

# The Matrix Stick-Breaking Process: Flexible Bayes Meta Analysis

David B. Dunson,<sup>1,\*</sup> Ya Xue<sup>2</sup> and Lawrence Carin<sup>2</sup>

<sup>1</sup>Biostatistics Branch, National Institute of Environmental Health Sciences, U.S. National  
Institutes of Health

<sup>2</sup>Department of Electrical & Computer Engineering, Duke University

\*dunson@stat.duke.edu

In analyzing data from multiple related studies, it is often of interest to borrow information across studies and to cluster similar studies. Although parametric hierarchical models are commonly used, a concern is sensitivity to the form chosen for the random effects distribution. A Dirichlet process (DP) prior can allow the distribution to be unknown, while clustering studies. However, the DP does not allow local clustering of studies with respect to a subset of the coefficients without making independence assumptions. Motivated by this problem, we propose a matrix stick-breaking process (MSBP) as a prior for a matrix of random probability measures. Theoretical properties of the MSBP are considered in detail, and methods are developed for posterior computation using MCMC. Using the MSBP as a prior for a matrix of study-specific regression coefficients, we demonstrate advantages over parametric modeling in simulated examples. The methods are further illustrated using a multinational uterotrophic bioassay study.

**Key Words:** Clustering; Dirichlet process; Hierarchical model; Mixture model; Nonparametric Bayes; Random effects.

## 1. Introduction

In many applications, data are collected for multiple related studies, and interest focuses on borrowing of information and clustering. Although we use the term *study* throughout, the same issues arise in analysis of data from multi-center trials, multiple platform experiments, and other settings. In fact, as high-throughput technology becomes more commonplace, there is increasing interest in automated methods for flexibly combining data collected from different sources, while accommodating heterogeneity.

For subject  $i$  ( $i = 1, \dots, n_m$ ) from study  $m$  ( $m = 1, \dots, M$ ), suppose that data consist of a response variable,  $y_{mi}$  and a vector of predictors,  $\mathbf{x}_{mi} = (x_{mi1}, \dots, x_{mip})'$ . A common strategy for meta analysis is to use a hierarchical model of the form:

$$\begin{aligned} y_{mi} &\sim f(\mathbf{x}_{mi}, \boldsymbol{\theta}_m, \boldsymbol{\phi}) \\ \boldsymbol{\theta}_m &\sim G, \end{aligned} \tag{1}$$

where  $f(\mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\phi})$  is the conditional distribution of  $Y$  given predictors  $\mathbf{x}$  and parameters  $\boldsymbol{\theta}, \boldsymbol{\phi}$ ,  $\boldsymbol{\phi} = (\phi_1, \dots, \phi_q)'$  is a vector of global parameters,  $\boldsymbol{\theta}_m = (\boldsymbol{\theta}_{m1}, \dots, \boldsymbol{\theta}_{mp})'$  is a vector of study-specific parameters, and  $G$  is the distribution of  $\boldsymbol{\theta}_m$  across studies. Typically,  $G$  is assumed to correspond to a normal distribution.

For concreteness, we focus on the case in which  $y_{mi} = \mathbf{x}_{mi}'\boldsymbol{\theta}_m + \epsilon_{mi}$ , with  $\epsilon_{mi} \sim N(0, \sigma^2)$ , so that  $\boldsymbol{\theta}_m$  is a vector of study-specific regression coefficients. Then, our interest focuses on borrowing of information across studies in estimating the study-specific coefficients, while also clustering studies separately for the different coefficients. Our motivation is drawn from a multinational validation study of the rodent uterotrophic bioassay, which is a system for identifying suspected agonists or antagonists of estrogen (Kanno et al., 2001). The Kanno et al. (2001) study collected data from 19 laboratories, with predictors including protocol type, dose of a known agonist (EE), and dose of a known antagonist (ZM). The primary goal of the study was to assess heterogeneity across labs in the different predictor effects, while

clustering labs with similar coefficients.

There is a rich Bayesian literature focusing on the case in which  $G$  corresponds to the normal distribution, so that  $\boldsymbol{\theta}_m \sim N_p(\boldsymbol{\theta}_0, \boldsymbol{\Omega})$ . Refer, for example, to Smith, Spiegelhalter and Thomas (1995). In addition, there is a growing literature on using nonparametric Bayes methodology to allow uncertainty in the random effects distribution,  $G$ . A natural choice of prior for  $G$  that induces clustering of studies having identical coefficients is the Ferguson (1973, 1974) Dirichlet process (DP). Refer to Bush and MacEachern (1996), Mukhopadhyay and Gelfand (1997), Kleinman and Ibrahim (1998) and Ishwaran and Takahara (2002) for articles on using the DP as a prior for random effects distributions. Burr and Doss (2005) recently applied the conditional Dirichlet process (Doss, 1985) to the problem of inference on the median of the random effects distribution.

To illustrate an important drawback of the DP in the bioassay application, suppose:

$$\boldsymbol{\theta}_m \stackrel{iid}{\sim} G, \quad G \sim DP(\alpha G_0), \quad (2)$$

where  $\alpha$  is a precision parameter and  $G_0$  is the base measure of the DP. Following the Pólya urn result of Blackwell and MacQueen (1973), expression (2) implies that studies  $m$  and  $m'$  are clustered together with prior probability,  $\Pr(\boldsymbol{\theta}_m = \boldsymbol{\theta}_{m'}) = 1/(1 + \alpha)$ . Hence, this formulation clusters the coefficients for all the predictors simultaneously and does not permit *local* predictor-specific clustering, which is of primary interest in the Kanno et al. (2001) study. The conditional DP approach of Burr and Doss (2005) also does not address this problem.

As an alternative approach to (2), we could specify independent DP priors for the coefficients as follows:

$$\theta_{mj} \stackrel{iid}{\sim} G_j, \quad G_j \sim DP(\alpha_j G_{0j}). \quad (3)$$

This approach allows differential clustering of the coefficients for different predictors, so is more flexible. However, independence is assumed across the predictors. This is unappealing,

because  $\theta_{mj} = \theta_{m'j}$  provides information that studies  $m$  and  $m'$  are similar, which should intuitively increase the probability that  $\theta_{mj'} = \theta_{m'j'}$ , for  $j \neq j'$ .

Motivated by borrowing of information across related nonparametric Bayesian models, Müller, Quintana and Rosner (2004) proposed an approach for incorporating dependence in  $G_j$  and  $G_{j'}$  by using a mixture specification in which  $G_j \sim \pi_0 F_0 + (1 - \pi_0) F_j$ , with  $0 \leq \pi \leq 1$  a mixture probability,  $F_0 \sim DP(\alpha F_0^*)$  a global distribution, and  $F_j \sim DP(\beta F_j^*)$  a local deviation. This approach allows dependence by incorporating shared atoms (contained in  $F_0$ ) within the distributions  $G_1, \dots, G_p$ . A hierarchical DP (HDP), which also shares atoms across dependent distributions, was proposed by Teh et al. (2006). An alternative strategy, which allows the atoms in the different distributions to be dependent but not identical, was proposed by De Iorio et al. (2004), relying on the dependent DP (DDP) of MacEachern (1999, 2000).

These methods are not appropriate for the problem we are focusing on, because the regression coefficients for the different predictors are intrinsically on different scales. Hence, it is not reasonable to borrow information by allowing identical coefficients for different predictors. Instead, we propose a novel approach for borrowing of information across predictors in the clustering process using a matrix stick-breaking process (MSBP). We follow recent authors (Griffin and Steel, 2006; Duan et al., 2006; Dunson, Pillai and Park, 2007) by incorporating dependency through the stick-breaking weights. However, our approach is fundamentally different from these approaches, and has different applications.

In Section 2, we provide a brief review of properties of the Dirichlet process and propose our prior structure, considering theoretical properties. In Section 3, we develop methods for efficient posterior computation. Section 4 contains simulation examples, Section 5 applies the method to data from the Kanno et al. (2001) study, and Section 6 discusses the results. Proofs are included in an Appendix.

## 2. Matrix Stick-Breaking Process

### 2.1 Background

The Dirichlet process (DP) prior has been increasingly used as a prior for clustering in regression applications. Dunson et al. (2006) used the DP to borrow information across functionally-related genes in assessing the health impact of single nucleotide polymorphisms (SNPs). MacLehose et al. (2006) applied a similar approach to epidemiologic applications involving multiple, correlated exposures. Kim, Tadesse and Vannucci (2006) used a DP for simultaneous variable selection and clustering.

