms in the Sethuraman (1994) stick-breaking characterization of the DP. This approach has been successfully applied in ANOVA (De Iorio et al., 2004) and spatial (Gelfand et al., 2004) applications. A limitation of these applications is the assumption of fixed weights on the atoms, which does not allow new features to appear as dose increases. To solve this problem, Dunson (2006) proposed a dynamic mixture of DPs (DMDP), which is related to a mixture structure originally proposed by Müller et al. (2004) to combine inferences across related nonparametric models. More flexible dependent nonparametric priors for continuous predictors have been proposed by Griffin and Steel (2006), Dunson et al. (2007), and Duan et al. (2005).

Motivated by tractability in addressing the problem of nonparametric Bayes testing of changes with dose, we focus here on generalizing the Dunson (2006) approach. Our goal is to obtain posterior probabilities for local and global null hypotheses corresponding to equivalence in an unknown distribution between groups. Instead of requiring exact equivalence, we treat the distributions in adjacent groups as *effectively* equivalent if the total variation norm between their probability measures is less than ϵ . We use a hierarchical modification of the DMDP to borrow information across groups while assigning probabilities to the local null and alternative hypotheses. Using an MCMC implementation, we obtain posterior model probabilities for the

local and global hypotheses from a single run, while also producing posterior distributions for thresholds and estimates of group-specific distributions. Multiple response data can be accommodated without complication.

In Section 2, we describe our generalization of the DMDP and hyperprior structure. In Section 3, we outline an MCMC methodology for testing and estimation. In Section 4, we evaluate our approach using simulated data. In Section 5 we apply our method to data from a genotoxicity experiment. In Section 6 we discuss the results and future work.

2. Nonparametric Model and Prior Structure

2.1 General Framework

Let $\mathbf{y}_{hi}$ be a vector of q continuous outcomes measured on subject i ($i = 1, \dots, n_h$) in group h ($h = 1, \dots, d$) and X_h be a score for group h , where $X_{h-1} < X_h < X_{h+1}$. In a toxicology study, X_h is the dose administered to each animal in group h with $X_1 = 0$ representing the control group. To relax distribution assumptions, we assume that $\mathbf{y}_{hi} \sim F_h$ which has the density function

$$f_h(\mathbf{y}_{hi}) = \int \mathbf{N}(\mathbf{y}_{hi}; \mu_{hi}, \Sigma_{hi}) dG_h(\mu_{hi}, \Sigma_{hi}), \quad (1)$$

where G_h is unknown. Hence, the density of $\mathbf{y}$ in group h is an infinite mixture of multivariate Gaussian densities, with the mixture distribution varying across groups. As the number of components increases without bound, a mixture of Gaussians converges uniformly to any smooth density function (see Sorenson and Alspach, 1971).

It is clear that differences in the mixture distributions, G_h and G_{h+1} , imply differences in the outcome distributions, F_h and F_{h+1} . Hence, we focus on testing for changes in the mixture distributions. In particular, our focus is on local null hypotheses characterizing differences in adjacent groups and on global hypotheses representing intersections of these local hypotheses. As in the Kolmogorov-Smirnov test, we could specify a point null hypothesis which requires $G_1, \dots, G_d$ to match. However, many have argued that point nulls are artificial and are rarely observed in practice (see, e.g., Berger and Delampady, 1987; Nickerson, 2000). In this paper, we use a more realistic null hypothesis under which the group-specific distributions *effectively* match. We formalize this hypothesis below.

Let $G_1, \dots, G_d$ denote probability measures corresponding to the mixture distributions for the different groups. In addition, let B denote any Borel set such that $B \in \mathcal{B}$, with $B \subset R^q$ and $\mathcal{B}$ the Borel sigma-algebra of subsets of R^q , and let the total variation norm of $G_{h+1} - G_h$ be defined as

$$\|G_{h+1} - G_h\|_{TV} = \max_{B \in \mathcal{B}} |G_{h+1}(B) - G_h(B)|.$$

When $\|G_{h+1} - G_h\|_{TV} = 0$ we say that G_{h+1} and G_h match. However, we are interested in local nulls which only require G_{h+1} and G_h to effectively match which we specify as the following:

$$H_{0h} : \|G_{h+1} - G_h\|_{TV} \leq \epsilon \quad (2)$$

for $h = 1, \dots, d-1$, where ϵ is some small constant such that when H_{0h} holds, there is no appreciable difference in the mixture distributions across groups h and $h+1$. The global null of no changes in the response distribution across groups (H_0) corresponds to the intersection of these local nulls. Note that in using the total variation norm to define our hypothesis tests, we treat differences across all sets as being potentially important, in terms of the application. This approach avoids the need to specify a response range which is of particular interest. Such a range may be difficult to choose objectively, particularly in toxicology studies as it is typically

not known how to exactly equate a response value in an animal to a corresponding value in humans.

2.2 DMDP Model

In addition to conducting hypothesis tests, we would like to estimate the density in each dose group. As mentioned in Section 1, a promising approach would be to assign a dependent nonparametric process to $G_1, \dots, G_d$ thereby allowing us to borrow information across groups. Using the method of Müller et al. (2004), we could let $G_h = \pi G_0 + (1 - \pi) G_h^*$ for $h = 1, \dots, d$, where G_0 is a global probability measure and G_h^* is an innovation measure specific to group h , with $G_0, G_1^*, \dots, G_d^*$ assigned independent nonparametric priors. However, this approach results in an over-specified model in that $d+1$ random measures are incorporated to characterize d unknown distributions. Teh et al. (2006) proposed an alternative model which assumes that the G_h are drawn from DPPs with a common DP-distributed base measure G_0 . However, these hierarchical priors treat the groups as exchangeable. To incorporate information on ordering in the groups, Dunson (2006) proposed a dynamic mixture of Dirichlet processes (DMDP), which is conceptually related to the Müller et al. (2004) approach, but avoids the over-specification problem.

We first consider the case of two groups ($d = 2$). To induce dependency, we model the mixture distributions using the following DMDP:

$$\begin{aligned} G_1 &\sim \text{DP}(\alpha_0 G_0) \\ G_2 &= (1 - \pi_1) G_1 + \pi_1 G_1^* G_1^* \sim \text{DP}(\alpha_1 G_{01}), \end{aligned} \quad (3)$$

where $0 \leq \pi_1 \leq 1$, G_1^* is an innovation distribution that characterizes changes in the mixture distribution of $\mathbf{y}$ caused by increasing the group score from X_1 (e.g., dose = 0) to X_2 (e.g., dose > 0), and G_{01} is a known distribution. Note that the DMDP implies the following densities for $\mathbf{y}$ in groups 1 and 2:

$$\begin{aligned} f_1(\mathbf{y}) &= \int N(\mathbf{y}; \mu, \Sigma) dG_1(\mu, \Sigma) \\ f_2(\mathbf{y}) &= (1 - \pi_1) f_1(\mathbf{y}) + \pi_1 \int N(\mathbf{y}; \mu, \Sigma) dG_1^*(\mu, \Sigma). \end{aligned} \quad (4)$$

Hence, the distribution of $\mathbf{y}$ in group 2 is a mixture of the distribution in group 1 and a different DP mixture of normals characterized by mixture distribution G_1^* . This is a natural model for toxicology data since we would expect that the distribution in the treatment group, F_2 , shares features with the distribution in the control group, F_1 , but that innovations may have occurred due to stress induced by the test chemical. The amount of borrowing of information from the control group and the level of innovation are represented by the weights $(1 - \pi_1)$ and π_1 respectively.

