



## On the Geometric Convergence of the Gibbs Sampler

Gareth O. Roberts, Nicholas G. Polson

*Journal of the Royal Statistical Society. Series B (Methodological)*, Volume 56, Issue 2 (1994), 377-384.

Stable URL:

<http://links.jstor.org/sici?sici=0035-9246%281994%2956%3A2%3C377%3AOTGCOT%3E2.0.CO%3B2-V>

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

*Journal of the Royal Statistical Society. Series B (Methodological)* is published by Royal Statistical Society. Please contact the publisher for further permissions regarding the use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/rss.html>.

---

*Journal of the Royal Statistical Society. Series B (Methodological)*  
©1994 Royal Statistical Society

JSTOR and the JSTOR logo are trademarks of JSTOR, and are Registered in the U.S. Patent and Trademark Office. For more information on JSTOR contact [jstor-info@umich.edu](mailto:jstor-info@umich.edu).

©2002 JSTOR

## On the Geometric Convergence of the Gibbs Sampler

By GARETH O. ROBERTS†  
*University of Cambridge, UK*

and

NICHOLAS G. POLSON  
*University of Chicago, USA*

[Received September 1990. Final revision February 1993]

### SUMMARY

The rate of convergence of the Gibbs sampler is discussed. The Gibbs sampler is a Monte Carlo simulation method with extensive application to computational issues in the Bayesian paradigm. Conditions for the geometric rate of convergence of the algorithm for discrete and continuous parameter spaces are derived, and an illustrative exponential family example is given.

*Keywords:* BAYESIAN INFERENCE; GEOMETRIC RATE OF CONVERGENCE; GIBBS SAMPLER; MARKOV CHAIN; MONTE CARLO SIMULATION

### 1. INTRODUCTION

This paper investigates conditions under which the Gibbs sampler (Gelfand and Smith, 1990; Tanner and Wong, 1987; Geman and Geman, 1984) converges at a geometric rate. The main results appear in Sections 2 and 3, where geometric convergence results are established, with respect to total variation and supremum norms under fairly natural conditions on the underlying distribution. For ease of exposition, we shall concentrate on the two most commonly encountered situations, where the state space is finite or continuous. All our results will establish uniform convergence, a strong form of geometric convergence, under appropriate regularity conditions. Uniform convergence is a useful property in its own right but also happens to be a sufficient condition for certain ergodic central limit theorems. Such results are important for estimation in Markov chain simulation but will not be considered in detail here.

Our approach is to apply the theory of Markov chains to the specific Gibbs sampler case. In the finite state space case, uniform ergodicity is automatic. However, the situation is more complicated for continuous state spaces where even well-behaved underlying distributions can give rise to Markov chains which converge slowly, or have unbounded kernels. We give two results in this context, corollary 2 and corollary 3 which establish uniform convergence under different sets of conditions on the underlying density. Finally we apply these results to an example of a Bayesian hierarchical model where regularity conditions for geometric convergence are naturally satisfied. Here the hierarchical structure of the model is crucial in permitting application of corollary 3.

Briefly, the Gibbs sampler is as follows. Consider a  $p$ -dimensional random variable  $(X_1, \dots, X_p)$ , together with its one-dimensional conditional densities,

†*Address for correspondence:* Statistical Laboratory, University of Cambridge, 16 Mill Lane, Cambridge, CB2 1SB, UK.

denoted by  $\pi(\theta_i|\theta_j, j \neq i), 1 \leq i \leq p$ . Suppose that the full joint distribution  $\pi(\theta_1, \dots, \theta_p)$  is characterized by these conditionals; then the Gibbs sampler algorithm is as follows. Given an arbitrary set of starting values  $X_1^{(0)}, \dots, X_p^{(0)}$ , draw  $X_1^{(1)}$  from  $\pi(\theta_1|X_2^{(0)}, \dots, X_p^{(0)})$ , then  $X_2^{(1)}$  from  $\pi(\theta_2|X_1^{(1)}, X_3^{(0)}, \dots, X_p^{(0)})$ , and so on up to  $X_p^{(1)}$  from  $\pi(\theta_p|X_1^{(1)}, \dots, X_{p-1}^{(1)})$ , completing one iteration of the scheme. After  $t$  such iterations we obtain  $X_1^{(t)}, \dots, X_p^{(t)}$ . As  $t \rightarrow \infty$ , under suitable regularity conditions, we achieve an approximate sample from the joint distribution  $\pi(\theta_1, \dots, \theta_p)$ .