We provide a brief background to illustrate some of the properties of the DP. Focusing on the  $p = 1$  special case of expression (3), we have  $\theta_m \stackrel{iid}{\sim} G$ , with  $G \sim DP(\alpha G_0)$ , for  $m = 1, \dots, M$ . The Blackwell and MacQueen (1973) result implies that the conditional distribution of  $\theta_m$  given  $\boldsymbol{\theta}_{(-m)} = \{\theta_{m'} : m' \neq m\}$  but marginalizing out the unknown  $G$  is:

$$(\theta_m | \boldsymbol{\theta}_{(-m)}) \sim \left( \frac{\alpha}{\alpha + M - 1} \right) G_0 + \left( \frac{1}{\alpha + M - 1} \right) \sum_{l \neq m} \delta_{\theta_l}, \quad (4)$$

where  $\delta_\theta$  is a probability measure concentrated at  $\theta$ . Hence, the coefficient for the predictor in study  $m$  is assigned to a new cluster with probability  $\alpha/(\alpha + M - 1)$ , and otherwise is assigned the same value as the coefficients in one or more of the other studies.

The stick-breaking formulation of the DP (Sethuraman, 1994) is also useful for inferring properties. Assuming  $G \sim DP(\alpha G_0)$ , we have

$$G = \sum_{h=1}^{\infty} \left\{ V_h \prod_{l < h} \bar{V}_l \right\} \delta_{\Theta_h}, \quad V_h \stackrel{iid}{\sim} \text{beta}(1, \alpha), \quad \Theta_h \stackrel{iid}{\sim} G_0, \quad (5)$$

where  $\mathbf{V} = \{V_h, h = 1, \dots, \infty\}$  is an infinite sequence of stick-breaking random probabilities,  $\bar{V}_h = 1 - V_h$ , and  $\boldsymbol{\Theta} = \{\Theta_h, h = 1, \dots, \infty\}$  is an infinite sequence of random atoms. Because the probability weights,  $\pi_h = V_h \prod_{l < h} \bar{V}_l$ , are stochastically decreasing in  $h$ , an accurate approximation to (5) can be obtained by truncating the infinite sum at  $N$  terms, with  $V_N = 1$  (Ishwaran and James, 2001, and references therein).

## 2.2 Proposed Formulation

The stick-breaking characterization in (5) has formed the starting point for a number of extensions of the DP to accommodate dependent collections of random probability measures. The DDP formulation of MacEachern (1999, 2000), which has been implemented by De Iorio et al. (2004) and Gelfand et al. (2005) in different settings, relies on assuming fixed probability weights,  $\boldsymbol{\pi} = \{\pi_h, h = 1, \dots, \infty\}$ , while allowing the atoms to vary across probability measures in the collection according to a stochastic process. Such an approach implicitly assumes that the allocation of individuals to clusters does not change, but the cluster-specific parameters may vary.

Griffin and Steel (2006) and Duan et al. (2006) instead propose models to allow changes in  $\boldsymbol{\pi}$ , while preserving the marginal DP property for any particular distribution in the collection. Griffin and Steel’s approach relies on incorporating order-dependence in the stick-breaking construction, while Duan et al. use a multivariate beta distribution for flexible modeling of spatial data. Here, we propose a *matrix stick-breaking process* (MSBP), which is motivated by the desire to borrow information across predictors and studies simultaneously.

In particular, we borrow information by incorporating dependence in the prior distributions for the coefficients  $\{\theta_{mj}\}$ . We start by assuming

$$\begin{aligned} \theta_{mj} &\stackrel{ind}{\sim} G_{mj}, \quad m = 1, \dots, M, j = 1, \dots, p \\ \mathcal{G} &\sim \mathcal{P}, \end{aligned} \tag{6}$$

where  $\mathcal{G} = \{G_{mj}, m = 1, \dots, M, j = 1, \dots, p\}$  is a matrix of random probability measures, and  $\mathcal{P}$  is a probability measure on  $(\Omega, \mathcal{F})$ , with  $\Omega$  the space of  $M \times p$  matrices with the  $(m, j)$  element a probability measure on  $(\mathcal{X}_j, \mathcal{B}_j)$ . Here,  $\mathcal{F}$  is a  $\sigma$ -algebra of subsets of  $\Omega$ ,  $\theta_{mj} \in \mathcal{X}_j$  (typically,  $\mathcal{X}_j = \mathfrak{R}$ ), and  $\mathcal{B}_j$  is a Borel  $\sigma$ -algebra of subsets of  $\mathcal{X}_j$ .

Our focus is on the specification for  $\mathcal{P}$ . Assuming each element in  $\mathcal{G}$  has a stick-breaking

representation, we let:

$$G_{mj} = \sum_{h=1}^N \left\{ V_{mjh} \prod_{l<h} \bar{V}_{mjl} \right\} \delta_{\Theta_{jh}}, \quad \Theta_{jh} \stackrel{ind}{\sim} G_{0j}, \quad (7)$$

where  $\mathbf{V} = \{V_{mjh}, m = 1, \dots, M, j = 1, \dots, p, h = 1, \dots, N\}$  is an array of random stick-breaking weights, and  $\Theta = \{\Theta_{jh}\}$  is a  $p \times N$  matrix of random atoms. The rows ( $j = 1, \dots, p$ ) of  $\Theta$  correspond to the different predictors, while the columns ( $h = 1, \dots, N$ ) correspond to the different clusters. Here,  $V_{mjN} = 1$ , for all  $m, j$ , to ensure that the elements of  $\pi_{mj} = \{V_{mjh} \prod_{l<h} \bar{V}_{mjl}, h = 1, \dots, N\}$  sum to one, for each  $m, j$ , so that (7) is a valid probability measure.

Dependence within rows and columns of  $\mathcal{G}$  will be incorporated through (1) dependent stick-breaking weights and (2) a common parametric prior,  $G_0 = \bigotimes_{j=1}^p G_{0j}$ , across the different studies. Focusing on the stick-breaking component, we let

$$V_{mjh} = U_{mh} W_{jh}, \quad U_{mh} \sim \text{beta}(1, \alpha), \quad W_{jh} \sim \text{beta}(1, \beta), \quad (8)$$

so that the probability,  $V_{mjh}$ , is decomposed into the product of  $U_{mh}$ , which measures the tendency to allocate study  $m$  to cluster  $h$ , and  $W_{jh}$ , which measures the tendency to allocate coefficients for predictor  $j$  to cluster  $h$ . We use  $\text{MSBP}_N(\alpha, \beta, G_0)$  to denote the choice of  $\mathcal{P}$  specified in (7) and (8), with  $\text{MSBP}(\alpha, \beta, G_0)$  denoting the limiting case as  $N \rightarrow \infty$ .

To motivate this choice, it is useful to consider some special cases. First, note that in the limit as  $\alpha \rightarrow 0$ ,  $U_{mh} \rightarrow 1$  a.s. and  $V_{mjh} = W_{jh}$ . Then, we have

$$G_{mj} = \sum_{h=1}^N \left\{ W_{jh} \prod_{l<h} \bar{W}_{jl} \right\} \delta_{\Theta_{jh}} = G_j,$$

where  $\Theta_{jh} \stackrel{ind}{\sim} G_{0j}$  and  $W_{jh} \stackrel{ind}{\sim} \text{beta}(1, \beta)$ . Note that this is the truncation approximation to a Dirichlet process prior, as described by Ishwaran and James (2001), and in the limit as  $N \rightarrow \infty$ ,  $G_j \sim DP(\beta G_{0j})$ . This special case corresponds to choosing independent DP priors for the distribution of the regression coefficients for the different predictors. Hence,

there is no borrowing of information across the predictors, only within a particular predictor across the studies. In the further special case in which  $\alpha \rightarrow 0$  and  $\beta \rightarrow \infty$ , we instead have  $\theta_{mj} \sim G_{0j}$ , which corresponds to a parametric hierarchical model. In addition, when  $\alpha \rightarrow 0$  and  $\beta \rightarrow 0$ , we instead have that  $\theta_{mj} = \theta_j$ , so that the coefficients for the different studies are identical and the data are pooled.

Borrowing of information across predictors occurs for  $\alpha > 0$ , with the random variables  $\mathbf{U}_m = \{U_{mh}, h = 1, \dots, N\}$  controlling the tendency of coefficients for study  $m$  to be allocated to particular clusters with high probability. The dependence structure and other properties are described in detail in the next subsection focusing on the case in which  $N \rightarrow \infty$ .

### 2.3 Basic Properties

Letting  $\pi_{mjh} = V_{mjh} \prod_{l < h} \bar{V}_{mjl}$ , for  $m = 1, \dots, M, j = 1, \dots, p, h = 1, \dots, \infty$ ,  $G_{mj}$  is a well defined probability measure if and only if the random weights  $\boldsymbol{\pi}_{mj} = \{\pi_{mjh}, h = 1, \dots, \infty\}$  sum to one almost surely.

