# 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 5th 2008

**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.

## 1 Introduction

Consider the classical one-way random effects model given by

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

where the random effects $\theta_1, \dots, \theta_q$ are iid $\mathrm{N}(\mu, \sigma_\theta^2)$, the $\varepsilon_{ij}$s are iid $\mathrm{N}(0, \sigma_e^2)$ and independent of the $\theta_i$s, and $(\mu, \sigma_\theta^2, \sigma_e^2)$ is an unknown parameter. There is a long history of Bayesian analysis using this

*Key words and phrases.* Asymptotic variance, Central limit theorem, Convergence rate, Drift condition, Geometric ergodicity, Minorization condition, One-way model, Standard error

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

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

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

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

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

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

Since the observed data are always conditioned upon, we have suppressed the notation of dependence on $y$ in the left-hand side of (2) 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 $(\mu, \sigma_e^2)$ 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 $\sigma_\theta^2$, so the choice of prior for this parameter is more important and subtle. A commonly used prior for $\sigma_\theta^2$ is a proper inverse gamma prior, which is a (conditionally) conjugate prior. When little or no prior information concerning $\sigma_\theta^2$ 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 $\big(\sigma_\theta^2\big)^{-\frac{1}{2}}$ works well unless $q$ is very small (say, below 5). Combining this prior with a uniform prior on $\big(\mu, \log(\sigma_e^2)\big)$ 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 all three of the following conditions hold:

$$a < 0\,, \qquad a + \frac{q}{2} > \frac{1}{2}\,, \qquad \text{and} \qquad a + b > \frac{1-M}{2}\,, \tag{3}$$

where $M$ is the total sample size; that is, $M = \sum_{i=1}^{q} m_i$. Note that (3) 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

$$\frac{\int_{\mathbb{R}^q} \int_{\mathbb{R}} \int_{\mathbb{R}_+} \int_{\mathbb{R}_+} g(\theta, \mu, \sigma_\theta^2, \sigma_e^2) f(y|\theta, \mu, \sigma_\theta^2, \sigma_e^2) f(\theta|\mu, \sigma_\theta^2, \sigma_e^2) \pi_{a,b}\big(\mu, \sigma_\theta^2, \sigma_e^2\big)\, d\sigma_e^2\, d\sigma_\theta^2\, d\mu\, d\theta}{\int_{\mathbb{R}^q} \int_{\mathbb{R}} \int_{\mathbb{R}_+} \int_{\mathbb{R}_+} f(y|\theta, \mu, \sigma_\theta^2, \sigma_e^2) f(\theta|\mu, \sigma_\theta^2, \sigma_e^2) \pi_{a,b}\big(\mu, \sigma_\theta^2, \sigma_e^2\big)\, d\sigma_e^2\, d\sigma_\theta^2\, d\mu\, d\theta}\,, \tag{4}$$

which is a ratio of two intractable integrals. Results in Section 2 show that it is actually possible to integrate $\theta$ and $\mu$ out of the integral in the denominator in closed form, so the denominator is really a 2-dimensional intractable integral. However, in general, the numerator is an intractable integral of dimension $q + 3$.

It is now well known that Markov chain Monte Carlo (MCMC) methods, and, in particular, the Gibbs sampler, can be used to approximate the posterior expectation in (4). In fact, one of the main examples in the seminal paper by Gelfand and Smith (1990) is an application of the Gibbs sampler to a Bayesian version of the one-way model with *proper* conjugate priors for $\mu$, $\sigma_\theta^2$ and $\sigma_e^2$. These authors considered the *simple* Gibbs sampler that 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 $\sigma^2 = \big(\sigma_\theta^2, \sigma_e^2\big)$, $\xi = (\mu, \theta)$ and suppose that the state of the chain at time $n$ is $\big(\sigma_n^2, \xi_n\big)$. One iteration of our sampler entails drawing $\sigma_{n+1}^2$ conditional on $\xi_n$, and then drawing $\xi_{n+1}$ conditional on $\sigma_{n+1}^2$. "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). Formally, the Markov transition density for the transition $\big(\sigma_n^2, \xi_n\big) \to \big(\sigma_{n+1}^2, \xi_{n+1}\big)$ is given by

$$k\big(\sigma_{n+1}^2, \xi_{n+1} \mid \sigma_n^2, \xi_n\big) = \pi\big(\sigma_{n+1}^2 \mid \xi_n\big)\, \pi\big(\xi_{n+1} \mid \sigma_{n+1}^2\big)\,.$$

(Recall that we are suppressing dependence on $y$.) Straightforward manipulation of (2) shows that, given $\xi$, $\sigma_\theta^2$ and $\sigma_e^2$ are independent random variables each with inverse gamma distributions, and given

$\sigma^2$, $\xi$ is multivariate normal. (The specific forms of these distributions are given in Section 2.) Thus, programming this block Gibbs sampler is straightforward.

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 (4). Let $L_1(\pi)$ denote the set of functions $h : \mathbb{R}^{q+1} \times \mathbb{R}_+^2 \to \mathbb{R}$ such that

$$\int_{\mathbb{R}^{q+1}} \int_{\mathbb{R}_+^2} \big|h(\sigma^2,\xi)\big| \pi(\sigma^2,\xi)\, d\sigma^2\, d\xi < \infty \;,$$

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 $\big(\sigma_0^{2*}, \xi_0^*\big), \big(\sigma_1^{2*}, \xi_1^*\big), \dots$ 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, in practice we could choose a preliminary value of $N$, say $N'$, draw a random sample of size $N'$ from $\pi$ 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 simply increase the sample size to a level that will ensure an acceptably small standard error. The main message here is that routine use of the CLT allows for straightforward determination of an appropriate Monte Carlo sample size.

Of course, 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, but the ergodic theorem (Meyn and Tweedie, 1993, Chapter 17) implies that it is a strongly consistent estimator of $\mathrm{E}_\pi g$. Thus, at this point in the comparison, all we've lost by using a Markov chain in place of the iid sample is unbiasedness! In other words, the iid sample can be replaced with a Markov chain sample (which is much easier to get) and the strong consistency of the classical Monte Carlo estimator still obtains. 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, the routine use of the CLT for choosing an appropriate sample size in the classical Monte Carlo context is far from routine when using MCMC. There are two reasons for this. The first is that, 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. Moreover, the standard method of establishing that there is a CLT is to prove that the underlying Markov chain is *geometrically ergodic* (see Appendix A for the definition), and this often requires difficult theoretical analysis (see, e.g., Jones and Hobert, 2001). The second reason is that, even when a CLT is known to hold, finding a consistent estimator of the the asymptotic variance is a challenging problem because this variance has a fairly complex form and because the dependence among the variables in the Markov chain complicates the asymptotic analysis of estimators based on the chain (see, e.g., Geyer (1992) and Chan and Geyer (1994)).

