Geometric Ergodicity of Random Scan Gibbs Samplers for Hierarchical One-Way Random Effects Models

Alicia A. Johnson
Department of Mathematics, Statistics, and Computer Science
Macalester College
ajohns24@macalester.edu

Galin L. Jones [*]
School of Statistics
University of Minnesota
galin@umn.edu

February 25, 2015

Abstract

We consider two Bayesian hierarchical one-way random effects models and establish geometric ergodicity of the corresponding random scan Gibbs samplers. Geometric ergodicity, along with a moment condition, guarantees a central limit theorem for sample means and quantiles. In addition, it ensures the consistency of various methods for estimating the variance in the asymptotic normal distribution. Thus our results make available the tools for practitioners to be as confident in inferences based on the observations from the random scan Gibbs sampler as they would be with inferences based on random samples from the posterior.

[*]Research supported by the the National Science Foundation and the National Institutes for Health.

1 Introduction

Suppose that for $i = 1, \ldots, K$

$$
\begin{aligned}
Y_i|\theta_i, \gamma_i &\stackrel{ind}{\sim} N(\theta_i, \gamma_i^{-1}) \\
\theta_i|\mu, \lambda_\theta, \lambda_i &\stackrel{ind}{\sim} \mathrm{N}(\mu, \lambda_\theta^{-1}\lambda_i^{-1}) \\
(\mu, \lambda_\theta, \gamma_1, \ldots, \gamma_K, \lambda_1, \ldots, \lambda_K) &\sim p(\mu, \lambda_\theta, \gamma_1, \ldots, \gamma_K, \lambda_1, \ldots, \lambda_K)
\end{aligned}
\tag{1}
$$

where p is a generic proper prior. Eventually we will consider two distinct choices for the prior p, but we leave the description until section 2. In both cases, the hierarchy in (1) yields a proper posterior which is intractable in the sense that the posterior quantities, such as expectations or quantiles, required for Bayesian inference are not available in closed form. Thus we will consider the use of Markov chain Monte Carlo (MCMC) methods.

Let y denote all of the data, $\lambda = (\lambda_1, \ldots, \lambda_K)^T$, $\xi = (\theta_1, \ldots, \theta_K, \mu)^T$, and $\gamma = (\gamma_1, \ldots, \gamma_K)^T$. In section 2 we will see that the specific forms of the posterior full conditional densities $f(\xi|\lambda_\theta, \lambda, \gamma, y)$, $f(\lambda_\theta|\xi, \lambda, \gamma, y)$, $f(\lambda|\xi, \lambda_\theta, \gamma, y)$ and $f(\gamma|\xi, \lambda_\theta, \lambda, y)$ are available and hence it is easy to construct Gibbs samplers to help perform posterior inference. Gibbs samplers can be implemented in either a deterministic scan or a random scan, among other variants (Johnson et al., 2013; Liu et al., 1994). Although deterministic scan MCMC algorithms are currently popular in the statistics literature, random scan algorithms were some of the first used in MCMC settings (Geman and Geman, 1984; Metropolis et al., 1953) and remain useful in applications (Lee et al., 2013; Richardson et al., 2010). Random scan Gibbs samplers can also be implemented adaptively while the deterministic scan version cannot. In addition, there has been recent interest in the theoretical properties of random scan algorithms (Diaconis et al., 2008; Johnson et al., 2013; Jones et al., 2014; Levine and Casella, 2006; Roberts and Rosenthal, 2015; Tan et al., 2013).

We will study the random scan Gibbs sampler which is now described. Let $r = (r_1, r_2, r_3, r_4)$ with $r_1 + r_2 + r_3 + r_4 = 1$ and each $r_i > 0$ where we call r the *selection probabilities*. If $(\xi^{(n)}, \lambda_\theta^{(n)}, \lambda^{(n)}, \gamma^{(n)})$ is the current state of the Gibbs sampler, then the $(n+1)$st state is obtained as follows.

Draw $U \sim \text{Uniform}(0, 1)$ and call the realized value u.

If $u \leq r_1$, draw $\xi' \sim f(\xi|\lambda_\theta^{(n)}, \lambda^{(n)}, \gamma^{(n)}, y)$ and set

$$(\xi^{(n+1)}, \lambda_\theta^{(n+1)}, \lambda^{(n+1)}, \gamma^{(n+1)}) = (\xi', \lambda_\theta^{(n)}, \lambda^{(n)}, \gamma^{(n)})$$

else if $r_1 < u \le r_1 + r_2$, draw $\lambda_\theta' \sim f(\lambda_\theta | \xi^{(n)}, \lambda^{(n)}, \gamma^{(n)}, y)$ and set

$$(\xi^{(n+1)}, \lambda_\theta^{(n+1)}, \lambda^{(n+1)}, \gamma^{(n+1)}) = (\xi^{(n)}, \lambda_\theta', \lambda^{(n)}, \gamma^{(n)})$$

else if $r_1 + r_2 < u \le r_1 + r_2 + r_3$, draw $\lambda' \sim f(\lambda | \xi^{(n)}, \lambda_\theta^{(n)}, \gamma^{(n)}, y)$ and set

$$(\xi^{(n+1)}, \lambda_\theta^{(n+1)}, \lambda^{(n+1)}, \gamma^{(n+1)}) = (\xi^{(n)}, \lambda_\theta^{(n)}, \lambda', \gamma^{(n)})$$

else if $r_1 + r_2 + r_3 < u \le 1$, draw $\gamma' \sim f(\gamma | \xi^{(n)}, \lambda_\theta^{(n)}, \lambda^{(n)}, y)$ and set

$$(\xi^{(n+1)}, \lambda_\theta^{(n+1)}, \lambda^{(n+1)}, \gamma^{(n+1)}) = (\xi^{(n)}, \lambda_\theta^{(n)}, \lambda^{(n)}, \gamma') \ .$$

Our goal is to investigate the conditions required for the random scan Gibbs sampler to produce reliable simulation results. Specifically, we will investigate conditions under which the Markov chain is geometrically ergodic, which we now define. Let $\mathsf{X} = \mathbb{R}^K \times \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+^K \times \mathbb{R}_+^K$ and $\mathcal{B}(\mathsf{X})$ denote the Borel sets. Let $P^n : \mathsf{X} \times \mathcal{B}(\mathsf{X}) \to [0,1]$ denote the n-step Markov kernel for the random scan Gibbs sampler so that if $A \in \mathcal{B}(\mathsf{X})$ and $n \ge 1$

$$P^n((\xi^{(1)}, \lambda_\theta^{(1)}, \lambda^{(1)}, \gamma^{(1)}), A) = \Pr((\xi^{(n+1)}, \lambda_\theta^{(n+1)}, \lambda^{(n+1)}, \gamma^{(n+1)}) \in A \,|\, (\xi^{(1)}, \lambda_\theta^{(1)}, \lambda^{(1)}, \gamma^{(1)})) \ .$$

Let F denote the posterior distribution associated with (1) and $\| \cdot \|$ denote the total variation norm. Then the random scan Gibbs sampler is geometrically ergodic if there exists $M : \mathsf{X} \to [0, \infty)$ and $t \in [0,1)$ such that for all $\xi, \lambda_\theta, \lambda, \gamma$ and $n = 1, 2, \ldots$

$$\| P^n((\xi, \lambda_\theta, \lambda, \gamma), \cdot) - F(\cdot) \| \le M(\xi, \lambda_\theta, \lambda, \gamma)\, t^n \ . \tag{2}$$

Geometric ergodicity is a useful stability property for MCMC samplers (Jones and Hobert, 2001; Roberts and Rosenthal, 2004) in that it ensures rapid convergence of the Markov chain since $t < 1$, the existence of a central limit theorem (CLT) (Chan and Geyer, 1994; Hobert et al., 2002; Jones, 2004; Roberts and Rosenthal, 1997), and consistency of various methods to estimate asymptotically valid Monte Carlo standard errors (Flegal et al., 2008; Flegal and Jones, 2010, 2011; Jones et al., 2006). To see the connection between (2) and the CLT let $g : \mathsf{X} \to \mathbb{R}$ and f be the posterior density and suppose we want to calculate

$$\mu_g := \int_{\mathsf{X}} g(\xi, \lambda_\theta, \lambda, \gamma) f(\xi, \lambda_\theta, \lambda, \gamma | y)\, d\xi\, d\lambda_\theta\, d\lambda\, d\gamma \ .$$

Assuming μ_g exists and the Markov chain is irreducible, aperiodic and Harris recurrent (see Meyn and Tweedie (2009) for definitions and section 2 for discussion of these conditions for our two random scan Gibbs samplers), then, as $n \to \infty$,

$$\mu_n := \frac{1}{n} \sum_{i=1}^{n} g(\xi^{(i)}, \lambda_\theta^{(i)}, \lambda^{(i)}, \gamma^{(i)}) \to \mu_g \quad \text{with probability 1.}$$

Of course, μ_n will be much more valuable if we can equip it with an asymptotically valid standard error. If the random scan Gibbs sampler is geometrically ergodic and

$$\int_{\mathsf{X}} g^2(\xi, \lambda_\theta, \lambda, \gamma) f(\xi, \lambda_\theta, \lambda, \gamma | y) \, d\xi \, d\lambda_\theta \, d\lambda \, d\gamma < \infty \, , \tag{3}$$

then there exists $\delta_g^2 < \infty$ such that, as $n \to \infty$, and for any initial distribution

$$\sqrt{n}(\mu_n - \mu_g) \xrightarrow{d} \mathrm{N}(0, \delta_g^2) \, . \tag{4}$$

The quantity δ_g^2 is complicated (Häggström and Rosenthal, 2007), but if the Markov chain is geometrically ergodic there are several methods for estimating it consistently (Flegal and Jones, 2010; Hobert et al., 2002; Jones et al., 2006). This then allows construction of asymptotically valid interval estimates of μ_g to describe the precision of μ_n (Flegal et al., 2008) and hence the reliability of the simulation. A similar approach is available for estimating posterior quantiles which, of course, are often useful for constructing posterior credible intervals; see the recent work of Doss et al. (2014).

There has been a fair amount of work on establishing geometric ergodicity of two-component deterministic scan Gibbs samplers (see, among others, Hobert and Geyer, 1998; Jones and Hobert, 2004; Marchev and Hobert, 2004; Papaspiliopoulos and Roberts, 2008; Roberts and Polson, 1994; Roberts and Rosenthal, 1999; Roy and Hobert, 2007; Tan and Hobert, 2012), while Doss and Hobert (2010) and Jones and Hobert (2004) considered geometric ergodicity of three-component deterministic scan Gibbs samplers. However, there has been almost no corresponding investigation for random scan Gibbs samplers. Liu et al. (1995) did study geometric convergence of random scan Gibbs samplers, but the required regularity conditions have so far precluded application of their results outside of simple settings. More recently, Johnson et al. (2013) and Tan et al. (2013) studied geometric ergodicity of some simple two-component random scan Gibbs samplers. In contrast, the random scan Gibbs samplers we consider have more components, making the required analysis substantially more complicated.

