Center-Adjusted Inference for a Nonparametric Bayesian Random Effect Distribution

BY YISHENG LI, PETER MÜLLER

Department of Biostatistics, Division of Quantitative Sciences

University of Texas M. D. Anderson Cancer Center, Houston, TX 77030, U.S.A.

ysli@mdanderson.org, pmueller@mdanderson.org

AND XIHONG LIN

Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115, U.S.A.

xlin@hsph.harvard.edu

SUMMARY

Dirichlet process (DP) prior models are a popular choice for semiparametric Bayesian random effect models. The base measure in the DP is often specified as a normal distribution with mean zero. The fact that this prior allows the random effect distribution to have a non-zero mean complicates interpretation of the inference for the fixed effects paired with the random effects. We show that the resulting inference on the fixed effects can be biased and poor. We propose a post-processing technique to adjust the inference. The approach uses a parametrization of the DP prior with a base measure centered at an unknown mean and is based on an analytic evaluation of the moments of the random moments of the DP. The adjustment for the moments of the DP can be conveniently incorporated into the Markov chain Monte Carlo simulations at essentially no additional computational cost. We conduct simulation studies to evaluate the performance of the proposed inference procedure in both a linear mixed model and a logistic linear mixed effect model. We illustrate the method by applying it to a prostate specific antigen dataset. We provide an R function that allows investigators to implement the proposed adjustment in a post-processing step of posterior simulation output, without any change to the posterior simulation itself.

Some key words: Bayesian nonparametric model; fixed effects; generalized linear mixed model; moments of a Dirichlet process; postprocessing; random probability measure.

1 Introduction

We propose an adjustment for inference in semiparametric Bayesian mixed effect models with a Dirichlet process (DP) prior on a random effect distribution G. We show that the (random) first two moments $\boldsymbol{\mu}_G$ and $\mathbf{Cov}_G$ of G are often confounded with some fixed effects and with the covariance matrix of the DP base measure. The latter is usually reported as the variance components of the random effects. We show that the Bayesian inference based on conventional DP prior is often biased. We derive easy-to-evaluate formula for the posterior moments of $\boldsymbol{\mu}_G$ and $\mathbf{Cov}_G$ and propose to use them in a straightforward post-processing step for Markov chain Monte Carlo (MCMC) output. In an application to inference for PSA profiles we show that the proposed adjustment can significantly change parameter estimates in a typical data analysis. Posterior means for some fixed effects change between 11 and 32%. The corresponding posterior standard deviations (SDs) and credible interval (CI) lengths change by more than 200%. The changes in the posterior means, SDs and lengths of CIs for the variance components are similarly large. We provide an R function for users to implement the proposed procedure.

Linear and generalized linear mixed models (LMMs & GLMMs) are an important and popular tool for analyzing correlated data. The random effects in such models are typically assumed normal, mainly for reasons of technical convenience. However, many applications require a more heterogeneous random effect distribution. For example, potentially relevant subject-specific covariates may not have been measured or are difficult to measure. Missing covariates can lead to a multimodal random effect distribution. In other applications, the distribution of the random effects may be skewed.

Estimation of the random effect distribution is important for predictive inference. Consider, for example, joint modeling of a primary endpoint and a longitudinal covariate. Unbiased estimates of the random effects are crucial. Inappropriately assuming normality can lead to excessive shrinkage towards zero and result in poor prediction.

These concerns lead many investigators to use nonparametric alternatives to normal random effect distributions. The DP is a popular choice as a nonparametric prior for the random effect distribution in mixed effect models within the Bayesian framework. For example, Kleinman and Ibrahim (1998b, a) modeled the random effect distribution as follows:

$$\boldsymbol{b}_i \mid G \overset{i.i.d.}{\sim} G,\ G \sim DP(M, G_0),\ G_0 = N(\mathbf{0}, \boldsymbol{D}), \tag{1}$$

where $DP(M, G_0)$ denotes a DP with a total mass parameter M and a base probability measure G_0 (Ferguson, 1973). We refer to a fixed effect as paired with a random effect if the columns in the design matrices of fixed effects and random effects match. See the discussion after equation (2) for a formal definition. In short, if the sampling model for the j-th repeated observation for the i-th subject involves a linear predictor $\eta_{ij} = \mathbf{x}_{ij}^T\boldsymbol{\beta} + \mathbf{z}_{ij}^T\boldsymbol{b}_i$ with fixed effects $\boldsymbol{\beta}$, subject-specific random effects $\boldsymbol{b}_i$ and known design vectors $\mathbf{x}_{ij}$ and $\mathbf{z}_{ij}$, then we refer to a subvector $\boldsymbol{\beta}^R$ of $\boldsymbol{\beta}$ as paired with $\boldsymbol{b}_i$ if the corresponding subvector of $\mathbf{x}_{ij}$ matches $\mathbf{z}_{ij}$, e.g., both contain an intercept. Posterior simulations in a LMM or GLMM based on model (1) for the random effects can be carried out using Gibbs sampling. A similar approach has been used by Bush and MacEachern (1996) in randomized block designs and many others. We argue that there is a difficulty in interpreting posterior inference for fixed effects that are paired with random effects in the above models. A similar difficulty arises in interpreting variance components of the random effects. In fact, the variance-covariance matrix $\boldsymbol{D}$ in (1) is confounded with the (random) second moment of G. As a result, inference on both the fixed effects and variance components is often biased.

In related work, Newton, Czado and Chappell (1996) proposed a centrally standardized Dirichlet process prior for the link function in a binary regression under which each realization of the link function has a median of zero. The approach is restricted to univariate distributions. Also, one may argue that in a DP random effect model, centering at the mean is more natural than centering at the median.

We propose a modified DP model and a post-processing procedure to address the aforementioned challenges. The model uses a DP prior for the sum of the random effects and their corresponding fixed effects with a base measure centered at an unknown mean. The post-processing technique is based on an analytic evaluation of the moments of the random moments of a random probability measure with a DP prior. Several recent references discussed the distribution of these random moments. For example, many authors have discussed the distribution of the mean of a DP random measure, including Hjort and Ongaro (2005) and Lijoi and Regazzini (2004). Epifani, Guglielmi and Melilli (2006) studied the distribution of the random variance of a DP random measure. Gelfand and Mukhopadhyay (1995) and Gelfand and Kottas (2002) used Monte Carlo integration to evaluate marginal posterior expectation of linear and nonlinear functionals of a nonparametric distribution whose prior is a DP mixture. They approximate the conditional expectation of the functional by a sample of the functional based on the predictive distribution of the parameters of the kernel. In this paper, we instead provide closed-form formula for the mean and covariance matrix of the (random) moments of a random measure with a DP prior. These closed-form expressions can be incorporated into the MCMC simulations and be used to adjust for inference for both the fixed effects paired with the nonparametric random effects and the second moments of the random effect distribution. We conduct simulation studies to evaluate the performance of the proposed moment-adjustment procedure and illustrate the method by analyzing a prostate specific antigen (PSA) dataset.

The remainder of this article is organized as follows. In Section 2 we discuss a difficulty with the näive inference in the DP random effect model and propose a modification to the conventional DP prior. In Section 3 we discuss posterior propriety under commonly used improper priors for the GLMM parameters. In Section 4 we derive the posterior mean and variance-covariance matrix for fixed effects that are paired with random effects, using results on the moments of the random first and second moments of a DP random measure. In Section 5, we derive new closed-form results on the expectation of the random third and fourth moments of a DP. We use these results to report the posterior summaries for the (random) covariance matrix of the random effects. In Section 6 we report results from simulation studies to show the performance of the proposed inference procedure in both a LMM and a logistic random effect model. In Section 7 we illustrate the method with inference for the PSA data. We provide concluding remarks in Section 8. Proofs of the propositions and lemmas are given in the Appendix.

2 A Hierarchically Centered Dirichlet Process Prior

For convenience, we use a nonparametric GLMM to illustrate our proposed method. However, unless indicated all results developed in the paper remain applicable for any nonparametric hierarchical model that contains the DP model (1) or (4) as a submodel.

Suppose y_{ij} arise independently from an exponential family with mean $\mu_{ij}^{\mathbf{b}}$ and variance $v_{ij}^{\mathbf{b}} = \phi v(\mu_{ij}^{\mathbf{b}})$ with a known dispersion parameter ϕ, conditional on the cluster-specific random effects $\boldsymbol{b}_i$ $(q \times 1)$, $i = 1, \ldots, m$, $j = 1, \ldots, n_i$. Consider the following conventional GLMM:

$$g(\mu_{ij}^{\mathbf{b}}) = \eta_{ij}^{\mathbf{b}}, \tag{2}$$

where $\eta_{ij}^{\mathbf{b}} = \mathbf{x}_{ij}^T\boldsymbol{\beta} + \mathbf{z}_{ij}^T\boldsymbol{b}_i$, $g(\cdot)$ is a monotone differentiable link function with inverse $h(\cdot)$, and $\boldsymbol{b}_i$ are independent and identically distributed with $E(\boldsymbol{b}_i) = 0$. Let $\mathbf{y}_i =$

$(y_{i1}, ..., y_{in_i})^T$ and $\mathbf{y} = (\mathbf{y}_1^T, \ldots, \mathbf{y}_m^T)^T$. Model (2) emcompasses the general LMM as a special case, with a normal distribution for y_{ij} given $\boldsymbol{b}_i$ and an identity link function $h(\cdot)$. Without loss of generality we assume that the fixed effects are partitioned into $(\boldsymbol{\beta}^F, \boldsymbol{\beta}^R)$ and similarly $\mathbf{x}_{ij} = (\mathbf{x}_{ij}^F, \mathbf{x}_{ij}^R)$, with $\mathbf{x}_{ij}^R = \mathbf{z}_{ij}$. We refer to $\boldsymbol{\beta}^R$ as fixed effects paired with the random effects $\boldsymbol{b}_i$. For example, $\boldsymbol{\beta}^R$ and $\boldsymbol{b}_i$ could be an intercept and a random intercept, respectively.

Consider the GLMM (2) with the DP prior model (1) for the random effects. The model includes the awkward feature that the unknown random effect distribution G can have a non-zero mean. This makes inference on the fixed effects $\boldsymbol{\beta}^R$ difficult to interpret. Let $\boldsymbol{\mu}_G = \int \boldsymbol{b}_i dG(\boldsymbol{b}_i)$ denote the random mean of G. We argue that instead of reporting inference on $\boldsymbol{\beta}^R$, it is more appropriate to report inference on $\boldsymbol{\beta}_{pair} \equiv \boldsymbol{\beta}^R + \boldsymbol{\mu}_G$.