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 (Hobert, Jones, Presnell and Rosenthal, 2002; 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, to apply our method, one simply runs the block Gibbs chain as usual, but after each iteration, a single Bernoulli variable is drawn. Specifically, suppose that the value of the chain at time $n$ is $\left(\sigma_n^2,\xi_n\right)$. We draw $\left(\sigma_{n+1}^2,\xi_{n+1}\right)$ as usual, and then draw a Bernoulli variable, call it $\delta_n$, whose success probability is a simple function of $\left(\sigma_n^2, \xi_n\right)$, $\left(\sigma_{n+1}^2, \xi_{n+1}\right)$ and a few constants (see equation (19)). Each time that $\delta_n = 1$ marks the end of a tour. (Note that the tours have random lengths.) Now consider running the block Gibbs sampler for $R$ tours; that is, we run the chain until the $R$th time that a $\delta_n = 1$. For $t = 1, \ldots, R$, let $N_t$ denote the (random) length of the $t$th tour and let $S_t = \sum g\left(\sigma_n^2, \xi_n\right)$ where the sum ranges over the values of $n$ that constitute the $t$th tour. The total number of iterations is $N = \sum_{t=1}^R N_t$. The obvious estimate of $\mathrm{E}_\pi g$ based on this simulation is

$$\widetilde{g}_R = \frac{1}{N} \sum_{n=0}^{N-1} g\left(\sigma_n^2, \xi_n\right) = \frac{\sum_{t=1}^R S_t}{\sum_{t=1}^R N_t} ,$$

which is the same as the usual estimator except that now the length of the simulation is random. The fact that the pairs $(N_1, S_1), \ldots, (N_R, S_R)$ are iid can be used to show that, if the block Gibbs Markov chain is geometrically ergodic and $\mathrm{E}_\pi |g|^{2+\alpha} < \infty$ for some $\alpha > 0$, then there exists a $\gamma^2 \in (0, \infty)$ such that, as $R \to \infty$, we have

$$\sqrt{R}\left(\widetilde{g}_R - \mathrm{E}_\pi g\right) \xrightarrow{\mathrm{d}} \mathrm{N}(0, \gamma^2) . \tag{5}$$

Furthermore, a simple, consistent estimator of $\gamma^2$ is given by

$$\hat{\gamma}^2 = \frac{R \sum_{t=1}^R \left(S_t - \widetilde{g}_R N_t\right)^2}{N^2} .$$

We conclude that, if the block Gibbs chain is geometrically ergodic and the "$2 + \alpha$" moment condition is satisfied, then we can calculate a valid asymptotic standard error for $\widetilde{g}_R$ and only stop the simulation when this standard error is acceptably small. In other words, we can mimic what is done in the classical Monte Carlo context.

Finally, our convergence rate analysis yields conditions under which the block Gibbs chain is geometrically ergodic. Loosely speaking, we are able to say that the chain is geometrically ergodic unless $M$ is very small and the within group sample sizes, $m_1, \ldots, m_q$, are highly unbalanced. Our convergence rate result (Proposition 2) applies to all priors $\pi_{a,b}$. Here is a special case.

**Corollary 1.** *Under the standard diffuse prior, the block Gibbs Markov 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.3 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.

Previous analyses 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 2 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 2 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 a detailed description of the block Gibbs Markov chain as well as the statement and proof of our convergence rate result. In Section 3, 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 4 using a real data set on styrene exposure. A short discussion appears in Section 5. Appendix A contains important background material on general state space Markov chains, and Appendix B contains some of the technical details from the proof of geometric ergodicity.

# 2 Geometric Ergodicity of the Block Gibbs Sampler

## 2.1 The block Gibbs sampler and the marginal chains

Recall that $\sigma^2 = \left(\sigma_\theta^2, \sigma_e^2\right)$, $\xi = (\mu, \theta)$, and that our block Gibbs Markov 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 $(\sigma^2, \xi)$ and $(\tilde{\sigma}^2, \tilde{\xi})$ denote the current and next states, respectively. Routine manipulation of (2) shows that $\sigma_\theta^2$ and $\sigma_e^2$ are conditionally independent given $\xi$; that is,

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

and that

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

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

Further manipulation of (2) (see Tan (2008)) shows that, given $\sigma^2$, $\xi$ 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

$$t = \sum_{i=1}^{q} \frac{m_i}{\sigma_e^2 + m_i \sigma_\theta^2} \;.$$

Using this notation, we have

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

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

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

The variances and covariances are given by

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

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 in Appendix A implies that the block

Gibbs Markov chain, $\{(\sigma_n^2, \xi_n)\}_{n=0}^{\infty}$, is Harris ergodic. As we now explain, our proof that this chain is also geometrically ergodic is indirect and rests upon an analysis of $\{\xi_n\}_{n=0}^{\infty}$.

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

$$k^*(\tilde{\xi} \mid \xi) = \int_{\mathbb{R}_+^2} \pi(\tilde{\xi} \mid \sigma^2)\, \pi(\sigma^2 \mid \xi)\, d\sigma^2 \,. \tag{6}$$

Clearly, $k^*$ is strictly positive on $\mathbb{R}^{q+1} \times \mathbb{R}^{q+1}$ so another application of Lemma 1 shows that the $\xi$-chain is Harris ergodic. We also conclude from Lemma 1 that the maximal irreducibility measure of the $\xi$-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 $\xi$-chain is a Feller chain. We are now in a position to use Proposition 3 in Appendix A to prove that the $\xi$-chain is geometric.

## 2.2 A proof of geometric ergodicity

A function $w : \mathbb{R}^{q+1} \to \mathbb{R}_+$ is said to be *unbounded off compact sets* if the level set $\{\xi : w(\xi) \le \gamma\}$ is compact for every $\gamma < \infty$. According to Proposition 3, we can prove that the $\xi$-chain is geometric by finding a $w$ that is unbounded off compact sets and satisfies the drift condition

$$\mathrm{E}\big(w(\tilde{\xi}) \mid \xi\big) \le \rho w(\xi) + L \qquad \text{for all } \xi \in \mathbb{R}^{q+1}, \tag{7}$$

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. It is easy to see that, for fixed $\epsilon > 0$ and $s \in (0,1]$, the function $w$ is unbounded off compact sets. Indeed, since $w$ is continuous, it is enough to show that, in the level set $\{\xi : w(\xi) \le \gamma\}$, $|\mu|$ 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 $\theta_i$s contained. Now, given that the $\theta_i$s are contained, $w_1 \to \infty$ as $|\mu| \to \infty$, so $\mu$ 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 (7) 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 6 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 $\xi \to \sigma^2 \to \tilde{\xi}$. This allows us to calculate the expectations above by conditioning on $\sigma^2$. Indeed, for $k \in \{1, 2\}$, we have

$$\mathrm{E}(\tilde{w}_k^s \mid \xi) = \mathrm{E}\big[\mathrm{E}(\tilde{w}_k^s \mid \sigma^2, \xi) \mid \xi\big] = \mathrm{E}\big[\mathrm{E}(\tilde{w}_k^s \mid \sigma^2) \mid \xi\big] \tag{8}$$

where the second equality follows from the fact that $\tilde{\xi}$ is conditionally independent of $\xi$ given $\sigma^2$.