*Lemma 1.* For  $\mathcal{G} = \{G_{mj}, m = 1, \dots, M, j = 1, \dots, p\}$ , with the elements defined in (7) and (8) for  $N \rightarrow \infty$ , we have  $\sum_{h=1}^{\infty} \pi_{mjh} = 1$  a.s. for all  $m, j$ .

Theorem 1 provides the prior mean and variance of the random measure  $G_{mj}$ .

*Theorem 1.* Letting  $G_{mj}$  denote the random measure defined in (7) and (8), for  $N \rightarrow \infty$ , and  $A \in \mathcal{B}_j$ , we have

$$\mathbb{E}\{G_{mj}(A)\} = G_{0j}(A), \quad \mathbb{V}\{G_{mj}(A)\} = \frac{2}{(\alpha + 2)(\beta + 2) - 2} G_{0j}(A) \{1 - G_{0j}(A)\}.$$

Note that the prior for  $G_{mj}$  is centered on  $G_{0j}$ , which corresponds to a probability measure obeying a parametric law. For example, a convenient choice is

$$G_{0j}(A) = \int_A (2\pi\psi_j^{-1})^{-1/2} \exp\{-\psi_j/2(z - \mu_j)^2\} dz.$$

In this case, the prior is centered on a normal hierarchical model having  $\theta_{mj} \sim N(\mu_j, \psi_j^{-1})$ . Confidence in this normal model is controlled by the precision parameters  $\alpha$  and  $\beta$ , with  $V\{G_{mj}(A)\} \rightarrow 0$  in the limit as either  $\alpha$  or  $\beta \rightarrow \infty$ .

Theorem 2 characterizes the correlation between the random measures,  $G_{mj}$  and  $G_{m'j}$ , corresponding to the priors on the  $j$ th coefficient for studies  $m$  and  $m'$ .

*Theorem 2.* Letting  $\mathcal{G}$  denote the array of random measures defined in (7) and (8), for  $N \rightarrow \infty$ , and  $A \in \mathcal{B}_j$ , we have

$$\rho = \text{corr}\{G_{mj}(A), G_{m'j}(A)\} = \frac{\alpha + \beta + \alpha\beta/2 + 1}{2\alpha + \beta + \alpha\beta + 1}.$$

The expression in theorem 2 is particularly useful in being free from the set  $A$ , so that it can be used as a general summary of correlation in the random measures. Focusing on limiting cases, we obtain (i)  $\lim_{\alpha \rightarrow 0} \rho = 1$ , (ii)  $\lim_{\beta \rightarrow 0} \rho = (1 + \alpha)/(1 + 2\alpha)$ , (iii)  $\lim_{\alpha \rightarrow \infty} \rho = 0$ ; and (iv)  $\lim_{\beta \rightarrow \infty} \rho = 0$ . In general,  $0 \leq \rho \leq 1$ , with the correlation coefficient increasing as  $\alpha$  decreases.

#### 2.4 Truncation Approximations

The  $N \rightarrow \infty$  formulation of the MSBP is appealing in avoiding the need to choose a bound  $N$  on the number of components. However, in practice, computation for the infinite-dimensional specification is infeasible and it is useful to consider finite  $N$  approximations. In this subsection, we assess the approximation error using a related approach to Ishwaran and James (2001).

*Theorem 3.* Let  $\boldsymbol{\pi}_{mj} = \{\pi_{mjh}, h = 1, \dots, \infty\}$  denote the random weights within the measure  $G_{mj}$ , where  $\mathcal{G} \sim \text{MSBP}(\alpha, \beta, G_0)$ . For any  $N, r \geq 1$ , let

$$\Gamma_{mj}(N, r) = \left( \sum_{h=N}^{\infty} \pi_{mjh} \right)^r, \quad \Upsilon_{mj}(N, r) = \sum_{h=N}^{\infty} (\pi_{mjh})^r.$$

Then

$$E\{\Gamma_{mj}(N, r)\} = \left\{ \sum_{k=0}^r C_r^k (-1)^{r-k} \mu_{r-k}(\alpha) \mu_{r-k}(\beta) \right\}^{N-1},$$

$$E\{\Upsilon_{mj}(N, r)\} = \frac{\mu_r(\alpha) \mu_r(\beta) E\{\Gamma_{mj}(N, r)\}}{1 - \sum_{k=0}^r C_r^k (-1)^{r-k} \mu_{r-k}(\alpha) \mu_{r-k}(\beta)},$$

where  $C_r^k = \frac{r!}{(r-k)!k!}$  and  $\mu_r(\lambda) = \prod_{l=1}^r l/(l + \lambda)$  is the  $r$ th non-central moment of  $\text{beta}(1, \lambda)$ , with  $\mu_0(a, b) \equiv 1$ .

Note that the expressions for  $E\{\Gamma_{mj}(N, r)\}$  and  $E\{\Upsilon_{mj}(N, r)\}$  are free of  $m, j$ , so the subscripts can be excluded in discussing these expectations. An accurate truncation approximation can be produced when  $E\{\Gamma(N, r)\} \approx 0$  and  $E\{\Upsilon(N, r)\} \approx 0$  for all  $r$ . In general, these quantities decay to 0 exponentially fast with increasing  $N$ , with the rate of decay increasing as  $\alpha$  and  $\beta$  decrease. For example,

$$E\{\Gamma(N, 1)\} = \left\{ 1 - \left( \frac{1}{1 + \alpha} \right) \left( \frac{1}{1 + \beta} \right) \right\}^{N-1}.$$

Values of  $\alpha$  and  $\beta$  less than one are typically recommended in applications, as discussed in the following sections. A reasonable strategy for choosing  $N$  is to plug in an upper bound on  $\alpha$  and  $\beta$ , and then choose  $N$  so that  $E\{\Gamma(N, 1)\} = \epsilon$ , for some arbitrarily small positive constant  $\epsilon$ .

## 2.5 Clustering Properties

As discussed in Section 2.2, a primary motivation for the MSBP over the DP is that the event,  $\theta_{mj} = \theta_{m'j}$ , provides information that studies  $m$  and  $m'$  are similar, which should lead to an increased prior probability of  $\theta_{mj'} = \theta_{m'j'}$ , for any  $j' \neq j$ . This property is apparent from Theorem 4, which also provides closed form expressions for marginal and conditional prior clustering probabilities in terms of the precision parameters,  $\alpha$  and  $\beta$ .

*Theorem 4.* Under the  $\mathcal{G} \sim \text{MSBP}(\alpha, \beta, G_0)$  prior, the probability that studies

$m$  and  $m'$  have identical coefficients for predictor  $j$  is

$$\Pr(\theta_{mj} = \theta_{m'j}) = \frac{1}{(\alpha + 1)(\beta + 2) - 1},$$

while the corresponding conditional probability given  $\theta_{mj'} = \theta_{m'j'}$  is

$$\begin{aligned} & \Pr(\theta_{mj} = \theta_{m'j'} \mid \theta_{mj'} = \theta_{m'j'}) \\ = & \frac{2^{(\beta+1)[(\alpha+2)^2(\beta+2)-4]}}{2^{(\alpha+2)(\beta+2)[(\alpha+1)(\beta+1)+\alpha][(\alpha+1)(\beta+1)+\beta]-[(\alpha+2)(\beta+2)-2][(\alpha+2)(\beta+2)-4]}. \end{aligned}$$

In addition, we have  $\Pr(\theta_{mj} = \theta_{m'j}) < \Pr(\theta_{mj} = \theta_{m'j'} \mid \theta_{mj'} = \theta_{m'j'})$ .

From the simple expression for  $\Pr(\theta_{mj} = \theta_{m'j})$ , it is clear that the clustering probability ranges between 0 and 1 depending on the values of  $\alpha$  and  $\beta$ , converging to 1 in the limit as  $\alpha, \beta \rightarrow 0$  and to 0 as either  $\alpha$  or  $\beta \rightarrow \infty$ . As expected given the results in Section 2.2,

$$\lim_{\alpha \rightarrow 0} \Pr(\theta_{mj} = \theta_{m'j'} \mid \theta_{mj'} = \theta_{m'j'}) = \frac{1}{\beta + 1},$$

which corresponds to the clustering probability in the special case of  $\theta_{mj} \stackrel{iid}{\sim} G_j$ , for  $m = 1, \dots, M$ , with  $G_j \sim DP(\beta G_{0j})$ , independently for  $j = 1, \dots, p$ . If  $\alpha \rightarrow \infty$  or  $\beta \rightarrow \infty$ ,  $\Pr(\theta_{mj} = \theta_{m'j} \mid \theta_{mj'} = \theta_{m'j'}) = \Pr(\theta_{mj} = \theta_{m'j}) = 0$ , and none of the studies are clustered together, so that borrowing of information relies entirely on the base parametric model.