Following the above arguments, we propose to model the distribution of $\boldsymbol{\beta}^R + \boldsymbol{b}_i$ as follows:

$$\boldsymbol{\beta}^R + \boldsymbol{b}_i \overset{i.i.d.}{\sim} G,\ G \sim DP(M, G_0),\ G_0 = N(\boldsymbol{\beta}_{\mathbf{b}}, \boldsymbol{D}), \tag{3}$$

where $\boldsymbol{\beta}_{\mathbf{b}}$ is an unknown vector of the mean parameters for the base probability measure. Given a lack of interpretation for inference on $\boldsymbol{\beta}^R$ and $\boldsymbol{\mu}_G$ separately, we propose to remove the paired fixed effects $\boldsymbol{\beta}^R$ from model (2). As a result, the random effect vector in the revised model, again denoted by $\boldsymbol{b}_i$, corresponds to $\boldsymbol{\beta}^R + \boldsymbol{b}_i$ in the original model. The prior model (3) now becomes

$$\boldsymbol{b}_i \overset{i.i.d.}{\sim} G,\ G \sim DP(M, G_0),\ G_0 = N(\boldsymbol{\beta}_{\mathbf{b}}, \boldsymbol{D}). \tag{4}$$

We call this prior a hierarchically centered DP prior. We further use $\boldsymbol{\beta} \equiv \boldsymbol{\beta}^F$ and $\mathbf{x}_{ij} \equiv \mathbf{x}_{ij}^F$ to denote the remaining fixed effect vector and corresponding design vector. Instead of inference on $\boldsymbol{\beta}^R$ in the original model, we report inference on $\boldsymbol{\beta}_{pair} = \boldsymbol{\mu}_G$ in the revised model. In the discussion that follows, we will still refer the revised centered model as model (2) except that $\mathbf{x}_{ij}$ only contains $\mathbf{x}_{ij}^F$ and $\boldsymbol{b}_i$ follows (4).

3 Prior Specification and Posterior Propriety in GLMM

We follow the conventional prior specification in the GLMM. That is, we assume a diffuse normal prior for each component of $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_{\mathbf{b}}$, a proper prior to be described below for $\boldsymbol{D}$, and a diffuse inverse Gamma (IG) prior for the residual variance if the GLMM (2) reduces to a LMM. All these priors are assumed independent. For a proper prior for $\boldsymbol{D}$, we consider both an inverse Wishart (IW) prior (or an IG prior if $\boldsymbol{D}$ reduces to a scalar) and a uniform shrinkage prior (USP) (Natarajan and Kass, 2000). For the latter, we define the USP as if the random effects were normally distributed. See Natarajan and Kass (2000) for corresponding detail.

The posterior propriety under model (2) when $\boldsymbol{b}_i \sim N(\mathbf{0}, \boldsymbol{D})$ along with a flat prior for $\boldsymbol{\beta}$ and a proper prior such as the USP for $\boldsymbol{D}$, was discussed by Natarajan and Kass (2000). They provide in general mild conditions for a sufficiently general class of GLMMs to have a proper posterior. We extend their results to model (2) with the hierarchically centered DP prior (4) for the distribution of $\boldsymbol{b}_i$.

Proposition 1. Consider model (2) with the centered DP prior (4) for the distribution of $\boldsymbol{b}_i$. We assume a flat prior for $(\boldsymbol{\beta}, \boldsymbol{\beta}_{\mathbf{b}})$ and a proper prior for $\boldsymbol{D}$. Under the conditions given in Theorem 3 or 4 of Natarajan and Kass (2000) (depending on the outcome type), the posterior is proper. In the special case of a LMM with an improper prior for σ^2 that is proportional to $1/\sigma^2$, the posterior is also proper.

See the proof of Proposition 1 in Appendix A.1. Proposition 1 justifies the common use of the diffuse normal prior for the fixed effects and the diffuse IG prior for the residual variance (when applicable) when the prior for the covariance matrix in the DP base measure is proper. Posterior simulation of the random effects follows the usual posterior MCMC scheme for DP mixture models. The simulation can include the total mass parameter M if the model is augmented with a gamma prior for M. See, for example, Neal (2000) for a review. Posterior simulation of the remaining model parameters may follow Kleinman and Ibrahim (1998a, b).

4 Adjustment for $\boldsymbol{\beta}^R$

In addition to $\boldsymbol{\mu}_G$, we denote by $\mathbf{Cov}_G$ the random covariance matrix of G. Let $\boldsymbol{b} = (\boldsymbol{b}_1^T, \ldots, \boldsymbol{b}_m^T)^T$ and $\boldsymbol{b}_{m+1}$ be the random effect for a future subject. Under the DP prior model (4), we have by conjugacy

$$[G \mid \boldsymbol{b}, \boldsymbol{\beta}_{\mathbf{b}}, \boldsymbol{D}, M] = DP(m+M, G_\star) \text{ and } [\boldsymbol{b}_{m+1} \mid \boldsymbol{b}, \boldsymbol{\beta}_{\mathbf{b}}, \boldsymbol{D}, M] = G_\star, \tag{5}$$

where $G_\star = \left\{M \cdot N(\boldsymbol{\beta}_{\mathbf{b}}, \boldsymbol{D}) + \sum_{i=1}^m \delta_{\mathbf{b}_i}\right\}/(m+M)$ with $\delta_{\mathbf{b_i}}$ denoting a point mass at $\boldsymbol{b}_i$. Here we use $[x \mid y]$ to generically denote a conditional distribution. Let $\boldsymbol{\mu}_{G_\star}$ and $\mathbf{Cov}_{G_\star}$ be the mean and covariance matrix of $G_\star$. We find:

$$\boldsymbol{\mu}_{G_\star} = \frac{M}{m+M}\boldsymbol{\beta}_{\mathbf{b}} + \frac{1}{m+M}\sum_{i=1}^m \boldsymbol{b}_i, \tag{6}$$

and

$$\mathbf{Cov}_{G_\star} = \left\{\frac{M}{m+M}(\boldsymbol{\beta}_{\mathbf{b}}\boldsymbol{\beta}_{\mathbf{b}}^T + \boldsymbol{D}) + \frac{1}{m+M}\sum_{i=1}^m \boldsymbol{b}_i{\boldsymbol{b}_i}^T\right\} - \boldsymbol{\mu}_{G_\star}{\boldsymbol{\mu}_{G_\star}}^T. \tag{7}$$

Based on (5), we develop a post-processing technique to compute the posterior mean and variance-covariance matrix of $\boldsymbol{\mu}_G$ and $\mathbf{Cov}_G$. Because (5), (6) and (7) are a result of the centered DP prior model (4) only, i.e., without any reference to the GLMM (2), all the results we derive in this and the next sections remain valid for any Bayesian hierarchical model with (4) as a nonparametric Bayesian submodel.

For inference on $\boldsymbol{\mu}_G$, we have the following proposition:

Proposition 2. (i) $E(\boldsymbol{\mu}_G \mid \mathbf{y}) = E(\boldsymbol{\mu}_{G_\star} \mid \mathbf{y}) = E\left(\frac{M}{m+M} \cdot \boldsymbol{\beta}_{\mathbf{b}} \mid \mathbf{y}\right) + E\left(\frac{1}{m+M}\sum_{i=1}^m \boldsymbol{b}_i \mid \mathbf{y}\right)$; (ii) $Cov(\boldsymbol{\mu}_G \mid \mathbf{y}) = E\left(\frac{\mathbf{Cov}_{G_\star}}{m+M+1} \mid \mathbf{y}\right) + Cov(\boldsymbol{\mu}_{G_\star} \mid \mathbf{y})$.

See the proof of Proposition 2 in Appendix A.2.

Proposition 2 suggests that the posterior mean and variance-covariance matrix of $\boldsymbol{\mu}_G$, or equivalently, $\boldsymbol{\beta}_{pair}$, can be computed based on the posterior samples of $(\boldsymbol{b}, \boldsymbol{\beta}_{\mathbf{b}}, \boldsymbol{D}, M)$. A CI for the i-th component of $\boldsymbol{\mu}_G$, denoted as $\mu_{G,i}$, can then be constructed. Specifically, the construction can be based on a normal approximation of the posterior distribution of $\mu_{G,i}$ using the estimated posterior mean $E(\mu_{G,i} \mid \mathbf{y})$ and the estimated posterior variance, the (i,i)-th element of $Cov(\boldsymbol{\mu}_G \mid \mathbf{y})$.

Corollary 1. Suppose $\boldsymbol{\theta}$ is a function of $(\boldsymbol{\beta}, \boldsymbol{b}, \boldsymbol{\beta}_b, \boldsymbol{D})$ and has the same dimension as $\boldsymbol{b}_i$. Then we have the following:

(i) $E(\boldsymbol{\theta} + \boldsymbol{\mu}_G \mid \mathbf{y}) = E(\boldsymbol{\theta} \mid \mathbf{y}) + E\left(\frac{M}{m+M} \cdot \boldsymbol{\beta}_{\mathbf{b}} \mid \mathbf{y}\right) + E\left(\frac{1}{m+M} \sum_{i=1}^{m} \boldsymbol{b}_i \mid \mathbf{y}\right)$;

(ii) $Cov(\boldsymbol{\theta} + \boldsymbol{\mu}_G \mid \mathbf{y}) = E\left(\frac{\mathbf{Cov}_{G_\star}}{m+M+1} \mid \mathbf{y}\right) + Cov(\boldsymbol{\theta} + \boldsymbol{\mu}_{G_\star} \mid \mathbf{y})$.

Proof. It is straightforward to compute $E(\boldsymbol{\theta} + \boldsymbol{\mu}_G \mid \mathbf{y})$ and $Cov(\boldsymbol{\theta} + \boldsymbol{\mu}_G \mid \mathbf{y})$ by first conditioning on $(\boldsymbol{\beta}, \boldsymbol{b}, \boldsymbol{\beta}_b, \boldsymbol{D}, \mathbf{y})$ and then marginalizing over $(\boldsymbol{\beta}, \boldsymbol{b}, \boldsymbol{\beta}_b, \boldsymbol{D})$.

5 Variance Components of the Random Effects

In addition to the inference on the fixed effects $\boldsymbol{\beta}_{pair}$, the centered DP GLMM (2) and (4) also allow us to make inference on the random variance-covariance matrix $\mathbf{Cov}_G$ of the random distribution G. In particular, we have the following proposition:

Proposition 3. (i) $E(\mathbf{Cov}_G \mid \mathbf{y}) = E\left(\frac{m+M}{m+M+1} \cdot \mathbf{Cov}_{G_\star} \mid \mathbf{y}\right)$.

See the proof of Proposition 3 (i) in Appendix A.3. In order to derive the posterior second moments for $\mathbf{Cov}_G$, we need the following two lemmas.

Lemma 1. Let $P \sim DP(M, \alpha)$, where $M > 0$. Suppose Z_1, Z_2 and Z_3 are random variables. If for all $i_1, i_2, i_3 \in \{0, 1\}$, $\int \mid Z_1^{i_1} Z_2^{i_2} Z_3^{i_3} \mid d\alpha < \infty$, then

$$E\int Z_1 dP\int Z_2 dP\int Z_3 dP=\mu_1\mu_2\mu_3+\frac{\sigma_{12}\mu_3+\sigma_{13}\mu_2+\sigma_{23}\mu_1}{M+1}+\frac{2\sigma_{123}}{(M+1)(M+2)}, \tag{8}$$

where $\mu_i = \int Z_i d\alpha$, $\sigma_{ij} = \int (Z_i - \mu_i)(Z_j - \mu_j)d\alpha$, $i, j = 1, 2, 3$, $i \neq j$, and $\sigma_{123} = \int (Z_1 - \mu_1)(Z_2 - \mu_2)(Z_3 - \mu_3)d\alpha$.

See the proof of Lemma 1 in Appendix A.4.

Lemma 2. Let P, α be as in Lemma 1. Let Z_1, Z_2, Z_3 and Z_4 be random variables. If for all $i_1, i_2, i_3, i_4 \in \{0, 1\}$, $\int \mid Z_1^{i_1} Z_2^{i_2} Z_3^{i_3} Z_4^{i_4} \mid d\alpha < \infty$, then

$$E\int Z_1 dP\int Z_2 dP\int Z_3 dP\int Z_4 dP=\mu_1\mu_2\mu_3\mu_4+$$
$$\frac{R_1}{M+1}+\frac{2R_2}{(M+1)(M+2)}+\frac{M\cdot R_3}{(M+1)(M+2)(M+3)}+\frac{6\sigma_{1234}}{(M+1)(M+2)(M+3)}, \tag{9}$$

where $R_1 = \sigma_{12}\mu_3\mu_4 + \sigma_{13}\mu_2\mu_4 + \sigma_{14}\mu_2\mu_3 + \sigma_{23}\mu_1\mu_4 + \sigma_{24}\mu_1\mu_3 + \sigma_{34}\mu_1\mu_2$, $R_2 = \sigma_{123}\mu_4 + \sigma_{124}\mu_3 + \sigma_{134}\mu_2 + \sigma_{234}\mu_1$, $R_3 = \sigma_{12}\sigma_{34} + \sigma_{13}\sigma_{24} + \sigma_{14}\sigma_{23}$, and μ_i, σ_{ij}, σ_{ijk} and σ_{1234} are defined in a similar manner as in Lemma 1.

See the proof of Lemma 2 in Appendix A.5.

Let $Cov_{G,ij}$ and $Cov_{G_\star,ij}$ be the (i, j)-th component of $\mathbf{Cov}_G$ and $\mathbf{Cov}_{G_\star}$ for $i \neq j$, respectively. Denote $Var_{G_\star,i}$ the (i, i)-th component of $\mathbf{Cov}_{G_\star}$. Based on Lemmas 1 and 2, we have the following proposition:

Proposition 3. (ii) Recall that $[\boldsymbol{b}_{m+1} \mid \boldsymbol{b}, \boldsymbol{\beta}_{\mathrm{b}}, \boldsymbol{D}, M] = G_\star$. Denote $b_{m+1}^{(i)}$ the i-th component of $\boldsymbol{b}_{m+1}$. Then

$$Cov(Cov_{G,i_1j_1}, Cov_{G,i_2j_2} \mid \mathbf{y}) =$$
$$E(L_1-L_2-L_3+L_4 \mid \mathbf{y})-E\left(\frac{m+M}{m+M+1}Cov_{G_\star,i_1j_1} \mid \mathbf{y}\right)E\left(\frac{m+M}{m+M+1}Cov_{G_\star,i_2j_2} \mid \mathbf{y}\right), \tag{10}$$

where L_i are defined as follows:

$$L_1 = \frac{E\left[b_{m+1}^{(i_1)} b_{m+1}^{(j_1)} b_{m+1}^{(i_2)} b_{m+1}^{(j_2)} \mid G_\star\right] + (m+M) E\left[b_{m+1}^{(i_1)} b_{m+1}^{(j_1)} \mid G_\star\right] E\left[b_{m+1}^{(i_2)} b_{m+1}^{(j_2)} \mid G_\star\right]}{m+M+1},$$

$$L_2 = \mu_1^{(L_2)} \mu_2^{(L_2)} \mu_3^{(L_2)} + \frac{\sigma_{12}^{(L_2)} \mu_3^{(L_2)} + \sigma_{13}^{(L_2)} \mu_2^{(L_2)} + \sigma_{23}^{(L_2)} \mu_1^{(L_2)}}{m+M+1} + \frac{2\sigma_{123}^{(L_2)}}{(m+M+1)(m+M+2)},$$

$$L_3 = \mu_1^{(L_3)} \mu_2^{(L_3)} \mu_3^{(L_3)} + \frac{\sigma_{12}^{(L_3)} \mu_3^{(L_3)} + \sigma_{13}^{(L_3)} \mu_2^{(L_3)} + \sigma_{23}^{(L_3)} \mu_1^{(L_3)}}{m+M+1} + \frac{2\sigma_{123}^{(L_3)}}{(m+M+1)(m+M+2)},$$

$$L_4 = \mu_1^{(L_4)} \mu_2^{(L_4)} \mu_3^{(L_4)} \mu_4^{(L_4)} + \frac{R_1^{(L_4)}}{m+M+1} + \frac{2R_2^{(L_4)}}{(m+M+1)(m+M+2)}$$
$$+ \frac{(m+M)R_3^{(L_4)}}{(m+M+1)(m+M+2)(m+M+3)} + \frac{6\,\sigma_{1234}^{(L_4)}}{(m+M+1)(m+M+2)(m+M+3)}.$$

In particular,

$$Var(Cov_{G,ij} \mid \mathbf{y}) = E(O_1 - 2O_2 + O_3 \mid \mathbf{y}) - \left[E\left(\frac{m+M}{m+M+1} Cov_{G_\star,ij} \mid \mathbf{y}\right)\right]^2, \quad (11)$$

where O_i are defined as follows:

$$O_1 = \frac{E\left([b_{m+1}^{(i)} b_{m+1}^{(j)}]^2 \mid G_\star\right) + (m+M)\left[E\left(b_{m+1}^{(i)} b_{m+1}^{(j)} \mid G_\star\right)\right]^2}{m+M+1},$$

$$O_2 = \mu_1^{(O_2)} \mu_2^{(O_2)} \mu_3^{(O_2)} + \frac{\sigma_{12}^{(O_2)} \mu_3^{(O_2)} + \sigma_{13}^{(O_2)} \mu_2^{(O_2)} + \sigma_{23}^{(O_2)} \mu_1^{(O_2)}}{m+M+1} + \frac{\sigma_{123}^{(O_2)}}{(m+M+1)(m+M+2)},$$

$$O_3 = \mu_1^{(O_3)} \mu_2^{(O_3)} \mu_3^{(O_3)} \mu_4^{(O_3)} + \frac{R_1^{(O_3)}}{m+M+1} + \frac{2\,R_2^{(O_3)}}{(m+M+1)(m+M+2)}$$
$$+ \frac{(m+M)R_3^{(O_3)}}{(m+M+1)(m+M+2)(m+M+3)} + \frac{6\,\sigma_{1234}^{(O_3)}}{(m+M+1)(m+M+2)(m+M+3)}.$$

The notations used for defining L_2 through L_4, O_2 and O_3 are given in Appendix A.6.

See the proof of Proposition 3 (ii) in Appendix A.7.

Remark. Proposition 3 allows us to compute the posterior mean and variance-covariance matrix of $\mathbf{Cov}_G$ (it is easiest to write $\mathbf{Cov}_G$ as a stacked column vector of its lower-diagonal elements). Noting the typical skewness of the posterior distribution of a variance, we construct a CI for $Var_{G,i}$ by matching its posterior mean and variance to those of a log-normal distribution. We choose the log-normal distribution because of its positive support. Similar to the approach to constructing a CI for $\mu_{G,i}$, we use a normal approximation for $Cov_{G,ij}$ with $i \neq j$.

Propositions 2 and 3 hold under model (4) for the random effects $\boldsymbol{b}_i$. In other words, as long as the posterior samples of $(\boldsymbol{b}, \boldsymbol{\beta}_{\mathbf{b}}, \boldsymbol{D}, M)$ can be obtained (e.g., through MCMC simulations), one can then post process the samples and draw adjusted inference for the mean, or known as the "fixed effects" paired with $\boldsymbol{b}_i$, and the variance components of $\boldsymbol{b}_i$.

6 Simulation Studies

6.1 A linear mixed model

We conduct a simulation study to examine the performance of the proposed center-adjusted inference in a LMM with nonparametric random intercept and slope. We generated 200 datasets from the following LMM:

$$Y_{ij} = \beta_0 + b_i^{(1)} + \left(\beta_1 + b_i^{(2)}\right) \cdot x_{ij} + \epsilon_{ij},\ i = 1, \ldots, 50,\ j = 1, \ldots, 10, \tag{12}$$

i.e., with $\boldsymbol{\beta} = (\beta_0, \beta_1)'$. We use $\beta_0 = 1$, $\beta_1 = 1$, $x_{ij} = j + 0.025 \cdot i - 5$, $\epsilon_{ij} \overset{i.i.d.}{\sim} N(0,\, \sigma^2 = 1)$, and $\boldsymbol{b}_i = (b_i^{(1)}, b_i^{(2)})' \overset{i.i.d.}{\sim} 1/3 \times N(\boldsymbol{\mu}^{(1)}, \boldsymbol{\Sigma}^{(1)}) + 2/3 \times N(\boldsymbol{\mu}^{(2)}, \boldsymbol{\Sigma}^{(2)})$, where $\boldsymbol{\mu}^{(1)} = (\mu_1^{(1)}, \mu_2^{(1)})' = (-2, 2)'$, $\boldsymbol{\Sigma}^{(1)} = [\sigma_{ij}^{(1)}]$ with $\sigma_{11}^{(1)} = \sigma_{22}^{(1)} = .1$ and $\sigma_{12}^{(1)} = -.09$; $\boldsymbol{\mu}^{(2)} = (\mu_1^{(2)}, \mu_2^{(2)})' = (1, -1)'$, $\boldsymbol{\Sigma}^{(2)}] = [\sigma_{ij}^{(2)}$ with $\sigma_{11}^{(2)} = \sigma_{22}^{(2)} = .5$ and $\sigma_{12}^{(2)} = -.45$. Under this bivariate bimodal normal mixture of $\boldsymbol{b}_i$, we have $E(\boldsymbol{b}_i) = (b_i^{(1)}, b_i^{(2)})' \equiv \boldsymbol{\mu} = (\mu_1, \mu_2)' = (0, 0)'$ and $Cov(\boldsymbol{b}_i) = Cov((b_i^{(1)}, b_i^{(2)})') \equiv \boldsymbol{\Sigma} = [\sigma_{ij}]$ with $\sigma_{11} = \sigma_{22} = 2.37$ and $\sigma_{12} = -2.33$.

We use the semiparametric LMM proposed in Section 2 for analyzing each simulated data. In particular, we use the centered DP prior model (4) for $\boldsymbol{b}_i$. We assume independent $N(0, 10^4)$ priors for β_{b0}, β_{b1}, and an IG prior $IG(10^{-2}, 10^{-2})$ for σ^2. Let I_2 denote the 2×2 identity matrix. Recall that $\boldsymbol{D}$ denotes the variance-covariance matrix of the base measure G_0. We assume an IW prior $IW(2, \boldsymbol{\Omega})$ for $\boldsymbol{D}$ with mean $E(\boldsymbol{D}^{-1}) = 2\boldsymbol{\Omega}$ where $\boldsymbol{\Omega} = 10^{-2} I_2$. The hyperparameters of the IW prior are chosen such that posterior inference is dominated by the data (c.f., Bernado and Smith, 1994). Posterior propriety follows by Proposition 1. Posterior simulations follow Kleinman and Ibrahim (1998b) with an additional step of sampling $\boldsymbol{\beta}_{\mathbf{b}}$. Inference for the fixed effects $\boldsymbol{\beta} \equiv (\beta_0\ \beta_1)'$ and the random effect covariance matrix $\boldsymbol{\Sigma}$ follow the moment-adjustment procedure proposed in Sections 4 and 5.

Table 1 reports relative bias, MSE, CI length (CIL) and coverage probability (CP) for the estimates of the fixed effect intercept and slope using both the traditional DP prior and the proposed centered DP prior approaches using both the IW and USP priors for variance components. The results show that the approach using the conventional DP prior gives larger biases and MSEs, much wider CIs, and either worse coverage probabilities with comparable CI lengths or slightly better coverage probabilities at the cost of doubled or even tripled CI lengths. In contrast, the proposed center-adjusted inference procedure leads to unbiased estimates of the fixed effects and the variance components. The 95% coverage probabilities for the fixed effects are close to the nominal value while those using the traditional DP prior are biased. Note that the variance components of the random effects using both procedures appear to be high when the IW prior is used.

In light of the documented difficulties with the use of an IG or IW prior for a random effect variance or covariance matrix (Natarajan and McCulloch, 1998; Natarajan and Kass, 2000, among others), we propose to extend the USP (Natarajan and Kass,

2000) to our semiparametric LMM and GLMM for the covariance matrix $\boldsymbol{D}$ in the DP base measure. While Natarajan and Kass (2000) show posterior propriety under mild conditions in their GLMMs with normal random effects, similar posterior propriety results hold in our semiparametric GLMMs as implied by Proposition 1. Posterior MCMC simulation can include a Metropolis step for sampling $\boldsymbol{D}$ with an IW density as the proposal. The corresponding simulation results are also reported in Table 1. The average CI lengths for the variance components now are considerably shorter than their IW counterparts, with the coverage probabilities preserved at a reasonable level (93-96%), being close to the nominal value. Similar results were obtained when varying the prior for M or fixing M to different constants.

6.2 A logistic random effect model

We use the following logistic linear mixed effect model as our simulation truth for the sampling model:

$$logit(p_{ij}) = \beta_0 + b_i^{(1)} + \left(\beta_1 + b_i^{(2)}\right) \cdot x_{ij},\ i = 1, \ldots, 100,\ j = 1, \ldots, 10, \qquad (13)$$

where x_{ij} are the same as in Section 6.1. We investigate the performance of the proposed adjustments in inference again using both an IW prior and a USP for the covariance matrix $\boldsymbol{D}$ in the DP base measure. The assumptions on the random effect distribution and the priors for the remaining parameters are similar to those in Section 6.1. We fixed $M = 5$. When a USP is used, the posterior conditional sampling of $\boldsymbol{D}$ follows the same strategy as for the LMM in Section 6.1. The corresponding results are summarized in Table 2. Note that when the IW prior is used for $\boldsymbol{D}$, even after the moment adjustments, the inference for the random effect covariance matrix is still poor and seriously biased. In contrast, the use of the USP has resulted in a good performance using the proposed inference on all model parameters, with a minimal bias and the coverage probability being close to the nominal value.

7 Application

We apply the proposed method to analyze data from a phase III clinical trial with prostate cancer patients. The trial was conducted at M. D. Anderson Cancer Center. The sample size was $n = 286$ patients. Patients were randomized to two treatment arms: a conventional androgen ablation (AA) therapy (149 patients) and the AA therapy plus three eight-week cycles of chemotherapy (CH) using ketoconazole and doxorubicin (KA) alternating with vinblastine and estramustine (VE) (137 patients). The outcome variable of interest is $y = \log(\text{PSA} + 1)$. PSA level is reported repeatedly over time starting with treatment initiation. The number of repeated measurements varies from 1 to 65 across patients. The investigators were interested in the PSA profiles post initialization of both treatments. For a more detailed description of the data, see Zhang, Müller and Do (2008).

We consider the following model for the log-transformed PSA level:

$$y_{vij} = \mu_0 + \theta_{0vi} + (\theta_{1vi} + v \cdot d_g)s_{vij} + (\theta_{2vi} + v \cdot d_\eta)\left(e^{-\phi_v s_{vij}} - 1\right) + \epsilon_{vij}, \qquad (14)$$

where $v = 0$ or 1 indicates treatment arm CH or AA, respectively, $i\ (= 1, \ldots, m_v)$ denotes the patient ID (in arm v), and $j\ (= 1, \ldots, n_{vi})$ indicates the measurement number for subject i in arm v, and s_{vij} is the time since treatment initiation (measured in years) at the jth repeated observation for patient i in arm v. The fixed effects d_g and d_η describe the effect of treatment on PSA slope and the size of the initial drop. We assume $\theta_{0vi} \overset{i.i.d.}{\sim} N(0, \sigma_0^2)$, $(\theta_{1vi}, \theta_{2vi}) \overset{i.i.d.}{\sim} G \equiv (G_1, G_2)^T$ with $G \sim DP\left(M, N\left(\boldsymbol{\beta} \equiv (\beta_1, \beta_2)', \boldsymbol{D} \equiv [d_{ij}]\right)\right)$, $\epsilon_{vij} \overset{i.i.d.}{\sim} N(0, \sigma^2)$, and θ_{0vi}, $(\theta_{1vi}, \theta_{2vi})$ and ϵ_{vij} are mutually independent.

Equation (14) models the typical features of PSA profiles for prostate cancer patients post treatment initiation. In particular, PSA levels tend to drop sharply after treatment initiation, and there is an additive increasing trend over time (linear in the log-transformed PSA level). Both, the initial drop and the trend, may differ between treatments.

We assume $\theta_{0vi} \sim N(0, \sigma_0^2)$ mainly for simplicity, assuming that neither the distribution of θ_{0vi} nor their estimates are of main scientific interest for the study. A scatterplot of the joint posterior means of $(\theta_{1vi}, \theta_{2vi})$ (Figure 1) suggests clear skewness and significant departure from normality (Verbeke and Lesaffre, 1996). This justifies the use of the centered DP prior model for the distribution of $(\theta_{1vi}, \theta_{2vi})$.

The prior used for the parameters in model (14) is independent across parameters with $p(\mu_0) = p(\beta_1) = p(\beta_2) = p(d_g) = p(d_\eta) = N(0, 10^4)$, $p(\phi_0) = p(\phi_1) = G(0.01, 0.001)$, $p(\sigma_0^2) = p(\sigma^2) = IG(.01, .01)$, and $p(\boldsymbol{D}) = IW(2,\ 0.01\, I_2)$. Here I_2 denotes a 2×2 identity matrix and the IW distribution is parametrized such that $E(\boldsymbol{D}^{-1}) = 0.02 I_2$. We fixed $M = 5$.

We implemented posterior simulation using a Gibbs sampler. An additional Metropolis step was used to define a transition probability to update ϕ_0 and ϕ_1, respectively. After a burn-in of 5,000 iterations, 20,000 samples were obtained with every 10th saved for posterior inference. Evaluation of Geweke's statistic (1992) suggested practical convergence of the Markov chains. We applied the adjustments for moments of the DP in posterior inference. Specifically, we report inference on $(\mu_{g_1}, \mu_{g_2}) \equiv (\int \theta_{1i} dG_1(\theta_{1i}), \int \theta_{2i} dG_2(\theta_{2i}))$ as inference on the slope of PSA and the initial drop for arm CH. Similarly, we report inference on $(\mu_{g_1} + d_g, \mu_{g_2} + d_\eta)$ as inference on the corresponding parameters for arm AA. Denote the 2×2 random covariance matrix of $(\theta_{1vi}, \theta_{2vi})$ by $\mathbf{Cov}_G = [\sigma_{ij}]$. We report posterior summaries for σ_{ij} as inference for the varinance components.

The posterior mean of d_η, i.e., the difference in the initial drop in PSA between the conventional AA and CH treatments, is -.15. The corresponding 95% CI is (-.32, .01), suggesting that the new CH treatment likely results in a larger initial drop. The difference in the rate of the drop, i.e., $\phi_1 - \phi_0$, has a posterior mean of -.41 and a 95% CI of (-1.00, .17). The difference in the increase in PSA, or d_g, has a posterior mean of -.01 and a 95% CI of (-.03, -0.0006). This significantly smaller rate of increase in PSA in the conventional AA arm (although the difference is small) might be related to its smaller initial drop.

For comparison, we report posterior inference with and without the proposed adjustment in Table 3. We report inference on the rate of initial drop in PSA as part of the treatment effect. This is an example of the inference that is not affected by the proposed adjustment. On the other hand, we report inference for all fixed effects that are paired with nonparametric random effects and for the variance components. The posterior mean of the average increase in PSA in each arm changes by approximately 10% between the proposed adjusted and unadjusted inferences. The posterior precision approximately tripled. For the average initial drop in PSA, the posterior mean changes by about 30% with the precision being more than tripled in both treatment arms, as a result of the adjusted inference. Even larger changes are seen in inference for the variance components σ_{ij}. For example, the posterior mean of the covariance between the two random effects flips sign under the proposed center-adjusted inference compared to the unadjusted inference. The reported positive covariance estimate is consistent with the scatterplot of the estimated random effects $(\hat{\theta}_{1vi}, \hat{\theta}_{2vi})$ under a normality assumption (Figure 1).

Finally, we investigated sensitivity of the proposed method with respect to M by considering alternatively a gamma prior for M, e.g., $p(M) = G(.8, .4)$ (with mean = 2 and variance = 5). The results (not shown) follow the same pattern as reported in Table 3.

8 Discussion

We have proposed a moment-adjustment procedure for inference on the fixed effects that are paired with random effects and the variance components of the random effects in a Bayesian hierarchical model where a hierarchically centered DP prior is assumed for the distribution of the random effects. The main results (Propositions 2 and 3) carry fully to any nonparametric Bayesian hierarchical model where a DP prior model (1) or (4) is assumed. In fact, this also applies to cases where the DP base measure is a parametric distribution other than normal, as long as the following are computable: 1) $\boldsymbol{\mu}_{G_\star}$ and $\mathbf{Cov}_{G_\star}$. This is for the evaluation of the posterior mean and covariance matrix of $\boldsymbol{\mu}_G$ and the posterior mean of $\mathbf{Cov}_G$; 2) Up to the fourth moments of $G_\star$. This is for the evaluation of the posterior second moments of $\mathbf{Cov}_G$. The only additional requirements for the proposed method to be applicable are: 1) the posterior samples of the parameters in the DP prior model (1) or (4) can be simulated; 2) the random mean and/or covariance matrix of the random effects are of scientific interest. In cases where only the predictive inference for the outcome variable is of interest, adjustments for the fixed effects and variance components may not be necessary. While the specific expressions for the proposed moment adjustments are lengthy, they are closed-form and easy to evaluate. Most importantly, we provide an R function (freely downloadable from http://odin.mdacc.tmc.edu/∼yishengli/DPPP.R) that allows easy implementation by the users.

We have demonstrated through simulations in the DP GLMMs that the proposed center-adjusted inference is effective in eliminating biases in the parameter estimates compared to the traditional DP prior-based inference. A rather appealing feature of the proposed procedure is that the method operates on a minimal introduction of new model structure and at essentially no additional computational cost. The implementation of the method requires essentially only post processing of the posterior samples of the model parameters.

In applying the proposed inference in the DP GLMMs, we also find that the USP leads to in general a more robust performance, while the IW prior may result in poor inference for the variance components of the random effects, an issue becoming even more prominent when the data to be analyzed are binary.

A. Appendix

A.1. Proof of Proposition 1

Consider the prior model (4) on the random effects. After marginalizing w.r.t. G, but conditional on all other hyperparameters, the prior is the Polya urn model (see, e.g., Walker *et al.*, 1999). Conditional on each configuration of ties in the Polya urn model the probability model on $\mathbf{Z}\boldsymbol{b}$ is a multivariate normal distribution with a block-diagonal variance-covariance matrix. By Theorem 3 or 4 in Natarajan and Kass (2000), the posterior conditional on the configuration is proper. In the special case of a LMM with an improper prior proprotional to $1/\sigma^2$ for σ^2, the posterior is proper by applying a modification of Theorem 1 in Chen *et al.* (2002). Since the Polya urn model defines a proper probability model on the configurations, which is a distribution on a finite sample space, the claim follows.

A.2. Proof of Proposition 2

(i) Since $[G \mid \boldsymbol{b}, \boldsymbol{\beta}_{\mathbf{b}}, \boldsymbol{D}, M] \sim DP(m+M, G_\star)$, by Theorem 3 of Ferguson (1973), $E(\boldsymbol{\mu}_G \mid \boldsymbol{b}, \boldsymbol{\beta}_{\mathbf{b}}, \boldsymbol{D}, M) = \boldsymbol{\mu}_{G_\star}$. Hence, $E(\boldsymbol{\mu}_G \mid \mathbf{y}) = E(E(\boldsymbol{\mu}_G \mid \boldsymbol{b}, \boldsymbol{\beta}_{\mathbf{b}}, \boldsymbol{D}, M) \mid \mathbf{y}) = E(\boldsymbol{\mu}_{G_\star} \mid \mathbf{y}) = E\left(\frac{M}{m+M} \cdot \boldsymbol{\beta}_{\mathbf{b}} \mid \mathbf{y}\right) + E\left(\frac{1}{m+M} \sum_{i=1}^{m} \boldsymbol{b}_i \mid \mathbf{y}\right).$

(ii) Let $\mu_{G_\star,i}$ be the i-th component of $\boldsymbol{\mu}_{G_\star}$. By Theorems 3 and 4 of Ferguson (1973), for all i, j,

$$
\begin{aligned}
&Cov(\mu_{G,i}, \mu_{G,j} \mid \boldsymbol{b}, \boldsymbol{\beta}_{\mathbf{b}}, \boldsymbol{D}, M) \\
&= E(\mu_{G,i} \cdot \mu_{G,j} \mid \boldsymbol{b}, \boldsymbol{\beta}_{\mathbf{b}}, \boldsymbol{D}, M) - E(\mu_{G,i} \mid \boldsymbol{b}, \boldsymbol{\beta}_{\mathbf{b}}, \boldsymbol{D}, M) E(\mu_{G,j} \mid \boldsymbol{b}, \boldsymbol{\beta}_{\mathbf{b}}, \boldsymbol{D}, M) \\
&= \tfrac{Cov_{G_\star,ij}}{m+M+1} + \mu_{G_\star,i}\mu_{G_\star,j} - \mu_{G_\star,i}\mu_{G_\star,j} = \tfrac{Cov_{G_\star,ij}}{m+M+1}.
\end{aligned}
$$

Thus, $Cov(\boldsymbol{\mu}_G \mid \mathbf{y}) = E(Cov(\boldsymbol{\mu}_G \mid \boldsymbol{b}, \boldsymbol{\beta}_{\mathbf{b}}, \boldsymbol{D}, M) \mid \mathbf{y}) + Cov(E(\boldsymbol{\mu}_G \mid \boldsymbol{b}, \boldsymbol{\beta}_{\mathbf{b}}, \boldsymbol{D}, M) \mid \mathbf{y})$ $= E\left(\frac{\mathbf{Cov}_{G_\star}}{m+M+1} \mid \mathbf{y}\right) + Cov(\boldsymbol{\mu}_{G_\star} \mid \mathbf{y})$.

A.3. Proof of Proposition 3 (i)

By Theorems 3 and 4 of Ferguson (1973),

$$
\begin{aligned}
&E(\mathbf{Cov}_G \mid \boldsymbol{b}, \boldsymbol{\beta}, \boldsymbol{D}, M) = E\left[\int \tilde{\boldsymbol{b}}\tilde{\boldsymbol{b}}^T d[G \mid \boldsymbol{b}, \boldsymbol{\beta}_{\mathbf{b}}, \boldsymbol{D}, M](\tilde{\boldsymbol{b}})\right] \\
&- E\left[\left\{\int \tilde{\boldsymbol{b}} d[G \mid \boldsymbol{b}, \boldsymbol{\beta}_{\mathbf{b}}, \boldsymbol{D}, M](\tilde{\boldsymbol{b}})\right\}\left\{\int \tilde{\boldsymbol{b}} d[G \mid \boldsymbol{b}, \boldsymbol{\beta}_{\mathbf{b}}, \boldsymbol{D}, M](\tilde{\boldsymbol{b}})\right\}^T\right] \\
&= E\left(\mathbf{Cov}_{G_\star} + \boldsymbol{\mu}_{G_\star}\boldsymbol{\mu}_{G_\star}{}^T - \tfrac{\mathbf{Cov}_{G_\star}}{m+M+1} - \boldsymbol{\mu}_{G_\star}\boldsymbol{\mu}_{G_\star}{}^T\right) = \tfrac{(m+M)\mathbf{Cov}_{G_\star}}{m+M+1}.
\end{aligned}
$$

Therefore, $E(\mathbf{Cov}_G \mid \mathbf{y}) = E(E(\mathbf{Cov}_G \mid \boldsymbol{b}, \boldsymbol{\beta}, \boldsymbol{D}, M) \mid \mathbf{y}) = E\left(\frac{(m+M)\mathbf{Cov}_{G_\star}}{m+M+1} \mid \mathbf{y}\right)$.

A.4. Proof of Lemma 1

Let $(\mathcal{X}, \mathcal{A})$ be the space and σ-field of subsets on which the probability measure α is defined. By Theorem 2 of Ferguson (1973), a Dirichlet process $DP(M, \alpha)$ can be alternatively constructed as $P(A) = \sum_{j=1}^{\infty} P_j \delta_{V_j}(A)$, for any $A \in \mathcal{A}$, where P_j are correlated random variables defined in Ferguson (1973) satisfying $P_j \geq 0$ and $\sum_{j=1}^{\infty} P_j = 1, a.s.$, V_j are i.i.d. random variables with values in $\mathcal{X}$ with probability measure α, and $\{P_j\}$ and $\{V_j\}$ are independent. Here $\delta_x(A) = 1$, if $x \in A$; and $\delta_x(A) = 0$ otherwise. Then we have

$$
\int Z_1 dP \int Z_2 dP \int Z_3 dP = \sum_i \sum_j \sum_k Z_1(V_i) Z_2(V_j) Z_3(V_k) P_i P_j P_k \tag{15}
$$

since all three series are absolutely convergent with probability one (see the proof of Theorem 3, Ferguson, 1973). The infinite summation (15) is bounded by

$$\sum_i \sum_j \sum_k \mid Z_1(V_i) Z_2(V_j) Z_3(V_k) \mid P_i P_j P_k. \tag{16}$$

If (16) is an integrable random variable, then the expectation of (15) can be taken inside the summation sign. Let $S(1,1,3) = \sum_{i \neq k} E[Z_1(V_i)Z_2(V_i)]E[Z_3(V_k)]E(P_iP_iP_k)$, $S(1,2,3) = \sum_{i \neq j \neq k} E[Z_1(V_i)]E[Z_2(V_j)]E[Z_3(V_k)]E(P_iP_jP_k)$, $S(1,1,1) = \sum_i E[Z_1(V_i)Z_2(V_i)Z_3(V_i)]E(P_iP_iP_i)$, etc. Then

$$\begin{aligned} E \int Z_1 dP \int Z_2 dP \int Z_3 dP &= \sum_i \sum_j \sum_k E[Z_1(V_i)Z_2(V_j)Z_3(V_k)]E(P_iP_jP_k) \\ &= S(1,2,3) + S(1,1,3) + S(1,2,1) + S(1,2,2) + S(1,1,1) \\ &= \mu_1\mu_2\mu_3 + (\sigma_{12}\mu_3 + \sigma_{13}\mu_2 + \sigma_{23}\mu_1) \left\{ \sum_{i \neq k} E(P_i^2 P_k) + \sum_i EP_i^3 \right\} + \sigma_{123} \sum_i EP_i^3. \end{aligned}$$

A similar equation shows that (16) is integrable. The distribution of the P_i depends on M, but not α, based on its definition (Ferguson, 1973). Hence, analogous to the proof of Theorem 4 of Ferguson (1973), we choose $\mathcal{X}$ to be the real line, α to give 2/3 probability to -1 and 1/3 probability to 2, and $Z_1(x) = Z_2(x) = Z_3(x) \equiv x$. Thus $\mu_1 = \mu_2 = \mu_3 = 0$ and $\sigma_{123} = 2$. Hence

$$\sum_i EP_i^3 = \frac{1}{2} E \left(\int x dP(x) \right)^3 = \frac{1}{2} E(3P(2) - 1)^3 = \frac{2}{(M+1)(M+2)},$$

since $P(2) \sim$ Beta$(M/3, 2M/3)$. A similar calculation gives us $\sum_{i \neq k} E(P_i^2 P_k) = M/[(M+1)(M+2)]$, by assuming α to give 1/2 probability to each of -1 and 1, and $Z_1(x) = Z_2(x) \equiv x$ and $Z_3 \equiv 1$. The equality (8) is thus proved.

A.5. Proof of Lemma 2

Define $S(i,j,k,\ell)$ like $S(i,j,k)$ in the proof of Lemma 1. By similar argument to that in the proof of Lemma 1, we have

$$E\int Z_1 dP \int Z_2 dP \int Z_3 dP \int Z_4 dP$$
$$= \sum_i \sum_j \sum_k \sum_l E[Z_1(V_i)Z_2(V_j)Z_3(V_k)Z_4(V_l)]E(P_iP_jP_kP_l)$$
$$= S(1,2,3,4)+S(1,1,3,4)+S(1,2,1,4)+S(1,2,3,1)+S(1,2,2,4)+S(1,2,3,2)+$$
$$S(1,2,3,3)+S(1,1,3,3)+S(1,2,1,2)+S(1,2,2,1)+S(1,1,1,4)+S(1,1,3,1)+$$
$$S(1,2,1,1)+S(1,2,2,2)+S(1,1,1,1)$$
$$= \mu_1\mu_2\mu_3\mu_4 + R_1 \sum_{i\neq k, k\neq l, l\neq i} E(P_i^2 P_k P_l) + (R_1+R_3)\sum_{i\neq k} E(P_i^2P_k^2)+$$
$$(2R_1+R_2)\sum_{i\neq l} E(P_i^3P_l) + (\sigma_{1234}+R_1+R_2)\sum_i EP_i^4$$

$$= \mu_1\mu_2\mu_3\mu_4 + R_1\left\{\sum_{i\neq k, k\neq l, l\neq i} E(P_i^2P_kP_l) + \sum_{i\neq k} E(P_i^2P_k^2) + 2\sum_{i\neq l} E(P_i^3P_l) + \sum_i EP_i^4\right\}+$$
$$R_2\left\{\sum_i EP_i^4 + \sum_{i\neq l} E(P_i^3P_l)\right\} + R_3\sum_{i\neq k} E(P_i^2P_k^2) + \sigma_{1234}\sum_i EP_i^4. \quad (17)$$

Assuming α to give 1/2 probability to each of -1 and 1, and $Z_1(x) = Z_2(x) = Z_3(x) = Z_4(x) \equiv x$, the left hand side (LHS) of (17) equals to

$$E\left(\int x dP(x)\right)^4 = E\{P(1) - P(-1)\}^4 = E\{2P(1)-1\}^4$$
$$= 2^4 EP(1)^4 - 4 * 2^3 EP(1)^3 + 6 * 2^2 EP(1)^2 - 4 * 2EP(1) + 1.$$

Since $P(1) \sim$ Beta$(M/2, M/2)$, we have

$$EP(1)^4 = \frac{\frac{M}{2}(\frac{M}{2}+1)(\frac{M}{2}+2)(\frac{M}{2}+3)}{M(M+1)(M+2)(M+3)} = \frac{(M+2)(M+6)}{2^4(M+1)(M+3)},$$
$$EP(1)^3 = \frac{\frac{M}{2}(\frac{M}{2}+1)(\frac{M}{2}+2)}{M(M+1)(M+2)} = \frac{(M+4)}{2^3(M+1)},\ EP(1)^2 = \frac{M+2}{2^2(M+1)},\ EP(1) = \frac{1}{2}.$$

The above is based on the moment formula for the beta distribution. Hence, the LHS of (17) equals $3/[(M+1)(M+3)]$. On the other hand, the right hand side (RHS) of (17) equals to $3\sum_{i\neq k} E(P_i^2 P_k^2) + \sum_i EP_i^4$. Thus, we have

$$3\sum_{i\neq k} E(P_i^2 P_k^2) + \sum_i EP_i^4 = \frac{3}{(M+1)(M+3)}. \tag{18}$$

Similarly, if we assume $Z_1(x) = Z_2(x) = Z_3(x) = Z_4(x) \equiv x$, α to assign 2/3 probabability to -1 and 1/3 probability to 2, (17) implies

$$2\sum_{i\neq k} E(P_i^2 P_k^2) + \sum_i EP_i^4 = \frac{2}{(M+1)(M+2)}, \tag{19}$$

since $P(2) \sim \text{Beta}(M/3, 2M/3)$. Equations (18) and (19) imply

$$\sum_{i\neq k} E(P_i^2 P_k^2) = \frac{M}{(M+1)(M+2)(M+3)} \tag{20}$$

and

$$\sum_i EP_i^4 = \frac{6}{(M+1)(M+2)(M+3)}. \tag{21}$$

Further assuming α to give 2/3 probability to -1 and 1/3 probability to 2, $Z_1(x) = Z_2(x) = Z_3(x) \equiv x$, and $Z_4(x) \equiv 1$, analogous calculation as above using (17) as above yields

$$\sum_{i\neq l} E(P_i^3 P_l) = \frac{2M}{(M+1)(M+2)(M+3)}. \tag{22}$$

Again, assuming α to give 1/2 probability to each of -1 and 1, $Z_1(x) = Z_2(x) \equiv x$, and $Z_3(x) = Z_4(x) \equiv 1$, we obtain

$$\sum_{i\neq k, i\neq l, k\neq l} E(P_i^2 P_k P_l) = \frac{M^2}{(M+1)(M+2)(M+3)}. \tag{23}$$

(9) is obtained by plugging (20), (21), (22) and (23) into (17).

A.6. Notations used for defining L_2 through L_4, O_2 and O_3 in Proposition 3 (ii)

In L_2:

$$\mu_1^{(L_2)} = Cov_{G_\star,i_1j_1} + \mu_{G_\star,i_1}\mu_{G_\star,j_1},\ \mu_2^{(L_2)} = \mu_{G_\star,i_2},\ \mu_3^{(L_2)} = \mu_{G_\star,j_2},\ \sigma_{23}^{(L_2)} = Cov_{G_\star,i_2j_2},$$

$$\sigma_{12}^{(L_2)} = \int b_{m+1}^{(i_1)} b_{m+1}^{(j_1)} b_{m+1}^{(i_2)} dG_\star(\boldsymbol{b}_{m+1}) - (Cov_{G_\star,i_1j_1} + \mu_{G_\star,i_1}\mu_{G_\star,j_1}) \times \mu_{G_\star,i_2},$$

$$\sigma_{13}^{(L_2)} = \int b_{m+1}^{(i_1)} b_{m+1}^{(j_1)} b_{m+1}^{(j_2)} dG_\star(\boldsymbol{b}_{m+1}) - (Cov_{G_\star,i_1j_1} + \mu_{G_\star,i_1}\mu_{G_\star,j_1}) \times \mu_{G_\star,j_2},$$

$$\sigma_{123}^{(L_2)} = \int \left(b_{m+1}^{(i_1)} b_{m+1}^{(j_1)} - \mu_1^{(L_2)}\right)\left(b_{m+1}^{(i_2)} - \mu_2^{(L_2)}\right)\left(b_{m+1}^{(j_2)} - \mu_3^{(L_2)}\right) dG_\star(\boldsymbol{b}_{m+1});$$

In L_3:

$$\mu_1^{(L_3)} = Cov_{G_\star,i_2j_2} + \mu_{G_\star,i_2}\mu_{G_\star,j_2},\ \mu_2^{(L_3)} = \mu_{G_\star,i_1},\ \mu_3^{(L_3)} = \mu_{G_\star,j_1},\ \sigma_{23}^{(L_3)} = Cov_{G_\star,i_1j_1},$$

$$\sigma_{12}^{(L_3)} = \int b_{m+1}^{(i_2)} b_{m+1}^{(j_2)} b_{m+1}^{(i_1)} dG_\star(\boldsymbol{b}_{m+1}) - (Cov_{G_\star,i_2j_2} + \mu_{G_\star,i_2}\mu_{G_\star,j_2}) \times \mu_{G_\star,i_1},$$

$$\sigma_{13}^{(L_3)} = \int b_{m+1}^{(i_2)} b_{m+1}^{(j_2)} b_{m+1}^{(j_1)} dG_\star(\boldsymbol{b}_{m+1}) - (Cov_{G_\star,i_2j_2} + \mu_{G_\star,i_2}\mu_{G_\star,j_2}) \times \mu_{G_\star,j_1},$$

$$\sigma_{123}^{(L_3)} = \int \left(b_{m+1}^{(i_2)} b_{m+1}^{(j_2)} - \mu_1^{(L_3)}\right)\left(b_{m+1}^{(i_1)} - \mu_2^{(L_3)}\right)\left(b_{m+1}^{(j_1)} - \mu_3^{(L_3)}\right) dG_\star(\boldsymbol{b}_{m+1});$$

In L_4:

$$\mu_1^{(L_4)} = \mu_{G_\star,i_1},\ \mu_2^{(L_4)} = \mu_{G_\star,j_1},\ \mu_3^{(L_4)} = \mu_{G_\star,i_2},\ \mu_4^{(L_4)} = \mu_{G_\star,j_2},$$

$$\sigma_{12}^{(L_4)} = Cov_{G_\star,i_1j_1},\ \sigma_{13}^{(L_4)} = Cov_{G_\star,i_1i_2},\ \sigma_{14}^{(L_4)} = Cov_{G_\star,i_1j_2},$$

$$\sigma_{23}^{(L_4)} = Cov_{G_\star,j_1i_2}, \sigma_{24}^{(L_4)} = Cov_{G_\star,j_1j_2},\ \sigma_{34}^{(L_4)} = Cov_{G_\star,i_2j_2},$$

$$\sigma_{123}^{(L_4)} = \int \left(b_{m+1}^{(i_1)} - \mu_1^{(L_4)}\right)\left(b_{m+1}^{(j_1)} - \mu_2^{(L_4)}\right)\left(b_{m+1}^{(i_2)} - \mu_3^{(L_4)}\right) dG_\star(\boldsymbol{b}_{m+1}),$$

$$\sigma_{124}^{(L_4)} = \int \left(b_{m+1}^{(i_1)} - \mu_1^{(L_4)}\right)\left(b_{m+1}^{(j_1)} - \mu_2^{(L_4)}\right)\left(b_{m+1}^{(j_2)} - \mu_4^{(L_4)}\right) dG_\star(\boldsymbol{b}_{m+1}),$$

$$\sigma_{134}^{(L_4)} = \int \left(b_{m+1}^{(i_1)} - \mu_1^{(L_4)}\right)\left(b_{m+1}^{(i_2)} - \mu_3^{(L_4)}\right)\left(b_{m+1}^{(j_2)} - \mu_4^{(L_4)}\right) dG_\star(\boldsymbol{b}_{m+1}),$$

$$\sigma_{234}^{(L_4)} = \int \left(b_{m+1}^{(j_1)} - \mu_2^{(L_4)}\right)\left(b_{m+1}^{(i_2)} - \mu_3^{(L_4)}\right)\left(b_{m+1}^{(j_2)} - \mu_4^{(L_4)}\right) dG_\star(\boldsymbol{b}_{m+1}),$$

$$R_1^{(L_4)} = \sum_{i<j,k<\ell,i\neq k,i\neq \ell,j\neq k,j\neq \ell} \sigma_{ij}^{(L_4)} \mu_k^{(L_4)} \mu_\ell^{(L_4)},$$

$$R_2^{(L_4)} = \sigma_{123}^{(L_4)}\mu_4^{(L_4)} + \sigma_{124}^{(L_4)}\mu_3^{(L_4)} + \sigma_{134}^{(L_4)}\mu_2^{(L_4)} + \sigma_{234}^{(L_4)}\mu_1^{(L_4)},$$

$$R_3^{(L_4)} = \sigma_{12}^{(L_4)}\sigma_{34}^{(L_4)} + \sigma_{13}^{(L_4)}\sigma_{24}^{(L_4)} + \sigma_{14}^{(L_4)}\sigma_{23}^{(L_4)},$$

$$\sigma_{1234}^{(L_4)} = \int \left(b_{m+1}^{(i_1)} - \mu_1^{(L_4)}\right)\left(b_{m+1}^{(j_1)} - \mu_2^{(L_4)}\right)\left(b_{m+1}^{(i_2)} - \mu_3^{(L_4)}\right)\left(b_{m+1}^{(j_2)} - \mu_4^{(L_4)}\right) dG_\star(\boldsymbol{b}_{m+1}).$$

In O_2:

$$\mu_1^{(O_2)} = Cov_{G_\star,ij} + \mu_{G_\star,i}\mu_{G_\star,j},\ \mu_2^{(O_2)} = \mu_{G_\star,i},\ \mu_3^{(O_2)} = \mu_{G_\star,j},\ \sigma_{23}^{(O_2)} = Cov_{G_\star,ij},$$

$$\sigma_{12}^{(O_2)} = \sigma_{13}^{(O_2)} = \int \left[b_{m+1}^{(i)}\right]^2 b_{m+1}^{(j)} dG_\star(\boldsymbol{b}_{m+1}) - (Cov_{G_\star,ij} + \mu_{G_\star,i}\mu_{G_\star,j}) \times \mu_{G_\star,i},$$

$$\sigma_{13}^{(O_2)} = \int b_{m+1}^{(i)}\left[b_{m+1}^{(j)}\right]^2 dG_\star(\boldsymbol{b}_{m+1}) - (Cov_{G_\star,ij} + \mu_{G_\star,i}\mu_{G_\star,j}) \times \mu_{G_\star,j},$$

$$\sigma_{123}^{(O_2)} = \int \left(b_{m+1}^{(i)} b_{m+1}^{(j)} - \mu_1^{(O_2)}\right)\left(b_{m+1}^{(i)} - \mu_2^{(O_2)}\right)\left(b_{m+1}^{(j)} - \mu_3^{(O_2)}\right) dG_\star(\boldsymbol{b}_{m+1});$$

In O_3:

$$\mu_1^{(O_3)} = \mu_2^{(O_3)} = \mu_{G_\star,i},\ \mu_3^{(O_3)} = \mu_4^{(O_3)} = \mu_{G_\star,j},$$

$$\sigma_{12}^{(O_3)} = Var_{G_\star,i},\ \sigma_{34}^{(O_3)} = Var_{G_\star,j},\ \sigma_{13}^{(O_3)} = \sigma_{14}^{(O_3)} = \sigma_{23}^{(O_3)} = \sigma_{24}^{(O_3)} = Cov_{G_\star,ij},$$

$$\sigma_{123}^{(O_3)} = \sigma_{124}^{(O_3)} = \int \left(b_{m+1}^{(i)} - \mu_1^{(O_3)}\right)^2\left(b_{m+1}^{(j)} - \mu_3^{(O_3)}\right) dG_\star(\boldsymbol{b}_{m+1}),$$

$$\sigma_{134}^{(O_3)} = \sigma_{234}^{(O_3)} = \int \left(b_{m+1}^{(i)} - \mu_1^{(O_3)}\right)\left(b_{m+1}^{(j)} - \mu_3^{(O_3)}\right)^2 dG_\star(\boldsymbol{b}_{m+1}),$$

$$R_1^{(O_3)} = \sum_{i<j,k<\ell,i\neq k,i\neq \ell,j\neq k,j\neq \ell} \sigma_{ij}^{(O_3)}\mu_k^{(O_3)}\mu_\ell^{(O_3)},$$

$$R_2^{(O_3)} = \sigma_{123}^{(O_3)}\mu_4^{(O_3)} + \sigma_{124}^{(O_3)}\mu_3^{(O_3)} + \sigma_{134}^{(O_3)}\mu_2^{(O_3)} + \sigma_{234}^{(O_3)}\mu_1^{(O_3)},$$

$$R_3^{(O_3)} = \sigma_{12}^{(O_3)}\sigma_{34}^{(O_3)} + \sigma_{13}^{(O_3)}\sigma_{24}^{(O_3)} + \sigma_{14}^{(O_3)}\sigma_{23}^{(O_3)},$$

$$\sigma_{1234}^{(O_3)} = \int \left(b_{m+1}^{(i)} - \mu_1^{(O_3)}\right)^2\left(b_{m+1}^{(j)} - \mu_3^{(O_3)}\right)^2 dG_\star(\boldsymbol{b}_{m+1}).$$

A.7. Proof of Proposition 3 (ii)

Assume $[\tilde{\boldsymbol{b}} \mid G] \sim G$ and $[\boldsymbol{b}_{m+1} \mid \boldsymbol{b}, \boldsymbol{\beta}_{\mathbf{b}}, \boldsymbol{D}, M] \sim G_\star$. Let $\tilde{b}^{(i)}$ and $b_{m+1}^{(i)}$ be the i-th component of $\tilde{\boldsymbol{b}}$ and $\boldsymbol{b}_{m+1}$, respectively. Define $I_1 = E(Cov_{G,i_1j_1} \cdot Cov_{G,i_2j_2} \mid \boldsymbol{b}, \boldsymbol{\beta}_{\mathbf{b}}, \boldsymbol{D}, M)$, $I_2 = E(Cov_{G,i_1j_1} \mid \boldsymbol{b}, \boldsymbol{\beta}_{\mathbf{b}}, \boldsymbol{D}, M)$ and $I_3 = E(Cov_{G,i_2j_2} \mid \boldsymbol{b}, \boldsymbol{\beta}_{\mathbf{b}}, \boldsymbol{D}, M)$. Then

$$Cov(Cov_{G,i_1j_1}, Cov_{G,i_2j_2} \mid \mathbf{y}) = E(I_1 \mid \mathbf{y}) - E(I_2 \mid \mathbf{y})E(I_3 \mid \mathbf{y}).$$

Based on Proposition 3 (i), $I_2 = (m+M)Cov_{G_\star,i_1j_1}/(m+M+1)$, and $I_3 = (m+M)Cov_{G_\star,i_2j_2}/(m+M+1)$.

To calculate I_1, we write $I_1 = J_1 - J_2 - J_3 + J_4$, where

$$J_1 = E\left[\int \tilde{b}^{(i_1)}\tilde{b}^{(j_1)} dG(\tilde{\boldsymbol{b}}) \int \tilde{b}^{(i_2)}\tilde{b}^{(j_2)} dG(\tilde{\boldsymbol{b}}) \mid \boldsymbol{b}, \boldsymbol{\beta}, \boldsymbol{D}, M\right],$$

$$J_2 = E\left[\int \tilde{b}^{(i_1)}\tilde{b}^{(j_1)} dG(\tilde{\boldsymbol{b}}) \int \tilde{b}^{(i_2)} dG(\tilde{\boldsymbol{b}}) \int \tilde{b}^{(j_2)} dG(\tilde{\boldsymbol{b}}) \mid \boldsymbol{b}, \boldsymbol{\beta}, \boldsymbol{D}, M\right],$$

$$J_3 = E\left[\int \tilde{b}^{(i_2)}\tilde{b}^{(j_2)} dG(\tilde{\boldsymbol{b}}) \int \tilde{b}^{(i_1)} dG(\tilde{\boldsymbol{b}}) \int \tilde{b}^{(j_1)} dG(\tilde{\boldsymbol{b}}) \mid \boldsymbol{b}, \boldsymbol{\beta}, \boldsymbol{D}, M\right],$$

and

$$J_4 = E\left[\int \tilde{b}^{(i_1)} dG(\tilde{\boldsymbol{b}}) \int \tilde{b}^{(j_1)} dG(\tilde{\boldsymbol{b}}) \int \tilde{b}^{(i_2)} dG(\tilde{\boldsymbol{b}}) \int \tilde{b}^{(j_2)} dG(\tilde{\boldsymbol{b}}) \mid \boldsymbol{b}, \boldsymbol{\beta}, \boldsymbol{D}, M\right].$$

By Theorem 4 of Ferguson (1973),

$$J_1 = \frac{Cov\left(b_{m+1}^{(i_1)}b_{m+1}^{(j_1)}, b_{m+1}^{(i_2)}b_{m+1}^{(j_2)} \mid G_\star\right)}{m+M+1} + \int b_{m+1}^{(i_1)}b_{m+1}^{(j_1)} dG_\star(\boldsymbol{b}_{m+1}) \int b_{m+1}^{(i_2)}b_{m+1}^{(j_2)} dG_\star(\boldsymbol{b}_{m+1})$$

$$= \frac{E\left[b_{m+1}^{(i_1)}b_{m+1}^{(j_1)}b_{m+1}^{(i_2)}b_{m+1}^{(j_2)} \mid G_\star\right]}{m+M+1} + \frac{m+M}{m+M+1} E\left[b_{m+1}^{(i_1)}b_{m+1}^{(j_1)} \mid G_\star\right] E\left[b_{m+1}^{(i_2)}b_{m+1}^{(j_2)} \mid G_\star\right].$$

To calculate J_2, we apply Lemma 1 for $Z_1 = \tilde{b}^{(i_1)}\tilde{b}^{(j_1)}$, $Z_2 = \tilde{b}^{(i_2)}$, and $Z_3 = \tilde{b}^{(j_2)}$. Following the notations in Lemma 1, we have

$$\mu_1 = Cov_{G_\star,i_1j_1} + \mu_{G_\star,i_1}\mu_{G_\star,j_1},\ \mu_2 = \mu_{G_\star,i_2},\ \mu_3 = \mu_{G_\star,j_2},\ \sigma_{23} = Cov_{G_\star,i_2j_2},$$

$$\sigma_{12} = \int (b_{m+1}^{(i_1)}b_{m+1}^{(j_1)} - \mu_1)(b_{m+1}^{(i_2)} - \mu_{G_\star,i_2}) dG_\star(\boldsymbol{b}_{m+1})$$

$$= \int b_{m+1}^{(i_1)} b_{m+1}^{(j_1)} b_{m+1}^{(i_2)} dG_\star(\boldsymbol{b}_{m+1}) - (Cov_{G_\star,i_1j_1} + \mu_{G_\star,i_1}\mu_{G_\star,j_1}) \times \mu_{G_\star,i_2},$$

$$\sigma_{13} = \int b_{m+1}^{(i_1)} b_{m+1}^{(j_1)} b_{m+1}^{(j_2)} dG_\star(\boldsymbol{b}_{m+1}) - (Cov_{G_\star,i_1j_1} + \mu_{G_\star,i_1}\mu_{G_\star,j_1}) \times \mu_{G_\star,j_2},$$

and

$$\sigma_{123} = \int \left(b_{m+1}^{(i_1)} b_{m+1}^{(j_1)} - \mu_1\right)\left(b_{m+1}^{(i_2)} - \mu_2\right)\left(b_{m+1}^{(j_2)} - \mu_3\right) dG_\star(\boldsymbol{b}_{m+1}).$$

Plugging the above expressions into (8), we obtain J_2. J_3 can be similarly computed.

To calculate J_4, we apply Lemma 2 for $Z_1 = \tilde{b}^{(i_1)}$, $Z_2 = \tilde{b}^{(j_1)}$, $Z_3 = \tilde{b}^{(i_2)}$, and $Z_4 = \tilde{b}^{(j_2)}$. Following the notations in Lemma 2, we then have

$$\mu_1 = \mu_{G_\star,i_1},\ \mu_2 = \mu_{G_\star,j_1},\ \mu_3 = \mu_{G_\star,i_2},\ \mu_4 = \mu_{G_\star,j_2},$$

$$\sigma_{12} = Cov_{G_\star,i_1j_1},\ \sigma_{13} = Cov_{G_\star,i_1i_2},\ \sigma_{14} = Cov_{G_\star,i_1j_2},$$

$$\sigma_{23} = Cov_{G_\star,j_1i_2}, \sigma_{24} = Cov_{G_\star,j_1j_2},\ \sigma_{34} = Cov_{G_\star,i_2j_2},$$

$$\sigma_{123} = \int \left(b_{m+1}^{(i_1)} - \mu_1\right)\left(b_{m+1}^{(j_1)} - \mu_2\right)\left(b_{m+1}^{(i_2)} - \mu_3\right) dG_\star(\boldsymbol{b}_{m+1}),$$

similarly for σ_{124}, σ_{134}, and σ_{234},

and

$$\sigma_{1234} = \int \left(b_{m+1}^{(i_1)} - \mu_1\right)\left(b_{m+1}^{(j_1)} - \mu_2\right)\left(b_{m+1}^{(i_2)} - \mu_3\right)\left(b_{m+1}^{(j_2)} - \mu_4\right) dG_\star(\boldsymbol{b}_{m+1}).$$

Plugging the above expressions into (9), we obtain J_4. Thus I_1 is computed, so is $Cov(Cov_{G,i_1j_1}, Cov_{G,i_2j_2} \mid \mathbf{y})$.

