Block Gibbs sampling for Bayesian random effects models with improper priors: Convergence and regeneration

Aixin Tan and James P. Hobert
Department of Statistics, University of Florida

May 14, 2009

Abstract

Bayesian versions of the classical one-way random effects model are widely used to analyze data. If the standard diffuse prior is adopted, there is a simple block Gibbs sampler that can be employed to explore the intractable posterior distribution. In this paper, theoretical and methodological results are developed that allow one to use this block Gibbs sampler with the same level of confidence that one would have using classical (iid) Monte Carlo. Indeed, a regenerative simulation method is developed that yields simple, asymptotically valid standard errors for the ergodic averages that are used to estimate intractable posterior expectations. These standard errors can be used to choose an appropriate (Markov chain) Monte Carlo sample size. The regenerative method rests on the assumption that the underlying Markov chain converges to its stationary distribution at a geometric rate. Another contribution of this paper is a result showing that, unless the data set is extremely small and unbalanced, the block Gibbs Markov chain is geometrically ergodic. We illustrate the use of the regenerative method with data from a styrene exposure study. R code for the simulation is posted as an online supplement.

Key Words: Asymptotic variance, Central limit theorem, Convergence rate, Geometric ergodicity, Minorization condition, One-way model.

1 Introduction

Consider the classical one-way random effects model given by

$$Y_{ij} = \theta_i + \varepsilon_{ij}\,,\ i = 1, \ldots, q\,,\ j = 1, \ldots, m_i\,, \tag{1}$$

where the random effects $\theta_1, \ldots, \theta_q$ are iid $\mathrm{N}(\mu, \sigma_\theta^2)$, the ε_{ij}s are iid $\mathrm{N}(0, \sigma_e^2)$ and independent of the θ_is, and $(\mu, \sigma_\theta^2, \sigma_e^2)$ is an unknown parameter. There is a long history of Bayesian analysis using this model starting with Hill (1965) and Tiao and Tan (1965). A Bayesian version of the model requires a prior distribution for $(\mu, \sigma_\theta^2, \sigma_e^2)$ and we consider the family of improper prior densities given by

$$\pi_{a,b}\big(\mu, \sigma_\theta^2, \sigma_e^2\big) = \big(\sigma_\theta^2\big)^{-(a+1)} \big(\sigma_e^2\big)^{-(b+1)} \ , \tag{2}$$

where a and b are known hyper-parameters. Letting $y = \{y_{ij}\}$ denote the vector of observed data and $\theta = \{\theta_i\}$ the vector of random effects, the $(q+3)$-dimensional posterior density is characterized by

$$\pi(\theta, \mu, \sigma_\theta^2, \sigma_e^2) \propto f(y|\theta, \mu, \sigma_\theta^2, \sigma_e^2) f(\theta|\mu, \sigma_\theta^2, \sigma_e^2) \pi_{a,b}\big(\mu, \sigma_\theta^2, \sigma_e^2\big) \ , \tag{3}$$

where

$$f(y|\theta, \mu, \sigma_\theta^2, \sigma_e^2) = \prod_{i=1}^{q} \prod_{j=1}^{m_i} \big(2\pi\sigma_e^2\big)^{-\frac{1}{2}} \exp\Big\{ -\frac{1}{2\sigma_e^2}(y_{ij} - \theta_i)^2 \Big\}$$

and

$$f(\theta|\mu, \sigma_\theta^2, \sigma_e^2) = \prod_{i=1}^{q} \big(2\pi\sigma_\theta^2\big)^{-\frac{1}{2}} \exp\Big\{ -\frac{1}{2\sigma_\theta^2}(\theta_i - \mu)^2 \Big\} \ .$$

Note that, in order to simplify the notation, the posterior density's dependence on the data is being suppressed, and we will adhere to this convention throughout the paper.

Our reasons for considering the family of improper priors $\pi_{a,b}$ stem from recommendations in the Bayesian literature. According to Gelman (2006), the choice of prior for (μ, σ_e^2) is not crucial since the data often contain a good deal of information about these parameters. On the other hand, there is typically relatively little information in the data concerning σ_θ^2, so the choice of prior for this parameter is more important and subtle. A commonly used prior for σ_θ^2 is a proper inverse gamma prior, which is a (conditionally) conjugate prior. When little or no prior information concerning σ_θ^2 is available, the (shape and scale) hyper-parameters of this prior are often set to very small values in an attempt to be "non-informative". However, in the limit, as the scale-parameter approaches 0 with the shape parameter either approaching 0 or fixed, not only does the prior become improper, but the corresponding posterior also becomes improper. Consequently, the posterior is not robust to small changes in these (somewhat arbitrarily chosen) hyper-parameters. This problem has led several authors, including Daniels

(1999) and Gelman (2006), to recommend that the proper inverse gamma prior not be used. In contrast, Gelman (2006) illustrates that the improper prior $\left(\sigma_\theta^2\right)^{-\frac{1}{2}}$ works well unless q is very small (say, below 5). Combining this prior with a uniform prior on $\left(\mu, \log(\sigma_e^2)\right)$ leads to $\pi_{-\frac{1}{2},0}$. van Dyk and Meng (2001) call $\pi_{-\frac{1}{2},0}$ the *standard diffuse prior* and we will refer to it as such throughout this paper.

Of course, whenever an improper prior is used, one must check that the resulting posterior is proper. Results in Hobert and Casella (1996) show that the posterior is proper if and only if

$$a < 0\,, \quad a + \frac{q}{2} > \frac{1}{2}\,, \quad \text{and} \quad a + b > \frac{1-M}{2}\,, \tag{4}$$

where M is the total sample size; that is, $M = \sum_{i=1}^{q} m_i$. Note that (4) implies $q > 1 - 2a > 1$ so a necessary condition for propriety is $q \geq 2$. Under the standard diffuse prior, the posterior is proper if and only if $q \geq 3$.

Making inference through the posterior distribution often boils down to computing expectations with respect to the posterior density. Unfortunately, despite the fact that $\pi_{a,b}$ has a very simple form, the posterior density is intractable. Indeed, letting $\mathbb{R}_+ = (0, \infty)$, the posterior expectation of $g(\theta, \mu, \sigma_\theta^2, \sigma_e^2)$ is given by

$$\int_{\mathbb{R}^q} \int_{\mathbb{R}} \int_{\mathbb{R}_+} \int_{\mathbb{R}_+} g(\theta, \mu, \sigma_\theta^2, \sigma_e^2)\, \pi\left(\theta, \mu, \sigma_\theta^2, \sigma_e^2\right) d\sigma_e^2\, d\sigma_\theta^2\, d\mu\, d\theta\,. \tag{5}$$

This is an intractable $q+3$ dimensional integral.

Markov chain Monte Carlo (MCMC) methods, and, in particular, the Gibbs sampler, can be used to approximate the posterior expectation in (5). The seminal paper by Gelfand and Smith (1990) illustrated an application of the *simple* Gibbs sampler to a Bayesian version of the one-way model with *proper* conjugate priors for μ, σ_θ^2 and σ_e^2. Their Gibbs sampler cycles through the $q+3$ components of the vector $(\theta_1, \ldots, \theta_q, \mu, \sigma_\theta^2, \sigma_e^2)$ one at a time and samples each one conditional on the most current values of the other $q+2$ components.

We study a *block* Gibbs sampler whose iterations have just two steps. Let $\sigma^2 = \left(\sigma_\theta^2, \sigma_e^2\right)$, $\xi = (\mu, \theta)$ and suppose that the state of the chain at time n is $\left(\sigma_n^2, \xi_n\right)$. One iteration of our sampler entails drawing σ_{n+1}^2 conditional on ξ_n, and then drawing ξ_{n+1} conditional on σ_{n+1}^2. "Blocking" variables together in this way and doing multivariate updates often leads to improved convergence properties relative to the simple (univariate) version of the Gibbs

sampler (see, e.g., Liu, Wong and Kong, 1994). Straightforward manipulation of (3) shows that, given ξ, σ_θ^2 and σ_e^2 are independent random variables each with inverse gamma distributions, and given σ^2, ξ is multivariate normal. (The specific forms of these distributions are given in Section 3.) Thus, programming this block Gibbs sampler poses no difficulty.

In this paper, we develop theoretical and methodological results that allow one to use the block Gibbs sampler with the same level of confidence that one would have using classical (iid) Monte Carlo. In order to explain this more carefully, let us briefly consider how classical Monte Carlo would be used to estimate intractable posterior expectations like (5). Let $L_1(\pi)$ denote the set of functions $h : \mathbb{R}^{q+1} \times \mathbb{R}_+^2 \to \mathbb{R}$ such that $\int_{\mathbb{R}^{q+1}} \int_{\mathbb{R}_+^2} \big|h\big(\sigma^2, \xi\big)\big| \pi\big(\sigma^2, \xi\big)\, d\sigma^2\, d\xi < \infty$, and define $L_2(\pi)$ analogously as the set of functions that are square integrable with respect to the posterior density. Also, let $\mathrm{E}_\pi h$ denote the posterior expectation of h. Suppose we wish to approximate $\mathrm{E}_\pi g$, where $g \in L_1(\pi)$, and imagine for a moment that we are able to make iid draws $\big(\sigma_0^{2*}, \xi_0^*\big), \big(\sigma_1^{2*}, \xi_1^*\big), \ldots$ from the posterior. With an iid sample in hand, we would estimate $\mathrm{E}_\pi g$ using the classical Monte Carlo estimator

$$\overline{g}_N^* = \frac{1}{N} \sum_{n=0}^{N-1} g\big(\sigma_n^{2*}, \xi_n^*\big) \; .$$

This estimator is unbiased and the strong law of large numbers (SLLN) implies that it converges almost surely to $\mathrm{E}_\pi g$; that is, it is also strongly consistent. In practice, we need to choose the sample size, N, and this is where the central limit theorem (CLT) comes in. Indeed, if $g \in L_2(\pi)$, then there is a CLT for g; that is, as $N \to \infty$, we have $\sqrt{N}\big(\overline{g}_N^* - \mathrm{E}_\pi g\big) \xrightarrow{\mathrm{d}} \mathrm{N}(0, v^2)$, where $v^2 = \mathrm{E}_\pi g^2 - \big(\mathrm{E}_\pi g\big)^2$. Thus, we could first draw a preliminary sample of size N' from π, and compute $\overline{g}_{N'}^*$ and

$$\hat{v}^2 = \frac{1}{N'-1} \sum_{n=0}^{N'-1} \big(g\big(\sigma_n^{2*}, \xi_n^*\big) - \overline{g}_{N'}^*\big)^2 \; .$$

Of course, $\hat{v}^2$ is a strongly consistent estimator of v^2. These quantities could then be used to assemble the asymptotic 95% confidence interval (CI) for $\mathrm{E}_\pi g$ given by $\overline{g}_{N'}^* \pm 2\hat{v}/\sqrt{N'}$. If we are satisfied with the width of this interval, we stop, whereas if the width is deemed too large, we continue the simulation. Indeed, given the pilot estimate $\hat{v}^2$, a CI of width l will require

a sample size of around $16\hat{v}^2/l^2$. The message here is that routine use of the CLT allows for determination of an appropriate Monte Carlo sample size.

In reality we are not able to make iid draws from the posterior, so we resort to an MCMC technique such as the block Gibbs sampler. Let $\left\{\left(\sigma_n^2, \xi_n\right)\right\}_{n=0}^{\infty}$ denote the block Gibbs Markov chain and consider applying the strategy outlined above with the Markov chain in place of the iid sample. First, the analogue of $\overline{g}_N^*$ is the ergodic average given by

$$\overline{g}_N = \frac{1}{N}\sum_{n=0}^{N-1} g\left(\sigma_n^2, \xi_n\right) .$$

If $\left(\sigma_0^2, \xi_0\right)$ is some fixed point, as it would usually be in practice, then $\overline{g}_N$ is not unbiased. Nevertheless, the ergodic theorem (Meyn and Tweedie, 1993, Chapter 17) implies that it is a strongly consistent estimator of $\mathrm{E}_\pi g$. Because of this, many view MCMC as a "free lunch" relative to classical Monte Carlo. Unfortunately, as we all know, *there is no free lunch*. Indeed, choosing an appropriate sample size in the MCMC context requires much more than routine use of the CLT. There are two reasons for this. First, when the iid sequence is replaced by a Markov chain, the second moment condition $\left(g \in L_2(\pi)\right)$ is no longer enough to guarantee that a CLT exists. Second, even when a CLT does hold, finding a consistent estimator of the asymptotic variance is challenging because this variance has a complex form and because the dependence among the variables in the Markov chain complicates the asymptotic analysis of the estimators (see, e.g., Geyer, 1992; Chan and Geyer, 1994; Jones, Haran, Caffo and Neath, 2006).

In this paper, we overcome the problems described above for the block Gibbs sampler through a convergence rate analysis and the development of a regenerative simulation method. In general, regeneration allows one to break a Markov chain up into iid segments (called "tours") so that asymptotic analysis can proceed using standard iid theory. While the theoretical details are fairly involved (Mykland, Tierney and Yu, 1995), the results of the theory and, more importantly, the application of the results, turns out to be quite straightforward. Indeed, results in Hobert, Jones, Presnell and Rosenthal (2002) show that, if the underlying Markov chain is geometrically ergodic and there exists an $\alpha > 0$ such that $\mathrm{E}_\pi|g|^{2+\alpha} < \infty$, then we can easily calculate an asymptotic standard error for $\overline{g}_N$ based on regenerative simulation

of the chain and only stop the simulation when this standard error is acceptably small.

Our convergence rate analysis for the block Gibbs chain (Proposition 3) indicates that the chain is geometrically ergodic unless M is very small and the group sizes, $m_1, \ldots, m_q$, are highly unbalanced. This result applies to all priors $\pi_{a,b}$ defined in (2). Here is a special case.

Corollary 1. *Under the standard diffuse prior, the block Gibbs chain is geometrically ergodic if*

1. $q \geq 4$ *and* $M \geq q + 3$, *or*
2. $q = 3$, $M \geq 6$ *and* $\min\left\{\left(\sum_{i=1}^{3} \frac{m_i}{m_i+1}\right)^{-1}, \frac{m^*}{M}\right\} < 2e^{-\gamma}$, *where* $m^* = \max\{m_1, m_2, m_3\}$ *and* $\gamma \doteq 0.577$ *is Euler's constant.*

Recall that, under $\pi_{-\frac{1}{2},0}$, the posterior is proper if and only if $q \geq 3$. When $q \geq 4$, our condition is satisfied for all reasonable data configurations. As for $q = 3$, it turns out that all balanced data sets with $\min\{m_1, m_2, m_3\} \geq 2$ satisfy the conditions, as do most reasonable unbalanced configurations. Appendix B.4 contains a table of all unbalanced configurations of (m_1, m_2, m_3) with $m^* \leq 12$ that satisfy the conditions of Corollary 1. (The appendices can be found in the online supplement to this article.)

Previous analysis of Gibbs samplers for Bayesian random effects models were performed by Hobert and Geyer (1998), Johnson and Jones (2008) and Jones and Hobert (2001, 2004). However, in each of these studies, the models that were considered have *proper priors* on all parameters. In fact, our Proposition 3 is the first of its kind for random effects models with *improper priors*, which, as we explained above, are the type of priors recommended in the Bayesian literature. It turns out that using improper priors complicates the analysis that is required to study the corresponding Markov chain. Indeed, Proposition 3 is much more than a straightforward extension of the existing results for proper priors. Another related paper is Papaspiliopoulos and Roberts (2008) who studied the convergence rates of Gibbs samplers for hierarchical linear models with different symmetric error distributions. What separates our results from theirs is that the variance components in our model are considered unknown parameters, while in their model the variance components are assumed known.

The remainder of the paper is organized as follows. Section 2 contains background material on general state space Markov chains. Section 3 contains a description of the block Gibbs Markov chain as well as the statement and proof of our convergence rate result. In Section 4, we develop our regenerative simulation method, which is based upon a *minorization* condition on the block Gibbs Markov chain. The regenerative method is illustrated in Section 5 using a real data set on styrene exposure. A short discussion appears in Section 6.

2 Background on General State Space Markov Chains

Let X be a set equipped with a countably generated σ-algebra $\sigma(\mathsf{X})$ and let $K : \mathsf{X} \times \sigma(\mathsf{X}) \to [0, 1]$ be a Markov transition function that defines a discrete time, time homogeneous Markov chain $X = \{X_n\}_{n=0}^{\infty}$. Thus, for $x \in \mathsf{X}$ and $A \in \sigma(\mathsf{X})$, $K(x, A) = \Pr(X_1 \in A \mid X_0 = x)$. Also, let $K^n : \mathsf{X} \times \sigma(\mathsf{X}) \to [0, 1]$, $n = 2, 3, \ldots$, denote the n-step Markov transition functions. Suppose that π is an invariant probability measure for the chain; i.e., $\int_{\mathsf{X}} K(x, dy)\pi(dx) = \pi(dy)$. The chain X is called *Harris ergodic* if it is ψ-irreducible, aperiodic and Harris recurrent, where ψ represents the maximal irreducibility measure of the chain. See Meyn and Tweedie (1993) for definitions.

Suppose that μ is a σ-finite measure on X and that the function $k : \mathsf{X} \times \mathsf{X} \to [0, \infty)$ satisfies $K(x, A) = \int_A k(y \mid x)\, \mu(dy)$ for any $x \in \mathsf{X}$ and any μ-measurable A. Then k is called the *Markov transition density* of X with respect to μ. The following result, established in Appendix A, shows that if X has a strictly positive density, then it is Harris ergodic.

Lemma 1. *Suppose X is a Markov chain with transition function K, transition density k (with respect to μ) and invariant probability measure π. If $k(y \mid x) > 0$ for all $x, y \in \mathsf{X}$, then X is Harris ergodic. Furthermore, μ is equivalent to the maximal irreducibility measure.*

If X is Harris ergodic, then, for any $x \in \mathsf{X}$,

$$\|K^n(x, \cdot) - \pi(\cdot)\| \downarrow 0 \quad \text{as } n \to \infty,$$

where $\|\cdot\|$ represents the total variation distance. Note that this tells us nothing about the *rate* of convergence. A Harris ergodic chain X is said to be *geometrically ergodic* if there exists a

function $c : \mathsf{X} \to [0, \infty)$ and a constant $0 < r < 1$ such that, for all $x \in \mathsf{X}$ and all $n = 0, 1, \ldots$

$$\|K^n(x, \cdot) - \pi(\cdot)\| \leq c(x)\, r^n \ .$$

Now suppose that X is topological and that $\sigma(\mathsf{X})$ is the Borel σ-field. If, for any open set $O \in \sigma(\mathsf{X})$, $K(\cdot, O)$ is a lower semicontinuous function, then X is called a *Feller chain* (Meyn and Tweedie, 1993, Chapter 6). The function $w : \mathsf{X} \to \mathbb{R}_+$ is said to be *unbounded off compact sets* if the level set $\{x \in \mathsf{X} : w(x) \leq d\}$ is compact for every $d > 0$. The following result is a combination of Meyn and Tweedie's (1993) Lemma 15.2.8 and Theorem 6.0.1.

Proposition 1. *Let X be a Markov chain on a topological space X. Assume that X is Harris ergodic and Feller and that the support of the maximal irreducibility measure has nonempty interior. If there exist $\rho < 1$, $L < \infty$ and a function $w : \mathsf{X} \to \mathbb{R}_+$ that is unbounded off compact sets such that*

$$E[w(X_1) \mid X_0 = x] \leq \rho w(x) + L \ , \tag{6}$$

then X is geometrically ergodic.

The inequality (6) is called a *drift condition* and the function w is called a *drift function*.

3 Geometric Ergodicity of the Block Gibbs Sampler

3.1 The block Gibbs sampler and the marginal chains

Recall that our block Gibbs sampler breaks the entire vector of variables into two blocks, $\sigma^2 = \left(\sigma_\theta^2, \sigma_e^2\right)$ and $\xi = (\mu, \theta)$, and updates them alternately. Formally, the block Gibbs chain, $\left\{\left(\sigma_n^2, \xi_n\right)\right\}_{n=0}^{\infty}$, has a Markov transition density (with respect to Lebesgue measure on $\mathbb{R}_+^2 \times \mathbb{R}^{q+1}$) given by

$$k\left(\tilde{\sigma}^2, \tilde{\xi} \mid \sigma^2, \xi\right) = \pi\left(\tilde{\sigma}^2 \mid \xi\right) \pi\left(\tilde{\xi} \mid \tilde{\sigma}^2\right) \ ,$$

where (σ^2, ξ) and $(\tilde{\sigma}^2, \tilde{\xi})$ denote the current and next states, respectively. Routine manipulation of (3) shows that σ_θ^2 and σ_e^2 are conditionally independent given ξ; that is,

$$\pi(\sigma^2 \mid \xi) = \pi(\sigma_\theta^2 \mid \xi)\, \pi(\sigma_e^2 \mid \xi) \ ,$$

and that

$$\sigma_\theta^2 \mid \xi \sim \text{IG}\left(\frac{q}{2}+a, \frac{1}{2}\sum_i(\theta_i-\mu)^2)\right) \quad \text{and} \quad \sigma_e^2 \mid \xi \sim \text{IG}\left(\frac{M}{2}+b, \frac{1}{2}\sum_{i,j}(y_{ij}-\theta_i)^2\right).$$

We say $X \sim \text{IG}(\alpha,\beta)$ if X is a random variable supported on $\mathbb{R}_+$ with density function proportional to $x^{-(\alpha+1)}e^{-\beta/x}$.

Further manipulation of (3) (see Tan (2008)) shows that, given σ^2, ξ has a multivariate normal density. To provide formulas for the elements of the mean vector and covariance matrix, we need a bit more notation. Let $\overline{y}_i = m_i^{-1}\sum_j y_{ij}$ and $t = \sum_{i=1}^q (\sigma_e^2 + m_i\sigma_\theta^2)^{-1} m_i$. Using this notation, we have

$$\text{E}\big(\mu\big|\sigma_\theta^2,\sigma_e^2\big) = \frac{1}{t}\sum_{i=1}^q \frac{m_i\overline{y}_i}{\sigma_e^2+m_i\sigma_\theta^2},$$

and for $k = 1, 2, \ldots, q$,

$$\text{E}\big(\theta_k\big|\sigma_\theta^2,\sigma_e^2\big) = \frac{\sigma_e^2}{\sigma_e^2+m_k\sigma_\theta^2}\left[\frac{1}{t}\sum_{i=1}^q \frac{m_i\overline{y}_i}{\sigma_e^2+m_i\sigma_\theta^2}\right] + \frac{\sigma_\theta^2 m_k\overline{y}_k}{\sigma_e^2+m_k\sigma_\theta^2}.$$

The variances and covariances are given by

$$\begin{aligned}
\text{Var}\big(\theta_i\big|\sigma_\theta^2,\sigma_e^2\big) &= \frac{\sigma_e^2}{\sigma_e^2+m_i\sigma_\theta^2}\left[\sigma_\theta^2 + \frac{\sigma_e^2}{\big(\sigma_e^2+m_i\sigma_\theta^2\big)t}\right] \\
\text{Cov}\big(\theta_i,\theta_j\big|\sigma_\theta^2,\sigma_e^2\big) &= \frac{\big(\sigma_e^2\big)^2}{\big(\sigma_e^2+m_i\sigma_\theta^2\big)\big(\sigma_e^2+m_j\sigma_\theta^2\big)t} \\
\text{Cov}\big(\theta_i,\mu\big|\sigma_\theta^2,\sigma_e^2\big) &= \frac{\sigma_e^2}{\big(\sigma_e^2+m_i\sigma_\theta^2\big)t} \\
\text{Var}\big(\mu\big|\sigma_\theta^2,\sigma_e^2\big) &= \frac{1}{t}.
\end{aligned}$$

Since $\pi(\xi \mid \sigma^2)$ and $\pi(\sigma^2 \mid \xi)$ are both strictly positive for $(\sigma^2,\xi) \in \mathbb{R}_+^2 \times \mathbb{R}^{q+1}$, it follows that k is a strictly positive Markov transition density. Thus, Lemma 1 implies that the block Gibbs Markov chain, $\big\{\big(\sigma_n^2,\xi_n\big)\big\}_{n=0}^\infty$, is Harris ergodic. As we now explain, our proof that this chain is also geometrically ergodic is indirect and rests upon an analysis of $\big\{\xi_n\big\}_{n=0}^\infty$.

It is well known that the two marginal sequences comprising a two-variable Gibbs chain are themselves Markov chains (Liu et al., 1994). Moreover, the Gibbs chain and its two marginal chains all converge at exactly the same rate (Diaconis, Khare and Saloff-Coste, 2008; Roberts and Rosenthal, 2001). Therefore, we can prove that the block Gibbs chain is geometric by

proving that the ξ-chain, $\{\xi_n\}_{n=0}^{\infty}$, is geometric. The ξ-chain has a Markov transition density (with respect to Lebesgue measure on $\mathbb{R}^{q+1}$) given by

$$k^*\big(\tilde{\xi} \mid \xi\big) = \int_{\mathbb{R}_+^2} \pi\big(\tilde{\xi} \mid \sigma^2\big)\, \pi\big(\sigma^2 \mid \xi\big)\, d\sigma^2 \,. \tag{7}$$

Clearly, k^* is strictly positive on $\mathbb{R}^{q+1} \times \mathbb{R}^{q+1}$ so another application of Lemma 1 shows that the ξ-chain is Harris ergodic. We also conclude from Lemma 1 that the maximal irreducibility measure of the ξ-chain is equivalent to Lebesgue measure on $\mathbb{R}^{q+1}$ and hence its support has non-empty interior. Finally, a simple application of Fatou's lemma shows that the ξ-chain is a Feller chain. We are now ready to use Proposition 1 to prove that the ξ-chain is geometric.

3.2 A proof of geometric ergodicity

According to Proposition 1, we can prove that the ξ-chain is geometric by finding a function $w : \mathbb{R}^{q+1} \to \mathbb{R}_+$ that is unbounded off compact sets and satisfies the drift condition

$$\mathrm{E}\big(w(\tilde{\xi}) \mid \xi\big) \leq \rho w(\xi) + L \qquad \text{for all } \xi \in \mathbb{R}^{q+1}, \tag{8}$$

where $\rho < 1$ and $L < \infty$. Our drift function takes the form $w(\xi) = \epsilon \left[w_1(\xi)\right]^s + \left[w_2(\xi)\right]^s$, where $w_1(\xi) = \sum_{i=1}^q (\theta_i - \mu)^2$, $w_2(\xi) = \sum_{i=1}^q m_i(\overline{y}_i - \theta_i)^2$ and $\epsilon > 0$ and $s \in (0, 1]$ are to be determined. For fixed $\epsilon > 0$ and $s \in (0, 1]$, the function w is unbounded off compact sets (but neither w_1 nor w_2 is by itself). Indeed, since w is continuous, it is enough to show that, in the level set $\{\xi : w(\xi) \leq d\}$, $|\mu|$ is bounded and $|\theta_i|$ is bounded for each $i \in \{1, 2, \ldots, q\}$. Note that $w_2 \to \infty$ as $|\theta_i| \to \infty$, and hence we have the θ_is contained. Now, given that the θ_is are contained, $w_1 \to \infty$ as $|\mu| \to \infty$, so μ is contained as well.

To keep the notation under control, we use w and $\tilde{w}$ to denote $w(\xi)$ and $w(\tilde{\xi})$, respectively. The left-hand side of (8) is

$$\mathrm{E}\big(w(\tilde{\xi}) \mid \xi\big) = \mathrm{E}(\tilde{w} \mid \xi) = \epsilon\, \mathrm{E}\big(\tilde{w}_1^s \mid \xi\big) + \mathrm{E}\big(\tilde{w}_2^s \mid \xi\big) \,.$$

Equation (7) shows that we can get the next state, $\tilde{\xi}$, by first drawing $\sigma^2 \sim \pi(\cdot \mid \xi)$, and then drawing $\tilde{\xi} \sim \pi(\cdot \mid \sigma^2)$, so graphically we have $\xi \to \sigma^2 \to \tilde{\xi}$. Therefore, we have

$$\mathrm{E}(\tilde{w}_k^s \mid \xi) = \mathrm{E}\big[\mathrm{E}(\tilde{w}_k^s \mid \sigma^2, \xi) \mid \xi\big] = \mathrm{E}\big[\mathrm{E}(\tilde{w}_k^s \mid \sigma^2) \mid \xi\big] \quad \text{for } k = 1, 2\,, \tag{9}$$

where the second equality is true because $\tilde{\xi}$ is conditionally independent of ξ given σ^2.

Since there are no restrictions on the constant L in (8), we do not have to keep track of any constants when calculating $\mathrm{E}(\tilde{w} \mid \xi)$. Hence, we will use the notation "$const$" to refer to a generic constant. Let $m^* = \max\{m_1, \ldots, m_q\}$. It is shown in Appendix B.1 that

$$\mathrm{E}(\tilde{w}_1 \mid \sigma^2) \leq \Delta_1 \sigma_\theta^2 + \Delta_2 \sigma_e^2 + const \quad \text{and} \quad \mathrm{E}(\tilde{w}_2 \mid \sigma^2) \leq (q+1)\sigma_e^2 + const$$

where

$$\Delta_1 = \min\left\{ q\Big(\sum_{i=1}^{q} \frac{m_i}{m_i+1} \Big)^{-1}, \frac{qm^*}{M} \right\}$$

and

$$\Delta_2 = \sum_{i=1}^{q} \frac{1}{m_i} - \sum_{i=1}^{q} \frac{1}{M(1+m_i)} + \max\left\{ q\Big(\sum_{i=1}^{q} \frac{m_i}{m_i+1} \Big)^{-1}, \frac{q}{M} \right\}.$$

For $s \in (0,1]$ and any $A, B > 0$, it is easy to see that $(A+B)^s \leq A^s + B^s$. Together with Jensen's inequality, this yields

$$\mathrm{E}(\tilde{w}_1^s \mid \sigma^2) \leq \big[\mathrm{E}(\tilde{w}_1 \mid \sigma^2)\big]^s \leq \big(\Delta_1\sigma_\theta^2 + \Delta_2\sigma_e^2 + const\big)^s \leq \Delta_1^s \big(\sigma_\theta^2\big)^s + \Delta_2^s \big(\sigma_e^2\big)^s + const\,, \quad (10)$$

and

$$\mathrm{E}(\tilde{w}_2^s \mid \sigma^2) \leq \big[\mathrm{E}(\tilde{w}_2 \mid \sigma^2)\big]^s \leq \big((q+1)\sigma_e^2 + const\big)^s \leq (q+1)^s \big(\sigma_e^2\big)^s + const\,. \quad (11)$$

To complete the calculation in (9), recall that

$$\sigma_\theta^2 \mid \xi \sim \mathrm{IG}\left(\frac{q}{2} + a, \frac{w_1}{2} \right) \quad \text{and} \quad \sigma_e^2 \mid \xi \sim \mathrm{IG}\left(\frac{M}{2} + b, \frac{w_2 + \mathrm{SSE}}{2} \right),$$

where $\mathrm{SSE} = \sum_{i,j} (y_{ij} - \overline{y}_i)^2$. This is where we have to make sure that $s \in (0,1]$ is not too large. Define the set

$$S = (0,1] \cap \left(0, \min\left\{ \frac{q}{2} + a, \frac{M}{2} + b \right\}\right).$$

Then, for any $s \in S$, $\mathrm{E}\Big(\big(\sigma_\theta^2\big)^s \mid \xi\Big)$ and $\mathrm{E}\Big(\big(\sigma_e^2\big)^s \mid \xi\Big)$ are both finite. In fact, routine calculations show that

$$\mathrm{E}\Big(\big(\sigma_\theta^2\big)^s \mid \xi\Big) = \frac{\Gamma(\frac{q}{2} + a - s)}{2^s \Gamma(\frac{q}{2} + a)} w_1^s\,, \quad (12)$$

and

$$\mathrm{E}\Big(\big(\sigma_e^2\big)^s \mid \xi\Big) = \frac{\Gamma(\frac{M}{2} + b - s)}{2^s \Gamma(\frac{M}{2} + b)} (w_2 + \mathrm{SSE})^s \leq \frac{\Gamma(\frac{M}{2} + b - s)}{2^s \Gamma(\frac{M}{2} + b)} w_2^s + const\,. \quad (13)$$

Define

$$\delta_1(s) = \frac{(q+1)^s \Gamma(\frac{M}{2}+b-s)}{2^s \Gamma(\frac{M}{2}+b)} \ , \quad \delta_2(s) = \frac{\Delta_2^s \Gamma(\frac{M}{2}+b-s)}{2^s \Gamma(\frac{M}{2}+b)} \quad \text{and} \quad \delta_3(s) = \frac{\Delta_1^s \Gamma(\frac{q}{2}+a-s)}{2^s \Gamma(\frac{q}{2}+a)} \ .$$

Combining (9)-(13), we have

$$\begin{aligned} \mathrm{E}(\tilde{w}_1^s \mid \xi) &\leq \mathrm{E}\Big(\Delta_1^s \big(\sigma_\theta^2\big)^s + \Delta_2^s \big(\sigma_e^2\big)^s + const \,\Big|\, \xi \Big) \\ &\leq \Delta_1^s \frac{\Gamma(\frac{q}{2}+a-s)}{2^s \Gamma(\frac{q}{2}+a)} w_1^s + \Delta_2^s \frac{\Gamma(\frac{M}{2}+b-s)}{2^s \Gamma(\frac{M}{2}+b)} w_2^s + const \qquad (14) \\ &= \delta_3(s) w_1^s + \delta_2(s) w_2^s + const \ . \end{aligned}$$

and

$$\begin{aligned} \mathrm{E}(\tilde{w}_2^s \mid \xi) &\leq \mathrm{E}\Big((q+1)^s \big(\sigma_e^2\big)^s + const \,\Big|\, \xi \Big) \\ &\leq (q+1)^s \frac{\Gamma(\frac{M}{2}+b-s)}{2^s \Gamma(\frac{M}{2}+b)} w_2^s + const \qquad (15) \\ &= \delta_1(s) w_2^s + const \ . \end{aligned}$$

The following result, which is established in Appendix B.2, provides conditions under which (14) and (15) can be combined to yield a valid drift condition for the block Gibbs sampler.

Proposition 2. *Fix $s \in S$. If $\delta_1(s) < 1$ and $\delta_3(s) < 1$, then there exist $\epsilon > 0$, $\rho < 1$ and $L < \infty$ such that $E\big(w(\tilde{\xi}) \mid \xi\big) \leq \rho w(\xi) + L$ for all $\xi \in \mathbb{R}^{q+1}$.*

In conjunction with Proposition 1, Proposition 2 shows that the ξ-chain (and hence the block Gibbs Markov chain) is geometrically ergodic as long as there exists an $s \in S$ such that both $\delta_1(s)$ and $\delta_3(s)$ are less than 1. Let $\Psi(x) = \frac{d}{dx} \log\big(\Gamma(x)\big)$ denote the digamma function. We show in Appendix B.3 that the desired s exists if $M + 2b \geq q + 3$ and $\Delta_1 < 2\exp\big(\Psi\big(\frac{q}{2}+a\big)\big)$. We can now state our main convergence rate result.

Proposition 3. *The block Gibbs chain is geometrically ergodic if*

1. $q \min\left\{ \left(\sum_{i=1}^q \frac{m_i}{m_i+1} \right)^{-1}, \frac{m^*}{M} \right\} < 2\exp\Big(\Psi\big(\frac{q}{2}+a\big)\Big)$, *and*
2. $M + 2b \geq q + 3$.

Loosely speaking, Proposition 3 shows that geometric ergodicity holds unless the data set is both small and unbalanced. Indeed, consider the first condition. The left-hand side will be large only if *both* $\left(\sum_{i=1}^{q} \frac{m_i}{m_i+1}\right)^{-1}$ and $\frac{m^*}{M}$ are large. The first term increases as the m_is get smaller and the second term increases as m^* gets larger relative to M; that is, as the data become more unbalanced. The second condition is a weak condition on the sample size. We show in Appendix B.4 that Corollary 1 from the Introduction follows easily from Proposition 3.

4 Minorization, Regeneration and the CLT

In this section, we first explain how regeneration of a generic Markov chain can be used to form a CLT whose asymptotic variance is easy to estimate. We then develop a regenerative simulation method specifically for the block Gibbs Markov chain by constructing a *minorization* condition for its transition density.

4.1 Regenerative simulation

Every ergodic Markov chain is a regenerative process (Meyn and Tweedie, 1993, Thm. 5.2.3). Regeneration times are easy to find for discrete state space Markov chains. Indeed, if we fix any particular state, then the chain probabilistically restarts at the random times immediately following returns to that fixed state. Unfortunately, identification of regeneration times is not as straightforward for general state space Markov chains. The most well known method of finding regeneration times for these general chains is based on establishing the following *minorization condition*:

$$k(\tilde{x} \mid x) \geq s(x)\, \nu(\tilde{x}) \quad \text{for all } \tilde{x}, x \in \mathsf{X}\,, \tag{16}$$

where $s : \mathsf{X} \to [0, 1)$ is a (non-trivial) function and $\nu : \mathsf{X} \to [0, \infty)$ is a density function. Equation (16) allows the transition density k to be expressed as a mixture of two other transition densities, one of which does not depend on the current state x. This mixture representation of k allows for the introduction of regenerations into the chain (see, e.g., Jones and Hobert, 2001; Mykland et al., 1995; Roberts and Rosenthal, 2004). We now describe a simple method of identifying the regeneration times in practice.

The Markov chain is simulated as usual, but after each iteration, an extra Bernoulli variable is drawn. To be more specific, suppose we start with $X_0 \sim \nu$ and then simulate the Markov chain according to k. After we have simulated $X_{n+1} = \tilde{x}$ using the current state, $X_n = x$, we generate an extra Bernoulli variable δ_n whose success probability is a function of x and $\tilde{x}$ given by

$$\Pr\big(\delta_n = 1 \mid X_n = x, X_{n+1} = \tilde{x}\big) = \frac{s(x)\,\nu(\tilde{x})}{k(\tilde{x} \mid x)}\,. \tag{17}$$

Then the regeneration times are $\tau_0 = 0$ and, for $t = 1, 2, 3, \ldots$, $\tau_t = \min\{n > \tau_{t-1} : \delta_{n-1} = 1\}$. Accordingly, the chain is broken up into "tours", $\big\{(X_{\tau_{(t-1)}}, \cdots, X_{\tau_t - 1}),\ t = 1, 2, \ldots\big\}$, that are independent stochastic replicas of each other.

Now suppose the chain X is Harris ergodic with invariant probability density π. Suppose that $g \in L_1(\pi)$ and consider using R tours of the Markov chain to estimate $\mathrm{E}_\pi g = \int_{\mathsf{X}} g(x)\pi(x)\,dx$. The total length of the simulation is τ_R, which is random. For $t = 1, \ldots, R$, define $N_t = \tau_t - \tau_{t-1}$, and $S_t = \sum_{n=\tau_{t-1}}^{\tau_t - 1} g(X_n)$. Then the (N_t, S_t) pairs are iid. We can write the obvious estimator of $\mathrm{E}_\pi g$ in terms of these pairs as follows

$$\widetilde{g}_R = \frac{1}{\tau_R}\sum_{n=0}^{\tau_R - 1} g(X_n) = \frac{\sum_{t=1}^{R} S_t}{\sum_{t=1}^{R} N_t}\,.$$

The estimator $\widetilde{g}_R$ is strongly consistent for $\mathrm{E}_\pi g$ as $R \to \infty$. Moreover, Hobert et al. (2002) show that, if the Markov chain X is geometrically ergodic and $\mathrm{E}_\pi |g|^{2+\alpha} < \infty$ for some $\alpha > 0$, then

$$\sqrt{R}\big(\widetilde{g}_R - \mathrm{E}_\pi g\big) \xrightarrow{\mathrm{d}} \mathrm{N}\big(0, \gamma^2\big) \quad \text{as } R \to \infty,$$

where

$$\gamma^2 = \frac{\mathrm{E}_\nu\left[(S_1 - N_1 \mathrm{E}_\pi g)^2\right]}{\left[\mathrm{E}_\nu N_1\right]^2}\,.$$

(The notation "E_ν" is meant to remind the reader that each tour is started with a draw from ν.) The entire motivation for using regeneration is that there is a simple, consistent estimator of γ^2. Indeed, Hobert et al. (2002) show that

$$\hat{\gamma}^2 = \frac{R\sum_{t=1}^{R}\big(S_t - \widetilde{g}_R N_t\big)^2}{\tau_R^2}$$

is a strongly consistent estimator of γ^2 as $R \to \infty$.

In practice, we choose a preliminary value of R that we believe will lead to a reasonable estimate of γ^2. We draw $X_0 \sim \nu$ and simulate iterations of the chain and the Bernoulli variables until the Rth time that $\delta = 1$. We then calculate $\widetilde{g}_R$ and $\hat{\gamma}^2$ and form the approximate 95% CI: $\widetilde{g}_R \pm 2\hat{\gamma}/\sqrt{R}$. If this interval is acceptably short, we stop. If not, we continue the simulation. Of course, given the pilot estimate $\hat{\gamma}^2$, we can calculate about how many tours will be required for a given level of accuracy. For example, if a CI of length l is desired, this will require about $16\hat{\gamma}^2/l^2$ total tours. Note that the chain is started with $X_0 \sim \nu$, and there is no need to throw away an initial portion of the simulation. In other words, when the regenerative method is employed, burn-in is a non-issue.

In the next subsection, we develop a minorization condition for our block Gibbs sampler.

4.2 Minorization for the block Gibbs sampler

The transition density of the block Gibbs chain is given by

$$k\big(\tilde{\sigma}^2, \tilde{\xi} \mid \sigma^2, \xi\big) = \pi\big(\tilde{\sigma}^2 \mid \xi\big)\, \pi\big(\tilde{\xi} \mid \tilde{\sigma}^2\big) \ .$$

We now construct a minorization condition for this transition density using a method outlined in Mykland et al. (1995). Fix $0 < d_1 < d_2 < \infty$ and $0 < d_3 < d_4 < \infty$ and let D denote the closed rectangle $[d_1, d_2] \times [d_3, d_4] \subset \mathbb{R}^2_+$. Also, fix a *distinguished point* $\xi^* \in \mathbb{R}^{q+1}$. Then

$$\begin{aligned}
k(\tilde{\sigma}^2, \tilde{\xi} \mid \sigma^2, \xi) &= \frac{\pi(\tilde{\sigma}^2 \mid \xi)}{\pi(\tilde{\sigma}^2 \mid \xi^*)} \pi(\tilde{\xi} \mid \tilde{\sigma}^2) \pi(\tilde{\sigma}^2 \mid \xi^*) \\
&\geq \left[\inf_{\sigma^2 \in D} \frac{\pi(\sigma^2 \mid \xi)}{\pi(\sigma^2 \mid \xi^*)} \right] \pi(\tilde{\xi} \mid \tilde{\sigma}^2) \pi(\tilde{\sigma}^2 \mid \xi^*) I_D(\tilde{\sigma}^2) \\
&= \left\{ c \frac{\pi(\underline{\sigma}^2 \mid \xi)}{\pi(\underline{\sigma}^2 \mid \xi^*)} \right\} \left\{ \frac{1}{c} \pi(\tilde{\xi} \mid \tilde{\sigma}^2) \pi(\tilde{\sigma}^2 \mid \xi^*) I_D(\tilde{\sigma}^2) \right\} \\
&=: s(\xi)\, \nu(\tilde{\sigma}^2, \tilde{\xi})
\end{aligned}$$

where $\underline{\sigma}^2 = (\underline{\sigma}^2_\theta, \underline{\sigma}^2_e)$ denotes the minimizer of $\pi(\sigma^2 \mid \xi)/\pi(\sigma^2 \mid \xi^*)$ as σ^2 ranges over the set D, and c is the normalizing constant. The value of c is not required in practice. Indeed, we can simulate from ν without knowledge of c by repeatedly drawing $\sigma^2 \sim \pi(\cdot \mid \xi^*)$ until the first time $\sigma^2 \in D$, and then drawing $\xi \sim \pi(\xi \mid \sigma^2)$. Furthermore, the probability (17), which must be calculated after each iteration of the Markov chain, involves s and ν only through their product. Thus, c cancels out.

We now develop a closed form expression for s. Since $\pi(\sigma^2 \mid \xi)$ factors into $\pi(\sigma_\theta^2 \mid \xi)$ and $\pi(\sigma_e^2 \mid \xi)$, the bivariate minimization problem becomes two separate univariate minimization problems. Let w_k^* stand for w_k evaluated at ξ^* for $k = 1, 2$. Then

$$\begin{aligned}\frac{\pi(\sigma_\theta^2 \mid \xi)}{\pi(\sigma_\theta^2 \mid \xi^*)} &= \frac{(\frac{1}{2}w_1)^{\frac{q}{2}+a}/\Gamma(\frac{q}{2}+a)}{(\frac{1}{2}w_1^*)^{\frac{q}{2}+a}/\Gamma(\frac{q}{2}+a)} \frac{\left(\sigma_\theta^2\right)^{-(\frac{q}{2}+a+1)} \exp[-\frac{1}{2}w_1/\sigma_\theta^2]}{\left(\sigma_\theta^2\right)^{-(\frac{q}{2}+a+1)} \exp[-\frac{1}{2}w_1^*/\sigma_\theta^2]} \\ &= \left(\frac{w_1}{w_1^*}\right)^{\frac{q}{2}+a} \exp\left[-\frac{1}{2}(w_1 - w_1^*)/\sigma_\theta^2\right],\end{aligned}$$

and

$$\begin{aligned}\frac{\pi(\sigma_e^2 \mid \xi)}{\pi(\sigma_e^2 \mid \xi^*)} &= \frac{\left[\frac{1}{2}(w_2 + \text{SSE})\right]^{\frac{M}{2}+b}/\Gamma(\frac{M}{2}+b)}{\left[\frac{1}{2}(w_2^* + \text{SSE})\right]^{\frac{M}{2}+b}/\Gamma(\frac{M}{2}+b)} \frac{\left(\sigma_e^2\right)^{-(\frac{M}{2}+b+1)} \exp\left[-\frac{1}{2}(w_2 + \text{SSE})/\sigma_e^2\right]}{\left(\sigma_e^2\right)^{-(\frac{M}{2}+b+1)} \exp\left[-\frac{1}{2}(w_2^* + \text{SSE})/\sigma_e^2\right]} \\ &= \left(\frac{w_2 + \text{SSE}}{w_2^* + \text{SSE}}\right)^{\frac{M}{2}+b} \exp\left[-\frac{1}{2}(w_2 - w_2^*)/\sigma_e^2\right].\end{aligned}$$

Hence, $\underline{\sigma}_\theta^2 = d_1$ when $w_1 > w_1^*$ and $\underline{\sigma}_\theta^2 = d_2$ when $w_1 \leq w_1^*$. Similarly, $\underline{\sigma}_e^2 = d_3$ when $w_2 > w_2^*$ and $\underline{\sigma}_e^2 = d_4$ when $w_2 \leq w_2^*$. Finally,

$$\begin{aligned}&\Pr\Big(\delta_n = 1 \;\Big|\; (\sigma_n^2, \xi_n) = (\sigma^2, \xi), (\sigma_{n+1}^2, \xi_{n+1}) = (\tilde{\sigma}^2, \tilde{\xi})\Big) \\ &= \frac{s(\xi)\,\nu(\tilde{\sigma}^2, \tilde{\xi})}{k\big(\tilde{\sigma}^2, \tilde{\xi} \mid \sigma^2, \xi\big)} = \frac{\pi(\underline{\sigma}^2 \mid \xi)\pi(\tilde{\sigma}^2 \mid \xi^*)}{\pi(\underline{\sigma}^2 \mid \xi^*)\pi(\tilde{\sigma}^2 \mid \xi)} I_D(\tilde{\sigma}^2) \\ &= \left(\frac{w_1}{w_1^*}\right)^{\frac{q}{2}+a} \exp\left[-\frac{1}{2}(w_1 - w_1^*)/\underline{\sigma}_\theta^2\right] \left(\frac{w_2 + \text{SSE}}{w_2^* + \text{SSE}}\right)^{\frac{M}{2}+b} \exp\left[-\frac{1}{2}(w_2 - w_2^*)/\underline{\sigma}_e^2\right] \\ &\times \left(\frac{w_1^*}{w_1}\right)^{\frac{q}{2}+a} \exp\left[-\frac{1}{2}(w_1^* - w_1)/\tilde{\sigma}_\theta^2\right] \left(\frac{w_2^* + \text{SSE}}{w_2 + \text{SSE}}\right)^{\frac{M}{2}+b} \exp\left[-\frac{1}{2}(w_2^* - w_2)/\tilde{\sigma}_e^2\right] I_D(\tilde{\sigma}^2) \\ &= \exp\left\{\frac{1}{2}\left[(w_1 - w_1^*)\left(\frac{1}{\tilde{\sigma}_\theta^2} - \frac{1}{\underline{\sigma}_\theta^2}\right) + (w_2 - w_2^*)\left(\frac{1}{\tilde{\sigma}_e^2} - \frac{1}{\underline{\sigma}_e^2}\right)\right]\right\} I_D(\tilde{\sigma}^2)\,.\end{aligned} \tag{18}$$

Theoretically, we could use any set $D = [d_1, d_2] \times [d_3, d_4]$ and any distinguished point ξ^* to run the regenerative simulation. However, the asymptotics for $\hat{\gamma}^2$ involve $R \to \infty$, so we would like for the chain to regenerate fairly often. Thus, we should choose D and ξ^* so that the probability in (18) is frequently close to one. Not surprisingly, there is trade-off between the size of the set D and the magnitude of the exponential term in (18) (when the indicator

is unity). Our strategy for choosing D and ξ^* is as follows. We run the block Gibbs sampler for an initial n_0 iterations (using starting value $\xi_0 = (\overline{y}, \overline{y}_1, \ldots, \overline{y}_q)$ for example). We take $[d_1, d_2]$ to be the shortest interval that contains 60% of the n_0 values of σ_θ^2, and we calculate $[d_3, d_4]$ similarly using the n_0 values of σ_e^2. The regeneration probability (18) involves ξ^* only through $w_1(\xi^*)$ and $w_2(\xi^*)$. Hence, instead of setting ξ^* equal to the median, say, of the n_0 values of ξ in the initial run of the chain, we calculate w_1 and w_2 for each of the n_0 values of ξ and we set w_1^* to be the median of the w_1 values and w_2^* to be the median of the w_2 values. There is one small caveat. This approach makes sense only if there happens to exist a $\hat{\xi} \in \mathbb{R}^{q+1}$ such that $\big(w_1(\hat{\xi}), w_2(\hat{\xi})\big) = (w_1^*, w_2^*)$. For balanced data, such a $\hat{\xi}$ exists if and only if $\sqrt{\frac{w_2^*}{m}} + \sqrt{w_1^*} \geq \sqrt{\frac{\text{SST}}{m}}$. See Tan (2008) for a proof of this result as well as guidelines for the unbalanced case.

5 An example: Styrene exposure data

In this section, we illustrate the regenerative simulation method using a real data set from Lyles, Kupper and Rappaport (1997). The computer simulation is coded in R and can be found in the online supplement to this article, available from the journal's website. The data set concerns thirteen workers who were randomly selected from a group within a boat manufacturing plant and each one's styrene exposure was measured on three separate occasions. So we have $q = 13$, $m_i \equiv m = 3$ and $M = m \times q = 39$. The data are summarized in Tables 1 and 2.

Consider modeling these data using the one-way model from Section 1 with the standard diffuse prior on the unknown parameters. The goal will be to explore the posterior distribution of (σ^2, ξ) given the data using the block Gibbs sampler. With $q = 13$ and $m_i \equiv m = 3$, the conditions of Corollary 1 are clearly satisfied so the block Gibbs sampler for the styrene data is geometrically ergodic. Suppose we want to approximate the posterior expectations of the two variance components, σ_θ^2 and σ_e^2, as well as the correlation between observations on the same worker, $\sigma_\theta^2/(\sigma_\theta^2+\sigma_e^2)$. Straightforward calculations show that all three of these functions satisfy the "$2 + \alpha$" moment condition. Therefore, all of the assumptions underlying the regenerative simulation method are satisfied.

Implementation of the regenerative simulation requires us to specify R, the total number

worker	1	2	3	4	5	6	7
$\overline{y}_i$	3.302	4.587	5.052	5.089	4.498	5.186	4.915
worker	8	9	10	11	12	13	
$\overline{y}_i$	4.876	5.262	5.009	5.602	4.336	4.813	

Table 1: Average styrene exposure level for each of the 13 workers.

$$\overline{y} = M^{-1} \sum_{i=1}^{13} \sum_{j=1}^{3} y_{ij} = 4.809$$
$$\text{SST} = 3 \sum_{i=1}^{13} (\overline{y}_i - \overline{y})^2 = 11.430$$
$$\text{SSE} = \sum_{i=1}^{13} \sum_{j=1}^{3} (y_{ij} - \overline{y}_i)^2 = 14.711$$

Table 2: Summary statistics for the styrene exposure data.

of regenerations. As we have mentioned in Section 4.1, the procedure to determine R has two steps. In the first step, we run the chain for an initial number of regenerations, call it R, that is believed to lead to a reasonable estimator of the asymptotic variance, γ^2. Here, we used $R = 5,000$ which took 87,169 iterations and consumed 20 seconds. The simulation results are summarized in Table 3. For each of the three parameters of interest, the table provides the estimate, $\tilde{g}_{\tau_R}$, the estimated asymptotic variance, $\hat{\gamma}^2$, the estimated standard error $\sqrt{\hat{\gamma}^2/R}$, and an approximate 95% CI, $\overline{g}_{\tau_R} \pm 2\sqrt{\hat{\gamma}^2/R}$. Mykland et al. (1995) recommended that $\hat{\gamma}^2$ not be used to estimate γ^2 unless the average tour length, $\overline{N} = R^{-1} \sum_{t=1}^{R} N_t$ has a coefficient of variation, $\text{CV}(\bar{N}) = \sqrt{\text{Var}(\bar{N})}/\text{E}(\bar{N})$, smaller than 0.1. A strongly consistent estimator of $\text{CV}(\bar{N})$ is given by $\widehat{\text{CV}}(\overline{N}) = \sqrt{\sum_{t=1}^{R}(N_t - \bar{N})^2/(R\bar{N})^2}$. For our simulation above, $\widehat{\text{CV}}(\overline{N}) = 0.018$ clearly meets the criteria. We also examined trace plots of $\hat{\gamma}^2$ for the parameters of interest, σ_θ^2, σ_e^2 and $\sigma_\theta^2/(\sigma_\theta^2 + \sigma_e^2)$, and all suggest that the variance estimators have stabilized by the 5,000th regeneration. Hence they are reasonable approximations of their respective estimands.

In the second step of the procedure, we decide how large R needs to be for the resulting CI to be shorter than a user-specified width based on the preliminary analysis above. Take the 95% CI of $\text{E}_\pi \sigma_\theta^2$ for example. Suppose that we desire its margin of error to be around 1% of the mag-

	$\tilde{g}_{\tau_R}$	$\hat{\gamma}^2$	$\sqrt{\hat{\gamma}^2/R}$	$\overline{g}_{\tau_R} \pm 2\sqrt{\hat{\gamma}^2/R}$
σ_θ^2	0.19003	0.03463	0.00263	(0.18477, 0.19529)
σ_e^2	0.61777	0.00883	0.00133	(0.61511, 0.62043)
$\frac{\sigma_\theta^2}{\sigma_\theta^2+\sigma_e^2}$	0.21288	0.03532	0.00266	(0.20757, 0.21820)

Table 3: Results based on $R = 5,000$ regenerations.

	$\tilde{g}_{\tau_R}$	$\hat{\gamma}^2$	$\sqrt{\hat{\gamma}^2/R}$	$\overline{g}_{\tau_R} \pm 2\sqrt{\hat{\gamma}^2/R}$
σ_θ^2	0.19023	0.03523	0.00094	(0.18835, 0.19210)
σ_e^2	0.61849	0.00966	0.00049	(0.61751, 0.61947)
$\frac{\sigma_\theta^2}{\sigma_\theta^2+\sigma_e^2}$	0.21304	0.03687	0.00096	(0.21112, 0.21496)

Table 4: Results based on $R = 40,000$ regenerations. Note that the margin of error of the 95% CI for $\mathrm{E}_\pi\sigma_\theta^2$ is $2 \times 0.00094 = 0.00188$, which is about 1% of the magnitude of the estimate of $\mathrm{E}_\pi\sigma_\theta^2$.

nitude of $\mathrm{E}_\pi\sigma_\theta^2$. Since $\tilde{g}_{\tau_{5000}} = 0.19003$, the desired width of the 95% CI, l, is approximately $2 \times 0.19003 \times 1\% \doteq 0.0038$ and will require about $16\hat{\gamma}^2/l^2 = 16 \times 0.03074/0.0038^2 \doteq 38,371$ regenerations. So we ran the chain for an additional 35,000 regenerations. The final chain with 40,000 regenerations accounted for 697,869 iterations and took 3 minutes to generate. The simulation results are summarized in Table 4. Figure 1 shows trace plots associated with the estimation of $\mathrm{E}_\pi\sigma_\theta^2$. These plots suggest that things have stabilized nicely by the 40,000th regeneration. We also examined the trace plots associated with σ_e^2 and $\sigma_\theta^2/(\sigma_\theta^2 + \sigma_e^2)$. They look very similar to the ones shown in Figure 1.

6 Discussion

Our block Gibbs sampler is a two-variable Gibbs sampler that updates $\sigma^2 = (\sigma_\theta^2, \sigma_e^2)$ and $\xi = (\mu, \theta)$ in turn. van Dyk and Meng (2001) studied an alternative two-variable Gibbs sampler in which the two groups of parameters are θ and $(\mu, \sigma_\theta^2, \sigma_e^2)$. Given $(\mu, \sigma_\theta^2, \sigma_e^2)$, θ has a multivariate normal distribution. On the other hand, the posterior density of $(\mu, \sigma_\theta^2, \sigma_e^2)$ given θ factors as

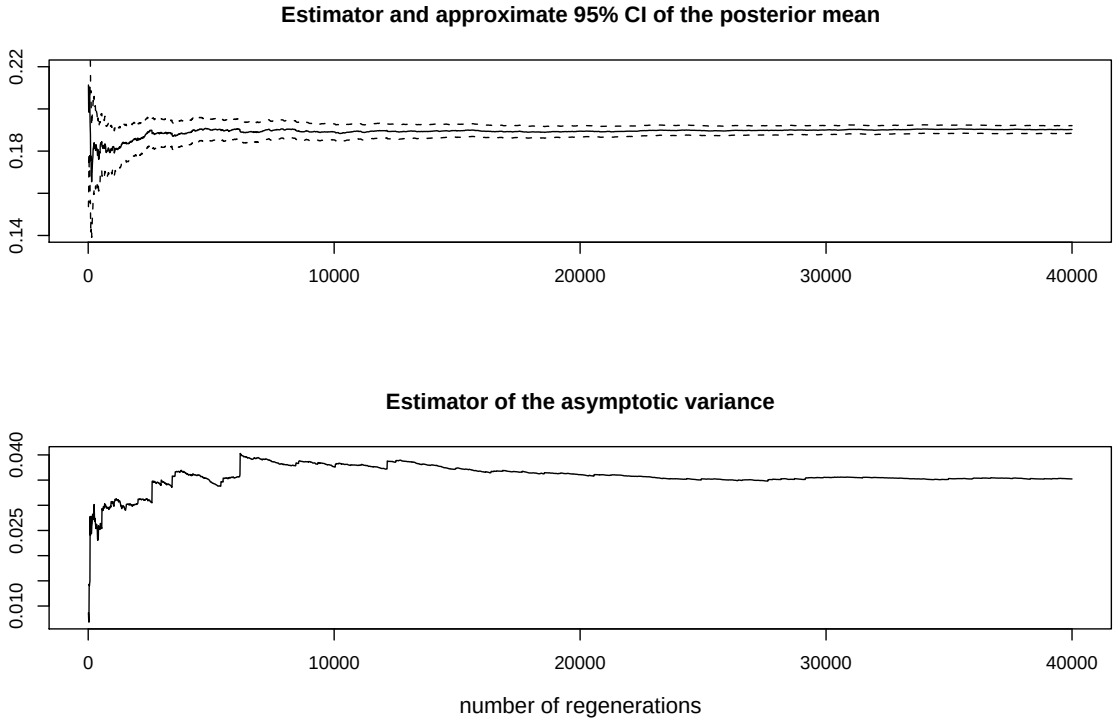

Figure 1: The solid line in the upper graph is the trace plot of the estimator of $\mathrm{E}_\pi \sigma_\theta^2$. That is, for $R = 1, \ldots, 40000$, we plotted $\widetilde{g}_R$ against R, where $\widetilde{g}_R$ is the estimator based on the first R tours out of the $40,000$ tours of our simulated chain. Similarly, the dashed lines are trace plots of the upper and lower bounds of the estimated asymptotic 95% CIs for $\mathrm{E}_\pi \sigma_\theta^2$. The lower graph is another trace plot that displays the convergence of the strongly consistent estimator, $\hat{\gamma}^2$, of the asymptotic variance, γ^2 .

$\pi(\mu, \sigma_\theta^2, \sigma_e^2 \mid \theta) = \pi(\sigma_e^2 \mid \theta)\, \pi(\mu, \sigma_\theta^2 \mid \theta)$. That is, given θ, σ_e^2 and (μ, σ_θ^2) are independent. It's easy to show that $\pi(\sigma_e^2 \mid \theta)$ has an inverse gamma density. Moreover, $\pi(\mu, \sigma_\theta^2 \mid \theta)$ can be factored as $\pi(\sigma_\theta^2 \mid \theta)\, \pi(\mu \mid \sigma_\theta^2, \theta)$ and routine calculations show that $\pi(\sigma_\theta^2 \mid \theta)$ and $\pi(\mu \mid \sigma_\theta^2, \theta)$ have inverse gamma and normal forms, respectively. Thus, it is just as easy to implement this two-variable Gibbs sampler as it is to implement our block Gibbs sampler. Unfortunately, our proof of geometric ergodicity cannot be easily adapted to van Dyk and Meng's chain because the drift condition we used is not appropriate for their blocking scheme. We strongly suspect that this alternative Markov chain is geometrically ergodic, but this remains an open question.

Supplemental materials

appendices.pdf This document contains the proofs for the results in the article and other technical details.

regenerate.R This file contains code to calculate the consistent estimator of the asymptotic variance using the regeneration method for the styrene exposure data set used in Section 5.

Acknowledgments

The authors thank the associate editor and two referees for helpful comments and suggestions. Hobert's research was supported by NSF grants DMS-0503648 and DMS-0805860.

References

CHAN, K. S. and GEYER, C. J. (1994). Discussion of "Markov chains for exploring posterior distributions". *The Annals of Statistics*, **22** 1747–1757.

DANIELS, M. J. (1999). A prior for the variance in hierarchical models. *The Canadian Journal of Statistics*, **27** 567–578.

DIACONIS, P., KHARE, K. and SALOFF-COSTE, L. (2008). Gibbs sampling, exponential families and orthogonal polynomials (with discussion). *Statistical Science*, **23** 151–200.

GELFAND, A. E. and SMITH, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, **85** 398–409.

GELMAN, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, **1** 515–533.

GEYER, C. J. (1992). Practical Markov chain Monte Carlo (with Discussion). *Statistical Science*, **7** 473–511.

HILL, B. M. (1965). Inference about variance components in the one-way model. *Journal of the American Statistical Association*, **60** 806–825.

HOBERT, J. P. and CASELLA, G. (1996). The effect of improper priors on Gibbs sampling in hierarchical linear mixed models. *Journal of the American Statistical Association*, **91** 1461–1473.

HOBERT, J. P. and GEYER, C. J. (1998). Geometric ergodicity of Gibbs and block Gibbs samplers for a hierarchical random effects model. *Journal of Multivariate Analysis*, **67** 414–430.

HOBERT, J. P., JONES, G. L., PRESNELL, B. and ROSENTHAL, J. S. (2002). On the applicability of regenerative simulation in Markov chain Monte Carlo. *Biometrika*, **89** 731–743.

JOHNSON, A. A. and JONES, G. L. (2008). Gibbs sampling for a Bayesian hierarchical version of the general linear mixed model. Tech. rep., University of Minnesota.

JONES, G. L., HARAN, M., CAFFO, B. S. and NEATH, R. (2006). Fixed-width output analysis for Markov chain Monte Carlo. *Journal of the American Statistical Association*, **101** 1537–1547.

JONES, G. L. and HOBERT, J. P. (2001). Honest exploration of intractable probability distributions via Markov chain Monte Carlo. *Statistical Science*, **16** 312–34.

JONES, G. L. and HOBERT, J. P. (2004). Sufficient burn-in for Gibbs samplers for a hierarchical random effects model. *The Annals of Statistics*, **32** 784–817.

LIU, J. S., WONG, W. H. and KONG, A. (1994). Covariance structure of the Gibbs sampler with applications to comparisons of estimators and augmentation schemes. *Biometrika*, **81** 27–40.

LYLES, R. H., KUPPER, L. L. and RAPPAPORT, S. M. (1997). Assessing regulatory compliance of occupational exposures via the balanced one-way random effects ANOVA model. *Journal of Agricultural, Biological, and Environmental Statistics*, **2** 417–439.

MEYN, S. P. and TWEEDIE, R. L. (1993). *Markov Chains and Stochastic Stability*. Springer Verlag, London.

MYKLAND, P., TIERNEY, L. and YU, B. (1995). Regeneration in Markov chain samplers. *Journal of the American Statistical Association*, **90** 233–41.

PAPASPILIOPOULOS, O. and ROBERTS, G. (2008). Stability of the Gibbs sampler for Bayesian hierarchical models. *The Annals of Statistics*, **36** 95–117.

ROBERTS, G. O. and ROSENTHAL, J. S. (2001). Markov chains and de-initializing processes. *Scandinavian Journal of Statistics*, **28** 489–504.

ROBERTS, G. O. and ROSENTHAL, J. S. (2004). General state space Markov chains and MCMC algorithms. *Probability Surveys*, **1** 20–71.

TAN, A. (2008). *Analysis of Markov chain Monte Carlo algorithms for random effects models*. Ph.D. thesis, Department of Statistics, University of Florida.

TIAO, G. C. and TAN, W. Y. (1965). Bayesian analysis of random-effect models in the analysis of variance. I. Posterior distribution of variance components. *Biometrika*, **52** 37–53.

VAN DYK, D. A. and MENG, X.-L. (2001). The art of data augmentation (with discussion). *Journal of Computational and Graphical Statistics*, **10** 1–50.