### 3. Posterior Computation

For posterior computation, we propose a modification of the blocked Gibbs sampling algorithm of Ishwaran and James (2001), relying on an  $MSBP_N(\alpha, \beta, G_0)$  prior and using a data augmentation scheme to facilitate efficient updating. We introduce several latent indicator vectors. First, let  $R_{mj} = h$  denote that  $\theta_{mj} = \Theta_{jh}$ , so that the coefficient for study  $m$  is allocated to the  $h$ th cluster. Under expressions (6) - (8),  $R_{mj}$  can be expressed as  $R_{mj} = \min\{l : S_{mjl} = T_{mjl} = 1\}$ , with  $S_{mjl} \stackrel{iid}{\sim} \text{Bernoulli}(U_{mh})$  and  $T_{mjl} \stackrel{iid}{\sim} \text{Bernoulli}(W_{jh})$ , for  $l = 1, \dots, N - 1$ , and  $S_{mjN} = T_{mjN} = 1$ . After augmenting the data in this manner, it is straightforward to update each of the unknowns based on their full conditional posterior distributions.

The conditional posterior distributions used in implementing the MCMC algorithm are:

1. The conditional for  $\Theta_{jh}$ , for  $j = 1, \dots, p$  and  $h = 1, \dots, N$ , is proportional to

$$G_{0j}(\Theta_{jh}) \prod_{m=1}^M \prod_{i=1}^{n_m} \{f(y_{mi}; \mathbf{x}_{mi}, \theta_{mj} = \Theta_{jh}, \boldsymbol{\theta}_{m(-j)}, \phi)\}^{1(R_{mj}=h)}, \quad (9)$$

where  $\boldsymbol{\theta}_{m(-j)}$  denotes the subvector of  $\boldsymbol{\theta}_m$  excluding element  $j$ . This conditional is normal for normal  $G_0$  and normal linear model likelihoods.

2. The conditional for  $R_{mj}$ , for  $j = 1, \dots, p, m = 1, \dots, M$ , is multinomial with

$$\Pr(R_{mj} = h | -) = \frac{\pi_{mjh} \prod_{i=1}^{n_m} f(y_{mi}; \mathbf{x}_{mi}, \theta_{mj} = \Theta_{jh}, \boldsymbol{\theta}_{m(-j)}, \phi)}{\sum_{l=1}^N \pi_{mjl} \prod_{i=1}^{n_m} f(y_{mi}; \mathbf{x}_{mi}, \theta_{mj} = \Theta_{jl}, \boldsymbol{\theta}_{m(-j)}, \phi)}. \quad (10)$$

3. The conditional for  $(S_{mjh}, T_{mjh})$ , for  $h = 1, \dots, R_{mj}$ , sets  $S_{mjh} = T_{mjh} = 1$ , for  $h = R_{mj}$ , and otherwise samples with probabilities:  $\kappa_{st} = \Pr(S_{mjh} = s, T_{mjh} = t)$ , where

$$\kappa_{00} = \frac{(1 - U_{mh})(1 - W_{jh})}{1 - U_{mh}W_{jh}}, \quad \kappa_{10} = \frac{U_{mh}(1 - W_{jh})}{1 - U_{mh}W_{jh}}, \quad \kappa_{01} = \frac{(1 - U_{mh})W_{jh}}{1 - U_{mh}W_{jh}}.$$

4. The conditional distributions for  $U_{mh}$ ,  $m = 1, \dots, M, h = 1, \dots, N - 1$ , and  $W_{jh}, j = 1, \dots, p, h = 1, \dots, N - 1$ , have the simple forms:

$$\begin{aligned} (U_{mh} | -) &\sim \text{beta}\left(1 + \sum_{j:R_{mj} \geq h} S_{mjh}, \alpha + \sum_{j:R_{mj} \geq h} (1 - S_{mjh})\right), \\ (W_{jh} | -) &\sim \text{beta}\left(1 + \sum_{m:R_{mj} \geq h} T_{mjh}, \beta + \sum_{m:R_{mj} \geq h} (1 - T_{mjh})\right). \end{aligned}$$

5. Finally, the conditional distribution of  $\phi$  is proportional to

$$f(\phi) \prod_{m=1}^M \prod_{i=1}^{n_m} f(y_{mi}; \mathbf{x}_{mi}, \boldsymbol{\theta}_m, \phi), \quad (11)$$

which is an gamma distribution when  $\phi$  is an error precision in a normal linear hierarchical model, and  $f(\phi)$  is a gamma prior.

Note that this algorithm can easily be adapted to allow categorical  $y_{mi}$  using the method of Albert and Chib (1993).

#### 4. Simulation Example

We first considered a simple simulation example to illustrate the approach. We assumed  $y_{mi} \sim N(\mathbf{x}'_{mi}\boldsymbol{\theta}_m, \phi^{-1})$ , with  $p = 5, n_m = 15, M = 8, \phi = 16$ , and

$$\begin{aligned}
 \boldsymbol{\theta}_1 &= (1 \quad 1 \quad 1 \quad 1 \quad 1)' \\
 \boldsymbol{\theta}_2 &= (1 \quad 1 \quad 1 \quad 1 \quad -1)' \\
 \boldsymbol{\theta}_3 &= (1 \quad 1 \quad -1 \quad -1 \quad 2)' \\
 \boldsymbol{\theta}_4 &= (1 \quad 1 \quad -1 \quad -1 \quad -2)' \\
 \boldsymbol{\theta}_5 &= (-1 \quad -1 \quad 2 \quad 2 \quad 3)' \\
 \boldsymbol{\theta}_6 &= (-1 \quad -1 \quad 2 \quad 2 \quad -3)' \\
 \boldsymbol{\theta}_7 &= (-1 \quad -1 \quad -2 \quad -2 \quad 4)' \\
 \boldsymbol{\theta}_8 &= (-1 \quad -1 \quad -2 \quad -2 \quad -4)'
 \end{aligned} \tag{12}$$

The  $MSBP_N(\alpha, \beta, G_0)$  model is implemented for this synthetic data set, with  $N = 20$ . The base distribution is specified as  $G_{0j} \sim N(\mu_j, \psi_j^{-1})$ , where  $\mu_j \sim N(0, 1)$  and  $\psi_j \sim Ga(1, 1)$  for all  $j$ . We place a Gamma prior  $Ga(0.1, 0.1)$  on  $\phi$ . As noted in Section 2.5, the precision parameters,  $\alpha$  and  $\beta$ , control the prior distribution on the number of clusters. In order for the data to inform more strongly about the clustering, we chose  $Ga(1, 1)$  hyperpriors on  $\alpha$  and  $\beta$ . The hyperparameter values were chosen to favor few clusters.

The Gibbs sampling algorithm described in Section 3 is used to obtain samples of the posteriors under the MSBP. The results shown below are based on 100,000 samples obtained after a burn-in period of 5,000 iterations. Rapid convergence has been observed in the diagnostic tests as described in Geweke (1992) and Raftery and Lewis (1992). In addition, mixing was good.

Figure 1 plots the posterior probability of two studies being assigned to the same cluster separately for each of the five predictors. The size of the shaded box is proportional to the posterior probability of pairwise clustering. It is apparent that the true clustering structure is well represented, with studies having the same coefficient for a predictor clustered together with high probability. Figure 2 shows the study-specific posterior means and 95% credible

intervals for each of the parameters in red, along with the true values (black) and results for a normal hierarchical analysis (green). The normal analysis corresponded to letting  $\alpha \rightarrow 0$  and  $\beta \rightarrow \infty$ , as described in Section 2.2.

It is clear that the posterior densities are concentrated around the true values. In addition, the 95% credible intervals from the MSBP analysis are narrower in each case, with the difference considerable when uncertainty in the parameters is high for the base parametric model. Although it is somewhat counter-intuitive that the parametric model would lead to higher estimates of uncertainty, this likely arises because there is less borrowing of information under the normal prior, as the variance needs to be high to reflect the heterogeneity in the data.

## 5. Multi-species Bioassay Application

### 5.1 Data Structure and Model

We illustrate the methodology using the uterotrophic bioassay study introduced in Section 1. To provide some additional information, data were collected for 2681 female rats from studies conducted in 19 laboratories from 8 nations. There were four protocols, with two relying on an immature female rat model and two using an adult ovariectomized rat model. Under each protocol, there were 11 treatment groups, with 6 animals per group and the groups including an untreated control, a vehicle control, and seven dose levels of EE, with the final two groups having both EE and ZM exposure. Basic information of participating laboratories can be found in Table 1. Refer to Kanno et al. (2001) for more details.