$Var(Cov_{G,ij} \mid \mathbf{y})$ can be obtained by replacing i_1 and i_2 by i, and j_1 and j_2 by j in (10). Proof is thus completed.

References

Bernado, J. M. and Smith, A. F. M. (1994), *Bayesian Theory*, John Wiley and Sons, New York.

Bush, C. A. and MacEachern, S. N. (1996), "A semiparametric Bayesian model for randomised block designs", *Biometrika* **83**, 275–285.

Chen, M.-H., Shao, Q.-M. and Xu, D. (2002), "Necessary and sufficient conditions on the propriety of posterior distributions for generalized linear mixed models", *Sankhyā A* **64**, 57–85.

Epifani, I., Guglielmi, A. and Melilli, E. (2006), "A stochastic equation for the law of the random Dirichlet variance", *Statistics and Probability Letters* **76**, 495–502.

Ferguson, T. S. (1973), "A Bayesian analysis of some non-parametric problems", *Annals of Statistics* **1**, 209–230.

Gelfand, A. E. and Kottas, A. (2002), "A computational approach for full nonparametric Bayesian inference under Dirichlet process mixture models", *Journal of Computational and Graphical Statistics* **11**, 289–305.

Gelfand, A. E. and Mukhopadhyay, S. (1995), "On nonparametric Bayesian inference for the distribution of a random sample", *The Canadian Journal of Statistics* **23**, 411–420.

Geweke, J. (1992), 'Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments', *in* J. M. Bernardo, J. Berger, A. P. Dawid, and A. F. M. Smith, eds, 'Bayesian Statistics 4', Oxford University Press, Oxford, U.K., pp. 169–193.

Hjort, N. L. and Ongaro, A. (2005), "Exact inference for random Dirichlet means", *Statistical Inference for Stochastic Processes* **8**, 227–254.

Kleinman, K. P. and Ibrahim, J. G. (1998*a*), "A semi-parametric Bayesian approach to generalized linear mixed models", *Statistics in Medicine* **17**, 2579–2596.

Kleinman, K. P. and Ibrahim, J. G. (1998*b*), "A semiparametric Bayesian approach to the random effects model", *Biometrics* **54**, 921–938.

Lijoi, A. and Regazzini, E. (2004), "Means of Dirichlet process and hypergeometric functions", *Annals of Probability* **32**, 1469–1495.

Natarajan, R. and Kass, R. E. (2000), "Reference Bayesian methods for generalized linear mixed models", *Journal of the American Statistical Association* **95**, 227–237.

Natarajan, R. and McCulloch, C. E. (1998), "Gibbs sampling with diffuse proper priors: a valid approach to data-driven inference?", *Journal of Computational and Graphical Statistics* **7**, 267–277.

Neal, R. M. (2000), "Markov chain sampling methods for Dirichlet process mixture models", *Journal of Computational and Graphical Statistics* **9**, 249–265.

Newton, M. A., Czado, C. and Chappell, R. (1996), "Bayesian inference for semiparametric binary regression", *Journal of the American Statistical Association* **91**, 142–153.

Verbeke, G. and Lesaffre, E. (1996), "A linear mixed-effects model with heterogeneity in the random-effects population", *Journal of the American Statistical Association* **433**, 217–221.

Walker, S. G., Damien, P., Laud, P. W. and Smith, A. F. M. (1999), "Bayesian nonparametric inference for random distributions and related functions (Disc: p510-527)", *Journal of the Royal Statistical Society B* **61**, 485–509.

Zhang, S., Müller, P. and Do, K.-A. (under revision), "A Bayesian semi-parametric survival model with longitudinal markers", *Biometrics* .

Table 1: Simulation results using center-adjusted vs conventional (i.e., non-centered and unadjusted) inference using DP prior with $M \sim G(2.5, .5)$ in model (12) based on 200 replicates. An IWP or USP was used for $\boldsymbol{D}$ in the DP base measure.

		Center-adjusted				Conventional			
Parameter	$\pi(\boldsymbol{D})$	Bias	MSE (SE)	CIL	CP	Bias	MSE (SE)	CIL	CP
β_0	IWP	.04	.04 (.004)	.85	.93	.16	.08 (.01)	2.88	.99
	USP	.03	.04 (.004)	.81	.93	.15	.09 (.01)	1.89	.97
β_1	IWP	-.04	.04 (.004)	.84	.93	-.15	.08 (.01)	1.88	.80
	USP	-.03	.04 (.004)	.80	.95	-.15	.09 (.01)	1.60	.85
σ^2	IWP	.04	.01 (.001)	.29	.94	.04	.01 (.001)	.29	.94
	USP	.04	.01 (.001)	.29	.92	.03	.01 (.001)	.29	.94
σ_{11}	IWP	.01	.10 (.01)	1.62	.99	.15	.33 (.03)	2.44	.99
	USP	-.07	.11 (.01)	1.29	.93	-.20	.31 (.02)	1.46	.45
σ_{22}	IWP	.01	.08 (.01)	1.53	1.00	.16	.30 (.03)	4.31	.96
	USP	-.06	.09 (.01)	1.19	.96	-.20	.31 (.02)	2.49	.98
σ_{12}	IWP	-.01	.09 (.01)	1.54	.99	-.15	.30 (.03)	4.16	1.00
	USP	.06	.09 (.01)	1.20	.94	.20	.31 (.02)	2.44	.99

Table 2: Simulation results using center-adjusted vs unadjusted inference using DP prior with $M = 5$ in model (13) based on 200 replicates. An IWP or USP was used for $\boldsymbol{D}$ in the DP base measure.

		Center-adjusted				Conventional			
Parameter	$\pi(\boldsymbol{D})$	Bias	MSE (SE)	CIL	CP	Bias	MSE (SE)	CIL	CP
β_0	IWP	.03	.06 (.01)	1.01	.97	.04	.13 (.02)	2.68	1.00
	USP	.07	.05 (.01)	.91	.94	.24	.14 (.01)	2.32	1.00
β_1	IWP	.03	.07 (.01)	1.16	.97	-.03	.13 (.02)	2.70	1.00
	USP	-.06	.06 (.01)	.96	.93	-.25	.14 (.01)	2.19	1.00
σ_{11}	IWP	.22	1.26 (.19)	4.77	.99	.51	3.83 (.66)	10.11	1.00
	USP	.02	.57 (.09)	3.31	.97	-.19	.58 (.04)	4.54	.99
σ_{22}	IWP	.34	2.04 (.31)	6.26	.99	.50	3.80 (.61)	10.34	1.00
	USP	-.02	.49 (.06)	3.67	.97	-.29	.77 (.04)	4.18	.97
σ_{12}	IWP	-.27	1.32 (.19)	5.26	.99	.50	3.33 (.54)	9.80	1.00
	USP	.03	.39 (.05)	3.13	.92	-.27	.67 (.04)	4.14	.99

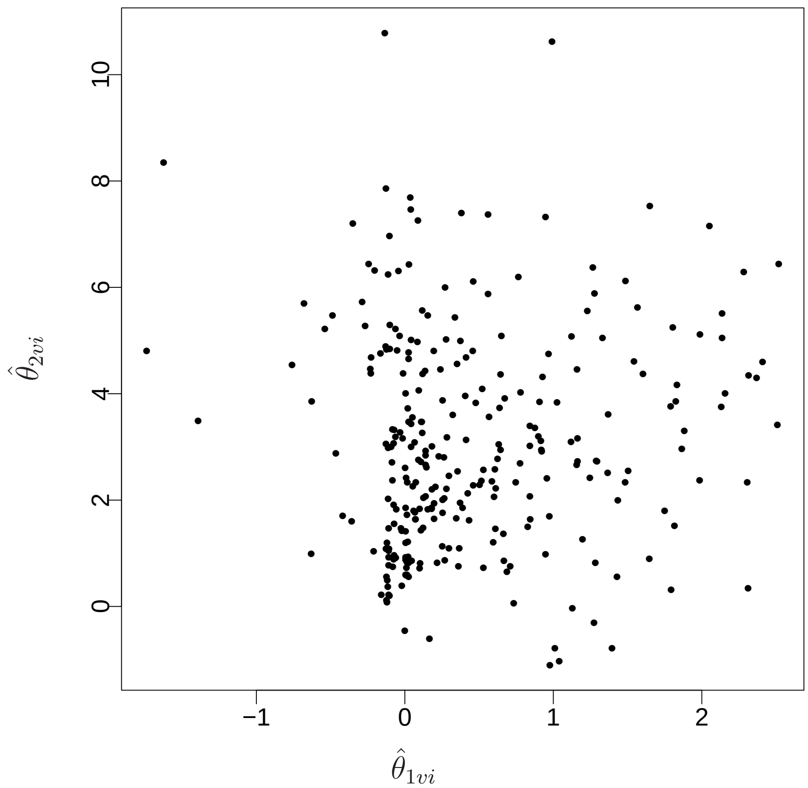

Figure 1: A scatterplot of the joint posterior means $(\hat{\theta}_{1vi}, \hat{\theta}_{2vi})$ assuming normally distributed $(\theta_{1vi}, \theta_{2vi})$ (with unknown means) in model (14)

Table 3: Posterior summaries with and without the proposed adjustment for rate of initial drop in PSA, increase in PSA per year, initial drop in PSA and variance components based on model (14) for the PSA data

Parameter	Adjustment	Posterior Mean	Posterior SD	95% CI
		Rate of initial drop in PSA		
		Arm CH		
ϕ_0	Cent-Adj/Unadj	8.44	.21	(8.04, 8.87)
		Arm AA		
ϕ_1	Cent-Adj/Unadj	8.03	.20	(7.63, 8.44)
		Increase in PSA per year		
		Arm CH		
μ_{g_1}	Cent-Adj	.63	.08	(.49, .78)
β_1	Unadj	.70	.25	(.24, 1.24)
		Arm AA		
$\mu_{g_1} + d_g$	Cent-Adj	.62	.08	(.47, .77)
$\beta_1 + d_g$	Unadj	.69	.25	(.21, 1.22)
		Initial drop in PSA		
		Arm CH		
μ_{g_2}	Cent-Adj	3.32	.14	(3.04, 3.59)
β_2	Unadj	4.33	.48	(3.37, 5.28)
		Arm AA		
$\mu_{g_2} + d_\eta$	Cent-Adj	3.17	.14	(2.89, 3.44)
$\beta_2 + d_\eta$	Unadj	4.18	.48	(3.23, 5.14)
		Variance components		
σ_{11}	Cent-Adj	1.17	.23	(.78, 1.68)
	Unadj	1.76	.68	(.89, 3.54)
σ_{22}	Cent-Adj	4.76	.51	(3.84, 5.84)
	Unadj	7.82	2.05	(4.65, 12.56)
σ_{12}	Cent-Adj	.35	.17	(.01, .68)
	Unadj	-.22	.60	(-1.70, 1.14)