The rest of the paper is organized as follows. In section 2 we will fully specify our two Bayesian hierarchical models and the random scan Gibbs samplers as well as our main results where we prove geometric ergodicity of both samplers. In section 3 we give some discussion and context for the results. Many technical details and calculations are deferred to the appendices.

2 Two Models and Random Scan Gibbs Samplers

We now turn our attention to describing the two hierarchical models, the associated random scan Gibbs samplers (RSGS), and our main results.

2.1 Two unknown variance components

Suppose for $i = 1, \dots, K$

$$\begin{aligned} Y_i|\theta_i, \gamma_i &\overset{ind}{\sim} N(\theta_i, \gamma_i^{-1}) \\ \theta_i|\mu, \lambda_\theta, \lambda_i &\overset{ind}{\sim} \mathrm{N}(\mu, \lambda_\theta^{-1}\lambda_i^{-1}) \\ \mu \sim \mathrm{N}(m_0, s_0^{-1}) \qquad & \gamma_i \overset{iid}{\sim} \mathrm{Gamma}(a_3, b_3) \\ \lambda_\theta \sim \mathrm{Gamma}(a_1, b_1) \qquad & \lambda_i \overset{iid}{\sim} \mathrm{Gamma}(a_2, b_2) \end{aligned} \tag{5}$$

where we assume the a_1, a_2, a_3, b_1, b_2, b_3 and s_0 are known positive constants while m_0 is a known scalar. (Note that if $X \sim \mathrm{Gamma}(a, b)$, then it has density proportional to $x^{a-1}e^{-bx}$ for $x > 0$.) The hierarchy results in a proper posterior density $f(\xi, \lambda_\theta, \lambda, \gamma|y)$ and four full conditional densities $f(\lambda_\theta|\xi, \lambda, \gamma)$, $f(\lambda|\xi, \lambda_\theta, \gamma)$, $f(\gamma|\xi, \lambda_\theta, \lambda)$ and $f(\xi|\lambda_\theta, \lambda, \gamma)$ which are given by

$$\begin{aligned} \lambda_\theta|\xi, \lambda, \gamma &\sim \mathrm{Gamma}\left(a_1 + \frac{K}{2}, b_1 + \frac{1}{2}\sum_{i=1}^{K} \lambda_i(\theta_i - \mu)^2\right) \\ \lambda_i|\xi, \lambda_\theta, \gamma &\overset{ind}{\sim} \mathrm{Gamma}\left(a_2 + \frac{1}{2}, b_2 + \frac{\lambda_\theta}{2}(\theta_i - \mu)^2\right) \\ \gamma_i|\xi, \lambda_\theta, \lambda &\overset{ind}{\sim} \mathrm{Gamma}\left(a_3 + \frac{1}{2}, b_3 + \frac{1}{2}(\theta_i - y_i)^2\right) \\ \xi|\lambda, \lambda_\theta, \gamma &\sim \mathrm{N}_K(\xi_0, V) \end{aligned} \tag{6}$$

where the components of ξ_0 and V are reported in Appendix C.

Let δ denote Dirac's delta and $r = (r_1, r_2, r_3, r_4)$ be the selection probabilities. Define

$$\begin{aligned} k(\xi', \lambda_\theta', \lambda', \gamma'|\xi, \lambda_\theta, \lambda, \gamma) = {} & r_1 f(\xi'|\lambda_\theta, \lambda, \gamma)\delta(\lambda_\theta' - \lambda_\theta)\delta(\lambda' - \lambda)\delta(\gamma' - \gamma) \\ & + r_2 f(\lambda_\theta'|\xi, \lambda, \gamma)\delta(\xi' - \xi)\delta(\lambda' - \lambda)\delta(\gamma' - \gamma) \\ & + r_3 f(\lambda'|\xi, \lambda_\theta, \gamma)\delta(\xi' - \xi)\delta(\lambda_\theta' - \lambda_\theta)\delta(\gamma' - \gamma) \\ & + r_4 f(\gamma|\xi, \lambda_\theta, \lambda)\delta(\xi' - \xi)\delta(\lambda' - \lambda)\delta(\lambda_\theta' - \lambda_\theta)\ . \end{aligned}$$

For $A \in \mathcal{B}(\mathsf{X})$ the Markov kernel for the RSGS is given by

$$
\begin{aligned}
P((\xi, \lambda_\theta, \lambda, \gamma), A) &= \int_A k(\xi', \lambda_\theta', \lambda', \gamma' | \xi, \lambda_\theta, \lambda, \gamma) d\xi' \, d\lambda_\theta' \, d\lambda' \, d\gamma' \\
&= r_1 P_\xi((\lambda_\theta, \lambda, \gamma), A) + r_2 P_{\lambda_\theta}((\xi, \lambda, \gamma), A) + r_3 P_\lambda((\xi, \lambda_\theta, \gamma), A) + r_4 P_\gamma((\xi, \lambda_\theta, \lambda), A)
\end{aligned}
$$

where

$$
P_\xi((\lambda_\theta, \lambda, \gamma), A) = \int_{\{\xi : (\xi, \lambda_\theta, \lambda, \gamma) \in A\}} f(\xi | \lambda_\theta, \lambda, \gamma) \, d\xi
$$

and similarly for the remaining Gibbs updates $P_{\lambda_\theta}((\xi, \lambda, \gamma), A)$, $P_\lambda((\xi, \lambda_\theta, \gamma), A)$, and $P_\gamma((\xi, \lambda_\theta, \lambda), A)$.

It is well known that the RSGS kernel P is reversible with respect to the posterior F (see e.g. Roberts and Rosenthal, 2004) and hence that the posterior is the invariant distribution for the RSGS. Let

$$
h(\xi', \lambda_\theta', \lambda', \gamma' | \xi, \lambda_\theta, \lambda, \gamma) = f(\xi' | \lambda_\theta, \lambda, \gamma) f(\lambda_\theta' | \xi', \lambda, \gamma) f(\lambda' | \xi', \lambda_\theta', \gamma) f(\gamma' | \xi', \lambda_\theta', \lambda') \; .
$$

Note that h is a strictly positive density on $\mathsf{X} = \mathbb{R}^K \times \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+^K \times \mathbb{R}_+^K$; in fact, h is the Markov transition density for the deterministic scan Gibbs sampler. Next observe that for $A \in \mathcal{B}(\mathsf{X})$

$$
P^4((\xi, \lambda_\theta, \lambda, \gamma), A) \geq r_1 r_2 r_3 r_4 \int_A h(\xi', \lambda_\theta', \lambda', \gamma' | \xi, \lambda_\theta, \lambda, \gamma) \, d\xi' \, d\lambda_\theta' \, d\lambda' \, d\gamma' \; .
$$

It follows immediately that the RSGS is irreducible, aperiodic and Harris recurrent.

We turn our attention to establishing geometric ergodicity of the RSGS. The tool we will use to do so is Lemma 15.2.8 in Meyn and Tweedie (2009). Geometric ergodicity will follow from finding a function $V : \mathbb{R}^{K+1} \times \mathbb{R}_+ \times \mathbb{R}_+^K \times \mathbb{R}_+^K \to \mathbb{R}_+$ which is unbounded off compact sets–that is, the level set

$$
\{(\xi, \lambda_\theta, \lambda, \gamma) \; : \; V(\xi, \lambda_\theta, \lambda, \gamma) \leq T\}
$$

is compact for every $T > 0$–and a $0 < \rho < 1$ and $L < \infty$ such that

$$
E[V(\xi', \lambda_\theta', \lambda', \gamma') | \xi, \lambda_\theta, \lambda, \gamma] \leq \rho V(\xi, \lambda_\theta, \lambda, \gamma) + L \; . \tag{7}
$$

See Jones and Hobert (2001) for an accessible treatment of the connection between drift conditions, such as the one in (7), and geometric ergodicity.

Theorem 1. *If $2a_1 + K - 2 > 0$ and $a_3 > 1$, then the RSGS is geometrically ergodic.*

Proof. Define $V = \sum_{i=1}^{18} A_i w_i$ where the w_i are defined as follows and the A_i are positive constants that will be determined later. For $0 < c < \min\{b_1, b_2, b_3\}$,

$$
\begin{array}{lll}
w_1(\lambda_\theta) = \lambda_\theta^{-1} & w_2(\lambda) = \displaystyle\sum_{i=1}^K \lambda_i^{-1/2} & w_3(\xi) = \displaystyle\sum_{i=1}^K (\theta_i - \mu)^2 \\
w_4(\xi, \gamma) = \displaystyle\sum_{i=1}^K \gamma_i (\theta_i - y_i)^2 & w_5(\lambda_\theta) = e^{c\lambda_\theta} & w_6(\lambda) = \displaystyle\sum_{i=1}^K e^{c\lambda_i} \\
w_7(\xi, \lambda_\theta) = \displaystyle\sum_{i=1}^K \left[\lambda_\theta(\theta_i - \mu)^2 + 2b_2\right]^{1/2} & w_8(\xi, \lambda) = \displaystyle\sum_{i=1}^K \lambda_i (\theta_i - \mu)^2 & w_9(\lambda_\theta) = \lambda_\theta \\
w_{10}(\gamma) = \displaystyle\sum_{i=1}^K e^{c\gamma_i} & w_{11}(\gamma) = \displaystyle\sum_{i=1}^K \gamma_i & w_{12}(\lambda_\theta, \gamma) = \lambda_\theta \displaystyle\sum_{i=1}^K \gamma_i^{-1} \\
w_{13}(\lambda, \gamma) = \displaystyle\sum_{i=1}^K \lambda_i \gamma_i^{-1} & w_{14}(\gamma) = \displaystyle\sum_{i=1}^K \gamma_i^{-1} & w_{15}(\xi, \lambda_\theta) = \lambda_\theta \displaystyle\sum_{i=1}^K (\theta_i - y_i)^2 \\
w_{16}(\lambda) = \displaystyle\sum_{i=1}^K \lambda_i & w_{17}(\xi, \lambda) = \displaystyle\sum_{i=1}^K \lambda_i (\theta_i - y_i)^2 & w_{18}(\xi) = \displaystyle\sum_{i=1}^K (\theta_i - y_i)^2 \; .
\end{array}
$$

We need to show that V is unbounded off compact sets. That is, for every $T > 0$ the level set

$$L_T = \{(\xi, \lambda_\theta, \lambda, \gamma) \; : \; V(\xi, \lambda_\theta, \lambda, \gamma) \leq T\}$$

is compact. Let $T > 0$ be arbitrary and observe that $w_1 \to \infty$ as $\lambda_\theta \to 0$ and $w_5 \to \infty$ as $\lambda_\theta \to \infty$; $w_2 \to \infty$ as $\lambda_i \to 0$ and $w_6 \to \infty$ as $\lambda_i \to \infty$; $w_{14} \to \infty$ as $\gamma_i \to 0$ and $w_{10} \to \infty$ as $\gamma_i \to \infty$; $w_{18} \to \infty$ as $|\theta_i| \to \infty$; and conditional on $\theta_i \in L_T$, $w_3 \to \infty$ as $|\mu| \to \infty$. It now follows from the continuity of V that L_T is closed and bounded, hence compact.