The outcomes of the bioassay were wet and blotted rat uterus weights. To reduce measurement error, we focus here on the blotted uterus weights. The bioassay data can be modeled using  $y_{mi} \sim N(\mathbf{x}'_{mi}\boldsymbol{\theta}_m, \phi^{-1})$ , where  $y_{mi}$  is the log-transformed blotted uterus weight for rat  $i$  in lab  $m$  and  $\mathbf{x}_{mi} = (x_{mi1}, \dots, x_{mi6})'$ , with  $x_{mi1} = 1$ ,  $x_{mi2}, x_{mi3}, x_{mi4}$  0/1 indicators of protocol B, C, and D, respectively,  $x_{mi5}$  dose of EE, and  $x_{mi6}$  dose of ZM.

Notation	Lab Name	Country	Protocols Conducted
F1	Citifrance	France	$A$
F2	Poulenc	France	$A$
K1	ChungKorea	Korea	$A, B$
K2	KoreaPark	Korea	$B, C$
G1	Berlin	Germany	$A$
G2	Basf	Germany	$A$
G3	Bayer	Germany	$A$
J1	Citijapan	Japan	$A, B, C$
J2	Hatano	Japan	$A, B, C, C'$
J3	InEnvTox	Japan	$A, B, C, C'$
J4	Mitsubishi	Japan	$A, B, C, C'$
J5	Nihon	Japan	$A, B, C, C'$
J6	Sumitomo	Japan	$A, B, C$
N1	Denmark	Netherlands	$B$
N2	TNO	Netherlands	$A, B$
B1	Huntingdon	UK	$C$
B2	Zeneca	UK	$A, B, C$
U1	Exxon	USA	$A$
U2	WIL	USA	$A, B$

Table 1: Information of 19 participating laboratories.

The primary focus of the study is on assessing heterogeneity among the labs in the effects of the different predictors, with a particular emphasis on assessing variability in the slopes,  $\theta_{m5}$  and  $\theta_{m6}$ . With this goal in mind, we repeated the analysis conducted in Section 4 for the simulation example, using the same priors and computational implementation.

## 5.2 Analysis and Results

For sake of brevity, we focus our discussion on the results for the intercept ( $\theta_{m1}$ ) and two dose effects ( $\theta_{m2}, \theta_{m3}$ ). Figure 3 presents pairwise posterior probabilities of labs being clustered together for these coefficients. Note that two labs being assigned to the same cluster implies an identical value for the regression coefficient, though this *soft* probabilistic clustering leads to posterior mean estimates that are different for the different labs.

From Figure 3, it is apparent that there is substantial evidence of heterogeneity among the labs in the intercept, as the posterior probabilities of certain labs being clustered together is small. For example, the clustering probability is close to zero for labs U1 and G2. However,

there is a large group of labs, which all have a moderate to high probability of being clustered together. These results are consistent with exploratory plots of the data and with our prior knowledge that rodent body weights can vary across labs. Such variability is a well known problem in carcinogenicity studies, as body weight is an important predictor of tumor response. However, in the current study, we are more interested in assessing heterogeneity in the estimated dose response across labs.

It is clear from Figure 3 that the EE slope is consistent across labs, as each of the labs has a moderate to high posterior probability of being clustered together with any of the other labs. These results are reassuring that different labs should obtain consistent results in future uterotrophic bioassay studies seeking to identify chemicals having estrogen agonist effects. The results for the estrogen antagonist, ZM, are somewhat less consistent. Although most pairs of labs having a high posterior probability of being clustered together for the ZM coefficient, there are a couple of labs that have slight divergent results.

Figure 3 only provides pairwise probabilities of being clustered together. If two labs have a low probability of being clustered together that does not necessarily imply that the coefficients for those labs have a biologically significant difference. To assess the magnitude of the difference, we plot the lab-specific coefficients and 95% credible intervals in Figure 4. The red lines are the results for the MSBP, while the green lines provide the results under the base parametric normal model. As in the simulation study, the normal model results tend to have wider credible intervals. From this figure, we can see that there are considerable differences in the intercepts, with labs F2 and G2 having control animals with low uterus weights, and labs F1, K1 and U1 having unusually high uterus weights. However, the variability across labs in the EM and ZM coefficients is not biologically significant, taking into account the level of uncertainty in the lab-specific coefficients.

Figure 5 plots the posterior densities of  $\alpha$  and  $\beta$ . As anticipated, the data did inform about the values and the posteriors deviated considerably from the priors. Although sensitiv-

ity analyses are excluded for reasons of space, the results were robust to reasonable changes in the hyperparameter values.

## 6. Discussion

Motivated by the problem of flexible borrowing information across predictors and studies, this article has proposed a new class of priors for a matrix of random probability measures. The proposed matrix stick-breaking process is a natural generalization of stick-breaking representations of the Dirichlet process. In addition to providing a flexible semiparametric approach for Bayesian meta analysis and modeling of data from multicenter studies, the MSBP should be broadly useful for borrowing strength across related semiparametric models. For example, we have obtained good results in machine learning applications involving multitask learning. The proposed computational implementation is efficient in cases we have considered and is no more difficult to implement than Gibbs samplers used for DP mixture models (Ishwaran and James, 2001).

## Appendix: Proofs of Lemmas and Theorems

### A.1 Proof of Lemma 1

Following a closely related approach to Ishwaran and James (2001), we first note that

$$1 - \sum_{h=1}^{N-1} \pi_{mjh} = \prod_{h=1}^{N-1} (1 - V_{mjh}).$$

Taking the limit as  $N \rightarrow \infty$ , we then have

$$1 - \sum_{h=1}^{\infty} \pi_{mjh} = \prod_{h=1}^{\infty} (1 - V_{mjh}),$$

so that Lemma 1 follows by showing  $\prod_{h=1}^{\infty} (1 - V_{mjh}) = 0$  or equivalently

$$\sum_{h=1}^{\infty} \log(1 - V_{mjh}) = \sum_{h=1}^{\infty} \log(1 - U_{mh}W_{jh}) = -\infty.$$

Because  $\{\log(1 - V_{mjh}), h = 1, \dots, \infty\}$  are independent random variables, this expression

holds iff  $\sum_{h=1}^{\infty} \mathbb{E}\{\log(1 - V_{mjh})\} = -\infty$  by the Kolmogorov three series theorem. Note that

$$\sum_{h=1}^{\infty} \mathbb{E}\{\log(1 - V_{mjh})\} \leq \sum_{h=1}^{\infty} \log \mathbb{E}(1 - V_{mjh}) = \sum_{h=1}^{\infty} \log \left(1 - \frac{1}{(\alpha + 1)(\beta + 1)}\right) = -\infty,$$

relying on Jensen's inequality and  $\mathbb{E}(V_{mjh}) = \mathbb{E}(U_{mh})\mathbb{E}(W_{jh}) = 1/\{(\alpha + 1)(\beta + 1)\}$ .

## A.2 Proof of Theorem 1 and 2

For any measurable set  $A \in \mathcal{B}_j$ , we have

$$\mathbb{E}\{G_{mj}(A)\} = \mathbb{E}\left\{\sum_{h=1}^{\infty} \pi_{mjh} \delta_{\Theta_{jh}^*}(A)\right\} = \sum_{h=1}^{\infty} \mathbb{E}(\pi_{mjh}) \Pr(\Theta_{j1} \in A) = G_{0j}(A),$$

where the series is absolutely convergent so that linearity of expectation can be used. In addition, the second moment can be expressed as:

$$\begin{aligned} \mathbb{E}\{G_{mj}^2(A)\} &= \mathbb{E}\left[\left\{\sum_{h=1}^{\infty} V_{mjh} \prod_{l < h} (1 - V_{mjl}) \delta_{\Theta_{jh}}(A)\right\}^2\right] \\ &= \mathbb{E}\left[\sum_{h=1}^{\infty} \left\{V_{mjh} \prod_{l < h} (1 - V_{mjl}) \delta_{\Theta_{jh}}(A)\right\}^2\right] \\ &\quad + 2\mathbb{E}\left[\sum_{h=1}^{\infty} \sum_{h'=h+1}^{\infty} \left\{V_{mjh} \prod_{l < h} (1 - V_{mjl}) \delta_{\Theta_{jh}}(A)\right\} \left\{V_{mjh'} \prod_{l' < h'} (1 - V_{mj'l'}) \delta_{\Theta_{jh'}}(A)\right\}\right]. \end{aligned}$$