2. FINITE CASE

Suppose that we have a  $p$ -dimensional distribution with state space  $\Theta$ , where each component  $\theta_i, 1 \leq i \leq p$ , has a finite state space and each of the conditional distributions is denoted by  $\pi(\theta_i|\theta_j, j \neq i), 1 \leq i \leq p$ . The Gibbs sampler applied to this situation gives rise to a  $p$ -dimensional Markov chain  $\{\mathbf{X}^{(n)}, n = 1, 2, \dots\}$  where  $\mathbf{X}^{(n)} = (X_1^{(n)}, X_2^{(n)}, \dots, X_p^{(n)})$ , defined by the transition kernel

$$\begin{aligned} \mathbf{P}[(\theta_1, \theta_2, \dots, \theta_p) \rightarrow (\phi_1, \dots, \phi_p)] &= K(\theta_1, \theta_2, \dots, \theta_p, \phi_1, \phi_2, \dots, \phi_p) \\ &= \prod_{i=1}^p \mathbf{P}[X_i^{(1)} = \phi_i | X_j^{(1)} = \phi_j, 1 \leq j < i, X_j^{(0)} = \theta_j, i < j \leq p]. \end{aligned} \tag{2.1}$$

Throughout, we shall write  $K^{(i)}$  for the iterated kernel:  $K^{(1)}(\theta, \phi) = K(\theta, \phi)$  and  $K^{(i)}(\theta, \phi) = \int K^{(i-1)}(\theta, \zeta) K(\zeta, \phi)$ . Also define the  $L^1$ -norm by  $\|g\|_1 = \int_{\Theta} g(\theta)$ .

To relate convergence theorems for finite Markov chains to our context, we require the following lemma, which first proves that the Gibbs sampler is well defined so long as its starting value is chosen to have a non-zero stationary mass, and secondly assures aperiodicity.

*Lemma 1.* For the Gibbs sampler, if  $\pi(\theta^{(0)}) > 0$  then

- (a)  $\pi(\theta^{(n)}) > 0$ , almost surely  $\forall n$ , and
- (b)  $\{\mathbf{X}^{(n)}, n > 0\}$  is aperiodic.

*Proof.*

- (a) By induction, suppose that  $\pi(\theta^{(n)}) > 0$ ; then  $\pi(\theta_2^{(n)}, \theta_3^{(n)}, \dots, \theta_p^{(n)}) > 0$  and so  $\pi(\theta_1^{(n+1)}|\theta_2^{(n)}, \dots, \theta_p^{(n)}) > 0$  almost surely. Similarly, by updating each parameter in turn, we obtain  $\pi(\theta_p^{(n+1)}|\theta_1^{(n+1)}, \dots, \theta_{p-1}^{(n+1)}) > 0$  and therefore  $\pi(\theta^{(n+1)}) > 0$ .
- (b) Aperiodicity will follow from  $K(\theta^{(n)}, \theta^{(n)}) > 0$ . However, we shall show that this will be true almost surely for all points of the Markov chain. Suppose otherwise that  $K(\theta^{(n)}, \theta^{(n)}) = 0$ ; then  $\pi(\theta_i^{(n)}|\theta_j^{(n)}, j \neq i) = 0$  for some  $i$ , and therefore  $\pi(\theta^{(n)}) = 0$  which is almost surely not the case by part (a).  $\square$

We say that any two states  $\theta, \phi \in \Theta$  communicate, if there is a chain of points  $\{\xi^{(i)}, 0 \leq i \leq n\}$ , such that  $\xi^{(0)} = \theta, \xi^{(n)} = \phi, \xi^{(i)}$  differs from  $\xi^{(i-1)}$  by one coordinate, and  $\pi(\xi^{(i)}) > 0, \forall i$ .

*Corollary 1.* Suppose that all states communicate. Then the finite state space Gibbs sampler of the subsection converges geometrically in  $L^1$  to  $\pi$ .

*Proof.* Suppose that  $\theta$  and  $\phi$  communicate in  $n(\theta, \phi)$  steps; then it is easy to check that  $K^{(n(\theta, \phi))}(\theta, \phi) > 0$ . Furthermore  $K^{(n)}(\theta, \phi) > 0$  for  $n \geq n(\theta, \phi)$  so that  $K^{(N)}(\cdot, \cdot)$  is a positive kernel where

$$N = \max \{n(\theta, \phi), (\theta, \phi) \in \Theta \times \Theta \text{ subject to } n(\theta, \phi) < \infty\}.$$

Therefore by Perron–Frobenius theory (see for example Seneta (1981)) geometric convergence in  $L^1$  is assured.  $\square$

Central limit theorems are also available for these ergodic averages; see for example theorem 4.6.9 of Kemeny and Snell (1976).

### 3. CONTINUOUS CASE

Let  $\mathbf{X}$  be a Markov chain on  $\mathbf{R}^p$ , with transition kernel  $K$  (assumed with respect to Lebesgue measure for simplicity, although this is not strictly necessary). Define also the iterated kernels:

$$K^{(n)}(\theta, \phi) = \int K(\theta, \xi) K^{(n-1)}(\xi, \phi) d\xi$$

with  $K^{(1)} \equiv K$ . We shall be concerned with density convergence rates, so we let  $f_n$  denote the density of  $\mathbf{X}^{(n)}$  for all  $n$ , suppressing the dependence on  $\mathbf{X}^{(0)}$ . Clearly  $f_n$  and  $f_{n+1}$  are connected via

$$f_{n+1}(\psi) = \int K(\theta, \psi) f_n(\theta) d\theta.$$

Suppose that  $\mathbf{X}$  is ergodic and let  $\pi$  be its stationary density. Let  $R = \{\theta, \pi(\theta) > 0\}$ , and suppose that  $K(\theta, \phi) > 0, \forall \theta, \phi \in \mathbf{R}^p$ .

Define the  $L^1$ -norm by

$$\|g\|_1 = \int_{\mathbf{R}^p} |g(\theta)| d\theta$$

and the  $L^\infty$ -norm by

$$\|g\|_\infty = \sup_{\theta \in \mathbf{R}^p} |g(\theta)|.$$

We shall consider the following strong form of geometric convergence. We say that

$$f_n \xrightarrow{L^1} \pi$$

uniformly, if  $\exists M < \infty, 0 < \rho < 1$  such that  $\|f_n - \pi\|_1 \leq M\rho^n$ .

Under uniform ergodicity of this type, it is possible to state ergodic central limit theorems, which are useful in estimation for the Gibbs sampler. We refer to Tierney (1991) for a summary of such available results, and alternative conditions ensuring uniform ergodicity.

*Lemma 2.* Suppose that there is a non-negative function  $K^*(\cdot)$  on  $\mathbf{R}^p$ , such that  $K^*(\phi) > 0$  on a set of positive Lebesgue measure, and for some  $N \in \mathbf{N}$

$$K^{(N)}(\theta, \phi) \geq K^*(\phi) \tag{3.1}$$

for all  $\theta$  in the domain of the Markov chain,  $\forall \phi \in \mathbf{R}^p$ . Then  $\mathbf{X}$  is aperiodic and positive recurrent and  $\mathbf{X}$  converges to  $\pi$ , uniformly in  $L^1$ , with rate  $\rho \leq \{1 - K^*(\mathbf{R}^p)\}^{1/N}$ , where  $K^*(\mathbf{R}^p) = \int_{\mathbf{R}^p} K^*(\phi) d\phi$ .

This can be proved by standard methods; see for example Nummelin (1984). Here we give a sketch of the proof.

*Sketch proof.* For  $n \geq N$ , and letting  $\bar{K}(\theta, \phi) = K^{(N)}(\theta, \phi) - K^*(\phi) \mathbf{I}[\phi \in A]$ , we have

$$\begin{aligned} g_n(\phi) &= f_n(\phi) - \pi(\phi) \\ &= \int \{f_{n-N}(\theta) - \pi(\theta)\} K^{(N)}(\theta, \phi) d\theta \\ &= \int \{f_{n-N}(\theta) - \pi(\theta)\} \bar{K}(\theta, \phi) d\theta. \end{aligned}$$

Therefore

$$\begin{aligned} \|g_n\|_1 &= \int |g_n(\phi)| d\phi = \int \left| \int \{f_{n-N}(\theta) - \pi(\theta)\} \bar{K}(\theta, \phi) d\theta \right| d\phi \\ &\leq \int \int |f_{n-N}(\theta) - \pi(\theta)| \bar{K}(\theta, \phi) d\theta d\phi \\ &= \{1 - K^*(\mathbf{R}^p)\} \int |g_{n-N}(\theta)| d\theta \\ &= \{1 - K^*(\mathbf{R}^p)\} \|g_{n-N}\|_1, \end{aligned}$$

as required. □

Lemma 2 can be a very powerful result in its own right. However, the minorization condition (3.1) may not hold directly for the Markov chain induced by single iterations of the Gibbs sampler. The following result shows how condition (3.1) can follow from smoothness conditions on  $K$  and  $\pi$ . This allows us to verify the condition for kernels of multiple iterations of the Gibbs sampler, where it cannot be directly checked. We make the following assumptions for the rest of the paper. Assume that  $R$  is bounded, and that the family of functions  $\{g; g(\cdot) = K(\theta, \cdot) \text{ for } \theta \in R\}$  is equicontinuous.

*Theorem 1.* Suppose that, in addition to the above conditions,  $K(\theta, \psi)$  is positive  $\forall (\theta, \psi) \in R \times R$ ; then

$$f_n \xrightarrow{L^1, L^\infty} \pi$$

uniformly.

*Proof.* Fix  $\psi \in R$ . Then  $K(\theta, \psi) > 0$  for all  $\theta, \psi \in R$ . Let  $K^*(\psi) = \inf_{\theta} \{K^{(2)}(\theta, \psi)\}$ , and suppose that  $K^*(\psi) = 0$  for that particular  $\psi$ . Then by the Arzela-Ascoli theorem (see for example Dunford and Schwartz (1958)) there is a sequence  $\{\theta_i\}$  such that  $K(\theta_i, \mu) \rightarrow h(\mu)$  uniformly in  $\mu$ , for some function  $h$ , and  $K^{(2)}(\theta_i, \psi) \rightarrow 0$  as  $i \rightarrow \infty$ . Also,

$$K^{(2)}(\theta_i, \psi) \rightarrow \int_R h(\mu) K(\mu, \psi) d\mu,$$

so that  $\int_R h(\mu) K(\mu, \psi) d\mu = 0$ .

Now  $h$  is non-negative and continuous (by uniform convergence), so that, since  $K(\mu, \psi) > 0, \forall \mu \in R$ , then  $h \equiv 0$ . However,

$$\int h(\mu) d\mu = \lim_{i \rightarrow \infty} \left\{ \int_R K(\theta_i, \mu) d\mu \right\} = 1$$

for a contradiction. Therefore, we can apply lemma 2 to  $K^{(2)}$  to complete the  $L^1$ -result. The  $L^\infty$ -result follows by verifying that  $\{K^{(n)}(\theta, \cdot), n \in \mathbf{N}\}$  are equicontinuous.  $\square$

Now, for the first time in this section, let  $\mathbf{X} = \{\mathbf{X}^{(n)}, n = 1, 2, \dots\}$  be the sample path of the Gibbs sampler. Then  $\mathbf{X}$  is a Markov chain with the transition kernel

$$K(\theta, \phi) = K(\theta_1, \dots, \theta_p, \phi_1, \dots, \phi_p) = \prod_{i=1}^n \pi(\phi_i | \phi_j, 1 \leq j < i, \theta_j, i < j \leq p) \tag{3.2}$$

with respect to  $p$ -dimensional Lebesgue measure. We shall assume that  $\pi$  is bounded and continuous. Suppose that the interior of the support of  $\pi$  is given by  $R$ .

*Lemma 3.* Suppose that  $K(\cdot, \cdot)$  is bounded and let

$$B_n = B_n(\theta^{(0)}) = \text{int supp}\{K^{(n)}(\theta^{(0)}, \cdot)\};$$

then  $B_n$  is open, and  $B_n \subseteq B_{n+1}, \forall n \in \mathbf{N}$ .

*Proof.* Since  $\pi$  is continuous,  $K(\cdot, \cdot)$  is continuous in both arguments. Moreover by the bounded convergence theorem  $K^{(n)}(\cdot, \cdot)$  is also continuous, and bounded. Therefore  $B_n$  is open in  $R$ . It remains to show that  $\theta \in B_n \Rightarrow \theta \in B_{n+1}$ . However,  $K(\theta, \theta) > 0$  by the form of equation (3.2), since  $\theta \in R$  and  $R$  is open. Therefore,  $\exists$  a ball  $A$  with centre  $\theta$ , such that  $K(\theta, \psi) > \epsilon > 0$  for  $\psi \in A$  by the continuity of  $K$ . Furthermore,  $\exists$  a ball  $B$  with centre  $\theta$  such that  $K^n(\theta^{(0)}, \phi) > \epsilon$  for  $\phi \in B$ . This implies that

$$\begin{aligned} K^{(n+1)}(\theta^{(0)}, \theta) &= \int K^{(n)}(\theta^{(0)}, \phi) K(\phi, \theta) d\theta \\ &\geq \min\{\text{meas}(A), \text{meas}(B)\} \epsilon^2 > 0, \end{aligned}$$

where  $\text{meas}(\cdot)$  denotes Lebesgue measure, as required.  $\square$

We say that any two states  $\theta, \phi \in R$  communicate if  $\exists$  a sequence of points  $\xi^{(0)}, \dots, \xi^{(n)}$  such that  $\xi^{(0)} = \theta^{(0)}, \xi^{(n)} = \phi$ , and where  $\xi^{(i+1)}$  and  $\xi^{(i)}$  differ by only one co-ordinate, and  $\pi(\xi^{(i)}) > 0, \forall i$ . It is easy to show that this is a symmetric relation. Suppose that  $\theta^{(0)} \in R$ .

*Lemma 4.* If all states in  $R$  communicate with  $\theta^{(0)}$ , then  $\cup_{n \geq 1} B_n = R$ .

*Proof.* Directly from the definition of communication, we show that, if  $\theta^{(0)}$  communicates with  $\phi$  in  $n$  steps, then  $K^{(n)}(\theta^{(0)}, \phi) > 0$ .

By induction, suppose that  $\theta^{(0)}$  communicates with  $\xi^{(i)}$  in  $i$  steps, and  $K^{(i)}(\theta^{(0)}, \xi^{(i)}) > 0$ . However, we can write

$$K(\xi^{(i)}, \xi^{(i+1)}) = \prod_{j=1}^p \pi(\xi_k^{(i+1)}, k < j, \xi_k^{(i)}, k \geq j) / \pi(\xi_k^{(i+1)}, k < j, \xi_k^{(i)}, k > j)$$

$$= \pi(\xi^{(i)})^{j^*-1} \pi(\xi^{(i+1)})^{m-j^*+1} / \prod_{j=1}^p \pi(\xi_k^{(i)}, k < j, \xi^{(i)}, k > j) > 0,$$

where  $j^*$  is the co-ordinate which differs between  $\xi^{(i)}$  and  $\xi^{(i+1)}$ . Therefore  $\exists$  a ball  $A$  with centre  $\xi^{(i)}$  such that  $K^{(i)}(\theta^{(0)}, \phi) > \epsilon > 0, \forall \phi \in A$ , and  $K(\theta, \xi^{(i+1)}) > \delta > 0$ , and

$$K^{(n+1)}(\theta^{(0)}, \xi^{(i+1)}) = \int K^{(n)}(\theta^{(0)}, \phi) K(\phi, \xi^{(i+1)}) d\phi \geq \text{meas}(A)\delta^2 > 0,$$

as required. □

*Corollary 2.* Suppose that  $\pi$  is continuous and bounded with compact support. Let  $\theta^{(0)} \in R$  and suppose that all states in  $R$  communicate with  $\theta^{(0)}$ . If  $B_N = R$  for some finite  $N$ , and  $K(\cdot, \cdot)$  is bounded, the Gibbs sampler converges uniformly to  $\pi$  in  $L^1$  and  $L^\infty$  for the initial value  $\theta^{(0)}$ .

*Proof.* Apply theorem 1 to  $K^{(N)}$ . □

*Remark 1.* It is interesting to consider conditions under which the kernel is bounded. However, general conditions are not easy to give. For example, a strange phenomenon can occur near the boundary  $\delta R$ , where the kernel can blow up in either argument. Curiously, unboundedness in the second argument occurs at parts of  $\delta R$  where the Gibbs sampler seems to mix particularly rapidly, i.e. where  $\theta$  achieves  $\sup(\theta_i)$  or  $\inf(\theta_i)$  for exactly one parameter  $\theta_i$ .

Whereas theorem 1 gives uniform ergodicity in any finite domain, a direct use of lemma 2 with the construction of a suitable minorization function is typically more appropriate in applications where  $R$  is rectangular (possibly unbounded in certain co-ordinates). This construction is often accomplished by a compactness argument. We give an applicable corollary and an illustrative example.

*Corollary 3.* The following holds for the Gibbs sampler:  $f_n \xrightarrow{L^1} \pi$  uniformly, if there exist lower semicontinuous non-negative functions  $g_i: \mathbf{R}^i \rightarrow \mathbf{R}^+, 1 \leq i \leq p$ , such that

$$\pi(\phi_i | \theta_j, 1 \leq j < i, \phi_j, i < j \leq p) \geq g_i(\phi_1, \dots, \phi_i)$$

and

$$\{\phi; g_i(\phi_1, \dots, \phi_i) > 0, 1 \leq i \leq p\} \text{ is non-empty.}$$

*Proof.* Just choose  $K^*(\phi) = \prod_{i=1}^p g_i(\phi_1, \dots, \phi_i)$ , and check that  $K^*(\phi)$  is lower semicontinuous, so that its support is (Lebesgue) non-null. □

The following example illustrates this result.

### 3.1. Example

Consider the following exponential family hierarchical model. Suppose that each data point  $y_i$  is drawn from an exponential family density with natural parameter  $\theta_i$ , i.e.

$$f(y_i|\theta_i) = \exp\{y_i\theta_i - a(\theta_i)\}, \quad \theta_i \in \Theta \subset \mathbf{R},$$

with respect to some dominating measure  $\mu_i(\cdot)$ . Here  $1 \leq i \leq k$  and  $a(\cdot)$  is a given function.

Suppose that the joint prior specification  $p(\theta_1, \dots, \theta_k)$  is a mixture of conjugate priors of the form

$$p(\theta_1, \dots, \theta_k) = \prod_{i=1}^k p(\theta_i|a_0, b_0) h(b_0) db_0,$$

$$p(\theta_i|a_0, b_0) \propto \exp[a_0\{b_0\theta_i - a(\theta_i)\}]$$

where the density for  $p(\theta_i|a_0, b_0)$  is with respect to some dominating measure. We assume that  $a_0$  is known, and that the prior density for the hyperparameter  $b_0$  is given by  $h(b_0)$ , which has compact support on the interval  $(\underline{b}_0, \bar{b}_0)$ . We shall show that this is enough to ensure the geometric convergence of the Gibbs sampler for simulating from the posterior distribution  $p(b_0, \theta_1, \dots, \theta_k|y)$ .

The relevant conditional densities required for the implementation of the Gibbs sampler are  $p(b_0|\theta, y)$  and  $p(\theta_i|b_0, \theta_j, j \neq i, y)$ . Now,

$$p(b_0|\theta, y) = \frac{\exp\left(a_0 b_0 \sum \theta_i\right) h(b_0)}{\int \exp\left(a_0 b \sum \theta_i\right) h(b) db}, \quad b_0 \in [\underline{b}_0, \bar{b}_0],$$

which is supported on the whole parameter space:  $(b_0, \theta) \in [\underline{b}_0, \bar{b}_0] \times \Theta^k$ .

By conditional independence,  $p(\theta_i|b_0, \theta_j, j \neq i, y) = p(\theta_i|b_0, y_i)$ , and, by assumption,

$$p(\theta_i|b_0, y_i) = \frac{\exp\{(a_0 b_0 + y_i)\theta_i - (1 + a_0) a(\theta_i)\}}{\int \exp\{(a_0 b_0 + y_i)\theta - (1 + a_0) a(\theta)\} d\theta}.$$

We make the assumption that the integral in the denominator exists for  $b_0 = \underline{b}_0$  and  $\bar{b}_0$ , so that this integral is uniformly bounded above by  $M$  say, for all  $b_0 \in [\underline{b}_0, \bar{b}_0]$ , say. Hence

$$p(\theta_i|y_i, b_0) \geq \frac{1}{M} \exp[-|a_0| \max\{|\underline{b}_0|, |\bar{b}_0|\} |\theta_i| + y_i \theta_i - (1 + a_0) a(\theta_i)].$$

Now, let the right-hand side be the minorization function  $g_i(\theta_1, \dots, \theta_i)$ ,  $1 \leq i \leq k$ . Note that this is supported on the whole parameter space  $\forall i$ . Finally let  $g_{n+1}(\theta, b) = p(b_0|y, \theta)$  and this completes the construction needed for the application of corollary 3. Therefore we have geometric convergence for this class of exponential family hierarchical models.

#### 4. CONCLUSION

We have considered the mechanics of the Gibbs sampler with kernel defined by inequality (3.1) and seen that its geometric convergence fits naturally into the framework of the associated Markov chain. Moreover, regularity conditions for geometric convergence are very naturally satisfied by the Gibbs sampler in many



contexts. However, there are many problems still to be resolved, both practically and theoretically. An important problem is that of obtaining workable estimates for rates of convergence, to construct sensible implementation strategies. Currently, these are only available in certain stylized problems, so that some kind of convergence diagnostic approach is invariably necessary.

It is important to realize that, although we have concentrated on the Gibbs sampler, applications to Metropolis-type constructions (see Metropolis *et al.* (1953), Hastings (1970) and Peskun (1973)) apply equally well.

#### ACKNOWLEDGEMENT

The authors would like to thank the referees for their insightful comments.

#### REFERENCES

- Dunford, N. and Schwartz, J. T. (1958) *Linear Operators: Part I, General Theory*. New York: Interscience.
- Gelfand, A. E. and Smith, A. F. M. (1990) Sampling based approaches to calculating marginal densities. *J. Am. Statist. Ass.*, **85**, 398-409.
- Geman, S. and Geman, D. (1984) Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. Pattn Anal. Mach. Intell.*, **6**, 721-751.
- Hastings, W. K. (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**, 97-109.
- Kemeny, J. G. and Snell, J. L. (1976) *Finite Markov Chains*. New York: Springer.
- Metropolis, N., Rosenbluth, A. W., Teller, A. H. and Teller, E. (1953) Equations of state calculations by fast computing machines. *J. Chem. Phys.*, **21**, 1087-1092.
- Nummelin, E. (1984) *General Irreducible Markov Chains and Non-negative Operators*. Cambridge: Cambridge University Press.
- Peskun, P. H. (1973) Optimum Monte Carlo sampling using Markov chains. *Biometrika*, **30**, 607-612.
- Seneta, E. (1981) *Non-negative Matrices and Markov Chains*. London: Allen and Unwin.
- Tanner, T. A. and Wong, W. H. (1987) The calculation of posterior distribution by data augmentation (with discussion). *J. Am. Statist. Ass.*, **82**, 528-550.
- Tierney, L. (1991) Exploring posterior distributions using Markov chains. *Technical Report*. University of Minnesota, Minneapolis.