In the above model, it is convenient to choose $G_0 \equiv G_{01}$, with these base measures having a conjugate normal-inverse Wishart form: $G_0(\mu, \Sigma) = N(\mu; \mu_0, \kappa \Sigma) \cdot W(\Sigma^{-1}; \nu_0, \nu_0^{-1} \mathbf{V}_0^{-1})$, where $W(\cdot; \nu_0, \nu_0^{-1} \mathbf{V}_0^{-1})$ denotes the Wishart density with degrees of freedom ν_0 and mean $\mathbf{V}_0^{-1}$. In this case, our nonparametric prior for F_1 is centered on a multivariate t -distribution, while the prior for F_2 is centered on a mixture of multivariate t -distributions. As α_0 and α_1 increase, we place more weight on this parametric base model for the unknown densities. By choosing hyperprior densities for α_0 and α_1 (as we describe in Section 2.4), the method can adapt flexibly to accommodate lack of fit in the multivariate t components.

Extending our model to d groups, for $h = 2, \dots, d$ we let

$$\begin{aligned}
G_h &= (1 - \pi_{h-1}) G_{h-1} + \pi_{h-1} G_{h-1}^* \\
&= \left\{ \prod_{l=1}^{h-1} (1 - \pi_l) \right\} G_1 + \sum_{l=1}^{h-2} \left\{ \prod_{m=l+1}^{h-1} (1 - \pi_m) \right\} \pi_l G_l^* + \pi_{h-1} G_{h-1}^* \\
&= \omega_{h1} G_1 + \sum_{l=1}^{h-1} \omega_{h,l+1} G_l^* \\
G_l^* &\sim DP(\alpha_l G_0), \quad \text{for } l = 1, \dots, h-1,
\end{aligned} \tag{5}$$

where $\omega_{hl} = \pi_{l-1} \prod_{m=1}^{h-1} (1 - \pi_m)$ for $l = 1, \dots, h-1$ and $\omega_{hl} = \pi_{h-1}$ for $l = h$ with $\pi_0 = 1$ and $\omega_h = (\omega_{h1}, \dots, \omega_{hh})'$ are group-specific probability weights on the different components in the mixture. Note that the above DMDP is not exchangeable on $(G_1, \dots, G_d)$. As we move from predictor level h to $h+1$ we decrease the weight assigned to $G_1, G_1^*, \dots, G_{h-1}^*$ and introduce a new unknown distribution to our mixture, G_h^* . As mentioned previously, this is a natural model for toxicology data; with every increase in dose level, h to $h+1$, individuals are subjected to higher levels of stress resulting in changes in the response distribution characterized by G_h^* . Hence, we would expect the response distributions at adjacent doses to be more similar than dose groups that are several levels apart, as is the case in our model.

Another important feature of our model is that it imposes a type of order restriction across groups. For instance, if the mixture distribution changes when we increase the predictor from the group h to the group $h+1$ level, all later groups ($h+2, \dots, d$) will also differ from group h . This is due to the following theorem:

THEOREM 1—For any G_h, G_{h+1} , and G_{h+j+1} distributed according to (5) ($h = 1, \dots, d-2$; $j = 1, \dots, d-h-1$)

$$\Pr(\|G_{h+j+1} - G_h\|_{TV} \leq \epsilon \|G_{h+1} - G_h\|_{TV} > \epsilon) = 0.$$

A proof of this theorem is provided in Web Appendix A. While it is possible, conceptually, for the distribution in group $h+j+1$ to shift back toward the distribution in group h , it is unlikely in any application for groups h and $h+j+1$ to effectively match when groups h and $h+1$ differ. For instance, in a toxicology experiment it is possible for a feedback mechanism to activate at a dose greater than $h+1$, causing G_{h+j+1} to resemble G_h . However, the feedback should not negate the effect of increasing dose from the group h to the group $h+1$ level. Biological plausibility motivated our formulation of the model hypotheses in this order-dependent manner.

2.3 Model Space Prior

In the two sample case of expression (3),

$$\begin{aligned}
H_0 : \|G_2 - G_1\|_{TV} &= \max_{B \in \mathcal{B}} |G_2(B) - G_1(B)| \\
&= \max_{B \in \mathcal{B}} |(1 - \pi_1) G_1(B) + \pi_1 G_1^*(B) - G_1(B)| \\
&= \pi_1 \|G_1^* - G_1\|_{TV} \leq \epsilon.
\end{aligned} \tag{6}$$

Note that H_0 holds for any fixed ϵ if: 1.) $\pi_1 \leq \epsilon$, regardless of G_1 and G_1^* or 2.)

$\|G_1^* - G_1\|_{TV} \leq \epsilon$, regardless of π_1 . Hence, the value of π_1 and the independent DPPs on G_1 and G_1^* control the prior probability allocated to H_0 . If probability close to zero is allocated to H_0 , then we may expect the posterior probability of H_0 to be small unless the sample size is large, clearly an unappealing property in most applications. Thus, we want to be able to control the prior probability allocated to H_0 .

We first consider the possibility of controlling $\Pr(H_0)$ through the priors for G_1 and G_1^* . The DPPs $G_1 \sim \text{DP}(\alpha_0 G_0)$ and $G_1^* \sim \text{DP}(\alpha_1 G_0)$ imply that

$$\|G_1 - G_1^*\|_{TV} = \max_{B \in \mathcal{B}} \left| \sum_{h=1}^{\infty} V_{1h} \prod_{l < h} (1 - V_{1l}) \delta_{\theta_{1h}}(B) - \sum_{h=1}^{\infty} V_{2h} \prod_{l < h} (1 - V_{2l}) \delta_{\theta_{2h}}(B) \right|,$$

where θ_{1h} and $\theta_{2h} \sim G_0$, $V_{1l} \sim \text{Be}(1, \alpha_0)$, $V_{2l} \sim \text{Be}(1, \alpha_1)$, and δ_θ denotes a degenerate distribution with mass at θ (Sethuraman, 1994). Suppose that we choose the Borel set B to contain $(\theta_{1h}, h = 1, \dots, \infty)$ but not $(\theta_{2h}, h = 1, \dots, \infty)$. Then, for a set B satisfying these conditions, $\delta_{\theta_{1h}}(B) = 1$ for $h = 1, \dots, \infty$ and $\delta_{\theta_{2h}}(B) = 0$ for $h = 1, \dots, \infty$. Hence, the above absolute value reduces to

$$\left| \sum_{h=1}^{\infty} V_{1h} \prod_{l < h} (1 - V_{1l}) \right| = 1.$$

The theorem below follows directly:

THEOREM 2—Let G_0 be a continuous base measure. For any $G_1 \sim \text{DP}(\alpha_0 G_0)$ independent of $G_1^* \sim \text{DP}(\alpha_1 G_0)$

$$\|G_1 - G_1^*\|_{TV} = 1.$$

From Theorem 2 we have $\Pr(\|G_1^* - G_1\|_{TV} < \epsilon) = 0$. Thus zero probability is allocated to H_0 by the total variation component, which implies that $\Pr(H_0)$ is entirely controlled by π_1 . Hence, one can effectively replace H_0 in (6) with $\pi_1 \leq \epsilon^*$, where ϵ^* is also close to zero. Extending this rationale to the d -group setting, we let

$$H_{0h} : \pi_h \leq \epsilon^*. \quad (7)$$

To control the prior probability allocated to H_{0h} , we propose a prior for π_h which is a mixture of its distribution under H_{0h} and $H_{1h} : \pi_h > \epsilon^*$. Let ζ_h be a latent variable that equals 1 when the local null H_{0h} is true and 0 otherwise and let $\Pr(\zeta_h = 1) = p_{0h}$. Given ζ_h , our prior for π_h has the following form:

$$\pi_h | \zeta_h \sim \zeta_h \cdot \text{Unif}(0, \epsilon^*) + (1 - \zeta_h) \cdot \text{Unif}(\epsilon^*, 1), h = 1, \dots, d - 1. \quad (8)$$

In the above mixture, $\mathbf{p}_0 = (p_{01}, \dots, p_{0h})'$ is chosen to reflect prior knowledge about the chances of H_0 being true. For example, to assign equal prior weight to H_0 and H_1 , $\mathbf{p}_0$ should be chosen

to satisfy $\Pr(H_0) = \prod_{h=1}^{d-1} p_{0h} = 0.5$. Under the Bayesian Bonferroni method of Westfall et al. (1997), this may be achieved by setting $p_{01} = \dots = p_{0,d-1} = 0.5^{1/(d-1)}$. However for large d , the probability of each local null hypothesis is nearly 1, which can make this approach overly conservative. We instead use a model space prior described by Hans and Dunson (2005). This approach induces dependency in the local null hypotheses by assuming $p_{0h} = p_0$ for $h = 1, \dots, d - 1$ where $p_0 \sim \text{Be}(a_p, b_p)$. As a consequence of this prior,

$\Pr(H_{0h} | \{\zeta_j : j \neq h\}) = (a_p + \sum_{j \neq h} \zeta_j) / (a_p + b_p + d - 2)$. Hence when several $H_{0j} (j \neq h)$ are true, we shrink toward no change between groups h and $h+1$, which is conceptually appealing in dose-response studies.

As in Hans and Dunson (2005), a_p and b_p are chosen so that on average $\Pr(H_0) = 0.5$, i.e.,

$E(p_0^{d-1}) = 0.5$. This implies that a_p and b_p must satisfy the following:

$$\begin{aligned} 0.5 &= \int_0^1 p_0^{d-1} \frac{\Gamma(a_p + b_p)}{\Gamma(a_p)\Gamma(b_p)} p_0^{a_p-1} (1 - p_0)^{b_p-1} dp_0 \\ &= \frac{\Gamma(a_p + b_p)\Gamma(a_p + d - 1)}{\Gamma(a_p)\Gamma(a_p + b_p + d - 1)}. \end{aligned} \quad (9)$$

We then impose the constraint $a_p + b_p = 1$ to represent unit information in the prior and solve (9) numerically for a_p and b_p .

2.4 Hyperpriors for DP Parameters

To decrease the sensitivity of our model to subjectively chosen hyperparameters, we specify the following hyperpriors:

$$\begin{aligned}\kappa^{-1} &\sim \text{Ga}(a_\kappa, b_\kappa) \quad \mu_0 \sim \text{N}(\beta, \Omega) \\ \alpha_h &\sim \text{Ga}(a_{\gamma h}, b_{\gamma h}) \quad \text{for } h = 0, 1, \dots, d-1.\end{aligned}$$

In practice, one may favor priors which assign high prior probability to low values of α_h since this implies that the number of mixture components is small (Antoniak, 1974) and a very flexible density estimator can be produced with few mixture components. Hence, a reasonable choice is $a_{\gamma h} = a_\gamma = 1$ and $b_{\gamma h} = b_\gamma = 1$. We also recommend following Escobar and West (1995) by choosing diffuse priors for κ^{-1} and μ_0 , with a_κ/b_κ small and, when possible, β chosen to reflect the mean of $\mathbf{y}$ in previous studies. Potentially, hyperpriors could also be assigned to $\mathbf{V}_0$ and v_0 , though we choose to fix these parameters to avoid over-parameterization. We recommend choosing $\mathbf{V}_0$ so that a $\text{N}(\beta, \mathbf{V}_0)$ distribution sufficiently covers an expected range of values of $\mathbf{y}$ and choosing v_0 to reflect one's confidence in this choice. When there is little prior information on $\mathbf{V}_0$, we strongly recommend performing a sensitivity analysis on v_0 as small values can often result in convergence or numerical problems, particularly in the case of a multivariate response.

3. Posterior Computation

3.1 Gibbs Sampling Methodology

We sample from the joint posterior of our model parameters using a Gibbs sampling methodology which is related to the Bush and MacEachern (1996) and West et al. (1994) approaches for the Dirichlet process mixture. Details may be found in Web Appendix B.

3.2 Hypothesis Testing

The global null may be formally evaluated using Rao-Blackwellization (Gelfand and Smith, 1990). For T iterations of the MCMC with a burn-in of T_b iterations, we compute

$$\begin{aligned}\widehat{\Pr}(H_0|\text{Data}) &= \frac{1}{T-T_b} \sum_{t=T_b+1}^T \Pr(\zeta_1 = \dots = \zeta_{d-1} = 1 | p_0^{(t)}, \mathbf{M}^{(t)}) \\ &= \frac{1}{T-T_b} \sum_{t=T_b+1}^T \prod_{h=1}^{d-1} p_{0h}^{*(t)},\end{aligned}\tag{10}$$

where the superscript (t) denotes a value sampled at iteration t ,

$\mathbf{M}^{(t)} = \{M_{hi}^{(t)}; h = 1, \dots, d; i = 1, \dots, n_h\}$, $M_{hi} \in \{1, \dots, h\}$ is the mixture indicator for subject i in group h , and $p_{0h}^{*(t)} = \Pr(\zeta_h = 1 | p_0^{(t)}, \mathbf{M}^{(t)})$. The explicit form for p_{0h}^* is given in Web Appendix B.

A common approach is to reject H_0 if (10) is less than some small pre-determined value (e.g., 0.05). However, it is appealing in a Bayesian analysis to report the posterior null hypothesis probability, or alternatively the Bayes factor, as a weight of evidence, avoiding the choice of an arbitrary cuto for determining significance. Hence, we also recommend reporting that there is evidence against H_0 when the posterior probabilities are only moderately small (i.e., between 0.05 and 0.1).

An attractive feature of our methodology is that we can test local hypotheses within the same Markov chain used to test H_0 . Since we properly calibrate our prior for p_0 to give equal prior probability to H_0 and H_1 , the posterior probabilities of the local nulls do not need to be adjusted (see Westfall et al., 1997). For instance, the posterior probability of a distribution change between groups h and $h + 1$ would be the average value of P_{0h}^* over $T - T_b$ iterations. In toxicology studies, there is also interest in comparing each dose group to control (group 1). The null hypothesis of no difference in the distributions of groups 1 and h is

$$H_{0,h-1}^* : \pi_j \leq \epsilon^* \quad j = 1, \dots, h-1 \quad (11)$$

and we estimate its posterior probability as

$$\widehat{\Pr}(H_{0,h-1}^* | \text{Data}) = \frac{1}{T - T_b} \sum_{t=T_b+1}^T \prod_{j=1}^{h-1} p_{0j}^{*(t)}. \quad (12)$$

3.3 Density Estimation

The predictive density of a future $\varphi_{h,n_h+1} = \{\mu_{h,n_h+1}, \Sigma_{h,n_h+1}\}$ can be obtained at each iteration of the MCMC:

$$\pi(\varphi_{h,n_h+1}) = \sum_{l=1}^h \left\{ \frac{\alpha_{l-1}}{\alpha_{l-1} + m_l} \cdot dG_0(\varphi_{h,n_h+1}) + \sum_{r=1}^{k_l} \left(\frac{m_{lr}}{\alpha_{l-1} + m_l} \right) \delta_{\theta_{lr}} \right\}, \quad (13)$$

where, as explained in Web Appendix B, m_l and k_l are the number of subjects and the number of unique means and covariances, respectively, in mixture component l , $\theta_{lr} = \{\mu_{lr}^*, \Sigma_{lr}^*\}$, and m_{lr} is the number of subjects whose mean and covariance equals θ_{lr} . This implies that the predictive density of $\mathbf{y}_{h,n_h+1}$ is

$$f_{h,n_h+1}(\mathbf{y}) = \omega_0 \cdot t\left(\mathbf{y}; v_0 - q + 1, \mu_0, \frac{v_0(1+\kappa)}{v_0 - q + 1} \mathbf{V}_0\right) + \sum_{l=1}^h \sum_{r=1}^{k_l} \omega_{hlr} \cdot N(\mathbf{y}; \mu_{lr}^*, \Sigma_{lr}^*), \quad (14)$$

where the multivariate t-density results from integrating $N(\cdot; \mu_{h,n_h+1}, \Sigma_{h,n_h+1})$ over G_0 and ω_0 and ω_{hlr} are the respective multipliers on $dG_0(\varphi_{h,n_h+1})$ and $\delta_{\theta_{lr}}$ in (13). Following T iterations, the Rao-Blackwellized estimate,

$$\bar{f}_{h,n_h+1}(\mathbf{y}) = \frac{1}{T - T_b} \sum_{t=T_b+1}^T f_{h,n_h+1}^{(t)}(\mathbf{y}), \quad (15)$$

may be computed over a grid of values.

In a toxicology study, one would also be interested in the posterior density of the lowest observed adverse effects level (LOAEL). Assuming that any change in the distribution relative to group 1 is an adverse effect, we can estimate the posterior density of the LOAEL as follows:

$$\begin{aligned} \widehat{\Pr}(\text{LOAEL} = h | \text{Data}) &= \frac{1}{T - T_b} \sum_{t=T_b+1}^T \left(1 - p_{0,h-1}^{*(t)} \right) \prod_{j=1}^{h-2} p_{0j}^{*(t)} \quad h = 2, \dots, d \\ \widehat{\Pr}(\text{LOAEL} = d + 1 | \text{Data}) &= \frac{1}{T - T_b} \sum_{t=T_b+1}^T \prod_{j=1}^{d-1} p_{0j}^{*(t)}, \end{aligned} \quad (16)$$

where a LOAEL of $d+1$ implies that the LOAEL is greater than the largest dose considered in the study. Hans and Dunson (2005) use a similar approach to estimate the posterior density of the LOAEL under umbrella orderings in parametric models.

4. Simulation Studies

4.1 Description of Data

We performed three simulation studies to evaluate the performance of our methodology. In each simulation case, we generated a vector of three responses for each subject, $\mathbf{y}_{hi} = (y_{hi1}, y_{hi2}, y_{hi3})'$, from a mixture of five multivariate normal distributions:

$$f_h(\mathbf{y}_{hi}) = \sum_{j=1}^5 w_{hj} \mathbf{N}(\mathbf{y}_{hi}; \boldsymbol{\mu}_j, \tau_j \cdot \mathbf{I}_3),$$

where $\boldsymbol{\mu}_j = (\mu_{j1}, \mu_{j2}, \mu_{j3})'$ and $w_{hj} \propto \exp(\gamma_{j0} + (h-1)\gamma_{j1})$ for $h = 1, \dots, 6$ and $i = 1, \dots, 30$. For each mixture component, $\mu_{jk} = x_j + k - 1$ where x_j is a constant.

The parameter values used in the simulations are provided in Table 1. Cases 1 and 2 differed only in their values of the mixture probabilities; in Case 1 we let $\gamma_{j0} = 1$ and $\gamma_{j1} = 0$ for $j = 1, \dots, 5$ to simulate under a null model, while in Case 2 the mixture probabilities varied with predictor level. As seen in Figure 1, the right tails of the marginal densities of y_1 become heavier with predictor level in Case 2 while, in Case 3, the modality of the distribution changes across h . The latter case may arise in toxicology experiments in which gene $\times$ dose interactions cause sub-populations to respond differently to treatment. The marginal densities of y_2 and y_3 follow a similar trend with dose, though their location parameters differ. Five data sets were simulated under each case.

4.2 Univariate Analyses

We first performed a univariate analysis on y_1 in order to compare our methodology to standard frequentist methods. In each case, we assumed $\mu_0 \sim N(0, 100)$, $\kappa^{-1} \sim (0.5, 50)$, and $\alpha_h \sim \text{Ga}(1, 1)$ for $h = 1, \dots, d$ a priori and let $v_0 = 10$ and $V_0 = 1$. In our mixture priors for $\pi_1, \dots, \pi_5$, we let $\epsilon^* = 0.05$ and $p_0 \sim \text{Be}(0.725, 0.275)$ to assign equal prior probability to H_0 and H_1 . We ran our MCMC for a total 55,000 iterations, discarding the first 5,000 as a burn-in, and saved every 10th iteration to thin the chain. For the purpose of comparing our method to the frequentist approaches, we decided to reject H_0 if $\widehat{\Pr}(H_0|\text{Data}) \leq 0.05$.

Our method correctly assigned high probability to H_0 in Case 1 (> 0.914 for each data set) and very low posterior probability to H_0 in Cases 2 and 3 (as seen in Table 2). These results were consistent with the p-values from a k -sample Kolmogorov-Smirnov test (Kiefer, 1959), while both Jonckheere's test and the Kruskal-Wallis test failed to reject the null in 5/5 Case 3 simulations. Since stochastic ordering was violated in Case 3, it is not surprising that Jonckheere's test was insignificant. Being sensitive to location but not shape changes, the Kruskal-Wallis test could detect distribution changes in case 2 but not 3, while our approach was sensitive to changes in each case.

In addition to assessing evidence of any change across the groups, it is of interest to identify dose groups that differ from control. We compared our approach to methods commonly used for non-Gaussian toxicology data, Dunn's method (1964) and Williams' (1986) modification of Shirley's method (1977). We also compared our method to $(d-1)$ 2-sample Kolmogorov-Smirnov tests performed under a Bonferroni correction. As seen in Table 2, our method found clear evidence that groups 5 and 6 differed from group 1 in 3/5 and 4/5 Case 2 simulations, respectively. These results were consistent with Shirley's method, while Dunn's method found a significant difference between groups 1 and 5 in only 2/5 simulations and the Kolmogorov-Smirnov tests could only identify differences between groups 1 and 6. In Case 3, our method revealed that groups 4–6 were all clearly different from group 1 and found significant differences between groups 1 and 3 in 3/5 simulations. The Kolmogorov-Smirnov tests were

more conservative and suggested group-related trends that were inconsistent with the simulated data; in two data sets a lower dose group was significantly different from group 1, while a higher dose group was not found to differ from baseline.

4.3 Multivariate Analyses

We next repeated our analyses using a multivariate approach. In addition to testing for a predictor effect on the distribution of $\mathbf{y}$, in Cases 2 and 3 we also tested for an effect on the joint distribution of y_1, z_1, z_2 and y_1, y_2, z_1 , where $(z_{hi1}, z_{hi2}) \sim N(\mathbf{0}, \mathbf{I}_2)$ for $h = 1, \dots, 5$ and $i = 1, \dots, 30$ independently of $\mathbf{y}_{hi}$. The purpose of these latter tests was to evaluate the performance of our multivariate method under varying numbers of affected outcomes. The implementation was very similar to that of the univariate analyses, with $\boldsymbol{\mu}_0 \sim N(\mathbf{0}, 100 \cdot \mathbf{I}_3)$ a priori and $\mathbf{V}_0 = \mathbf{I}_3$.

Figure 2 summarizes the results. As in the univariate analyses, our method gave high posterior probability to H_0 in Case 1; the posterior probability was greater than 0.989 in each of the five cases. In Case 2, the posterior probabilities of H_0, H_{04}^* , and H_{05}^* were small in each analysis. However, in most cases, the posterior probabilities of H_{01}^*, H_{02}^* and H_{03}^* increased as we decreased the number of y 's (the affected outcomes) in our model. Thus our method tends to be conservative when only one outcome is moderately affected, which is a desirable feature for a multivariate method. A similar trend was seen in the posterior probabilities of H_{01}^* and H_{02}^* in Case 3.

4.4 Sensitivity Analyses

As recommended in Section 2.4 we performed sensitivity analyses on the choice of v_0 . The results are provided in Web Figures 1 and 2. The results of our univariate analyses were relatively unaffected by an increase ($v_0 = 20$) and decrease ($v_0 = 5$) in the degrees of freedom. In our multivariate analyses of $\mathbf{y}$, increasing v_0 to 20 substantially changed the posterior probability of H_{01}^* and H_{02}^* in 2/5 Case 3 simulations, while all other results were fairly robust. In one of these data sets, the MCMC mixed poorly which could be reflective of a poor choice of $\mathbf{V}_0$. When $v_0 = 5$, the prior is quite diffuse and computational instabilities arise due to difficulties in inversion of ill-conditioned covariance matrices, arising in some of the MCMC samples.

5. Genotoxicity Example

5.1 Data and Methods

We considered data from a genotoxicity study analyzed previously by Dunson (2006), Dunson et al. (2003), and Dunson and Taylor (2005). The study assessed the effect of oxidative stress, induced by H_2O_2 , on the frequency of DNA strand breaks using the comet assay (single cell electrophoresis). Human lymphoblast cells were exposed to 5 different concentrations of hydrogen peroxide (0, 5, 20, 50, and 100 μM H_2O_2). Several surrogate measures of DNA damage were obtained for 100 cells in each group, though we focused on the two best surrogates: $y_1 = \% \text{ tail DNA}$ and $y_2 = \text{Olive tail moment (OTM)}$. More information on these surrogates, including diagnostics demonstrating non-normality, may be found in Duez et al. (2003) and Dunson et al. (2003).

Our implementation was identical to that described in the simulation studies, with a few exceptions. Based on the range of surrogate values reported in Duez et al. (2003), we chose $\boldsymbol{\beta} = (50, 10)'$, $V_{011} = 181$, $V_{022} = 10$, and $V_{012} = V_{021} = 38.2$, where V_{0ij} denotes element (i, j) of $\mathbf{V}_0$. Note that we chose a large value for the covariance because $\% \text{ tail DNA}$ and OTM are likely to be highly correlated. However, since we are uncertain about how well these mean and

variance values represent the current data, we chose a relatively diffuse prior for μ_0 , $N(\beta, 6 \cdot V_0)$, and let $v_0 = 10$. Also, using the guidelines in Section 2.3, we assumed $p_0 \sim \text{Be}(0.7, 0.3)$ a priori. We ran our MCMC for 105,000 iterations, otherwise implementing as in Section 4.

5.2 Results

Our method provided strong evidence in favor of an effect of H_2O_2 on the frequency of DNA strand breaks as $\widehat{\Pr}(H_0|\text{Data}) < 0.001$. As demonstrated by the posterior predictive densities in Figure 3, treatment with H_2O_2 changes the joint distribution of the % tail DNA and OTM from a unimodal distribution favoring small values to a multi-modal distribution supporting large levels of DNA damage. While genotoxic effects are evident at the smallest dose of H_2O_2 (the posterior probability that 5 μM is the LOAEL is > 0.999), they appear to level off at the higher doses; the outcome distribution clearly does not differ between the 50 and 100 μM groups ($\widehat{\Pr}(\text{No difference } 50 - 100 \mu\text{M}|\text{Data}) = 0.990$) and there is little evidence that the distribution changes at doses higher than 5 μM ($\widehat{\Pr}(\text{No difference } 5 - 100 \mu\text{M}|\text{Data}) = 0.281$). Figure 3 demonstrates that these densities are in agreement with the observed data, supporting goodness of fit.

Although previous analyses of these data have also demonstrated a relationship between H_2O_2 and the frequency of DNA strand breaks, we have provided several new pieces of information. For instance, neither the latent response models fit by Dunson et al. (2003) and Dunson (2006) nor the quantile regression approach of Dunson and Taylor (2005) provided a formal method for testing for changes across dose groups and for estimating threshold doses. In genetic epidemiology studies of DNA repair it is important to use minimal doses of the test agent to avoid high levels of cell death, which make the comet assay unreliable. Hence our finding that 5 μM is a potential threshold level for DNA damage has important implications for future experiments. Another advantage of our method is our ability to obtain smooth, nonparametric estimates of the joint density of the surrogates.

5.3 Sensitivity Analyses

As in Section 4.4, we performed a sensitivity analysis on our choice of v_0 . Increasing the degrees of freedom to $v_0 = 20$ did not change the posterior probability of H_0 or the LOAEL, but did increase the probability of no change across doses 5–100 μM to 0.826. When $v_0 = 5$ mixing was poor and the model did not converge after 105,000 iterations.

6. Discussion

We have proposed a nonparametric method which models distribution changes across an ordinal predictor and provides a formal approach for Bayesian testing of local and global changes across groups. In addition, we can perform inferences on multivariate distributions and identify thresholds. Our method provides more informative results than many frequentist k -sample tests and, as demonstrated in simulation studies, may be more sensitive to changes in distributional shape. We acknowledge, however, that these results are based on a small number data sets and larger simulation studies are needed to adequately characterize the sensitivity of our method.

Although our method is relatively simple to program, the computation can be intensive for large data sets especially those with several responses and dose groups. For example, it took approximately 33 hours to complete the MCMC for our analysis of the comet assay data using a Matlab (version 7.1) program compiled in C and run on a Dell Optiplex SX270 (2.8 GHz Pentium 4 processor). Thus, a more efficient computational method would likely be necessary

for massive, high dimensional data sets. One promising approach may be to develop a Variational Bayes (VB) implementation (see Blei and Jordan, 2006).

Some extensions of our method may also serve as exciting areas of future research. For example, a formal method of incorporating historical data would be useful in toxicology studies with extensive historical control databases. It may also be of interest to incorporate order restrictions into our method which ensure that the the quantiles of the response are non-decreasing with dose. This extension would be related to the nonparametric approaches for stochastic ordering described by Gelfand and Kottas (2001) and Hoff (2003).

Acknowledgements

We would like to thank Jack Taylor, NIEHS, for providing the data used in our example. This research was supported in part by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences.

Web Appendix A: Proof of Theorem 1

Note that

$$\|G_{h+j+1} - G_h\|_{TV} = \max_{B \in \mathcal{B}} \left| \prod_{l=h}^{h+j} (1 - \pi_l) G_h(B) + \sum_{l=h}^{h+j-1} \pi_l \prod_{m=l+1}^{h+j} (1 - \pi_m) G_l^*(B) + \pi_{h+j} G_{h+j}^*(B) - G_h(B) \right|.$$

Under Sethuraman (1994) stick-breaking representation of the Dirichlet process

$$\begin{aligned} G_1(B) &= \sum_{j=1}^{\infty} V_{1j} \prod_{k < j} (1 - V_{1k}) \delta_{\theta_{1k}}(B), \\ G_l^*(B) &= \sum_{j=1}^{\infty} V_{l+1,j} \prod_{k < j} (1 - V_{l+1,k}) \delta_{\theta_{l+1,k}}(B) \quad l = 1, \dots, h+j, \end{aligned}$$

$\theta_{lk} \stackrel{iid}{\sim} G_0, V_{lj} \stackrel{ind}{\sim} \text{Be}(1, \alpha_{l-1})$, and δ_{θ} denotes a point mass at θ . Suppose we chose B to contain $(\theta_{lk}, k = 1, \dots, \infty)$ for $l = 1, \dots, h$ but not for $l > h$. Then, for a set B satisfying these conditions, $G_h(B) = 1$ and $G_l^*(B) = 0$ for $l = h, h+1, \dots, h+j$ which causes the absolute value to reduce to

$$\begin{aligned} \left| \prod_{l=h}^{h+j} (1 - \pi_l) - 1 \right| &= 1 - \prod_{l=h}^{h+j} (1 - \pi_l) \\ &= \pi_h + (1 - \pi_h) \sum_{l=h+1}^{h+j-1} \pi_l \prod_{k=l+1}^{h+j} (1 - \pi_k) + \pi_{h+j} (1 - \pi_h) \\ &\geq \pi_h \\ &> \epsilon, \end{aligned}$$

since $\pi_h > \epsilon$ when $\|G_{h+1} - G_h\|_{TV} > \epsilon$. Hence

$$\Pr(\|G_{h+j+1} - G_h\|_{TV} \leq \epsilon | \|G_{h+1} - G_h\|_{TV} > \epsilon) = 0.$$

Web Appendix B: Gibbs Sampling Methodology

To simplify notation, let $\phi_{hi} = \{\mu_{hi}, \Sigma_{hi}\}$. As in Dunson (2006), the derivation of our full conditional posterior distributions involves re-writing the model in (5) as

$$\begin{aligned} \varphi_{hi} &\sim \sum_{l=1}^h 1(M_{hi} = l) G_l^{**} \\ M_{hi} &\sim \text{Multinomial}(1, \dots, h; \omega_{h1}, \dots, \omega_{hh}), \quad \text{for } i = 1, \dots, n_h \\ G_l^{**} &\sim \text{DP}(\alpha_{l-1} G_0), \quad \text{for } l = 1, \dots, h, \end{aligned} \tag{B-1}$$

where $G_l^{**} = G_1$ for $l = 1$ and $G_l^{**} = G_{l-1}^*$ for $l = 2, \dots, h$. As mentioned in Section 2.2, ω_{hi} is distinct for each h . For $l = 1, \dots, d$, let $\phi_l = \{\phi_{hi} : M_{hi} = l; h = l, \dots, d, i = 1, \dots, n_h\}$

and $m_l = \sum_{h=l}^d \sum_{i=1}^{n_h} 1(M_{hi} = l)$. Due to the discrete nature of the DP measure G_l^{**} , the elements of ϕ_l are allocated to $k_l \leq m_l$ distinct values (or clusters) $\theta_l = (\theta_{l1}, \dots, \theta_{lk_l})'$ where $\theta_{lr} = \{\mu_{lr}^*, \Sigma_{lr}^*\}$ and we let $m_{lr} = \sum_{h,i: M_{hi}=l} 1(\varphi_{hi} = \theta_{lr})$.

Let $\varphi_l^{(hi)}$ be the subset of ϕ_l excluding subject h, i ($l = 1, \dots, d$). The full conditional prior of ϕ_{hi} given $\varphi^{(hi)} = \{\varphi_1^{(hi)}, \dots, \varphi_d^{(hi)}\}$ is

$$\sum_{l=1}^h \omega_{hl} \left\{ \left(\frac{\alpha_{l-1}}{\alpha_{l-1} + m_l^{(hi)}} \right) G_0 + \sum_{r=1}^{k_l^{(hi)}} \left(\frac{m_{lr}^{(hi)}}{\alpha_{l-1} + m_l^{(hi)}} \right) \delta_{\theta_{lr}^{(hi)}} \right\}, \quad (\text{B-2})$$

where $m_l^{(hi)}$ is the number of latent $\phi_{h',i'}$ ($h', i' \neq h, i$) in $\varphi_l^{(hi)}$ whose unique values are $\theta_{lr}^{(hi)} = (\theta_{lr}^{(hi)}, r = 1, \dots, k_l^{(hi)})$ and $m_{lr}^{(hi)}$ denotes the number of $\phi_{h',i'}$ having value $\theta_{lr}^{(hi)}$. To simplify notation we will henceforth let

$$\omega_{hl0}^{(hi)} = \frac{\omega_{hl} \cdot \alpha_{l-1}}{\alpha_{l-1} + m_l^{(hi)}}$$

and $\omega_{hlr}^{(hi)}$ denote the multiplier on $\delta_{\theta_{lr}^{(hi)}}$ in (B-2). Updating the conditional prior with the data, we have the following full conditional posterior for ϕ_{hi} :

$$\sum_{l=1}^h \left\{ \omega_{hl0}^{(hi)} G_{0hi} + \sum_{r=1}^{k_l^{(hi)}} \omega_{hlr}^{(hi)} \delta_{\theta_{lr}^{(hi)}} \right\}, \quad (\text{B-3})$$

where

$$\begin{aligned} \tilde{\omega}_{hl0}^{(hi)} &= c \cdot \omega_{hl0}^{(hi)} \cdot t(\mathbf{y}_{hi}; v_0 - q + 1, \mu_0, \mathbf{V}_0^*) \\ \mathbf{V}_0^* &= \frac{v_0(1+\kappa)}{v_0 - q + 1} \cdot \mathbf{V}_0 \\ \omega_{hlr}^{(hi)} &= c \cdot \omega_{hlr}^{(hi)} \cdot N(\mathbf{y}_{hi}; \mu_{lr}^{*(hi)}, \Sigma_{lr}^{*(hi)}) \\ G_{0hi}(\varphi_{hi}) &= W(\Sigma_{hi}^{-1}; v_0 + 1, \mathbf{V}_{0hi}^{-1}) \cdot N(\mu_{hi}; \frac{\kappa}{\kappa+1}(\mathbf{y}_{hi} + \kappa^{-1}\mu_0), \frac{\kappa}{\kappa+1}\Sigma_{hi}) \\ \mathbf{V}_{0hi} &= v_0 \cdot \mathbf{V}_0 + \frac{1}{\kappa+1}(\mathbf{y}_{hi} - \mu_0)(\mathbf{y}_{hi} - \mu_0)' \end{aligned}$$

and $t(\cdot; v, \mu, \mathbf{V})$ denotes the multivariate t density with v degrees of freedom, location parameter μ , and scale $\mathbf{V}$.

Gibbs sampling may proceed by sampling directly from (B-3). However to improve mixing, we apply an algorithm which parallels those described by Bush and MacEachern (1996) and West et al. (1994). Let $S_{hi} = (0, l)$ if ϕ_{hi} is allocated to a new cluster in mixture component l and $S_{hi} = (r, l)$ if $\varphi_{hi} = \theta_{rl}^{(hi)}$. At each iteration of our MCMC, we update the parameters in our model using the following steps:

1. Sample each S_{hi} from its full conditional posterior distribution which is multinomial with $\Pr(S_{hi} = (r, l) | \cdot) = \tilde{\omega}_{hlr}$ for $l = 1, \dots, h, r = 0, 1, \dots, k_l^{(hi)}$. Whenever $S_{hi} = (0, l)$ sample a new value for ϕ_{hi} from G_{0hi} and assign subject h, i to their own cluster in component l .
2. Given the updated values of $\mathbf{S} = \{S_{hi}, h = 1, \dots, d; i = 1, \dots, n_h\}$, for $l = 1, \dots, d$ and $r = 1, \dots, k_l$ sample θ_{lr} from

$$W\left(\Sigma_{lr}^{*-1}; v_0 + m_{lr}, \mathbf{V}_{0lr}^{*-1}\right) \cdot N\left(\mu_{lr}^*; \kappa^* \left(\kappa^{-1} \mu_0 + m_{lr} \bar{\mathbf{y}}_{lr}\right), \kappa^* \Sigma_{lr}^*\right), \quad (\text{B-4})$$

where

$$\mathbf{V}_{0lr}^* = v_0 \mathbf{V}_0 + \sum_{h=1}^d \sum_{i: S_{hi}=(l,r)} \left(\mathbf{y}_{hi} - \bar{\mathbf{y}}_{lr}\right) \left(\mathbf{y}_{hi} - \bar{\mathbf{y}}_{lr}\right)' + \frac{m_{lr}}{\kappa \cdot m_{lr} + 1} \left(\mu_0 - \bar{\mathbf{y}}_{lr}\right) \left(\mu_0 - \bar{\mathbf{y}}_{lr}\right)',$$

$\kappa^* = \kappa / (1 + \kappa \cdot m_{lr})$, and $\bar{\mathbf{y}}_{lr}$ is the mean response vector in cluster (l, r) .

The priors for our remaining model parameters are conjugate, and thus may be updated by adding the following Gibbs steps to our MCMC algorithm:

3. For $h = 1, \dots, d-1$ sample:

a.) ζ_h from $\text{Bin}(1, p_{0h}^*)$ where

$$p_{0h}^* = \frac{p_0 \cdot \Phi_h(\epsilon^*)}{p_0 \cdot \Phi_h(\epsilon^*) + \frac{\epsilon^*}{1-\epsilon^*} (1-p_0)(1-\Phi_h(\epsilon^*))},$$

with

$$\begin{aligned} \Phi_h(\epsilon^*) &= \int_0^{\epsilon^*} \text{Be}(x; a_h^*, b_h^*) dx \\ a_h^* &= 1 + \sum_{j=h+1}^d \sum_{i=1}^{n_j} 1(M_{ji} = h+1) \\ b_h^* &= 1 + \sum_{j=h+1}^d \sum_{i=1}^{n_j} 1(M_{ji} < h+1). \end{aligned}$$

b.) π_h from

$$\pi(\pi_h | \cdot) = \zeta_h \cdot \frac{\text{Be}(\pi_h; a_h^*, b_h^*)}{\Phi_h(\epsilon^*)} 1(0 \leq \pi_h \leq \epsilon^*) + (1 - \zeta_h) \cdot \frac{\text{Be}(\pi_h; a_h^*, b_h^*)}{1 - \Phi_h(\epsilon^*)} 1(\epsilon^* < \pi_h \leq 1), \quad (\text{B-5})$$

where the notation $a|\cdot$ denotes a given all other variables.

4. Sample p_0 from

$$\pi(p_0 | \cdot) = \text{Be}\left(p_0; a_p + \sum_{h=1}^{d-1} \zeta_h, b_p + d - 1 - \sum_{h=1}^{d-1} \zeta_h\right). \quad (\text{B-6})$$

5. For $h = 0, 1, \dots, d-1$ sample a latent variable, u_h , from $\pi(u_h | a_{lp}, m_{h+1}) = \text{Be}(u_h; a_h + 1, m_{h+1})$ and then sample α_h from its full conditional posterior

$$\pi(\alpha_h | u_h, k_{h+1}, m_{h+1}) = p_{uh} \text{Ga}(\alpha_h; a_{\gamma_h} + k_{h+1}, b_{\gamma_h} - \log(u_h)) + (1 - p_{uh}) \text{Ga}(\alpha_h; a_{\gamma_h} + k_{h+1} - 1, b_{\gamma_h} - \log(u_h)), \quad (\text{B-7})$$

where

$$\frac{p_{uh}}{1 - p_{uh}} = \frac{a_{\gamma_h} + k_{h+1} - 1}{m_{h+1} (b_{\gamma_h} - \log(u_h))}.$$

6. Sample μ_0 from $\pi(\mu_0 | \cdot) = N(\mu_0; \beta^*, \Omega^*)$ where

$$\begin{aligned} \beta^* &= \Omega^* \left(\Omega^{-1} \beta + \frac{1}{\kappa} \sum_{h=1}^d \sum_{r=1}^{k_h} \Sigma_{hr}^{*-1} \mu_{hr}^* \right) \\ \Omega^* &= \left\{ \Omega^{-1} + \frac{1}{\kappa} \sum_{h=1}^d \sum_{r=1}^{k_h} \Sigma_{hr}^{*-1} \right\}^{-1}. \end{aligned}$$

7. Sample κ^{-1} from

$$\pi(\kappa^{-1}|\cdot) = \text{Ga}\left(\kappa^{-1}; a_{\kappa} + \frac{q}{2} \sum_{h=1}^d k_h, b_{\kappa} + \frac{1}{2} \sum_{h=1}^d \sum_{r=1}^{k_h} (\mu_{hr}^* - \mu_0)' \Sigma_{hr}^{*-1} (\mu_{hr}^* - \mu_0)\right). \quad (\text{B-8})$$

References

- Ahmad R. Multivariate k-sample problem and generalization of the Kolmogorov-Smirnov test. *Annals of Statistical Mathematics* 1976;28:259–265.
- Antoniak CE. Mixtures of Dirichlet processes with applications to nonparametric problems. *Annals of Statistics* 1974;2:1152–1174.
- Basu S, Chib S. Marginal likelihood and Bayes factors for Dirichlet process mixture models. *Journal of the American Statistical Association* 2003;98:224–235.
- Berger JO, Delampady M. Testing precise hypotheses. *Statistical Science* 1987;2:317–352.
- Berger JO, Guglielmi A. Bayesian and conditional frequentist testing of a parametric model versus nonparametric alternatives. *Journal of the American Statistical Association* 2001;96:174–184.
- Blei DM, Jordan MI. Variational inference for Dirichlet process mixtures. *Bayesian Analysis* 2006;1:1–23.
- Bush CA, MacEachern SN. A semiparametric Bayesian model for randomized block designs. *Biometrika* 1996;83:175–185.
- Carota, C.; Parmigiani, G. On Bayes factors for nonparametric alternatives.. In: Bernardo, JM.; Berger, JO.; Dawid, AP.; Smith, AFM., editors. *Bayesian Statistics 5*. Oxford University Press; London: 1996. p. 507–511.
- Carota C, Parmigiani G. Semiparametric regression for count data. *Biometrika* 2002;89:265–281.
- Cifarelli, D.; Regazzini, E. Technical report. Quaderni Istituto Matematica Finanziaria; Turin, Italy: 1978. Non parametrici in condizioni di scambiabilit a parziale e impiego di medie associative..
- De Iorio M, Mller P, Rosner GL, MaEachern SN. An ANOVA model for dependent random measures. *Journal of the American Statistical Association* 2004;99:205–215.
- Dietz EJ. Multivariate generalizations of Jonckheere's test for ordered alternatives. *Communications in Statistics, Part A* 1989;18:3763–3783.
- Dietz EJ, Killeen TJ. A nonparametric multivariate test for monotone trend with pharmaceutical applications. *Journal of the American Statistical Association* 1981;76:169–174.
- Duan, JA.; Guindani, M.; Gelfand, AE. ISDS Discussion Paper 2005–23. Duke University; 2005. Generalized spatial Dirichlet process models..
- Duez P, Dehon G, Kumps A, Dubois J. Statistics of the comet assay: a key to discriminate between genotoxic effects. *Journal of the American Statistical Association* 2003;76:169–174.
- Dunn O. Multiple comparisons using rank sums. *Technometrics* 1964;6:241–252.
- Dunson DB. Bayesian dynamic modeling of latent trait distributions. *Biostatistics* 2006;7:551–568. [PubMed: 16488893]
- Dunson DB, Pillai D, Park J-H. Bayesian density regression. *Journal of the Royal Statistical Society, Series B* 2007;69:163–183.
- Dunson DB, Taylor JA. Approximate Bayesian inference for quantiles. *Journal of Nonparametric Statistics* 2005;17:385–400.
- Dunson DB, Watson M, Taylor JA. Bayesian latent variable models for median regression on multiple outcomes. *Biometrics* 2003;59:296–304. [PubMed: 12926714]
- Escobar MD, West M. Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association* 1995;90:578–588.
- Ferguson TS. A Bayesian analysis of some non-parametric problems. *Annals of Statistics* 1973;1:209–230.
- Ferguson TS. Prior distributions on spaces of probability measures. *Annals of Statistics* 1974;2:615–629.
- Gelfand AE, Kottas A. Nonparametric Bayesian modeling for stochastic order. *Annals of the Institute of Statistical Mathematics* 2001;53:865–876.

- Gelfand, AE.; Kottas, A.; MacEachern, SN. Technical Report AMS 2004–5, Department of Applied Math and Statistics. University of California; Santa Cruz: 2004. Bayesian nonparametric spatial modeling with Dirichlet process mixing..
- Gelfand AE, Smith AFM. Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Society* 1990;85:398–409.
- Giudici P, Mezzetti M, Muliere P. Mixtures of Dirichlet process priors for variable selection in survival analysis. *Journal of Statistical Planning and Inference* 2003;111:101–115.
- Gopalan R, Berry DA. Bayesian multiple comparisons using Dirichlet process priors. *Journal of the American Statistical Association* 1998;93:1130–1139.
- Griffin JE, Steel MFJ. Order-based dependent Dirichlet process. *Journal of the American Statistical Association* 2006;101:179–194.
- Hans C, Dunson DB. Bayesian inferences on umbrella orderings. *Biometrics* 2005;61:1018–1026. [PubMed: 16401275]
- Hoff PD. Bayesian methods for partial stochastic orderings. *Biometrika* 2003;90:303–317.
- Huang P, Tilley BC, Woolson RF, Lipsitz S. Adjusting O'Brien's test to control type I error for the generalized nonparametric Behrens-Fisher problem. *Biometrics* 2005;61:532–539. [PubMed: 16011701]
- Jonckheere AR. A distribution free k-sample test against ordered alternatives. *Biometrika* 1954;41:133–145.
- Kiefer J. *k*-sample analogues of the Kolmogorov-Smirnov and Cramér-v. Mises tests. *Annals of Mathematical Statistics* 1959;30:420–447.
- Kleinman KP, Ibrahim JG. A semiparametric Bayesian approach to the random effects model. *Biometrics* 1998;54:921–938. [PubMed: 9750242]
- MacEachern, SN. *ASA Proceedings of the Section on Bayesian Statistical Science*. American Statistical Association; Alexandria, VA: 1999. Dependent nonparametric processes..
- MacEachern, SN. Dependent Dirichlet processes. Unpublished manuscript, Department of Statistics. The Ohio State University; 2000.
- Mira, A.; Petrone, S. Bayesian hierarchical nonparametric inference for change-point problems. In: Berger, JO.; Bernardo, JM.; Dawid, AP.; Smith, AFM., editors. *Bayesian Statistics 5*. Oxford University Press; London: 1996. p. 609-620.
- Mukhopadhyay S, Gelfand AE. Dirichlet process mixed generalized linear models. *Journal of the American Statistical Association* 1997;92:633–639.
- Müller P, Quintana FA. Nonparametric Bayesian data analysis. *Statistical Science* 2004;19:95–110.
- Müller P, Quintana FA, Rosner G. A method for combining inference across related nonparametric Bayesian models. *Journal of the Royal Statistical Society, Series B* 2004;66:735–749.
- Nickerson RS. Null hypothesis significance testing: a review of an old and continuing controversy. *Psychological Methods* 2000;5:241–301. [PubMed: 10937333]
- O'Brien PC. Procedures for comparing samples with multiple endpoints. *Biometrics* 1984;40:1079–1087. [PubMed: 6534410]
- Pennell ML, Dunson DB. Bayesian semiparametric dynamic frailty models for multiple event time data. *Biometrics* 2006;62:1044–1052. [PubMed: 17156278]
- Sethuraman J. A constructive definition of Dirichlet priors. *Statistica Sinica* 1994;4:639–650.
- Shirley E. A non-parametric equivalent of Williams' test for contrasting increasing dose levels of a treatment. *Biometrics* 1977;33:386–389. [PubMed: 884197]
- Sorenson HW, Alspach DL. Recursive Bayesian estimation using Gaussian sums. *Automatica* 1971;7:465–479.
- Teh YW, Jordan MI, Beal MJ, Blei DM. Hierarchical Dirichlet processes. *Journal of the American Statistical Association* 2006;101:1566–1581.
- Terpstra TJ. The asymptotic normality and consistency of Kendall's test against trend, when ties are present in one ranking. *Indagationes Mathematicae* 1952;14:327–333.
- Tomlinson, G.; Escobar, M. Analysis of densities. Technical report. University of Toronto; 1999.
- Verdinelli I, Wasserman L. Bayesian goodness of fit testing using infinite dimensional exponential families. *Annals of Statistics* 1998;20:1215–1241.

- West, M.; Müller, P.; Escobar, MD. Hierarchical priors and mixture models with application in regression and density estimation.. In: Smith, A.; Freeman, P., editors. Aspects of Uncertainty: A Tribute to D. V. Lindley. Wiley; New York: 1994. p. 363-386.
- Westfall PH, Johnson WO, Utts JM. A Bayesian perspective on the Bonferroni adjustment. *Biometrika* 1997;84:419–427.
- Williams D. A note on Shirley's nonparametric test for comparing several dose levels with a zero-dose control. *Biometrics* 1986;42:183–186. [PubMed: 3719054]

Figure 1. Marginal density of y_1 at each predictor level in simulation cases 1–3.

Figure 2.

Posterior probabilities of global and local null hypotheses for the multivariate analyses in simulation cases 1–3. Labels for y-axes denote the different outcomes used in each analysis. Data points correspond to results from data sets 1 (+), 2 (o), 3 (*), 4 (·), and 5 (x). As a point of reference for evaluating each posterior probability, we have provided a dashed line at 0.1.

Figure 3. Posterior predictive density of % tail DNA and OTM in each dose group. ‘.’ denotes an observation.

Web Figure 1.

Sensitivity analysis of v_0 in our univariate analysis of y_1 . Values on the vertical axis correspond to posterior probabilities of the different null hypotheses. Data points correspond to results from data sets 1 (+), 2 (o), 3 (*), 4 (·) and 5 (x).

Web Figure 2.

Sensitivity analysis of v_0 in our multivariate analysis of y . Values on the vertical axis correspond to posterior probabilities of the different null hypotheses. Data points correspond to results from data sets 1 (+), 2 (o), 3 (*), 4 (·), and 5(×). Mixing was poor for Case 3, data set 2 when $v_0 = 20$.

Table 1

Parameter values for the mixture components in simulation cases 1–3.

Parameter	Case 1				
	1	2	3	4	5
x_j	-0.6	-0.5	-0.4	0.5	1
τ_j	0.16	0.25	0.36	0.903	1.323
γ_{j0}	1	1	1	1	1
γ_{j1}	0	0	0	0	0
Case 2					
x_j	-0.6	-0.5	-0.4	0.5	1
τ_j	0.16	0.25	0.36	0.903	1.323
γ_{j0}	1	1	1	-3	-3
γ_{j1}	-1	-0.25	-0.25	0.25	0.75
Case 3					
x_j	-2	-1	0	1	2
τ_j	0.16	0.25	0.36	0.903	1.323
γ_{j0}	-3	-3	1	-3	-3
γ_{j1}	1	0.5	-1	0.5	1

Table 2

Posterior probabilities of global and local null hypotheses for the univariate analysis of y_1 in simulation cases 1–3. The letters in superscript indicate that the following frequentist tests were significant at the 0.05 level: J=Jonckheere, KW=Kruskal-Wallis, KS=Kolmogorov-Smirnov, S=Shirley, D=Dunn. Since Shirley's test assumes a monotone trend in location, it was not performed when Jonckheere's test was insignificant.

Case 1						
Data set	H_0	H_{01}^*	H_{02}^*	H_{03}^*	H_{04}^*	H_{05}^*
1	0.914	0.987	0.927	0.925	0.918	0.914
2	0.977	0.998	0.996	0.989	0.986	0.977
3	0.979	0.998	0.995	0.990	0.988	0.979
4	0.971	0.988	0.980	0.978	0.976	0.971
5	0.971	0.987	0.983	0.981	0.978	0.971
Case 2						
1	0.066 ^{J,KW,KS}	0.987	0.980	0.952	0.080	0.066 ^{J,KW,KS}
2	<0.001 ^{J,KW,KS}	0.990	0.945	0.888	<0.001 ^{S,D}	<0.001 ^{S,D,KS}
3	0.001 ^{J,KW,KS}	0.990	0.944	0.075	0.004 ^S	0.001 ^{S,D,KS}
4	<0.001 ^{J,KW,KS}	0.994	0.959	0.253	<0.001 ^{S,D}	<0.001 ^{S,D}
5	0.002 ^{J,KW,KS}	0.963	0.914	0.843	0.202	0.002 ^{S,D,KS}
Case 3						
1	<0.001 ^{KS}	0.260	0.193	<0.001 ^{KS}	<0.001 ^{KS}	<0.001 ^{KS}
2	<0.001 ^{KS}	0.799	0.027	<0.001 ^{KS}	<0.001 ^{KS}	<0.001 ^{KS}
3	<0.001 ^{KS}	0.838	<0.001 ^{KS}	<0.001	<0.001 ^{KS}	<0.001 ^{KS}
4	<0.001 ^{KS}	0.637	0.097	<0.001	<0.001 ^{KS}	<0.001 ^{KS}
5	<0.001 ^{KS}	0.864	0.048	<0.001 ^{KS}	<0.001 ^{KS}	<0.001 ^{KS}