Expressing the right hand side of this equation as **I** + **II**, we obtain:

$$\begin{aligned}
\mathbf{I} &= \mathbb{E} \left\{ \sum_{h=1}^{\infty} U_{mh}^2 W_{jh}^2 \prod_{l < h} (1 - 2U_{ml}W_{jl} + U_{ml}^2 W_{jl}^2) \delta_{\Theta_{jh}}(A)^2 \right\} \\
&= \sum_{h=1}^{\infty} \frac{4}{(\alpha+1)(\alpha+2)(\beta+1)(\beta+2)} \left\{ 1 - \frac{2}{(\alpha+1)(\beta+1)} \right. \\
&\quad \left. + \frac{4}{(\alpha+1)(\alpha+2)(\beta+1)(\beta+2)} \right\}^{h-1} G_{0j}(A) \\
&= \frac{2}{(\alpha+2)(\beta+2) - 2} G_{0j}(A). \\
\mathbf{II} &= 2\mathbb{E} \left[ \sum_{h=1}^{\infty} \left\{ V_{mjh}(1 - V_{mjh}) \prod_{l < h} (1 - V_{mjl})^2 \delta_{\Theta_{jh}}(A) \right\} \sum_{h'=h+1}^{\infty} \left\{ V_{mjh'} \prod_{l=h+1}^{h'-1} (1 - V_{mjl}) \delta_{\Theta_{jh'}}(A) \right\} \right] \\
&= 2 \sum_{h=1}^{\infty} \left\{ \frac{1}{(\alpha+1)(\beta+1)} - \frac{4}{(\alpha+1)(\alpha+2)(\beta+1)(\beta+2)} \right\} \\
&\quad \times \left\{ 1 - \frac{2}{(\alpha+1)(\beta+1)} + \frac{4}{(\alpha+1)(\alpha+2)(\beta+1)(\beta+2)} \right\}^{h-1} G_{0j}^2(A) \\
&= \left\{ 1 - \frac{2}{(\alpha+2)(\beta+2) - 2} \right\} G_{0j}^2(A).
\end{aligned}$$

Expressions for  $\mathbb{E}\{G_{mj}^2(A)\}$  and  $\mathbb{V}\{G_{mj}(A)\}$  follow from straightforward algebra. Following a similar approach, we obtain:

$$\mathbb{E}\{G_{mj}(A)G_{m'j}(A)\} = \frac{1}{(\alpha+1)(\beta+2) - 1} G_{0j}(A) \{1 - G_{0j}(A)\},$$

and theorem 2 follows from straightforward algebra.

### A.3 Proof of Theorem 3

$$\begin{aligned}
\mathbb{E}\{\Gamma_{mj}(N, r)\} &= \mathbb{E} \left\{ \left( \sum_{h=N}^{\infty} \pi_{mjh} \right)^2 \right\} = \mathbb{E} \left\{ \prod_{h=1}^{N-1} (1 - V_{mjh})^r \right\} = \prod_{h=1}^{N-1} \mathbb{E}\{(1 - V_{mjh})^r\} \\
&= \left[ \sum_{k=0}^r C_r^k (-1)^{r-k} \mathbb{E}(V_{mjh}^{r-k}) \right]^{N-1} = \left[ \sum_{k=0}^r C_r^k (-1)^{r-k} \mu_{r-k}(\alpha) \mu_{r-k}(\beta) \right]^{N-1}.
\end{aligned}$$

The expression for  $\Upsilon_{mj}(N, r)$  can be expanded as:

$$\Upsilon_{mj}(N, r) = \sum_{h=N}^{\infty} (\pi_{mjh})^r = \pi_{mjN}^r + \sum_{h=N+1}^{\infty} (\pi_{mjh})^r = (\pi_{mjN})^r + (1 - V_{mjN})^r \sum_{h'=N}^{\infty} (\pi_{mjh'})^r.$$

Taking expectations on both sides yields

$$\mathbb{E}\{\Upsilon_{mj}(N, r)\} = \mathbb{E}\{(\pi_{mjN})^r\} + \mathbb{E}\{(1 - V_{mjN})^r\} \mathbb{E}\{\Upsilon_{mj}(N, r)\}.$$

Hence, we have

$$\begin{aligned} \mathbb{E}\{\Upsilon_{mj}(N, r)\} &= \frac{\mathbb{E}\left\{V_{mjN}^r \prod_{l < N} (1 - V_{mjl})^r\right\}}{1 - \mathbb{E}\{(1 - V_{mjN})^r\}} \\ &= \frac{\mu_r(\alpha)\mu_r(\beta)\mathbb{E}\{\Gamma_{mj}(N, r)\}}{1 - \sum_{k=0}^r C_k^r (-1)^{r-k} \mu_{r-k}(\alpha)\mu_{r-k}(\beta)}. \end{aligned}$$

#### A.4 Proof of Theorem 4

$$\begin{aligned} \Pr(\theta_{mj} = \theta_{m'j}) &= \sum_{h=1}^{\infty} \Pr(\theta_{mj} = \Theta_{jh}) \Pr(\theta_{m'j} = \Theta_{jh}) \\ &= \mathbb{E}\left\{\sum_{h=1}^{\infty} U_{mh} U_{m'h} W_{jh}^2 \prod_{l < h} (1 - U_{ml} W_{jl} - U_{m'l} W_{jl} + U_{ml} U_{m'l} W_{jl}^2)\right\} \\ &= \frac{2}{(\alpha + 1)^2(\beta + 1)(\beta + 2)} \sum_{h=1}^{\infty} \left\{1 - \frac{2}{(\alpha + 1)(\beta + 1)} + \frac{2}{(\alpha + 1)^2(\beta + 1)(\beta + 2)}\right\}^{h-1} \\ &= \frac{1}{(\alpha + 1)(\beta + 2) - 1}. \end{aligned}$$

The derivation of  $\Pr(\theta_{mj} = \theta_{m'j} | \theta_{mj'} = \theta_{m'j'})$  follows along similar lines. To show that  $\Pr(\theta_{mj} = \theta_{m'j} | \theta_{mj'} = \theta_{m'j'}) > \Pr(\theta_{mj} = \theta_{m'j})$ , we first multiply the numerator and denominator in the expression for  $\Pr(\theta_{mj} = \theta_{m'j})$  by the numerator in the expression for  $\Pr(\theta_{mj} = \theta_{m'j} | \theta_{mj'} = \theta_{m'j'})$ . It is then straightforward to demonstrate monotonicity in the denominators.

## References

- Albert, J.H. and Chib, S. (1993), ‘‘Bayesian Analysis of Binary and Polychotomous Response Data,’’ *Journal of the American Statistical Association*, **88**, 669-679.
- Burr, D. and Doss, H. (2005), ‘‘A Bayesian Semiparametric Model for Random-Effects Meta-Analysis,’’ *Journal of the American Statistical Association*, 100, 242-251.

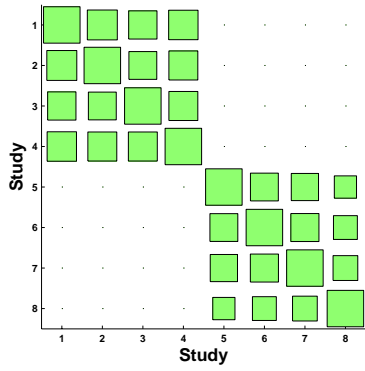
- Bush, C.A. and MacEachern, S.N. (1996), “A Semiparametric Bayesian Model for Randomised Block Designs,” *Biometrika*, 83, 275-285.
- De Iorio, M., Müller, P., Rosner, G.L. and MacEachern, S.N. (2002), “An ANOVA Model for Dependent Random Measures,” *Journal of the American Statistical Association*, 99, 205-215.
- Doss, H. (1985), “Bayesian Nonparametric Estimation of the Median: Part I: Computation of the Estimates,” *The Annals of Statistics*, 13, 1432-1444.
- Duan, J., Guindani, M. and Gelfand, A.E. (2005), “Generalized Spatial Dirichlet Process Models,” *ISDS Discussion Paper*, Duke University.
- Dunson, D.B. (2006), “Bayesian Dynamic Modeling of Latent Trait Distributions,” *Biostatistics*, 7, 551-568.
- Dunson, D.B. and Herring, A.H. (2006), “Bayesian Selection and Clustering of Polymorphisms in Functionally Related Genes,” *ISDS Discussion Paper*, Duke University.
- Dunson, D.B., Pillai, N. and Park, J-H. (2007), “Bayesian Density Regression,” *Journal of the Royal Statistical Society B*, in press.
- Ferguson, T.S. (1973), “A Bayesian Analysis of Some Nonparametric Problems,” *Annals of Statistics*, 1, 209-230.
- Ferguson, T.S. (1974), “Prior Distributions on Spaces of Probability Measures,” *Annals of Statistics*, 2, 615-629.
- Gelfand, A.E., Kottas, A. and MacEachern, S.N. (2005), “Bayesian Nonparametric Spatial Modeling with Dirichlet Process Mixing,” *Journal of the American Statistical Association*, 100, 1021-1035.