Now all that is left is to establish (7). It is sufficient to consider only $r_1 = r_2 = r_3 = r_4 = 1/4$, since if the RSGS is geometrically ergodic for some selection probabilities it is geometrically ergodic for all selection probabilities (Proposition 19 Jones et al., 2014). Next observe that due to the form of $k(\xi', \lambda_\theta', \lambda', \gamma' | \xi, \lambda_\theta, \lambda, \gamma)$ the required expectation can be expressed as

$$
\begin{aligned}
E[V(\xi', \lambda_\theta', \lambda', \gamma') \,|\, \xi, \lambda_\theta, \lambda, \gamma] = \frac{1}{4} \int & V(\xi', \lambda_\theta, \lambda, \gamma) f(\xi' | \lambda_\theta, \lambda, \gamma) \, d\xi' \\
+ \frac{1}{4} \int & V(\xi, \lambda_\theta', \lambda, \gamma) f(\lambda_\theta' | \xi, \lambda, \gamma) \, d\lambda_\theta' \\
+ \frac{1}{4} \int & V(\xi, \lambda_\theta, \lambda', \gamma) f(\lambda' | \xi, \lambda_\theta, \gamma) \, d\lambda' \\
+ \frac{1}{4} \int & V(\xi, \lambda_\theta, \lambda, \gamma') f(\gamma' | \xi, \lambda_\theta, \lambda) \, d\gamma' \; .
\end{aligned}
$$

Hence we need to calculate expectations of the w_i with respect to the posterior full conditional densities. These calculations are done in appendix A yielding

$$E[V(\xi', \lambda_\theta', \lambda', \gamma') | \xi, \lambda_\theta, \lambda, \gamma] \leq c_1 w_1(\lambda_\theta) + \cdots + c_{18} w_{18}(\xi) + L$$

where $c_i = 3A_i/4$ for $i = 1, 2, 5, 6, 10$, $c_4 = A_4/2$,

$$c_3 = \frac{3}{4}A_3 + \frac{2a_1 + K}{8b_1}A_7 + \frac{2a_2 + 1}{8b_2}A_8 \qquad c_7 = \frac{1}{2}A_7 + \frac{1}{4\sqrt{2}}\frac{\Gamma(a_2)}{\Gamma(a_2 + 1/2)}A_2$$

$$c_8 = \frac{1}{2}A_8 + \frac{1}{4(2a_1 + K - 2)}A_1 \qquad c_{11} = \frac{3}{4}A_{11} + \frac{s_0^{-1} + \Delta^2}{4}A_4$$

$$c_{12} = \frac{1}{2}A_{12} + \frac{1}{4}A_7 + \frac{1}{4}A_{15} \qquad c_{13} = \frac{1}{2}A_{13} + \frac{1}{4}A_8 + \frac{1}{4}A_{17}$$

$$c_{15} = \frac{1}{2}A_{15} + \frac{1}{4(2a_3 - 1)}A_{12} \qquad c_{17} = \frac{1}{2}A_{17} + \frac{1}{4(2a_3 - 1)}A_{13}$$

$$\begin{aligned}
c_9 &= \frac{3}{4}A_9 + \frac{K(2s_0^{-1} + \Delta^2)}{4}A_7 + \frac{Kb_3}{2(2a_3 - 1)}A_{12} + \frac{K(s_0^{-1} + \Delta^2)}{4}A_{15} \\
c_{14} &= \frac{3}{4}A_{14} + \frac{1}{4}A_3 + \frac{2a_1 + K}{8b_1}A_{12} + \frac{2a_2 + 1}{8b_2}A_{13} + \frac{1}{4}A_{18} \\
c_{16} &= \frac{3}{4}A_{16} + \frac{2s_0^{-1} + \Delta^2}{4}A_8 + \frac{b_3}{2(2a_3 - 1)}A_{13} + \frac{s_0^{-1} + \Delta^2}{4}A_{17} \\
c_{18} &= \frac{3}{4}A_{18} + \frac{2a_3 + 1}{8b_3}A_4 + \frac{1}{4(2a_3 - 1)}A_{14} + \frac{2a_1 + K}{8b_1}A_{15} + \frac{2a_2 + 1}{8b_2}A_{17}
\end{aligned}$$

and

$$\begin{aligned}
L = {} & \frac{b_1}{2(2a_1 + K - 2)}A_1 + \frac{K}{4}\left(\frac{2}{s_0} + \Delta^2\right)A_3 + \frac{K}{4}A_4 + \frac{1}{4}\left(\frac{b_1}{b_1 - c}\right)^{a_1 + K/2} A_5 + \frac{K}{4}\left(\frac{b_2}{b_2 - c}\right)^{a_2 + 1/2} A_6 \\
& + \frac{K(2b_2 + 1)}{2}A_7 + \frac{2a_1 + K}{8b_1}A_9 + \frac{K}{4}\left(\frac{b_3}{b_3 - c}\right)^{a_3 + 1/2} A_{10} + \frac{K(2a_3 + 1)}{8b_3}A_{11} + \frac{Kb_3}{2(2a_3 - 1)}A_{14} \\
& + \frac{K(2a_2 + 1)}{8b_2}A_{16} + \frac{K}{4}\left(\frac{1}{s_0} + \Delta^2\right)A_{18} \; .
\end{aligned}$$

Notice our assumption that $2a_1 + K - 2 > 0$ and $a_3 > 1$ ensures that all of the above quantities are well-defined.

To establish (7) we need to establish the existence of $\rho < 1$. Recall that we still have not chosen the values of the A_i. Let $A_4 > 0$, $A_5 > 0$, $A_6 > 0$, $A_7 = A_8 = 1$, $A_{10} > 0$,

$$0 < A_1 < 2(2a_1 + K - 2) \qquad 0 < A_2 < \frac{2\sqrt{2}\Gamma(a_2 + 1/2)}{\Gamma(a_2)}$$

$$A_3 > \frac{2a_1 + K}{2b_1} + \frac{2a_2 + 1}{2b_2} \qquad A_9 > K\left[2s_0^{-1} + \Delta^2 + \frac{2b_3}{2a_3 - 1}A_{12} + (s_0^{-1} + \Delta^2)A_{15}\right]$$

$$A_{11} > [s_0^{-1} + \Delta^2]A_4 \qquad \frac{1}{2} + \frac{1}{2}A_{15} < A_{12} < 2(2a_3 - 1)A_{15}$$

$$\frac{1}{2} + \frac{1}{2}A_{17} < A_{13} < 2(2a_3 - 1)A_{17} \qquad A_{15} > \frac{1}{4(2a_3 - 1) - 1}$$

$$A_{16} > \frac{2}{s_0} + \Delta^2 + \frac{2b_3}{2a_3 - 1}A_{13} + (s_0^{-1} + \Delta^2)A_{17} \qquad A_{17} > \frac{1}{4(2a_3 - 1) - 1}$$

$$A_3 + \frac{2a_1 + K}{2b_1}A_{12} + \frac{2a_2 + 1}{2b_2}A_{13} + A_{18} < A_{14} < (2a_3 - 1)\left[A_{18} - \frac{2a_3 + 1}{2b_3}A_4 - \frac{2a_1 + K}{2b_1}A_{15} - \frac{2a_2 + 1}{2b_2}A_{17}\right]$$

and

$$\begin{aligned} A_{18} > \frac{1}{2a_3 - 2}&\left[A_3 + \frac{2a_1 + K}{2b_1}A_{12} + \frac{2a_2 + 1}{2b_2}A_{13}\right] \\ &+ \frac{2a_3 - 1}{2a_3 - 2}\left[\frac{2a_3 + 1}{2b_3}A_4 + \frac{2a_1 + K}{2b_1}A_{15} + \frac{2a_2 + 1}{2b_2}A_{17}\right] . \end{aligned}$$

The conditions on $A_1, \ldots, A_{18}$ ensure the existence of ρ such that

$$\max\left\{\frac{c_1}{A_1}, \frac{c_2}{A_2}, \ldots, \frac{c_{18}}{A_{18}}\right\} \leq \rho < 1$$

and hence we have shown the existence of $0 < \rho < 1$ and an $L < \infty$ such that

$$\begin{aligned} E[V(\xi', \lambda_\theta', \lambda', \gamma')|\xi, \lambda_\theta, \lambda, \gamma] &\leq c_1 w_1(\lambda_\theta) + \cdots + c_{18}w_{18}(\xi) + L \\ &= \frac{c_1}{A_1}A_1 w_1(\lambda_\theta) + \cdots + \frac{c_{18}}{A_{18}}A_{18}w_{18}(\xi) + L \\ &\leq \max\left\{\frac{c_1}{A_1}, \frac{c_2}{A_2}, \ldots, \frac{c_{18}}{A_{18}}\right\} V(\xi, \lambda_\theta, \lambda, \gamma) + L \\ &\leq \rho V(\xi, \lambda_\theta, \lambda, \gamma) + L . \end{aligned}$$

We conclude that the RSGS is geometrically ergodic. □

2.2 One unknown variance component

Doss and Hobert (2010) consider a hierarchical model which is a special case of the model presented in the previous section, that is, (5). Let $\lambda = (\lambda_1, \ldots, \lambda_K)^T$ and suppose that for $i = 1, \ldots, K$

$$
\begin{aligned}
Y_i|\theta_i &\overset{ind}{\sim} N(\theta_i, \gamma_i^{-1}) \\
\theta_i|\mu, \lambda_\theta, \lambda &\overset{ind}{\sim} \mathrm{N}(\mu, \lambda_\theta^{-1}\lambda_i^{-1}) \\
\mu &\sim \mathrm{N}(m_0, s_0^{-1}) \\
\lambda_\theta \sim \mathrm{Gamma}(a_1, b_1) &\qquad \lambda_i \overset{iid}{\sim} \mathrm{Gamma}(a_2, b_2)
\end{aligned}
\tag{8}
$$

where the γ_i are known and positive as are s_0, a_1, a_2, b_1 and b_2. Also, $m_0 \in \mathbb{R}$ is known. Doss and Hobert (2010) use this model in the context of meta-analysis where it can be reasonable to assume the γ_i are known. The posterior density is $f(\xi, \lambda_\theta, \lambda|y)$, yielding full conditional distributions

$$
\begin{aligned}
&\lambda_\theta|\xi, \lambda \sim \mathrm{Gamma}\left(a_1 + \frac{K}{2}, b_1 + \frac{1}{2}\sum_{i=1}^{K} \lambda_i(\theta_i - \mu)^2\right) \\
&\lambda_i|\xi, \lambda_\theta \overset{ind}{\sim} \mathrm{Gamma}\left(a_2 + \frac{1}{2}, b_2 + \frac{\lambda_\theta}{2}(\theta_i - \mu)^2\right) \\
&\xi|\lambda, \lambda_\theta \sim \mathrm{N}_K(\xi_0, V)
\end{aligned}
\tag{9}
$$

where the components of ξ_0 and V are reported in Appendix C. Notice that, due to the conditional independence assumptions, these are the same as the full conditionals in (6).

Let δ denote Dirac's delta and $r = (r_1, r_2, r_3)$ denote the selection probabilities. Define

$$
\begin{aligned}
k(\xi', \lambda_\theta', \lambda'|\xi, \lambda_\theta, \lambda) = {}& r_1 f(\xi'|\lambda_\theta, \lambda)\delta(\lambda_\theta' - \lambda_\theta)\delta(\lambda' - \lambda) + r_2 f(\lambda_\theta'|\xi, \lambda)\delta(\xi' - \xi)\delta(\lambda' - \lambda) \\
&+ r_3 f(\lambda'|\xi, \lambda_\theta)\delta(\xi' - \xi)\delta(\lambda_\theta' - \lambda_\theta)
\end{aligned}
$$

so that if $A \in \mathcal{B}(\mathsf{X})$ then the Markov kernel for the RSGS is given by

$$
\begin{aligned}
P((\xi, \lambda_\theta, \lambda), A) &= \int_A k(\xi', \lambda_\theta', \lambda'|\xi, \lambda_\theta, \lambda) d\xi' \, d\lambda_\theta' \, d\lambda' \\
&= r_1 P_\xi((\lambda_\theta, \lambda), A) + r_2 P_{\lambda_\theta}((\xi, \lambda), A) + r_3 P_\lambda((\xi, \lambda_\theta), A)
\end{aligned}
$$

where

$$
P_\xi((\lambda_\theta, \lambda), A) = \int_{\{\xi : (\xi, \lambda_\theta, \lambda) \in A\}} f(\xi|\lambda_\theta, \lambda) \, d\xi
$$

and similarly for the remaining Gibbs updates $P_{\lambda_\theta}((\xi, \lambda), A)$, and $P_\lambda((\xi, \lambda_\theta), A)$.

It is well known that the RSGS kernel P is reversible with respect to the posterior (see e.g. Roberts and Rosenthal, 2004) and hence that the posterior is the invariant distribution for the RSGS. Let

$$
h(\xi', \lambda_\theta', \lambda'|\xi, \lambda_\theta, \lambda) = f(\xi'|\lambda_\theta, \lambda) f(\lambda_\theta'|\xi', \lambda) f(\lambda'|\xi', \lambda_\theta') \; ,
$$

which is a strictly positive density on $\mathsf{X} = \mathbb{R}^K \times \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+^K$. Next observe that for $A \in \mathcal{B}(\mathsf{X})$

$$P^3((\xi, \lambda_\theta, \lambda), A) \geq r_1 r_2 r_3 \int_A h(\xi', \lambda_\theta', \lambda' | \xi, \lambda_\theta, \lambda)\, d\xi'\, d\lambda_\theta'\, d\lambda'$$

and it follows that the RSGS is irreducible, aperiodic and Harris recurrent.

As in section 2.1 we will establish geometric ergodicity using Lemma 15.2.8 from Meyn and Tweedie (2009). Specifically, we construct a function $V : \mathbb{R}^{K+1} \times \mathbb{R}_+ \times \mathbb{R}_+^K \to \mathbb{R}_+$ which is unbounded off compact sets and a $0 < \rho < 1$ and $L < \infty$ such that

$$E[V(\xi', \lambda_\theta', \lambda') | \xi, \lambda_\theta, \lambda] \leq \rho V(\xi, \lambda_\theta, \lambda) + L\,. \tag{10}$$

Theorem 2. *If $2a_1 + K - 2 > 0$, then the RSGS is geometrically ergodic.*

Proof. Define $V = \sum_{i=1}^{10} A_i w_i$ where the constants A_i are to be determined below and the functions w_i are as follows. Letting $0 < c < b_1 \wedge b_2$, and $\alpha_i = 1/\gamma_i + 2/s_0 + \Delta^2$ and $\alpha = \sum_{i=1}^K \alpha_i$, set

$$w_1(\lambda_\theta) = \lambda_\theta^{-1} \qquad w_2(\lambda) = \sum_{i=1}^K \lambda_i^{-1/2} \qquad w_3(\xi) = \sum_{i=1}^K (\theta_i - \mu)^2$$

$$w_4(\xi) = \sum_{i=1}^K \gamma_i (\theta_i - y_i)^2 \qquad w_5(\lambda_\theta) = e^{c\lambda_\theta} \qquad w_6(\lambda) = \sum_{i=1}^K e^{c\lambda_i}$$

$$w_7(\xi, \lambda_\theta) = \sum_{i=1}^K \left[\lambda_\theta (\theta_i - \mu)^2 + 2b_2\right]^{1/2} \qquad w_8(\xi, \lambda) = \sum_{i=1}^K \lambda_i (\theta_i - \mu)^2 \qquad w_9(\lambda_\theta) = \lambda_\theta$$

$$w_{10}(\lambda) = \sum_{i=1}^K \alpha_i \lambda_i\,.$$

We begin by showing that V is unbounded off compact sets, that is, the set

$$L_T = \{\xi, \lambda_\theta, \lambda \,:\, V(\xi, \lambda_\theta, \lambda) \leq T\}$$

is compact for every $T > 0$. The argument is the same as one given by Doss and Hobert (2010) in their proof of geometric ergodicity of the deterministic scan Gibbs sampler, but is included here for the sake of completeness since their drift function did not have our w_8, w_9 and w_{10}. Let $T > 0$ be arbitrary and observe that $w_5 \to \infty$ and $w_9 \to \infty$ as $\lambda_\theta \to \infty$ while $w_1 \to \infty$ as $\lambda_\theta \to 0$; $w_6 \to \infty$ and $w_{10} \to \infty$ as $\lambda_i \to \infty$ while $w_2 \to \infty$ as $\lambda_i \to 0$; $w_4 \to \infty$ as $|\theta_i| \to \infty$; and conditional on $\theta_i \in L_T$ we see that $w_3 \to \infty$ and $w_8 \to \infty$ as $|\mu| \to \infty$. It now follows from the continuity of V that L_T is closed and bounded, hence compact.

Now we need to establish (10). It is sufficient to consider only $r_1 = r_2 = r_3 = 1/3$ since if the RSGS is geometrically ergodic for some selection probabilities it is geometrically ergodic for all selection probabilities (Proposition 19 Jones et al., 2014). Notice that

$$E[V(\xi',\lambda'_\theta,\lambda')|\xi,\lambda_\theta,\lambda] = r_1\int V(\xi',\lambda_\theta,\lambda)f(\xi'|\lambda_\theta,\lambda)\,d\xi' + r_2\int V(\xi,\lambda'_\theta,\lambda)f(\lambda'_\theta|\xi,\lambda)\,d\lambda'_\theta + r_3\int V(\xi,\lambda_\theta,\lambda')f(\lambda'|\xi,\lambda_\theta)\,d\lambda' \,.$$

Hence we need to calculate expectations of the w_i with respect to the full conditional distributions. We calculate the required expectations in appendix B. If

$$L = \frac{2b_1}{3(2a_1+K-2)}A_1 + \frac{\alpha}{3}A_3 + \frac{1}{3}\left[K + (\Delta^2 + s_0^{-1})\sum_{i=1}^{K}\gamma_i\right]A_4 + \frac{1}{3}\left(\frac{b_1}{b_1-c}\right)^{a_1+K/2}A_5 + \frac{K}{3}\left(\frac{b_2}{b_2-c}\right)^{a_2+1/2}A_6 + \frac{2K(2b_2+1)}{3}A_7 + \frac{2a_1+K}{6b_1}A_9 + \frac{(2a_2+1)\alpha}{6b_2}A_{10}$$

and

$$\begin{aligned} c_1 &= \frac{2}{3}A_1 & c_2 &= \frac{2}{3}A_2 \\ c_3 &= \frac{2}{3}A_3 + \frac{2a_1+K}{6b_1}A_7 + \frac{2a_2+1}{6b_2}A_8 & c_4 &= \frac{2}{3}A_4 \\ c_5 &= \frac{2}{3}A_5 & c_6 &= \frac{2}{3}A_6 \\ c_7 &= \frac{1}{3\sqrt{2}}\frac{\Gamma(a_2)}{\Gamma(a_2+1/2)}A_2 + \frac{1}{3}A_7 & c_8 &= \frac{1}{3(2a_1+K-2)}A_1 + \frac{1}{3}A_8 \\ c_9 &= \frac{\alpha}{3}A_7 + \frac{2}{3}A_9 & c_{10} &= \frac{1}{3}A_8 + \frac{2}{3}A_{10}\,, \end{aligned}$$

then our calculations yield

$$E[V(\xi',\lambda'_\theta,\lambda')|\xi,\lambda_\theta,\lambda] \le c_1 w_1(\lambda_\theta) + \cdots + c_{10}w_{10}(\lambda) + L\,.$$

Now we need to choose the A_i so as to ensure the existence of $\rho < 1$ in (10). To this end let $A_4 > 0$, $A_5 > 0$, $A_6 > 0$, $A_7 = A_8 = 1$, $A_9 > \alpha$, $A_{10} > 1$,

$$0 < A_1 < 2(2a_1+K-2), \quad 0 < A_2 < \frac{2\sqrt{2}\Gamma(a_2+1/2)}{\Gamma(a_2)} \quad \text{and} \quad A_3 > \frac{2a_1+K}{2b_1} + \frac{2a_2+1}{2b_2}\,.$$

Then there exists ρ such that

$$\max\left\{\frac{c_1}{A_1},\ldots,\frac{c_{10}}{A_{10}}\right\} \le \rho < 1\,.$$

Hence we have shown that

$$\begin{aligned} E[V(\xi',\lambda'_\theta,\lambda')|\xi,\lambda_\theta,\lambda] &\le c_1 w_1(\lambda_\theta) + \cdots + c_{10}w_{10}(\lambda) + L \\ &= \frac{c_1}{A_1}A_1 w_1(\lambda_\theta) + \cdots + \frac{c_{10}}{A_{10}}A_{10}w_{10}(\lambda) + L \\ &\le \max\left\{\frac{c_1}{A_1},\ldots,\frac{c_{10}}{A_{10}}\right\}V(\xi,\lambda_\theta,\lambda) + L \\ &\le \rho V(\xi,\lambda_\theta,\lambda) + L\,. \end{aligned}$$

We conclude that the RSGS is geometrically ergodic. □

3 Discussion

MCMC methods can produce asymptotically valid point estimates of posterior quantities such as means and quantiles under benign regularity conditions. Indeed these conditions often hold by construction of the algorithm. However, ensuring reliability of the simulation requires some idea of the quality of the estimation and this is best addressed through the construction of an asymptotically valid Monte Carlo standard error (Flegal et al., 2008; Jones and Hobert, 2001). This requires a central limit theorem as in (4) and methods to consistently estimate the variance of the asymptotic normal distribution (Jones et al., 2006). In general, it is much more difficult to establish a central limit theorem for Markov chains than for a random sample. However, when the Markov chain is reversible and geometrically ergodic, then the same moment condition is required for both settings; recall (3). By establishing geometric ergodicity the practitioner has all of the tools available to perform rigorous output analysis to describe the quality of the estimation procedure. See Flegal and Hughes (2012) for an R package which allows easy implementation of the suggested methodology.

We have focused on establishing geometric ergodicity of random scan Gibbs samplers for two Bayesian one-way random effects models. While the random scan Gibbs sampler is as easy to implement as the deterministic scan version, there has been little work in comparing the two methods. However, Roberts and Rosenthal (2015) recently have shown in some examples the random scan version is more efficient while in other examples the deterministic scan is more efficient. Unfortunately, their results do not apply to the Gibbs samplers considered here. Generally, comparison of random scan with deterministic scan is difficult. There seem to be two technical reasons for this. Notice that the computational cost to complete one sweep of the deterministic scan algorithm is essentially the same as that required to complete 4 steps of the random scan version in section 2.1 and 3 steps of the random scan version in 2.2. Also, the random scan Gibbs sampler is reversible with respect to the posterior while the deterministic scan Gibbs sampler is not. However, there has been speculation (Johnson et al., 2013) that random scan Gibbs samplers and deterministic scan Gibbs samplers converge at the same qualitative rate for the two-component setting. There has been some progress in this direction (Tan et al., 2013), but little is known in the case where there are more than two components.

The RSGS considered in section 2.1 has four components and it is unknown whether the deterministic scan Gibbs sampler is geometrically ergodic. However, Hobert and Geyer (1998) and Jones and Hobert (2004) considered deterministic scan Gibbs samplers for the same hierarchy in (5) except that they assumed each $\lambda_i = 1$ and hence our work required much new analysis. Interestingly, the conditions on the sample size in our Theorem 1 are slightly weaker than those obtained by either Hobert and Geyer (1998) or Jones and Hobert (2004) for the deterministic scan Gibbs sampler.

Doss and Hobert (2010) consider the setting of section 2.2 and prove that the deterministic scan Gibbs sampler is geometrically ergodic when $a_2 = b_2 = d/2$, $d \geq 1$ and $2a_1 + K - 2 > 0$. However, we think that their restrictions on a_2 and b_2 play no role in their argument. That is, the sufficient conditions for geometric ergodicity of RSGS are the same as those Doss and Hobert (2010) obtained for geometric ergodicity of the deterministic scan Gibbs sampler.

We note that establishing geometric ergodicity of random scan Gibbs samplers also has implications for other MCMC algorithms. For example, Latuszynski et al. (2013) require the standard RSGS to be geometrically ergodic to implement asymptotically valid adaptive RSGS algorithms. Thus our work allows rigorous implementation of adaptive RSGS algorithms. Finally, our results imply geometric ergodicity of the class of random scan Metropolis-Hastings-within-Gibbs samplers where each ratio of the proposal density to the target full conditional is bounded; see Theorem 9 in Jones et al. (2014) for the details.

A Conditional Expectations for the Proof of Theorem 1

This section contains the calculation of the conditional expectations required for the proof of Theorem 1. Some supplementary calculations are given in appendix C. Note that

$$\begin{aligned} E[w_1(\lambda_\theta')|\xi, \lambda_\theta, \lambda, \gamma] &= \frac{3}{4} w_1(\lambda_\theta) + \frac{1}{4} E[(\lambda_\theta')^{-1}|\xi, \lambda, \gamma] \\ &= \frac{3}{4} w_1(\lambda_\theta) + \frac{1}{4(2a_1 + K - 2)} w_8(\xi, \lambda) + \frac{b_1}{2(2a_1 + K - 2)} , \end{aligned}$$

$$\begin{aligned} E[w_2(\lambda')|\xi, \lambda_\theta, \lambda, \gamma] &= \frac{3}{4} w_2(\lambda) + \frac{1}{4} \sum_{i=1}^{K} E[(\lambda_i')^{-1/2}|\xi, \lambda_\theta, \gamma] \\ &= \frac{3}{4} w_2(\lambda) + \frac{1}{4} \sum_{i=1}^{K} \frac{\Gamma(a_2)}{\Gamma(a_2 + 1/2)} \left(b_2 + \frac{\lambda_\theta}{2} (\theta_i - \mu)^2 \right)^{1/2} \\ &= \frac{3}{4} w_2(\lambda) + \frac{1}{4\sqrt{2}} \frac{\Gamma(a_2)}{\Gamma(a_2 + 1/2)} w_7(\xi, \lambda_\theta) \end{aligned}$$

and

$$\begin{aligned}
E[w_3(\xi')|\xi,\lambda_\theta,\lambda,\gamma] &= \frac{3}{4}w_3(\xi) + \frac{1}{4}\sum_{i=1}^{K} E[(\theta_i' - \mu')^2|\lambda_\theta,\lambda,\gamma] \\
&\leq \frac{3}{4}w_3(\xi) + \frac{1}{4}\sum_{i=1}^{K}\left[\frac{1}{\gamma_i} + \frac{2}{s_0} + \Delta^2\right] \quad \text{by (12)} \\
&= \frac{3}{4}w_3(\xi) + \frac{1}{4}w_{14}(\gamma) + \frac{K}{4}\left(\frac{2}{s_0} + \Delta^2\right) .
\end{aligned}$$

Now

$$E[w_4(\xi',\gamma')|\xi,\lambda_\theta,\lambda,\gamma] = \frac{1}{2}w_4(\xi,\gamma) + \frac{1}{4}E[w_4(\xi',\gamma)|\lambda_\theta,\lambda,\gamma] + \frac{1}{4}E[w_4(\xi,\gamma')|\xi,\lambda_\theta,\lambda]$$

and hence we consider $E[w_4(\xi',\gamma)|\lambda_\theta,\lambda,\gamma]$ and $E[w_4(\xi,\gamma')|\xi,\lambda_\theta,\lambda]$ individually before returning to the calculation for $E[w_4(\xi',\gamma')|\xi,\lambda_\theta,\lambda,\gamma]$:

$$\begin{aligned}
E[w_4(\xi',\gamma)|\lambda_\theta,\lambda,\gamma] &= \sum_{i=1}^{K}\gamma_i E\left[(\theta_i' - y_i)^2|\lambda_\theta,\lambda,\gamma\right] \\
&\leq \sum_{i=1}^{K}\gamma_i\left(\frac{1}{\gamma_i} + \frac{1}{s_0} + \Delta^2\right) \quad \text{by (13)} \\
&= K + (s_0^{-1} + \Delta^2)w_{11}(\gamma)
\end{aligned}$$

and

$$\begin{aligned}
E[w_4(\xi,\gamma')|\xi,\lambda_\theta,\lambda] &= \sum_{i=1}^{K}(\theta_i - y_i)^2\, E\left[\gamma_i'|\xi,\lambda_\theta,\lambda\right] \\
&= \sum_{i=1}^{K}(\theta_i - y_i)^2\frac{2a_3+1}{2b_3 + (\theta_i - y_i)^2} \\
&\leq \frac{2a_3+1}{2b_3}w_{18}(\xi) .
\end{aligned}$$

Therefore,

$$\begin{aligned}
E[w_4(\xi',\gamma')|\xi,\lambda_\theta,\lambda,\gamma] &\leq \frac{1}{2}w_4(\xi,\gamma) + \frac{1}{4}\left[K + (s_0^{-1} + \Delta^2)w_{11}(\gamma)\right] + \frac{1}{4}\left[\frac{2a_3+1}{2b_3}w_{18}(\xi)\right] \\
&= \frac{1}{2}w_4(\xi,\gamma) + \frac{s_0^{-1}+\Delta^2}{4}w_{11}(\gamma) + \frac{2a_3+1}{8b_3}w_{18}(\xi) + \frac{K}{4} .
\end{aligned}$$

Recall that $0 < c < \min\{b_1,b_2,b_3\}$. Thus

$$\begin{aligned}
E[w_5(\lambda_\theta')|\xi,\lambda_\theta,\lambda,\gamma] &= \frac{3}{4}w_5(\lambda_\theta) + \frac{1}{4}E\left[e^{c\lambda_\theta'}|\xi,\lambda,\gamma\right] \\
&= \frac{3}{4}w_5(\lambda_\theta) + \frac{1}{4}\left(\frac{2b_1 + \sum_{i=1}^{K}\lambda_i(\theta_i-\mu)^2}{2b_1 + \sum_{i=1}^{K}\lambda_i(\theta_i-\mu)^2 - 2c}\right)^{a_1+K/2} \\
&\leq \frac{3}{4}w_5(\lambda_\theta) + \frac{1}{4}\left(\frac{b_1}{b_1-c}\right)^{a_1+K/2} .
\end{aligned}$$

Similarly,

$$E[w_6(\lambda')|\xi,\lambda_\theta,\lambda,\gamma] \le \frac{3}{4}w_6(\lambda) + \frac{K}{4}\left(\frac{b_2}{b_2-c}\right)^{a_2+1/2} .$$

Next,

$$E[w_7(\xi',\lambda'_\theta)|\xi,\lambda_\theta,\lambda,\gamma] = \frac{1}{2}w_7(\xi,\lambda_\theta) + \frac{1}{4}E[w_7(\xi',\lambda_\theta)|\lambda_\theta,\lambda,\gamma] + \frac{1}{4}E[w_7(\xi,\lambda'_\theta)|\xi,\lambda,\gamma]$$

where

$$\begin{aligned}
E[w_7(\xi',\lambda_\theta)|\lambda_\theta,\lambda,\gamma] &= \sum_{i=1}^K E\left(\left[\lambda_\theta(\theta'_i-\mu')^2+2b_2\right]^{1/2}|\lambda_\theta,\lambda,\gamma\right)\\
&\le \sum_{i=1}^K \left(\lambda_\theta\, E[(\theta'_i-\mu')^2|\lambda_\theta,\lambda,\gamma]+2b_2\right)^{1/2} \qquad \text{by Jensen's inequality}\\
&\le \sum_{i=1}^K \left[\lambda_\theta\left(\frac{1}{\gamma_i}+\frac{2}{s_0}+\Delta^2\right)+2b_2\right]^{1/2} \quad \text{by (12)}\\
&\le \sum_{i=1}^K \left[\lambda_\theta\left(\frac{1}{\gamma_i}+\frac{2}{s_0}+\Delta^2\right)+2b_2+1\right]\\
&= K(2s_0^{-1}+\Delta^2)w_9(\lambda_\theta)+w_{12}(\lambda_\theta,\gamma)+K(2b_2+1)
\end{aligned}$$

and

$$\begin{aligned}
E[w_7(\xi,\lambda'_\theta)|\xi,\lambda,\gamma] &= \sum_{i=1}^K E\left(\left[\lambda'_\theta(\theta_i-\mu)^2+2b_2\right]^{1/2}|\xi,\lambda,\gamma\right)\\
&\le \sum_{i=1}^K \left[(\theta_i-\mu)^2E(\lambda'_\theta|\xi,\lambda,\gamma)+2b_2\right]^{1/2} \qquad \text{by Jensen's inequality}\\
&= \sum_{i=1}^K \left[(\theta_i-\mu)^2\frac{2a_1+K}{2b_1+\sum_{i=1}^K\lambda_i(\theta_i-\mu)^2}+2b_2\right]^{1/2}\\
&\le \sum_{i=1}^K \left[\frac{2a_1+K}{2b_1}(\theta_i-\mu)^2+2b_2+1\right]\\
&= \frac{2a_1+K}{2b_1}w_3(\xi)+K(2b_2+1) .
\end{aligned}$$

Hence

$$\begin{aligned}
E[w_7(\xi',\lambda'_\theta)|\xi,\lambda_\theta,\lambda,\gamma] = \frac{2a_1+K}{8b_1}w_3(\xi)+\frac{1}{2}w_7(\xi,\lambda_\theta)+\frac{K(2s_0^{-1}+\Delta^2)}{4}w_9(\lambda_\theta)+\frac{1}{4}w_{12}(\lambda_\theta,\gamma)\\
+\frac{K(2b_2+1)}{2} .
\end{aligned}$$

Further,

$$E[w_8(\xi', \lambda')|\xi, \lambda_\theta, \lambda, \gamma] = \frac{1}{2} w_8(\xi, \lambda) + \frac{1}{4} E[w_8(\xi', \lambda)|\lambda_\theta, \lambda, \gamma] + \frac{1}{4} E[w_8(\xi, \lambda')|\xi, \lambda_\theta, \gamma]$$

where

$$\begin{aligned} E[w_8(\xi', \lambda)|\lambda_\theta, \lambda, \gamma] &= \sum_{i=1}^{K} \lambda_i E[(\theta_i' - \mu')^2|\lambda_\theta, \lambda, \gamma] \\ &\leq \sum_{i=1}^{K} \lambda_i \left(\frac{1}{\gamma_i} + \frac{2}{s_0} + \Delta^2 \right) \quad \text{by (12)} \\ &= w_{13}(\lambda, \gamma) + \left(\frac{2}{s_0} + \Delta^2 \right) w_{16}(\lambda) \end{aligned}$$

and

$$\begin{aligned} E[w_8(\xi, \lambda')|\xi, \lambda_\theta, \gamma] &= \sum_{i=1}^{K} (\theta_i - \mu)^2 E[\lambda_i'|\lambda_\theta, \lambda, \gamma] \\ &= \sum_{i=1}^{K} (\theta_i - \mu)^2 \frac{2a_2 + 1}{2b_2 + \lambda_\theta(\theta_i - \mu)^2} \\ &\leq \frac{2a_2 + 1}{2b_2} w_3(\xi) \ . \end{aligned}$$

Hence

$$E[w_8(\xi', \lambda')|\xi, \lambda_\theta, \lambda, \gamma] \leq \frac{2a_2 + 1}{8b_2} w_3(\xi) + \frac{1}{2} w_8(\xi, \lambda) + \frac{1}{4} w_{13}(\lambda, \gamma) + \frac{2s_0^{-1} + \Delta^2}{4} w_{16}(\lambda) \ .$$

Continuing with the calculations,

$$\begin{aligned} E[w_9(\lambda_\theta')|\xi, \lambda_\theta, \lambda, \gamma] &= \frac{3}{4} w_9(\lambda_\theta) + \frac{1}{4} E(\lambda_\theta'|\xi, \lambda, \gamma) \\ &= \frac{3}{4} w_9(\lambda_\theta) + \frac{1}{4} \frac{2a_1 + K}{2b_1 + \sum_{i=1}^{K} \lambda_i(\theta_i - \mu)^2} \\ &\leq \frac{3}{4} w_9(\lambda_\theta) + \frac{2a_1 + K}{8b_1} \ , \end{aligned}$$

$$\begin{aligned} E[w_{10}(\gamma')|\xi, \lambda_\theta, \lambda, \gamma] &= \frac{3}{4} w_{10}(\gamma) + \frac{1}{4} \sum_{i=1}^{K} E\left[e^{c\gamma_i'}|\xi, \lambda_\theta, \lambda \right] \\ &= \frac{3}{4} w_{10}(\gamma) + \frac{1}{4} \sum_{i=1}^{K} \left(\frac{2b_3 + (\theta_i - y_i)^2}{2b_3 + (\theta_i - y_i)^2 - 2c} \right)^{a_3 + 1/2} \\ &\leq \frac{3}{4} w_{10}(\gamma) + \frac{K}{4} \left(\frac{b_3}{b_3 - c} \right)^{a_3 + 1/2} \end{aligned}$$

and

$$\begin{aligned} E[w_{11}(\gamma')|\xi,\lambda_\theta,\lambda,\gamma] &= \frac{3}{4}w_{11}(\gamma) + \frac{1}{4}\sum_{i=1}^K E\left[\gamma_i'|\xi,\lambda_\theta,\lambda\right] \\ &= \frac{3}{4}w_{11}(\gamma) + \frac{1}{4}\sum_{i=1}^K \frac{2a_3+1}{2b_3+(\theta_i-y_i)^2} \\ &\le \frac{3}{4}w_{11}(\gamma) + \frac{K(2a_3+1)}{8b_3} \ . \end{aligned}$$

Further,

$$E[w_{12}(\lambda_\theta',\gamma')|\xi,\lambda_\theta,\lambda,\gamma] = \frac{1}{2}w_{12}(\lambda_\theta,\gamma) + \frac{1}{4}E[w_{12}(\lambda_\theta',\gamma)|\xi,\lambda,\gamma] + \frac{1}{4}E[w_{12}(\lambda_\theta,\gamma')|\xi,\lambda_\theta,\lambda]$$

where

$$\begin{aligned} E[w_{12}(\lambda_\theta',\gamma)|\xi,\lambda,\gamma] &= w_{14}(\gamma)E[\lambda_\theta'|\xi,\lambda,\gamma] \\ &= w_{14}(\gamma)\frac{2a_1+K}{2b_1+\sum_{i=1}^K\lambda_i(\theta_i-\mu)^2} \\ &\le \frac{2a_1+K}{2b_1}w_{14}(\gamma) \end{aligned}$$

and

$$\begin{aligned} E[w_{12}(\lambda_\theta,\gamma')|\xi,\lambda_\theta,\lambda] &= w_9(\lambda_\theta)\sum_{i=1}^K E\left(\frac{1}{\gamma_i'}|\xi,\lambda_\theta,\lambda\right) \\ &= w_9(\lambda_\theta)\sum_{i=1}^K \frac{2b_3+(\theta_i-y_i)^2}{2a_3-1} \\ &= \frac{2b_3K}{2a_3-1}w_9(\lambda_\theta) + \frac{1}{2a_3-1}w_{15}(\xi,\lambda_\theta) \ . \end{aligned}$$

Hence

$$\begin{aligned} E[w_{12}(\lambda_\theta',\gamma')|\xi,\lambda_\theta,\lambda,\gamma] \le{}& \frac{b_3K}{2(2a_3-1)}w_9(\lambda_\theta) + \frac{1}{2}w_{12}(\lambda_\theta,\gamma) + \frac{2a_1+K}{8b_1}w_{14}(\gamma) \\ &+ \frac{1}{4(2a_3-1)}w_{15}(\xi,\lambda_\theta) \ . \end{aligned}$$

Similarly,

$$E[w_{13}(\lambda',\gamma')|\xi,\lambda_\theta,\lambda,\gamma] = \frac{1}{2}w_{13}(\lambda,\gamma) + \frac{1}{4}E[w_{13}(\lambda',\gamma)|\xi,\lambda_\theta,\gamma] + \frac{1}{4}E[w_{13}(\lambda,\gamma')|\xi,\lambda_\theta,\lambda]$$

where

$$\begin{aligned} E[w_{13}(\lambda',\gamma)|\xi,\lambda_\theta,\gamma] &= \sum_{i=1}^K \frac{1}{\gamma_i}E[\lambda_i'|\xi,\lambda_\theta,\gamma] \\ &= \sum_{i=1}^K \frac{1}{\gamma_i}\frac{2a_2+1}{2b_2+\lambda_\theta(\theta_i-\mu)^2} \\ &\le \frac{2a_2+1}{2b_2}w_{14}(\gamma) \end{aligned}$$

and

$$\begin{aligned}
E[w_{13}(\lambda,\gamma')|\xi,\lambda_\theta,\lambda] &= \sum_{i=1}^{K} \lambda_i E\left(\frac{1}{\gamma_i'}|\xi,\lambda_\theta,\lambda\right) \\
&= \sum_{i=1}^{K} \lambda_i \frac{2b_3 + (\theta_i - y_i)^2}{2a_3 - 1} \\
&= \frac{2b_3}{2a_3-1} w_{16}(\lambda) + \frac{1}{2a_3-1} w_{17}(\xi,\lambda) \ .
\end{aligned}$$

Hence

$$E[w_{13}(\lambda',\gamma')|\xi,\lambda_\theta,\lambda,\gamma] \le \frac{1}{2} w_{13}(\lambda,\gamma) + \frac{2a_2+1}{8b_2} w_{14}(\gamma) + \frac{b_3}{2(2a_3-1)} w_{16}(\lambda) + \frac{1}{4(2a_3-1)} w_{17}(\xi,\lambda) \ .$$

Next

$$\begin{aligned}
E[w_{14}(\gamma')|\xi,\lambda_\theta,\lambda,\gamma] &= \frac{3}{4} w_{14}(\gamma) + \frac{1}{4}\sum_{i=1}^{K} E\left(\frac{1}{\gamma_i'}|\xi,\lambda_\theta,\lambda\right) \\
&= \frac{3}{4} w_{14}(\gamma) + \frac{1}{4}\sum_{i=1}^{K} \frac{2b_3 + (\theta_i - y_i)^2}{2a_3 - 1} \\
&= \frac{3}{4} w_{14}(\gamma) + \frac{1}{4(2a_3-1)} w_{18}(\xi) + \frac{Kb_3}{2(2a_3-1)} \ .
\end{aligned}$$

Further,

$$E[w_{15}(\xi',\lambda_\theta')|\xi,\lambda_\theta,\lambda,\gamma] = \frac{1}{2} w_{15}(\xi,\lambda_\theta) + \frac{1}{4} E[w_{15}(\xi',\lambda_\theta)|\lambda_\theta,\lambda,\gamma] + \frac{1}{4} E[w_{15}(\xi,\lambda_\theta')|\xi,\lambda,\gamma]$$

where

$$\begin{aligned}
E[w_{15}(\xi',\lambda_\theta)|\lambda_\theta,\lambda,\gamma] &= \lambda_\theta \sum_{i=1}^{K} E\left[(\theta_i' - y_i)^2|\lambda_\theta,\lambda,\gamma\right] \\
&\le \lambda_\theta \sum_{i=1}^{K} \left[\frac{1}{\gamma_i} + \frac{1}{s_0} + \Delta^2\right] \qquad \text{by (13)} \\
&\le K\left(\Delta^2 + \frac{1}{s_0}\right) w_9(\lambda_\theta) + w_{12}(\lambda_\theta,\gamma)
\end{aligned}$$

and

$$\begin{aligned}
E[w_{15}(\xi,\lambda_\theta')|\xi,\lambda,\gamma] &= w_{18}(\xi) E[\lambda_\theta'|\xi,\lambda,\gamma] \\
&= w_{18}(\xi) \frac{2a_1 + K}{2b_1 + \sum_{i=1}^{K} \lambda_i(\theta_i - \mu)^2} \\
&\le \frac{2a_1+K}{2b_1}\, w_{18}(\xi) \ .
\end{aligned}$$

Hence

$$E[w_{15}(\xi', \lambda'_\theta)|\xi, \lambda_\theta, \lambda, \gamma] \le \frac{K}{4}\left(\Delta^2 + \frac{1}{s_0}\right) w_9(\lambda_\theta) + \frac{1}{4}w_{12}(\lambda_\theta, \gamma) + \frac{1}{2}w_{15}(\xi, \lambda) + \frac{2a_1 + K}{8b_1}\, w_{18}(\xi)\ .$$

Now

$$\begin{aligned} E[w_{16}(\lambda')|\xi, \lambda_\theta, \lambda, \gamma] &= \frac{3}{4}w_{16}(\lambda) + \frac{1}{4}\sum_{i=1}^{K} E[\lambda'_i|\xi, \lambda_\theta, \gamma] \\ &= \frac{3}{4}w_{16}(\lambda) + \frac{1}{4}\sum_{i=1}^{K} \frac{2a_2 + 1}{2b_2 + \lambda_\theta(\theta_i - \mu)^2} \\ &\le \frac{3}{4}w_{16}(\lambda) + \frac{K(2a_2+1)}{8b_2}\ . \end{aligned}$$

Note that

$$E[w_{17}(\xi', \lambda')|\xi, \lambda_\theta, \lambda, \gamma] = \frac{1}{2}w_{17}(\xi, \lambda) + \frac{1}{4}E[w_{17}(\xi', \lambda)|\lambda_\theta, \lambda, \gamma] + \frac{1}{4}E[w_{17}(\xi, \lambda')|\xi, \lambda_\theta, \gamma]$$

where

$$\begin{aligned} E[w_{17}(\xi', \lambda)|\lambda_\theta, \lambda, \gamma] &= \sum_{i=1}^{K} \lambda_i E\left[(\theta'_i - y_i)^2|\lambda_\theta, \lambda, \gamma\right] \\ &\le \sum_{i=1}^{K} \lambda_i \left(\frac{1}{\gamma_i} + \frac{1}{s_0} + \Delta^2\right) \quad \text{by (13)} \\ &\le \left(\frac{1}{s_0} + \Delta^2\right) w_{16}(\lambda) + w_{13}(\lambda, \gamma) \end{aligned}$$

and

$$\begin{aligned} E[w_{17}(\xi, \lambda')|\xi, \lambda_\theta, \gamma] &= \sum_{i=1}^{K} (\theta_i - y_i)^2 E[\lambda'_i|\xi, \lambda_\theta, \gamma] \\ &= \sum_{i=1}^{K} (\theta_i - y_i)^2 \frac{2a_2 + 1}{2b_2 + \lambda_\theta(\theta_i - \mu)^2} \\ &\le \frac{2a_2+1}{2b_2}\, w_{18}(\xi)\ . \end{aligned}$$

Hence

$$E[w_{17}(\xi', \lambda')|\xi, \lambda_\theta, \lambda, \gamma] \le \frac{1}{4}w_{13}(\lambda, \gamma) + \frac{1}{4}\left(\frac{1}{s_0} + \Delta^2\right) w_{16}(\lambda) + \frac{1}{2}w_{17}(\xi, \lambda) + \frac{2a_2 + 1}{8b_2}\, w_{18}(\xi)\ .$$

Finally,

$$\begin{aligned} E[w_{18}(\xi')|\xi, \lambda_\theta, \lambda, \gamma] &= \frac{3}{4}w_{18}(\xi) + \frac{1}{4}\sum_{i=1}^{K} E\left[(\theta'_i - y_i)^2|\lambda_\theta, \lambda, \gamma\right] \\ &\le \frac{3}{4}w_{18}(\xi) + \frac{1}{4}\sum_{i=1}^{K}\left[\frac{1}{\gamma_i} + \frac{1}{s_0} + \Delta^2\right] \quad \text{by (13)} \\ &\le \frac{1}{4}w_{14}(\gamma) + \frac{3}{4}w_{18}(\xi) + \frac{K}{4}\left[\frac{1}{s_0} + \Delta^2\right]\ . \end{aligned}$$

B Conditional Expectations for the Proof of Theorem 2

This section contains the calculation of the conditional expectations required for the proof of Theorem 2. Some supplementary calculations are given in appendix C. Note that

$$\begin{aligned} E[w_1(\lambda_\theta')|\xi, \lambda_\theta, \lambda] &= \frac{2}{3} w_1(\lambda_\theta) + \frac{1}{3} E[w_1(\lambda_\theta')|\xi, \lambda] \\ &= \frac{2}{3} w_1(\lambda_\theta) + \frac{1}{3(2a_1 + K - 2)} w_8(\xi, \lambda) + \frac{2b_1}{3(2a_1 + K - 2)} , \end{aligned}$$

$$\begin{aligned} E[w_2(\lambda')|\xi, \lambda_\theta, \lambda] &= \frac{2}{3} w_2(\lambda) + \frac{1}{3} \sum_{i=1}^{K} E[(\lambda_i')^{-1/2}|\xi, \lambda_\theta] \\ &= \frac{2}{3} w_2(\lambda) + \frac{1}{3\sqrt{2}} \frac{\Gamma(a_2)}{\Gamma(a_2 + 1/2)} w_7(\xi, \lambda_\theta) , \end{aligned}$$

$$\begin{aligned} E[w_3(\xi')|\xi, \lambda_\theta, \lambda] &= \frac{2}{3} w_3(\xi) + \frac{1}{3} \sum_{i=1}^{K} E[(\theta_i' - \mu')^2|\lambda_\theta, \lambda] \\ &\leq \frac{2}{3} w_3(\xi) + \frac{1}{3} \sum_{i=1}^{K} \left[\frac{1}{\gamma_i} + \frac{2}{s_0} + \Delta^2 \right] \qquad \text{by (12)} \\ &= \frac{2}{3} w_3(\xi) + \frac{\alpha}{3} \end{aligned}$$

and

$$\begin{aligned} E[w_4(\xi')|\xi, \lambda_\theta, \lambda] &= \frac{2}{3} w_4(\xi) + \frac{1}{3} \sum_{i=1}^{K} \gamma_i E[(\theta_i' - y_i)^2|\lambda_\theta, \lambda] \\ &\leq \frac{2}{3} w_4(\xi) + \frac{1}{3} \sum_{i=1}^{K} \gamma_i \left[\frac{1}{\gamma_i} + \frac{1}{s_0} + \Delta^2 \right] \qquad \text{by (13)} \\ &\leq \frac{2}{3} w_4(\xi) + \frac{1}{3} \left[K + (\Delta^2 + s_0^{-1}) \sum_{i=1}^{K} \gamma_i \right] . \end{aligned}$$

Further

$$\begin{aligned} E[w_5(\lambda_\theta')|\xi, \lambda_\theta, \lambda] &= \frac{2}{3} w_5(\lambda_\theta) + \frac{1}{3} E(e^{c\lambda_\theta'}|\xi, \lambda) \\ &= \frac{2}{3} w_5(\lambda_\theta) + \frac{1}{3} \left(\frac{b_1 + w_8(\xi, \lambda)/2}{b_1 - c + w_8(\xi, \lambda)/2} \right)^{a_1 + K/2} \\ &\leq \frac{2}{3} w_5(\lambda_\theta) + \frac{1}{3} \left(\frac{b_1}{b_1 - c} \right)^{a_1 + K/2} \end{aligned}$$

and, similarly,

$$E[w_6(\lambda')|\xi, \lambda_\theta, \lambda] \leq \frac{2}{3} w_6(\lambda) + \frac{K}{3} \left(\frac{b_2}{b_2 - c} \right)^{a_2 + 1/2} .$$

Next,

$$E[w_7(\xi',\lambda_\theta')|\xi,\lambda_\theta,\lambda] = \frac{1}{3}w_7(\xi,\lambda_\theta) + \frac{1}{3}E[w_7(\xi,\lambda_\theta')|\xi,\lambda] + \frac{1}{3}E[w_7(\xi',\lambda_\theta)|\lambda_\theta,\lambda]$$

where

$$\begin{aligned}
E[w_7(\xi,\lambda_\theta')|\xi,\lambda] &= \sum_{i=1}^K E\{[\lambda_\theta'(\theta_i-\mu)^2 + 2b_2]^{1/2}\,|\xi,\lambda\} \\
&\le \sum_{i=1}^K [(\theta_i-\mu)^2 E(\lambda_\theta'|\xi,\lambda) + 2b_2]^{1/2} \qquad \text{by Jensen's inequality} \\
&= \sum_{i=1}^K \left[(\theta_i-\mu)^2 \frac{2a_1+K}{2b_1+\sum_{i=1}^K \lambda_i(\theta_i-\mu)^2} + 2b_2\right]^{1/2} \\
&\le \sum_{i=1}^K \left[(\theta_i-\mu)^2 \frac{2a_1+K}{2b_1} + 2b_2\right]^{1/2} \\
&\le \sum_{i=1}^K \left[(\theta_i-\mu)^2 \frac{2a_1+K}{2b_1} + 2b_2 + 1\right] \\
&= \frac{2a_1+K}{2b_1} w_3(\xi) + K(2b_2+1)
\end{aligned}$$

and

$$\begin{aligned}
E[w_7(\xi',\lambda_\theta)|\lambda_\theta,\lambda] &= \sum_{i=1}^K E\{[\lambda_\theta(\theta_i'-\mu')^2 + 2b_2]^{1/2}\,|\lambda_\theta,\lambda\} \\
&\le \sum_{i=1}^K \{\lambda_\theta E[(\theta_i'-\mu')^2|\lambda_\theta,\lambda] + 2b_2\}^{1/2} \qquad \text{by Jensen's inequality} \\
&\le \sum_{i=1}^K \left\{\lambda_\theta\left[\frac{1}{\gamma_i} + \frac{2}{s_0} + \Delta^2\right] + 2b_2\right\}^{1/2} \qquad \text{by (12)} \\
&\le \sum_{i=1}^K \{\alpha_i\lambda_\theta + 2b_2 + 1\}^{1/2} \\
&\le \alpha w_9(\lambda_\theta) + K(2b_2+1)\ .
\end{aligned}$$

It follows that

$$E[w_7(\xi',\lambda_\theta')|\xi,\lambda_\theta,\lambda] \le \frac{1}{3}w_7(\xi,\lambda_\theta) + \frac{2a_1+K}{6b_1}w_3(\xi) + \frac{\alpha}{3}w_9(\lambda_\theta) + \frac{2K(2b_2+1)}{3}\ .$$

Further,

$$E[w_8(\xi',\lambda')|\xi,\lambda_\theta,\lambda] = \frac{1}{3}w_8(\xi,\lambda) + \frac{1}{3}E[w_8(\xi',\lambda)|\lambda_\theta,\lambda] + \frac{1}{3}E[w_8(\xi,\lambda')|\xi,\lambda_\theta]$$

where

$$E[w_8(\xi', \lambda)|\lambda_\theta, \lambda] = \sum_{i=1}^{K} \lambda_i E[(\theta_i' - \mu')^2|\lambda_\theta, \lambda] \leq \sum_{i=1}^{K} \alpha_i \lambda_i = w_{10}(\lambda)$$

and

$$E[w_8(\xi, \lambda')|\xi, \lambda_\theta] = \sum_{i=1}^{K} (\theta_i - \mu)^2 E[\lambda_i'|\xi, \lambda_\theta] = \sum_{i=1}^{K} (\theta_i - \mu)^2 \frac{2a_2 + 1}{2b_2 + \lambda_\theta(\theta_i - \mu)^2} \leq \frac{2a_2 + 1}{2b_2} w_3(\xi) \ .$$

Putting this together we obtain

$$E[w_8(\xi', \lambda')|\xi, \lambda_\theta, \lambda] = \frac{1}{3} w_8(\xi, \lambda) + \frac{1}{3} w_{10}(\lambda) + \frac{2a_2 + 1}{6b_2} w_3(\xi) \ .$$

Continuing the calculations:

$$\begin{aligned} E[w_9(\lambda_\theta')|\xi, \lambda_\theta, \lambda] &= \frac{2}{3} w_9(\lambda_\theta) + \frac{1}{3} E[w_9(\lambda_\theta')|\xi, \lambda] \\ &= \frac{2}{3} w_9(\lambda_\theta) + \frac{1}{3} \frac{2a_1 + K}{2b_1 + \sum_{i=1}^{K} \lambda_i(\theta_i - \mu)^2} \\ &\leq \frac{2}{3} w_9(\lambda_\theta) + \frac{2a_1 + K}{6b_1} \end{aligned}$$

and

$$\begin{aligned} E[w_{10}(\lambda')|\xi, \lambda_\theta, \lambda] &= \frac{2}{3} w_{10}(\lambda) + \frac{1}{3} \sum_{i=1}^{10} \alpha_i E[\lambda_i'|\xi, \lambda_\theta] \\ &= \frac{2}{3} w_{10}(\lambda) + \frac{1}{3} \sum_{i=1}^{10} \alpha_i \frac{2a_2 + 1}{2b_2 + \lambda_\theta(\theta_i - \mu)^2} \\ &\leq \frac{2}{3} w_{10}(\lambda) + \frac{(2a_2 + 1)\alpha}{6b_2} \ . \end{aligned}$$

C Components of ξ_0 and V

Doss and Hobert (2010) derive the components of ξ_0 and V and upper bounds for them. Letting

$$t = \sum_{i=1}^{K} \frac{\gamma_i \lambda_\theta \lambda_i}{\lambda_\theta \lambda_i + \gamma_i} \ ,$$

ξ_0 has components

$$E[\theta_i|\lambda_\theta, \lambda] = \frac{\lambda_\theta \lambda_i}{\lambda_\theta \lambda_i + \gamma_i} \left[\frac{1}{s_0 + t} \left[\sum_{j=1}^{K} \frac{\gamma_j \lambda_\theta \lambda_j y_j}{\lambda_\theta \lambda_j + \gamma_j} + m_0 s_0 \right] \right] + \frac{\gamma_i y_i}{\lambda_\theta \lambda_i + \gamma_i}$$

and

$$E[\mu|\lambda_\theta, \lambda] = \frac{1}{s_0 + t} \left[\sum_{j=1}^{K} \frac{\gamma_j \lambda_\theta \lambda_j y_j}{\lambda_\theta \lambda_j + \gamma_j} + m_0 s_0 \right] \ .$$

Let Δ denote the length of the convex hull of $\{y_1, \dots, y_K, m_0\}$. Note that $E[\mu|\lambda_\theta, \lambda]$ is a convex combination of $y_1, \dots, y_K, m_0$ and that $E[\theta_i|\lambda_\theta, \lambda]$ is a convex combination of $E[\mu|\lambda_\theta, \lambda]$ and $y_1, \dots, y_K$. Hence

$$(E[\theta_i|\lambda_\theta, \lambda] - E[\mu|\lambda_\theta, \lambda])^2 \le \Delta^2 \quad \text{and} \quad (E[\theta_i|\lambda_\theta, \lambda] - y_i)^2 \le \Delta^2 \quad \text{for} \quad i = 1, \dots, K \,. \tag{11}$$

The components of V are given by

$$\begin{aligned}
\text{Var}(\theta_i|\lambda_\theta, \lambda) &= \frac{1}{\lambda_\theta \lambda_i + \gamma_i} \left[1 + \frac{\lambda_\theta^2 \lambda_i^2}{(\lambda_\theta \lambda_i + \gamma_i)(s_0 + t)} \right] \le \frac{1}{\gamma_i} + \frac{1}{s_0} \\
\text{Cov}(\theta_i, \theta_j|\lambda_\theta, \lambda) &= \frac{\lambda_\theta^2 \lambda_i \lambda_j}{(\lambda_\theta \lambda_i + \gamma_i)(\lambda_\theta \lambda_j + \gamma_j)(s_0 + t)} \le \frac{1}{s_0} \\
\text{Cov}(\theta_i, \mu|\lambda_\theta, \lambda) &= \frac{\lambda_\theta \lambda_i}{(\lambda_\theta \lambda_i + \gamma_i)(s_0 + t)} \le \frac{1}{s_0} \\
\text{Var}(\mu|\lambda_\theta, \lambda) &= \frac{1}{s_0 + t} \le \frac{1}{s_0} \,.
\end{aligned}$$

Now we can see that for $i = 1, \dots, K$

$$\begin{aligned}
E[(\theta_i - \mu)^2|\lambda_\theta, \lambda] &= \text{Var}(\theta_i|\lambda_\theta, \lambda) + \text{Var}(\mu|\lambda_\theta, \lambda) - 2\text{Cov}(\theta_i, \mu|\lambda_\theta, \lambda) + [E(\theta_i|\lambda_\theta, \lambda) - E(\mu|\lambda_\theta, \lambda)]^2 \\
&\le \frac{1}{\gamma_i} + \frac{2}{s_0} + \Delta^2 \qquad (12)
\end{aligned}$$

and

$$\begin{aligned}
E[(\theta_i - y_i)^2|\lambda_\theta, \lambda] &= \text{Var}(\theta_i|\lambda_\theta, \lambda) + [E(\theta_i|\lambda_\theta, \lambda) - y_i]^2 \\
&\le \frac{1}{\gamma_i} + \frac{1}{s_0} + \Delta^2 \,. \qquad (13)
\end{aligned}$$

References

Chan, K. S. and Geyer, C. J. (1994). Comment on "Markov chains for exploring posterior distributions". *The Annals of Statistics*, 22:1747–1758.

Diaconis, P., Khare, K., and Saloff-Coste, L. (2008). Gibbs sampling, orhtogonal polynomials and orthogonal families. *Statistical Science*, 23:151–178.

Doss, C. R., Flegal, J. M., Jones, G. L., and Neath, R. C. (2014). Markov chain Monte Carlo estimation of quantiles. *Electronic Journal of Statistics*, 8:2448–2478.

Doss, H. and Hobert, J. P. (2010). Estimation of Bayes factors in a class of hierarchical random effects models using a geometrically ergodic MCMC algorithm. *Journal of Computational and Graphical Statistics*, 19:295–312.

Flegal, J. M., Haran, M., and Jones, G. L. (2008). Markov chain Monte Carlo: Can we trust the third significant figure? *Statistical Science*, 23:250–260.

Flegal, J. M. and Hughes, J. (2012). mcmcse: Monte Carlo standard errors for MCMC R package version 1.0-1. http://cran.r-project.org/web/packages/mcmcse/index.html.

Flegal, J. M. and Jones, G. L. (2010). Batch means and spectral variance estimators in Markov chain Monte Carlo. *The Annals of Statistics*, 38:1034–1070.

Flegal, J. M. and Jones, G. L. (2011). Implementing Markov chain Monte Carlo: Estimating with confidence. In Brooks, S., Gelman, A., Jones, G., and Meng, X.-L., editors, *Handbook of Markov Chain Monte Carlo*. CRC Press, Boca Raton, FL.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6:721–741.

Häggström, O. and Rosenthal, J. S. (2007). Variance conditions for Marlov chain CLTs. *Electronic Communications in Probability*, 12:454–464.

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., Jones, G. L., and Neath, R. C. (2013). Component-wise Markov chain Monte Carlo. *Statistical Science*, 28:360–375.

Jones, G. L. (2004). On the Markov chain central limit theorem. *Probability Surveys*, 1:299–320.

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*, 16:312–334.

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.

Jones, G. L., Roberts, G. O., and Rosenthal, J. S. (2014). Convergence of conditional Metropolis-Hastings samplers. *Advances in Applied Probability*, 46:422–445.

Latuszynski, K., Roberts, G. O., and Rosenthal, J. S. (2013). Adaptive Gibbs samplers and related MCMC methods. *Annals of Applied Probability*, 23:66–98.

Lee, K.-J., Jones, G. L., Caffo, B. S., and Bassett, S. (2013). Spatial Bayesian variable selection models for functional magnetic resonance imaging data. *Bayesian Analysis*, 9:699–732.

Levine, R. A. and Casella, G. (2006). Optimizing random scan Gibbs samplers. *Journal of Multivariate Analysis*, 97:2071–2100.

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

Liu, J. S., Wong, W. H., and Kong, A. (1995). Covariance structure and convergence rate of the Gibbs sampler with various scans. *Journal of the Royal Statistical Society,* Series B, 57:157–169.

Marchev, D. and Hobert, J. P. (2004). Geometric ergodicity of van Dyk and Meng's algorithm for the multivariate Student's t model. *Journal of the American Statistical Association*, 99:228–238.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equations of state calculations by fast computing machine. *Journal of Chemical Physics*, 21.

Meyn, S. P. and Tweedie, R. L. (2009). *Markov Chains and Stochastic Stability.* Cambridge University Press, Cambridge.

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

Richardson, S., Bottolo, L., and Rosenthal, J. S. (2010). Bayesian models for sparse regression analysis of high dimensional data. In Bernardo, J. M., Bayarri, M. J., Berger, J. O., Dawid, A. P., Heckerman, D., Smith, A. F. M., and West, M., editors, *Bayesian Statistics 9*. Oxford University Press, Oxford.

Roberts, G. O. and Polson, N. G. (1994). On the geometric convergence of the Gibbs sampler. *Journal of the Royal Statistical Society,* Series B, 56:377–384.

Roberts, G. O. and Rosenthal, J. S. (1997). Geometric ergodicity and hybrid Markov chains. *Electronic Communications in Probability*, 2:13–25.

Roberts, G. O. and Rosenthal, J. S. (1999). Convergence of slice sampler Markov chains. *Journal of the Royal Statistical Society,* Series B, 61:643–660.

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

Roberts, G. O. and Rosenthal, J. S. (2015). Surprising convergence properties of some simple Gibbs samplers under various scans. *Preprint*.

Roy, V. and Hobert, J. P. (2007). Convergence rates and asymptotic standard errors for Markov chain Monte Carlo algorithms for Bayesian probit regression. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 69:607–623.

Tan, A. and Hobert, J. P. (2012). Block Gibbs sampling for Bayesian random effects models with improper priors: Convergence and regeneration. *Journal of Computational and Graphical Statistics*, 18:861–878.

Tan, A., Jones, G. L., and Hobert, J. P. (2013). On the geometric ergodicity of two-variable Gibbs samplers. In Jones, G. L. and Shen, X., editors, *Advances in Modern Statistical Theory and Applications: A Festschrift in honor of Morris L. Eaton*. Institute of Mathematical Statistics, Beachwood, Ohio, USA.