Since there are no restrictions on the constant $L$ in (7), 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

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

where

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

and

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

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

$$\mathrm{E}(\tilde{w}_1^s \mid \sigma^2) \le \big[\mathrm{E}(\tilde{w}_1 \mid \sigma^2)\big]^s \le \big(\Delta_1 \sigma_\theta^2 + \Delta_2 \sigma_e^2 + \mathit{const}\big)^s \le \Delta_1^s \big(\sigma_\theta^2\big)^s + \Delta_2^s \big(\sigma_e^2\big)^s + \mathit{const} \ , \tag{9}$$

and

$$\mathrm{E}(\tilde{w}_2^s \mid \sigma^2) \le \big[\mathrm{E}(\tilde{w}_2 \mid \sigma^2)\big]^s \le \big((q+1)\sigma_e^2 + \mathit{const}\big)^s \le (q+1)^s \big(\sigma_e^2\big)^s + \mathit{const} \ . \tag{10}$$

To complete the calculation in (8), 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

$$\mathrm{E}\Big(\big(\sigma_\theta^2\big)^s \mid \xi\Big) = \frac{\Gamma(\frac{q}{2}+a-s)}{2^s\Gamma(\frac{q}{2}+a)} w_1^s \;, \tag{11}$$

and

$$\mathrm{E}\Big(\big(\sigma_e^2\big)^s \mid \xi\Big) = \frac{\Gamma(\frac{M}{2}+b-s)}{2^s\Gamma(\frac{M}{2}+b)} (w_2 + \mathrm{SSE})^s \le \frac{\Gamma(\frac{M}{2}+b-s)}{2^s\Gamma(\frac{M}{2}+b)} w_2^s + const \;. \tag{12}$$

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 (8)-(12), we have

$$\begin{aligned} \mathrm{E}(\tilde{w}_1^s \mid \xi) &\le \mathrm{E}\Big(\Delta_1^s\big(\sigma_\theta^2\big)^s + \Delta_2^s\big(\sigma_e^2\big)^s + const \,\Big|\, \xi\Big) \\ &\le \Delta_1^s \frac{\Gamma(\frac{q}{2}+a-s)}{2^s\Gamma(\frac{q}{2}+a)} w_1^s + \Delta_2^s \frac{\Gamma(\frac{M}{2}+b-s)}{2^s\Gamma(\frac{M}{2}+b)} w_2^s + const \\ &= \delta_3(s) w_1^s + \delta_2(s) w_2^s + const \;. \end{aligned} \tag{13}$$

and

$$\begin{aligned} \mathrm{E}(\tilde{w}_2^s \mid \xi) &\le \mathrm{E}\Big((q+1)^s\big(\sigma_e^2\big)^s + const \,\Big|\, \xi\Big) \\ &\le (q+1)^s \frac{\Gamma(\frac{M}{2}+b-s)}{2^s\Gamma(\frac{M}{2}+b)} w_2^s + const \\ &= \delta_1(s) w_2^s + const \;. \end{aligned} \tag{14}$$

**Proposition 1.** *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*

$$E\big(w(\tilde{\xi}) \mid \xi\big) \le \rho w(\xi) + L \quad \textit{for all } \xi \in \mathbb{R}^{q+1}.$$

*Proof.* It follows from (13) and (14) that

$$\begin{aligned} \mathrm{E}(\epsilon\tilde{w}_1^s + \tilde{w}_2^s \mid \xi) &\le \epsilon\delta_3(s) w_1^s + (\delta_1(s) + \epsilon\delta_2(s)) w_2^s + const \\ &= \rho(\epsilon, s)(\epsilon w_1^s + w_2^s) + \epsilon(\delta_3(s) - \rho(\epsilon, s)) w_1^s + const \end{aligned}$$

where $\rho(\epsilon, s) = \delta_1(s) + \epsilon\delta_2(s)$. Therefore, we will have a viable drift condition if

$$\rho(\epsilon, s) < 1 \;\text{ and }\; \delta_3(s) - \rho(\epsilon, s) \le 0 \;. \tag{15}$$

Clearly, (15) requires that $\delta_1(s) < 1$ and $\delta_3(s) < 1$. We now show that these conditions are also sufficient for the existence of $\epsilon > 0$ such that (15) is satisfied.

There are two cases. In the first case, $\delta_1(s) \le \delta_3(s) < 1$. If we take $\epsilon = (\delta_3(s) - \delta_1(s))/\delta_2(s)$, then $\rho(\epsilon, s) = \delta_3(s) < 1$ and $\delta_3(s) - \rho(\epsilon, s) = 0$. In the second case, $\delta_3(s) < \delta_1(s) < 1$. Now take $\epsilon = \big(1 - \delta_1(s)\big)/\big(2\delta_2(s)\big)$. Then

$$\rho(\epsilon, s) = \delta_1(s) + \frac{1 - \delta_1(s)}{2} = \frac{1 + \delta_1(s)}{2} < 1 \ ,$$

and

$$\delta_3(s) - \rho(\epsilon, s) = \delta_3(s) - \frac{1 + \delta_1(s)}{2} < 0 \ .$$

Hence, if $\delta_1(s) < 1$ and $\delta_3(s) < 1$, then there is a viable drift condition. $\square$

In conjunction with Proposition 3, Proposition 1 shows that the $\xi$-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.2 that the desired $s$ exists if $M + 2b \ge q + 3$ and $\Delta_1 < 2\exp\big(\Psi\big(\frac{q}{2} + a\big)\big)$. We can now state our main convergence rate result.

**Proposition 2.** *The Markov chain underlying the block Gibbs sampler is geometrically ergodic if*

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

Loosely speaking, Proposition 2 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 (and the condition will fail) only if *both* $\big(\sum_{i=1}^{q} \frac{m_i}{m_i+1}\big)^{-1}$ and $\frac{m^*}{M}$ are large. The first term increases as the $m_i$s 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.3 that Corollary 1 from the Introduction follows straightforwardly from Proposition 2.

## 3 Minorization, Regeneration and the CLT

In this section, we provide a general description of how a minorization condition can be used to introduce regenerations into a Markov chain. We then explain how these regenerations can be used to form a CLT whose asymptotic variance is easy to estimate. Finally, we establish a minorization condition for our block Gibbs Markov chain and provide a formula for the success probabilities of the extra Bernoulli draws described in the Introduction.

## 3.1 The method

Let $\mathsf{X} \subset \mathbb{R}^p$ and suppose that $k : \mathsf{X} \times \mathsf{X} \to [0, \infty)$ is a Markov transition density that defines a Markov chain $X = \{X_n\}_{n=0}^{\infty}$. Thus, for $x \in \mathsf{X}$ and $A \subset \mathsf{X}$, $\Pr(X_{n+1} \in A \mid X_n = x) = \int_A k(\tilde{x} \mid x)\, d\tilde{x}$. Suppose we can find a (non-trivial) function $s : \mathsf{X} \to [0, 1)$ and a density function $\nu : \mathsf{X} \to [0, \infty)$ such that,

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

Equation (16) is called a *minorization condition* (Jones and Hobert, 2001; Meyn and Tweedie, 1993; Roberts and Rosenthal, 2004). It can be used to express the transition density $k$ as a mixture of two other transition densities, one of which does not depend on the current state. Indeed, define

$$r(\tilde{x} \mid x) = \frac{k(\tilde{x} \mid x) - s(x)\, \nu(\tilde{x})}{1 - s(x)} \,,$$

and note that, for each fixed $x \in \mathsf{X}$, $r(\tilde{x} \mid x)$ is a density in $\tilde{x}$. We can now write

$$k(\tilde{x} \mid x) = s(x)\, \nu(\tilde{x}) + \big(1 - s(x)\big) r(\tilde{x} \mid x) \,. \tag{17}$$

Thinking of $s(x)$ and $1 - s(x)$ as two fixed numbers in $(0, 1)$ whose sum is unity, the right-hand side of (17) is a mixture of two transition densities, $\nu(\tilde{x})$ and $r(\tilde{x} \mid x)$. This mixture provides an alternative method of simulating the Markov chain. Given the current state $X_n = x$, we can draw $X_{n+1}$ by flipping a coin, $\delta_n \sim \text{Bernoulli}(s(x))$, and then drawing $X_{n+1} \sim \nu(\cdot)$ if $\delta_n = 1$ or $X_{n+1} \sim r(\cdot \mid x)$ if $\delta_n = 0$. Since $\nu$ does not depend on the current state, this method of simulation results in a regeneration every time $\delta_n = 1$. To make things clearer, suppose that we start the chain with $X_0 \sim \nu$ and then proceed to simulate the chain using the sequential method just described. Every time that $\delta_n = 1$, $X_{n+1}$ is drawn from $\nu$ and the process probabilistically restarts itself; i.e., we have a regeneration. 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"

$$\Big\{ \big(X_{\tau_{(t-1)}}, \cdots, X_{\tau_t - 1}\big),\ t = 1, 2, \ldots \Big\}$$

that are independent stochastic replicas of each other. Thus, asymptotic analysis of an estimator based on these tours can proceed using standard techniques from iid theory (such as the SLLN and the CLT).

Now suppose the chain $X$ is Harris ergodic with invariant probability density $\pi$. 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 $\tau_R$, which is random. For $t = 1, \ldots, R$, define $N_t = \tau_t - \tau_{t-1}$, which is the length of the $t$th tour, and $S_t = \sum_{n=\tau_{t-1}}^{\tau_t - 1} g(X_n)$. Since each is based on a separate tour, 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 "$\mathrm{E}_\nu$" is meant to remind the reader that each tour is started with a draw from $\nu$.) The entire motivation for using regeneration is that there is a simple, consistent estimator of $\gamma^2$. Indeed, Hobert et al. (2002) show that

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

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

This regenerative simulation method allows us to use the Markov chain $X$ to approximate $\mathrm{E}_\pi g$ and have the same level of confidence in our answers that we would have had if classical (iid) Monte Carlo had been used. Indeed, assume that $X$ is geometrically ergodic and that there exists an $\alpha > 0$ such that $\mathrm{E}_\pi |g|^{2+\alpha} < \infty$. To implement the method, we choose a preliminary value of $R$ that we believe will lead to a reasonable estimate of $\gamma^2$. We then draw $X_0 \sim \nu$ and use (17) to simulate $R$ tours; that is, we simulate iterations of the chain using (17) until the $R$th 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.

It is often the case that drawing from $r(\cdot|\cdot)$ is problematic. Fortunately, Mykland et al. (1995) noticed a clever way to circumvent this difficulty. Instead of simulating $\delta_n | X_n$ and then $X_{n+1} | \delta_n, X_n$, which results in a draw from $(\delta_n, X_{n+1}) | X_n$, we simply draw $X_{n+1} | X_n$, in the usual way, and then draw $\delta_n | X_n, X_{n+1}$. This alternative sampling scheme provides a draw from the same joint density, but avoids having to draw from $r$. Moreover, given $(X_n, X_{n+1})$, $\delta_n$ is a Bernoulli variable with success probability given by

$$\Pr\left(\delta_n = 1 \mid X_n = x, X_{n+1} = \tilde{x}\right) = \frac{s(x)\, \nu(\tilde{x})}{k(\tilde{x} \mid x)} \; . \tag{18}$$

Hence, in practice, we just simulate the chain as usual, but after each iteration we simulate an extra Bernoulli. In the next subsection, we develop a minorization condition for our block Gibbs sampler.

### 3.2 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 $\sigma^2$ ranges over the set $D$, and $c$ is the normalizing constant; that is,

$$c = \int_{\mathbb{R}^2_+} \int_{\mathbb{R}^{q+1}} \pi(\xi \mid \sigma^2)\pi(\sigma^2 \mid \xi^*) I_D(\sigma^2)\, d\xi\, d\sigma^2 = \left[\int_{d_1}^{d_2} \pi(\sigma^2_\theta \mid \xi^*)\, d\sigma^2_\theta\right] \left[\int_{d_3}^{d_4} \pi(\sigma^2_e \mid \xi^*)\, d\sigma^2_e\right] .$$

The value of $c$ is actually not required in practice. Indeed, we can simulate from $\nu$ 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 (18), which must be calculated after each iteration of the Markov chain, involves $s$ and $\nu$ 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 $\pi(\sigma^2_\theta \mid \xi)$ and $\pi(\sigma^2_e \mid \xi)$, the bivariate minimization problem becomes two separate univariate minimization problems. Let $w_k^*$ stand for $w_k$ evaluated at $\xi^*$ for $k = 1, 2$. Then

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

and

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

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

$$\begin{aligned}
&\Pr\Big(\delta_n = 1 \;\Big|\; (\sigma_n^2, \xi_n) = (\sigma^2, \xi), (\sigma_{n+1}^2, \xi_{n+1}) = (\tilde{\sigma}^2, \tilde{\xi})\Big) \\
&= \frac{s(\xi)\, \nu(\tilde{\sigma}^2, \tilde{\xi})}{k\big(\tilde{\sigma}^2, \tilde{\xi} \mid \sigma^2, \xi\big)} = \frac{\pi(\underline{\sigma}^2 \mid \xi)\pi(\tilde{\sigma}^2 \mid \xi^*)}{\pi(\underline{\sigma}^2 \mid \xi^*)\pi(\tilde{\sigma}^2 \mid \xi)} I_D(\tilde{\sigma}^2) \\
&= \left(\frac{w_1}{w_1^*}\right)^{\frac{q}{2}+a} \exp\left[-\frac{1}{2}(w_1 - w_1^*)/\underline{\sigma}_\theta^2\right] \left(\frac{w_2 + \text{SSE}}{w_2^* + \text{SSE}}\right)^{\frac{M}{2}+b} \exp\left[-\frac{1}{2}(w_2 - w_2^*)/\underline{\sigma}_e^2\right] \\
&\times \left(\frac{w_1^*}{w_1}\right)^{\frac{q}{2}+a} \exp\left[-\frac{1}{2}(w_1^* - w_1)/\tilde{\sigma}_\theta^2\right] \left(\frac{w_2^* + \text{SSE}}{w_2 + \text{SSE}}\right)^{\frac{M}{2}+b} \exp\left[-\frac{1}{2}(w_2^* - w_2)/\tilde{\sigma}_e^2\right] I_D(\tilde{\sigma}^2) \\
&= \exp\left\{\frac{1}{2}\left[(w_1 - w_1^*)\left(\frac{1}{\tilde{\sigma}_\theta^2} - \frac{1}{\underline{\sigma}_\theta^2}\right) + (w_2 - w_2^*)\left(\frac{1}{\tilde{\sigma}_e^2} - \frac{1}{\underline{\sigma}_e^2}\right)\right]\right\} I_D(\tilde{\sigma}^2)\,.
\end{aligned} \tag{19}$$

Theoretically, we could use any set $D = [d_1, d_2] \times [d_3, d_4]$ and any distinguished point $\xi^*$ 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 $\xi^*$ so that the probability in (19) 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 (19) (when the indicator is unity). Our strategy for choosing $D$ and $\xi^*$ is as follows. We run the block Gibbs sampler for an initial $n_0$ iterations (using starting value $\xi_0 = (\overline{y}_1, \dots, \overline{y}_q, \overline{y})$ for example). We take $[d_1, d_2]$ to be the shortest interval that contains 60% of the $n_0$ values of $\sigma_\theta^2$, and we calculate $[d_3, d_4]$ similarly using the $n_0$ values of $\sigma_e^2$. The regeneration probability (19) involves $\xi^*$ only through $w_1(\xi^*)$ and $w_2(\xi^*)$. Hence, instead of setting $\xi^*$ equal to the median, say, of the $n_0$ values of $\xi$ in the initial run of the chain, we calculate $w_1$ and $w_2$ for each of the $n_0$ values of $\xi$ 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 $\sqrt{\frac{w_2^*}{m}} + \sqrt{w_1^*} \ge \sqrt{\frac{\text{SST}}{m}}$. See Tan (2008) for a proof of this result as well as guidelines for the unbalanced cases.

# 4 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). Thirteen workers 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.

| 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

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 $(\sigma^2, \xi)$ 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, $\sigma_\theta^2$ and $\sigma_e^2$, 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 of regenerations, or say, the number of iid tours in the chain. As we have mentioned in Section 3.1, the procedure to determine $R$ has two steps. In the first step, we run the chain for a preliminary $R$ regenerations that is believed to lead to a reasonable estimator of the asymptotic variance, $\gamma^2$. Here, we used $R = 5,000$ which took 87,169 iterations and consumed 20 seconds (coded in R). 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 $\gamma^2$ 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, $\sigma_\theta^2$, $\sigma_e^2$ 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.

| | $\tilde{g}_{\tau_R}$ | $\hat{\gamma}^2$ | $\sqrt{\hat{\gamma}^2/R}$ | $\overline{g}_{\tau_R} \pm 2\sqrt{\hat{\gamma}^2/R}$ |
|---|---|---|---|---|
| $\sigma_\theta^2$ | 0.19003 | 0.03463 | 0.00263 | (0.18477, 0.19529) |
| $\sigma_e^2$ | 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

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 magnitude of $\text{E}_\pi\sigma_\theta^2$. 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. To take into account the possibility that the asymptotic variance can be slightly underestimated by the $\hat{\gamma}^2$ obtained in the preliminary analysis, we run the chain a little longer than what the calculation suggests. Here, we decide to produce a chain with $R = 40,000$ regenerations. Actually, only 35,000 more iid tours need to be simulated, and then combined with the 5,000 iid tours we already have. 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. The trace plots of the estimated asymptotic 95% CIs for $E_\pi\sigma_\theta^2$ and $\hat{\gamma}^2$ are provided in Figure 1. These plots suggest that things have stabilized quite nicely by the 40,000th regeneration.

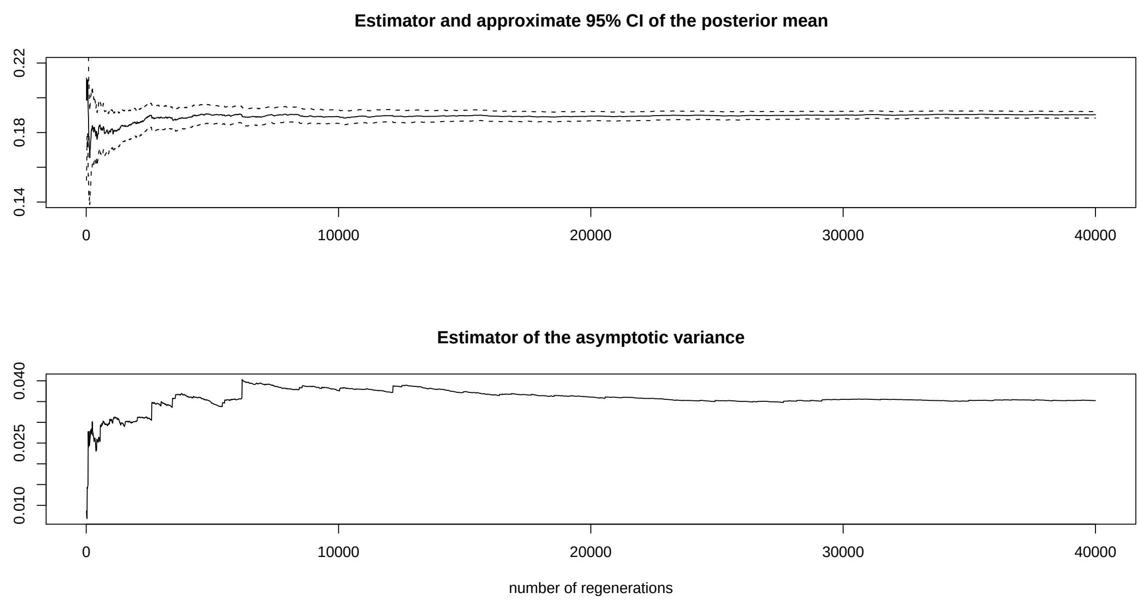


Figure 1: The solid line in the upper graph is the trace plot of the estimator of $\mathrm{E}_\pi \sigma_\theta^2$. That is, for $R = 1, \dots, 40000$, we plotted $\widetilde{g}_R$ against $R$, where $\widetilde{g}_R$ is the estimator based on the first $R$ tours out of the $40,000$ tours of our simulated 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, $\gamma^2$.

| | $\tilde{g}_{\tau_R}$ | $\hat{\gamma}^2$ | $\sqrt{\hat{\gamma}^2/R}$ | $\overline{g}_{\tau_R} \pm 2\sqrt{\hat{\gamma}^2/R}$ |
|---|---|---|---|---|
| $\sigma_\theta^2$ | 0.19023 | 0.03523 | 0.00094 | (0.18835, 0.19210) |
| $\sigma_e^2$ | 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 $\mathrm{E}_\pi \sigma_\theta^2$.

## 5 Discussion

Our block Gibbs sampler is a two-variable Gibbs sampler that updates $\sigma^2 = (\sigma_\theta^2, \sigma_e^2)$ and $\xi = (\theta, \mu)$ in turn. There is another equally valid way to block the variables that results in a different two-variable Gibbs sampler, which is also easy to implement. Indeed, van Dyk and Meng (2001) use a blocking scheme in which the two groups of parameters are $\theta$ and $(\mu, \sigma_\theta^2, \sigma_e^2)$. Given $(\mu, \sigma_\theta^2, \sigma_e^2)$, $\theta$ has a multivariate normal distribution (and our results in Subsection 2.1 can be used to write down its exact form). On the other hand, the posterior density of $(\mu, \sigma_\theta^2, \sigma_e^2)$ given $\theta$ factors as follows:

$$\pi(\mu, \sigma_\theta^2, \sigma_e^2 \mid \theta) = \pi(\sigma_e^2 \mid \theta)\, \pi(\mu, \sigma_\theta^2 \mid \theta) \ ;$$

that is, given $\theta$, $\sigma_e^2$ and $(\mu, \sigma_\theta^2)$ 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 this alternative chain because the drift condition we used is not appropriate for the alternative blocking scheme. We strongly suspect that this alternative Markov chain is geometrically ergodic, but this remains an open question.

# APPENDICES

## A Background on General State Space Markov Chains

Let $\mathsf{X}$ be a set equipped with a countably generated $\sigma$-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 $\pi$ 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 $\psi$-irreducible, aperiodic and Harris recurrent. See Meyn and Tweedie (1993) for definitions.

Suppose that $\mu$ is a non-trivial, $\sigma$-finite measure on $\mathsf{X}$ and that the function $k : \mathsf{X} \times \mathsf{X} \to [0, \infty)$ satisfies

$$K(x, A) = \int_A k(y \mid x)\, \mu(dy) \quad \text{for any } x \in \mathsf{X} \text{ and any } \mu\text{-measurable } A\ .$$

Then $k$ is called the *Markov transition density* of $X$ with respect to $\mu$. The following result, established in Tan (2008), 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 $\mu$) and invariant probability measure $\pi$. If $k(y \mid x) > 0$ for all $x, y \in \mathsf{X}$, then $X$ is Harris ergodic. Furthermore, $\mu$ is equivalent to the maximal irreducibility measure.*

If $X$ is Harris ergodic, then, no matter what the distribution of $X_0$,

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

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 $\mathsf{X}$ is topological and that $\sigma(\mathsf{X})$ is the Borel $\sigma$-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 $\{x \in \mathsf{X} : w(x) \leq \gamma\}$ is compact for every $\gamma > 0$. The following result is a combination of Meyn and Tweedie's (1993) Lemma 15.2.8 and Theorem 6.0.1.

**Proposition 3.** *Let $X$ be a Markov chain on a topological space $\mathsf{X}$. 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*

$$E[w(X_1) \mid X_0 = x] \leq \rho w(x) + L \ , \tag{20}$$

*then $X$ is geometrically ergodic.*

The inequality (20) is called a *drift condition* and the function $w$ is called a *drift function*.

# B Technical Details

## B.1 Upper bounds for conditional expectations

We begin by establishing some inequalities involving $t$, which will help us evaluate conditional expectations. First, note that

$$t = \sum_{i=1}^{q} \frac{m_i}{m_i\sigma_\theta^2 + \sigma_e^2} \geq \sum_{i=1}^{q} \frac{m_i}{(1+m_i)\max\{\sigma_e^2, \sigma_\theta^2\}} = \left(\sum_{i=1}^{q} \frac{m_i}{m_i+1}\right) \frac{1}{\max\left\{\sigma_\theta^2, \sigma_e^2\right\}} \ . \tag{21}$$

**Lemma 2.** *Let $m^* = \max\{m_1, \ldots, m_q\}$. Then for each $i = 1, ..., q$, we have*

$$\frac{m^*\sigma_\theta^2 + \sigma_e^2}{M} \geq \frac{1}{t} \geq \frac{m_i\sigma_\theta^2 + \sigma_e^2}{M(1+m_i)} \ .$$

*Proof.* The first inequality holds because

$$\frac{1}{t} = \left(\sum_{i=1}^{q} \frac{m_i}{m_i\sigma_\theta^2 + \sigma_e^2}\right)^{-1} \leq \left(\sum_{i=1}^{q} \frac{m_i}{m^*\sigma_\theta^2 + \sigma_e^2}\right)^{-1} = \frac{m^*\sigma_\theta^2 + \sigma_e^2}{M} \ .$$

For the second inequality, first note that

$$t = \sum_{i=1}^{q} \frac{m_i}{m_i\sigma_\theta^2 + \sigma_e^2} \leq \sum_{i=1}^{q} \frac{m_i}{m_i\sigma_\theta^2} = \frac{q}{\sigma_\theta^2} \quad \text{and} \quad t = \sum_{i=1}^{q} \frac{m_i}{m_i\sigma_\theta^2 + \sigma_e^2} \leq \sum_{i=1}^{q} \frac{m_i}{\sigma_e^2} = \frac{M}{\sigma_e^2} \ .$$

If $\sigma_\theta^2 \leq \sigma_e^2$, then

$$\frac{1}{m_i\sigma_\theta^2 + \sigma_e^2} \frac{1}{t} \geq \frac{\sigma_e^2}{M(m_i\sigma_\theta^2 + \sigma_e^2)} \geq \frac{1}{M(1+m_i)} \ ,$$

else if $\sigma_\theta^2 > \sigma_e^2$, then

$$\frac{1}{m_i\sigma_\theta^2 + \sigma_e^2} \frac{1}{t} \geq \frac{\sigma_\theta^2}{q(m_i\sigma_\theta^2 + \sigma_e^2)} > \frac{1}{q(1+m_i)} \geq \frac{M}{\sigma_e^2} \ .$$

$\square$

Note that $\mathrm{E}\big(\mu\big|\sigma_\theta^2,\sigma_e^2\big)$ is a convex combination of the $\overline{y}_i$. Hence, as a function of $\sigma_\theta^2$ and $\sigma_e^2$, this conditional expectation is uniformly bounded by a constant. Along the same lines, for each fixed $k$, $\mathrm{E}\big(\theta_k\big|\sigma_\theta^2,\sigma_e^2\big)$ is a convex combination of $\mathrm{E}\big(\mu\big|\sigma_\theta^2,\sigma_e^2\big)$ and $\overline{y}_k$, so it too is uniformly bounded by a constant. Using these facts along with the forms of the conditional densities given in Subsection 2.1, we have

$$\begin{aligned}
\mathrm{E}(\tilde{w}_1|\sigma^2) &= \mathrm{E}\bigg[\sum_i(\tilde{\theta}_i-\tilde{\mu})^2\Big|\sigma_\theta^2,\sigma_e^2\bigg] \\
&= \sum_i\Big[\mathrm{Var}\big[(\tilde{\theta}_i-\tilde{\mu})\big|\sigma_\theta^2,\sigma_e^2\big]+\Big(\mathrm{E}\big[(\tilde{\theta}_i-\tilde{\mu})\big|\sigma_\theta^2,\sigma_e^2\big]\Big)^2\Big] \\
&= \sum_i\frac{\sigma_\theta^2\sigma_e^2}{\sigma_e^2+m_i\sigma_\theta^2}+\sum_i\frac{\big(\sigma_e^2\big)^2}{(m_i\sigma_\theta^2+\sigma_e^2)^2t}-2\sum_i\frac{\sigma_e^2}{(m_i\sigma_\theta^2+\sigma_e^2)t}+\frac{q}{t}+const \\
&\le \sum_i\frac{\sigma_\theta^2\sigma_e^2}{m_i\sigma_\theta^2+\sigma_e^2}-\sum_i\frac{\sigma_e^2}{(m_i\sigma_\theta^2+\sigma_e^2)t}+\frac{q}{t}+const \\
&\le \sum_i\frac{\sigma_e^2}{m_i}-\sum_i\frac{\sigma_e^2}{M(1+m_i)}+\frac{q}{t}+const
\end{aligned}$$

where the final inequality uses Lemma 2. We now bound $\frac{q}{t}$ in two different ways. First, applying (21), we have

$$\frac{q}{t}\le q\bigg(\sum_{i=1}^q\frac{m_i}{m_i+1}\bigg)^{-1}\max\left\{\sigma_\theta^2,\sigma_e^2\right\}\le q\bigg(\sum_{i=1}^q\frac{m_i}{m_i+1}\bigg)^{-1}\big(\sigma_\theta^2+\sigma_e^2\big)\;.$$

For the other way, apply Lemma 2 and

$$\frac{q}{t}\le\frac{qm^*\sigma_\theta^2}{M}+\frac{q\sigma_e^2}{M}\;.$$

Hence,

$$\begin{aligned}
\mathrm{E}(\tilde{w}_1|\sigma^2)\le q&\bigg(\sum_{i=1}^q\frac{m_i}{m_i+1}\bigg)^{-1}\sigma_\theta^2 \\
&+\Bigg[\sum_i\frac{1}{m_i}-\sum_i\frac{1}{M(1+m_i)}+q\bigg(\sum_{i=1}^q\frac{m_i}{m_i+1}\bigg)^{-1}\Bigg]\sigma_e^2+const\;,
\end{aligned}$$

and

$$\mathrm{E}(\tilde{w}_1|\sigma^2)\le\frac{qm^*}{M}\sigma_\theta^2+\bigg[\sum_i\frac{1}{m_i}-\sum_i\frac{1}{M(1+m_i)}+\frac{q}{M}\bigg]\sigma_e^2+const\;.$$

Now,

$$
\begin{aligned}
\mathrm{E}(\tilde{w}_2|\sigma^2) &= \sum_i m_i \mathrm{E}\big[(\overline{y}_i - \tilde{\theta}_i)^2 \big| \sigma_\theta^2, \sigma_e^2\big] \\
&= \sum_i m_i \Big[\mathrm{Var}(\tilde{\theta}_i|\sigma_\theta^2, \sigma_e^2) + \Big(\mathrm{E}\big[(\tilde{\theta}_i - \overline{y}_i)\big|\sigma_\theta^2, \sigma_e^2\big]\Big)^2\Big] \\
&= \sum_i \frac{m_i \sigma_\theta^2 \sigma_e^2}{m_i \sigma_\theta^2 + \sigma_e^2} + \sum_i \frac{m_i \big(\sigma_e^2\big)^2}{(m_i\sigma_\theta^2 + \sigma_e^2)^2 t} + const \\
&\le \sum_i \sigma_e^2 + \sum_i \frac{m_i \sigma_e^2}{(m_i \sigma_\theta^2 + \sigma_e^2) t} \sigma_e^2 = (q+1)\sigma_e^2 + const\ .
\end{aligned}
$$

## B.2 Finding an $s \in S$ such that $\delta_1(s) < 1$ and $\delta_3(s) < 1$

This section contains a proof of the following result.

**Proposition 4.** *If $M + 2b \ge q + 3$ and $\Delta_1 < 2\exp\big(\Psi\big(\frac{q}{2} + a\big)\big)$, then there exists $s \in S$ such that $\delta_1(s) < 1$ and $\delta_3(s) < 1$.*

We will prove Proposition 4 by establishing that

- $\delta_1(s) < 1$ for any $s \in (0,1)$ if $M + 2b \ge q + 3$, and
- $\delta_3(s_0) < 1$ for some small positive $s_0$ if $\Delta_1 < 2\exp\big(\Psi\big(\frac{q}{2} + a\big)\big)$.

It is well known that $\Psi'(x) > 0$ for all $x > 0$ and that $\Psi(x+1) = \Psi(x) + \frac{1}{x}$. A couple of common values of the digamma function that we will encounter later are $\Psi(1) = -\gamma$ and $\Psi(\frac{1}{2}) = -\gamma - 2\log(2)$, where $\gamma := \lim_{p\to\infty}\big(1 + \frac{1}{2} + \cdots + \frac{1}{p} - \log(p)\big) \doteq 0.577$ is Euler's constant. Also, the recurrence formula yields: $\Psi(\frac{3}{2}) = -\gamma - 2\log(2) + 2$.

### B.2.1 Bounding $\delta_1(s)$

Recall that $\delta_1(s)$ is actually a function of $s$, $q$ and $\frac{M}{2} + b$. Indeed,

$$
\delta_1\bigg(s, q, \frac{M}{2} + b\bigg) = \frac{\Gamma\big(\frac{M}{2} + b - s\big)}{\Gamma\big(\frac{M}{2} + b\big)} \Big(\frac{q+1}{2}\Big)^s \ .
$$

Now, for any fixed $s > 0$, $\Gamma(x-s)/\Gamma(x)$ is decreasing in $x$ for $x > s$, because

$$
\frac{d}{dx}\Big[\log(\Gamma(x-s)) - \log(\Gamma(x))\Big] = \Psi(x-s) - \Psi(x) < 0 \quad \text{for all } x > s > 0\ .
$$

Therefore, with $(s,q)$ fixed, $\delta_1\big(s, q, \frac{M}{2} + b\big)$ is decreasing in $\big(\frac{M}{2} + b\big)$ as long as $\frac{M}{2} + b > s$. Consequently, to show that $\delta_1\big(s, q, \frac{M}{2} + b\big) < 1$ for all $s \in (0,1)$ if $M + 2b \ge q + 3$, we need only prove the following.

**Lemma 3.** $\delta_1\big(s, q, \frac{q+3}{2}\big) < 1$ *for all* $s \in (0,1)$ *and* $q \geq 2$.

*Proof.* Fix $s \in (0,1)$ and define

$$T(x) = \frac{\Gamma(x+s)}{\Gamma(x)}(x+s-1)^{-s} \quad \text{for } x > 1-s\ .$$

We claim that

1. $T(x)$ is strictly decreasing in $x$, and
2. $\lim_{x\to\infty} T(x) = 1$.

To prove claim 1, we will show that $Q(x) = \log(T(x))$ is decreasing in $x$. First, for $x > 0$

$$\Psi(x) = -\gamma + \sum_{p=1}^{\infty}\left(\frac{1}{p} - \frac{1}{x+p-1}\right)$$

(Abramowitz and Stegun, 1964, p.259). Note that $\big(\frac{1}{p} - \frac{1}{x+p-1}\big)$ is nonnegative for all $p$ when $x \geq 1$ and negative for all $p$ when $x < 1$. Hence, the above series is absolutely convergent for all $x > 0$. Clearly, $Q(x) = -s\log(x+s-1) + \log(\Gamma(x+s)) - \log(\Gamma(x))$ and its derivative can be expressed as follows

$$Q'(x) = -s\frac{1}{x+s-1} + \sum_{p=1}^{\infty}\left(\frac{1}{p} - \frac{1}{x+s+p-1}\right) - \sum_{p=1}^{\infty}\left(\frac{1}{p} - \frac{1}{x+p-1}\right). \tag{22}$$

The fraction in the first term of (22) can be written as the following absolutely convergent telescoping series

$$\frac{1}{x+s-1} = \sum_{p=1}^{\infty}\left(\frac{1}{x+s+p-2} - \frac{1}{x+s+p-1}\right).$$

Therefore,

$$\begin{aligned} Q'(x) &= s\sum_{p=1}^{\infty}\left(\frac{1}{x+s+p-1} - \frac{1}{x+s+p-2}\right) + \sum_{p=1}^{\infty}\left(\frac{1}{x+p-1} - \frac{1}{x+s+p-1}\right) \\ &= \sum_{p=1}^{\infty}\left[-(1-s)\frac{1}{x+s+p-1} - s\frac{1}{x+s+p-2} + \frac{1}{x+p-1}\right]. \end{aligned}$$

The convexity of the function $h(z) = \frac{1}{z}$ on $\mathbb{R}_+$ combined with the fact that $(1-s)(x+s+p-1) + s(x+s+p-2) = x+p-1$ can be used to show that every term in the series above is negative. It follows that $Q(x)$ and $T(x)$ are both decreasing in $x$ for $x > 1-s$.

We now prove claim 2. Fix $s \in (0,1)$ and define $S(x) = x^{-s}\Gamma(x+s)/\Gamma(x)$. As $x \to \infty$, $S(x) \to 1$ (Abramowitz and Stegun, 1964, p.257). As a consequence,

$$\lim_{x\to\infty} T(x) = \lim_{x\to\infty} S(x)\left(\frac{x}{x+s-1}\right)^{s} = 1\ .$$

Finally, for fixed $s \in (0,1)$, note that $\frac{q+3}{2} - s > 1 - s$ and

$$\delta_1\big(s, q, \frac{q+3}{2}\big) = \big(\frac{q+1}{2}\big)^s \frac{\Gamma\big(\frac{q+3}{2} - s\big)}{\Gamma\big(\frac{q+3}{2}\big)} = \Big(T\big(\frac{q+3}{2} - s\big)\Big)^{-1} .$$

It follows from claims 1 and 2 that $T\big(\frac{q+3}{2} - s\big) > 1$ and hence $\delta_1\big(s, q, \frac{q+3}{2}\big) < 1$. □

### B.2.2 Bounding $\delta_3(s)$

Recall that $\delta_3(s)$ is actually a function of $s$, $m$ and $a$. If we define

$$A(s,q,a) = \frac{\Gamma(\frac{q}{2} + a - s)}{2^s \Gamma(\frac{q}{2} + a)} ,$$

then we have $\delta_3(s,m,a) = A(s,q,a)\Delta_1^s(m)$. Note that there exists an $s_0 \in S$ such that $\delta_3(s_0,m,a) < 1$ if and only if $\Delta_1(m) < A^*(q,a)$, where

$$A^*(q,a) := \sup_{s \in S} A^{-\frac{1}{s}}(s,q,a) = 2 \sup_{s \in S} \left( \frac{\Gamma(\frac{q}{2} + a)}{\Gamma(\frac{q}{2} + a - s)} \right)^{\frac{1}{s}} .$$

We now establish a lower bound for $A^*(q,a)$. Define

$$g(s,q,a) = \log\left(\frac{1}{2} A^{-\frac{1}{s}}(s,q,a)\right) = \frac{1}{s}\left[\log\left(\Gamma\Big(\frac{q}{2} + a\Big)\right) - \log\left(\Gamma\Big(\frac{q}{2} + a - s\Big)\right)\right] .$$

Then

$$\lim_{s \to 0} g(s,q,a) = \frac{d \log(\Gamma(x))}{dx}\bigg|_{x = \frac{q}{2} + a} = \Psi\Big(\frac{q}{2} + a\Big) .$$

Hence,

$$A^*(q,a) \geq \lim_{s \to 0} 2 \exp\big[g(s,q,a)\big] = 2 \exp\left(\Psi\Big(\frac{q}{2} + a\Big)\right) . \tag{23}$$

We conclude that there exists $s \in S$ such that $\delta_3(s,m,a) < 1$ if $\Delta_1(m) < 2\exp\left(\Psi\Big(\frac{q}{2} + a\Big)\right)$.

*Remark* 1. It is easy to show that $\frac{\partial g(s,q,a)}{\partial s}\big|_{s=0} < 0$. In other words, for fixed $q$ and $a$, $g(s,q,a)$ is decreasing in $s$ in a neighborhood of $s = 0$. Furthermore, numerical calculations suggest that $g(s,q,a)$ is decreasing on the entire set $S$ for any fixed $q$ and $a$. Hence, we believe the lower bound on $A^*(q,a)$ in (23) is sharp.

## B.3 Specializing Proposition 2 to the case where $a = -\frac{1}{2}$

Here we study the condition $\Delta_1(m) < 2\exp\left(\Psi\Big(\frac{q}{2} + a\Big)\right)$ in the special case where $a = -\frac{1}{2}$. It follows from (3) that, when $a = -\frac{1}{2}$, the posterior is improper if $q \leq 2$. Hence, we can restrict attention to the case $q \geq 3$. When $q \geq 4$, we have

$$2\exp\left(\Psi\Big(\frac{q}{2} + a\Big)\right) \geq 2\exp\left(\Psi\Big(\frac{3}{2}\Big)\right) = 2\exp(-\gamma - 2(\log 2 - 1)) \doteq 2.074 .$$

Now recall that

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

Since $\frac{m_i}{1+m_i} \geq \frac{1}{2}$, we have

$$\Delta_1(m) \leq \frac{q}{\sum_i \frac{m_i}{m_i+1}} \leq \frac{q}{\sum_i \frac{1}{2}} = 2 < 2\exp\left( \Psi\left( \frac{q}{2} + a \right) \right) ,$$

so, when $a = -\frac{1}{2}$ and $q \geq 4$, the condition $\Delta_1(m) < 2\exp\left( \Psi\left( \frac{q}{2} + a \right) \right)$ is *always satisfied*.

Now, when $q = 3$, we have

$$2\exp\left( \Psi\left( \frac{q}{2} + a \right) \right) = 2\exp\left( \Psi(1) \right) = 2\exp(-\gamma) \doteq 1.123 .$$

For balanced data,

$$\Delta_1(m) \leq \frac{qm^*}{M} = 1 < 2\exp\left( \Psi(1) \right) .$$

Hence, when $a = -\frac{1}{2}$ and $q = 3$ and the data are balanced, the condition $\Delta_1(m) < 2\exp\left( \Psi\left( \frac{q}{2} + a \right) \right)$ is satisfied. Finally, if $q = 3$ and the data are unbalanced, then $\Delta_1(m) < 2\exp(-\gamma)$ if and only if

$$\sum_i \frac{m_i}{m_i+1} > \frac{3}{2\exp(-\gamma)} \doteq 2.67 \quad \text{or} \quad m^* < \frac{2\exp(-\gamma)}{3} M \doteq 0.374M . \tag{24}$$

Table 5 displays all unbalanced data configurations with $q = 3$ and $m^* \leq 12$ that satisfy (24).

# References


ABRAMOWITZ, M. and STEGUN, I. A. (1964). *Handbook of Mathematical Tables*. Dover.

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. *Statistical Sciences,* to appear.

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.

| $m_1$ | $m_2$ | $m_3$ | $m_1$ | $m_2$ | $m_3$ | $m_1$ | $m_2$ | $m_3$ | $m_1$ | $m_2$ | $m_3$ | $m_1$ | $m_2$ | $m_3$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 | 4 | 4 | 7 | 8 | 10 | 7 | 10 | 11 | 6 | 9 | 12 | 8 | 11 | 12 |
| 4 | 5 | 5 | 7 | 9 | 10 | 7 | 11 | 11 | 6 | 10 | 12 | 8 | 12 | 12 |
| 5 | 6 | 6 | 7 | 10 | 10 | 8 | 8 | 11 | 6 | 11 | 12 | 9 | 9 | 12 |
| 5 | 7 | 7 | 8 | 8 | 10 | 8 | 9 | 11 | 6 | 12 | 12 | 9 | 10 | 12 |
| 6 | 6 | 7 | 8 | 9 | 10 | 8 | 10 | 11 | 7 | 7 | 12 | 9 | 11 | 12 |
| 6 | 7 | 7 | 8 | 10 | 10 | 8 | 11 | 11 | 7 | 8 | 12 | 9 | 12 | 12 |
| 6 | 8 | 8 | 9 | 9 | 10 | 9 | 9 | 11 | 7 | 9 | 12 | 10 | 10 | 12 |
| 7 | 7 | 8 | 9 | 10 | 10 | 9 | 10 | 11 | 7 | 10 | 12 | 10 | 11 | 12 |
| 7 | 8 | 8 | 6 | 9 | 11 | 9 | 11 | 11 | 7 | 11 | 12 | 10 | 12 | 12 |
| 7 | 9 | 9 | 6 | 10 | 11 | 10 | 10 | 11 | 7 | 12 | 12 | 11 | 11 | 12 |
| 8 | 8 | 9 | 6 | 11 | 11 | 10 | 11 | 11 | 8 | 8 | 12 | 11 | 12 | 12 |
| 8 | 9 | 9 | 7 | 8 | 11 | 5 | 11 | 12 | 8 | 9 | 12 | | | |
| 6 | 10 | 10 | 7 | 9 | 11 | 5 | 12 | 12 | 8 | 10 | 12 | | | |

Table 5: A complete list of all unbalanced configurations $(m_1, m_2, m_3)$ with $m^* \leq 12$ that satisfy $\Delta_1(m) < 2\exp(-\gamma)$.


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

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.