- Griffin, J.E. and Steel, M.F.J. (2006), "Order-based Dependent Dirichlet Process," *Journal of the American Statistical Association*, to appear.
- Geweke, J. (1992), "Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments," *Bayesian Statistics*, 4, 169-193.
- Kanno, J., Onyon, L., Haseman, J., Fenner-Crisp, P., Ashby, J. and Owens, W. (2001), "The OECD program to validate the rat uterotrophic bioassay to screen compounds for *in vivo* estrogenic responses: Phase 1," *Environmental Health Perspectives*, 109, 785-794.
- Kim, S., Tadesse, M.G. and Vannucci, M. (2006), "Variable Selection in Clustering via Dirichlet Process Mixture Models," *Biometrika*, in press.
- Kleinman, K.P. and Ibrahim, J.G. (1998), "A Semiparametric Bayesian Approach to the Random Effects Model," *Biometrics*, 54, 921-938.
- Ishwaran, H. and James, L.F. (2001), "Gibbs Sampling for Stick-Breaking Priors," *Journal of the American Statistical Association*, **96**, 161-73.
- Ishwaran, H. and Takahara, G. (2002), "Independent and identically distributed Monte Carlo algorithms for semiparametric linear mixed models. *Journal of the American Statistical Association* **97**, 1154-1166.
- MacEachern, S.N. (1999), "Dependent Nonparametric Process," *ASA Proceeding of the Section on Bayesian Statistical Science*, American Statistical Association, Alexandria, VA.
- MacEachern, S.N. (2000), "Dependent Dirichlet Processes," Unpublished manuscript, Department of Statistics, The Ohio State University.

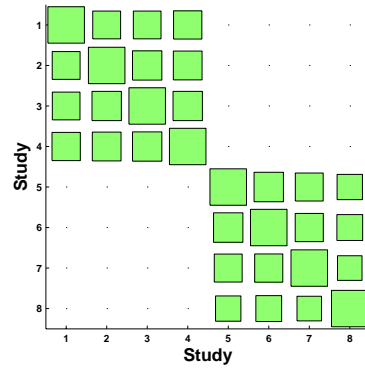
- MacLehose, R.F., Dunson, D.B., Herring, A.H. and Hoppin, J.A. (2006), “Bayesian Methods for Highly Correlated Exposure Data,” *Epidemiology*, in press.
- Mukhopadhyay, S. and Gelfand, A.E. (1997), “Dirichlet Process Mixed Generalized Linear Models,” *Journal of the American Statistical Association* **92**, 633-639.
- Müller, P., Quintana, F.A. and Rosner, G. (2004), “A Method for Combining Inference Across Related Nonparametric Bayesian Models,” *Journal of the Royal Statistical Society B*, **66**, 735-749.
- Papaspiliopoulos, O. and Roberts, G.O. (2006), “Retrospective MCMC for Inference for Dirichlet Mixture Models,” *Biometrika*, in press.
- Raftery, A. E. and Lewis, S. (1992), “How many iterations in the Gibbs sampler?” *Bayesian Statistics*, 4, 763-773.
- Rodriguez, A., Dunson, D.B., and Gelfand, A.E. (2006), “The Nested Dirichlet Process,” *ISDS Discussion Paper*, Duke University.
- Sethuraman, J. (1994), “A Constructive Definition of Dirichlet Priors,” *Statistica Sinica*, **4**, 639-650.
- Smith, T.C., Spiegelhalter, D.J. and Thomas, A. (1995), “Bayesian Approaches to Random-Effects Meta-Analysis: A Comparative Study,” *Statistics in Medicine*, 14, 2685-2699.

### Figure Captions:

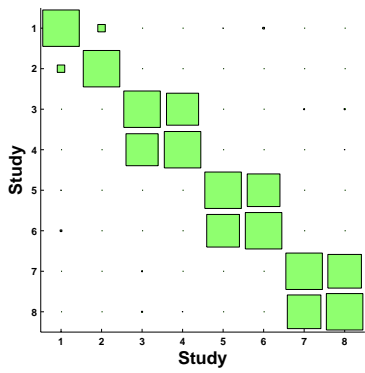
1. Pairwise posterior probabilities of two studies being assigned to the same cluster for the simulation example analyzed using the MSBP model.
2. Posterior means and 95% credible intervals for the study-specific coefficients from the simulation example. The black solid bar indicates the true value, the red shows the estimates for the MSBP model, and the green shows the estimates based on the normal hierarchical base model.
3. Pairwise posterior probabilities of two laboratories being assigned to the same cluster within the rat uterotrophic bioassay application based on the MSBP analysis. Results shown are for the intercept and slopes for EE and ZM.
4. Posterior means and 95% credible intervals for the lab-specific coefficients in the rat uterotrophic bioassay application. The red lines show the results for the MSBP analysis, while the green lines show the results based on the base normal hierarchical model.
5. Posterior densities of  $\alpha$  and  $\beta$  based on the MSBP analysis for the rat uterotrophic bioassay application.



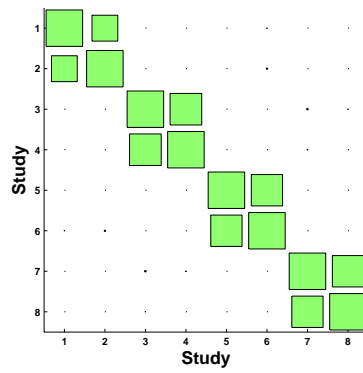
(a) Predictor 1



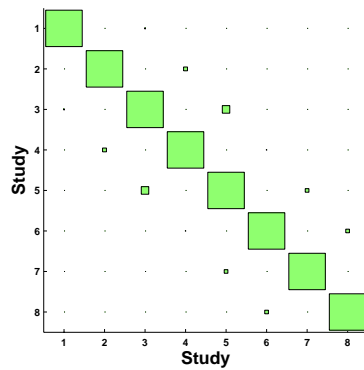
(b) Predictor 2



(c) Predictor 3

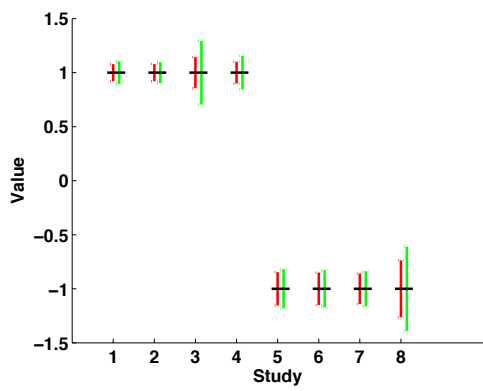


(d) Predictor 4

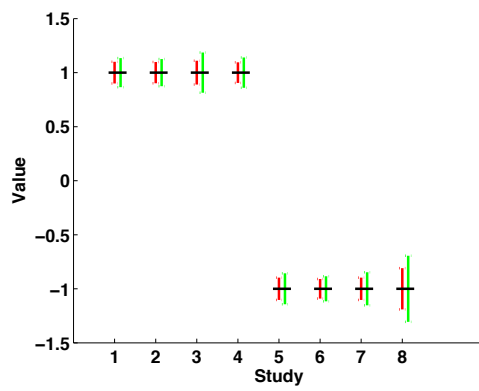


(e) Predictor 5

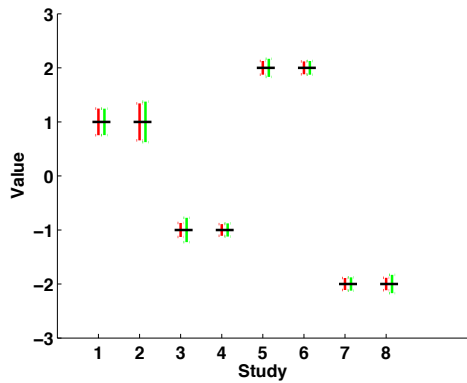
Figure 1:



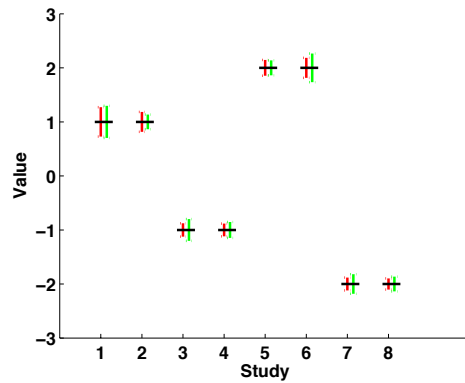
(a) Predictor 1



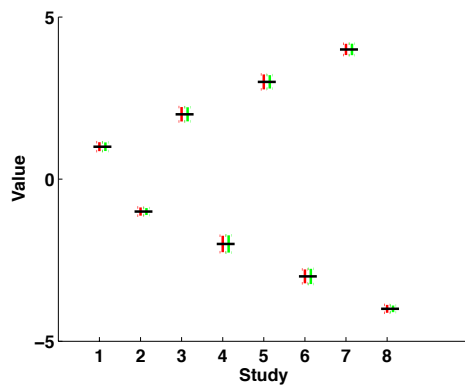
(b) Predictor 2



(c) Predictor 3

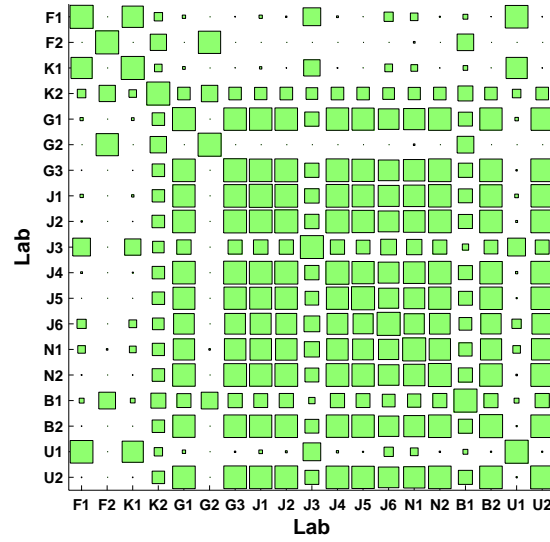


(d) Predictor 4

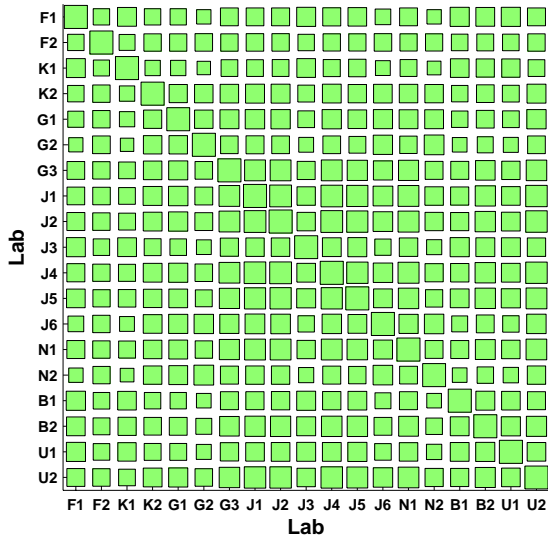


(e) Predictor 5

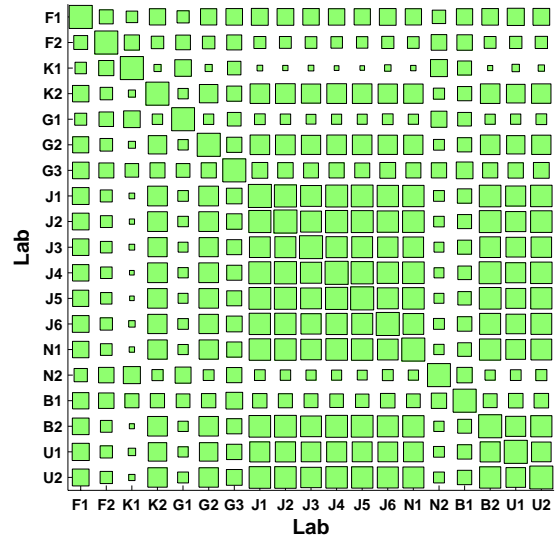
Figure 2:



(a) Intercept

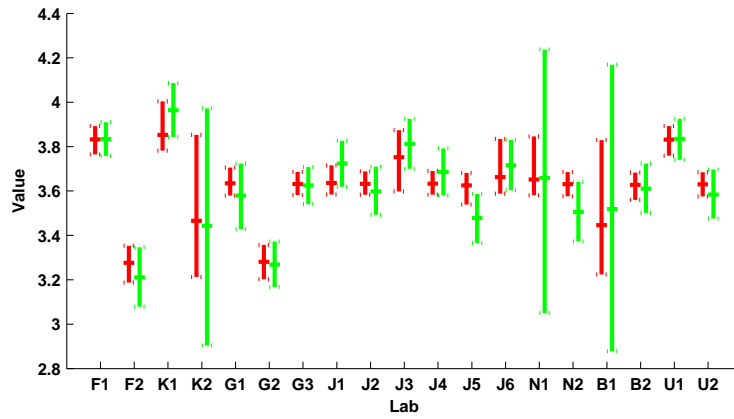


(b) EE coefficient

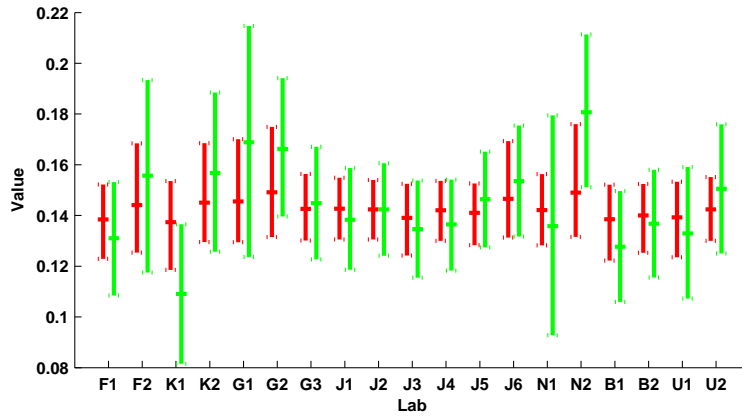


(c) ZM coefficient

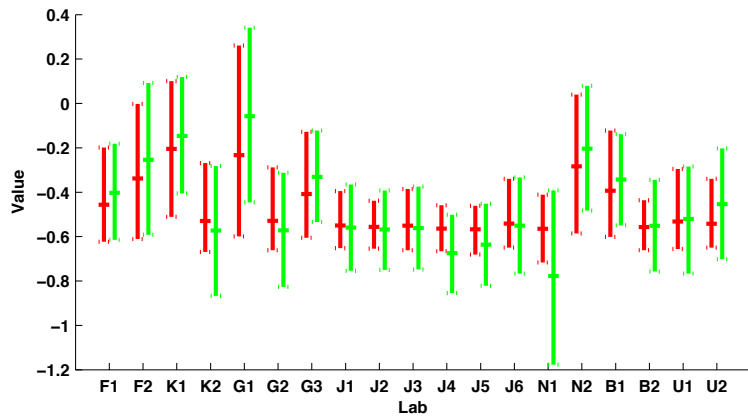
Figure 3:



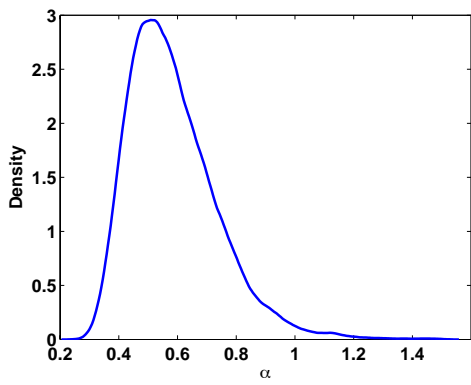
(a) Intercept



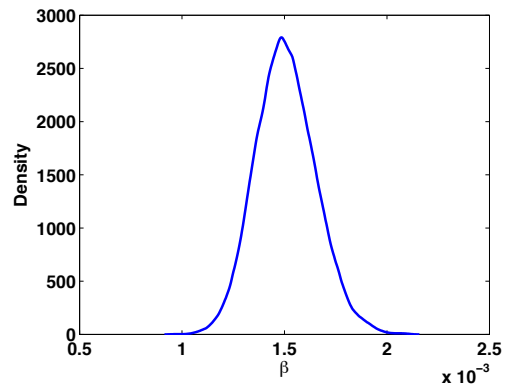
(b) EE coefficient



(c) ZM coefficient



(a)  $\alpha$



(b)  $\beta$

Figure 5: