# Variable-at-a-time Implementations of Metropolis-Hastings

Ronald C. Neath
Department of Statistics and Computer Information Systems
Zicklin School of Business
Baruch College, City University of New York
ronald.neath@baruch.cuny.edu

Galin L. Jones
School of Statistics
University of Minnesota
galin@stat.umn.edu

March 4, 2009

**Abstract**

It is common practice in Markov chain Monte Carlo to update a high-dimensional chain one variable (or sub-block of variables) at a time, rather than conduct a single block update. While this modification can make the choice of proposal easier, the theoretical convergence properties of the associated Markov chain have received limited attention. We present conditions under which the chain converges uniformly to its stationary distribution at a geometric rate. Also, we develop a recipe for performing regenerative simulation in this setting and demonstrate its application for estimating Markov chain Monte Carlo standard errors. In both our investigation of convergence rates and in Monte Carlo standard error estimation we pay particular attention to the case with state-independent component-wise proposals. We illustrate our results in two examples, a toy Bayesian inference problem and a practically relevant example involving maximum likelihood estimation for a generalized linear mixed model.

# 1 Introduction

Let $\pi$ be a probability distribution having support $\mathsf{X}$. The canonical Markov chain Monte Carlo (MCMC) method for making draws from $\pi$ is the Metropolis-Hastings algorithm, described here. Let $X^{(k)} = x$ denote the current state, and suppose the stationary distribution $\pi$ has a density with respect to some reference measure $\mu$ (often Lebesgue or counting measure or their product), which we will also denote by $\pi$. Let $p(\cdot, \cdot)$ denote the user-defined proposal kernel density. The updated state $X^{(k+1)}$ is obtained via

1. Simulate $x^*$ from proposal density $p(x, \cdot)$
2. Calculate acceptance probability $\alpha(x, x^*)$, where
$$\alpha(x, y) = \min\left\{1, \frac{\pi(y)}{\pi(x)}\frac{p(y, x)}{p(x, y)}\right\}$$
3. Set
$$X^{(k+1)} = \begin{cases} x^* & \text{with probability} \quad \alpha(x, x^*) \\ x & \text{with probability} \quad 1 - \alpha(x, x^*) \end{cases}$$

Thus the choice of a Metropolis-Hastings sampler boils down to choosing a proposal density kernel $p(\cdot, \cdot)$. One common choice is to use a proposal kernel that satisfies $p(x, y) = p(y, x)$ in which case this is called a Metropolis algorithm. If, further, $p(x, y) = p(x - y) = p(y - x)$ for all $x$ and $y$, the sampler is called a Metropolis random walk. Another common choice of candidate is a proposal $p(\cdot)$ that does not depend on the current state of the chain, that is, the Metropolis-Hastings independence sampler (MHIS).

In practice, the selection of a candidate distribution can be a challenging proposition, particularly in problems where the state space has high dimension. This has led to investigation of optimal scaling of Metropolis algorithms and so-called adaptive algorithms which allow the proposal kernel to change over the course of the simulation (see, for example, Bédard and Rosenthal, 2008; Rosenthal, 2008, and the references therein). An alternative approach is to, rather than update the chain as a single block, update one variable (or sub-block of variables) at a time. The choice between these two strategies is frequently unclear (see, e.g., Roberts and Sahu, 1997), although a general guideline seems to be that updating as a single block may not be advantageous if the components of $\pi$ are only weakly correlated. Further, by breaking a high-dimensional simulation problem into several smaller-dimensional problems, the component-wise approach can make the choice of candidate distributions much easier.

One hopes that whatever version of Metropolis-Hastings is used, the simulation will relatively quickly produce a representative sample from the target population $\pi$. Thus an important consideration in MCMC is the rate of convergence of the chain to its stationary distribution. Specificity requires notation. Let $\mathcal{B}$ be the Borel $\sigma$-algebra on $\mathsf{X}$ and $P^n(x, dy)$ denote the $n$-step Markov transition kernel, that is, for any $x \in \mathsf{X}$, $A \in \mathcal{B}$, and $n \in \mathbb{Z}^+$, $P^n(x, A) = \Pr(X^{(n+j)} \in A | X^{(j)} = x)$ for the Markov chain $\Phi = \left\{X^{(0)}, X^{(1)}, X^{(2)}, \ldots\right\}$. If the chain is Harris ergodic (i.e., aperiodic, $\pi$-irreducible, and positive Harris recurrent), then, as $n \to \infty$, $||P^n(x, \cdot) - \pi(\cdot)|| \to 0$ where $\|\cdot\|$ denotes the total variation norm. We will consider the rate of this convergence in the following way. Suppose there exist a real-valued function $M(x)$ on $\mathsf{X}$ and $0 < t < 1$ such that

$$||P^n(x, \cdot) - \pi(\cdot)|| \leq M(x)t^n \ . \tag{1}$$

When $M$ is bounded say $\Phi$ is uniformly ergodic and otherwise say it is geometrically ergodic.

Harris ergodicity is well-understood for Metropolis-Hastings chains. Tierney (1994) showed that a Metropolis-Hastings chain with block updates is Harris ergodic under weak conditions while Chan and Geyer (1994) and Roberts and Rosenthal (2006) found conditions under which variable-at-a-time chains are Harris recurrent. Also, much work has been done on establishing (1) for various versions of Metropolis-Hastings. For example, Tierney (1994) and Mengersen and Tweedie (1996) showed that the MHIS sampler is uniformly ergodic if there exists $\epsilon > 0$ such that $p(x) \geq \epsilon\pi(x)$ for all $x \in \mathsf{X}$. Mengersen and Tweedie (1996) further proved that if $\operatorname{ess\,inf}\{p(x)/\pi(x)\} = 0$ in $\pi$-measure, the resulting MHIS is not even geometrically ergodic. Moreover, Mengersen and Tweedie (1996) proved that the Metropolis random walk on $\mathbb{R}^d$ cannot be uniformly ergodic. However, Christensen et al. (2001); Jarner and Hansen (2000); Mengersen and Tweedie (1996); Roberts and Tweedie (1996) have established conditions under which Metropolis yields a geometrically ergodic chain.

Note well that none of these convergence rate results apply to variable-at-a-time implementations (though Fort et al., 2003; Roberts and Rosenthal, 1998, have shown that random scan Metropolis random walks can be geometrically ergodic). Thus we begin by considering uniform ergodicity of such samplers in Section 2, eventually focusing on a variable-at-a-time version of the Metropolis-Hastings algorithm with state-independent candidate distributions. In particular, we find conditions under which the resulting chain converges uniformly to its stationary distribution at a geometric rate and apply them to a toy Bayesian example and a practically relevant example involving maximum likelihood estimation for a generalized linear mixed model. We also provide an empirical comparison of the variable-at-a-time sampler with the Metropolis random walk and the MHIS.

In addition to generally ensuring the rapid convergence required for useful simulation, (1) is also a key sufficient condition for calculating asymptotically valid Monte Carlo standard errors (Flegal et al., 2008; Flegal and Jones, 2008; Hobert et al., 2002; Jones et al., 2006). One of the most common goals in MCMC is to evaluate the quantity $E_\pi g = \int_{\mathsf{X}} g(x)\pi(dx)$ where $\pi$ is a probability distribution with support $\mathsf{X}$ and $E_\pi|g| < \infty$. In Markov chain Monte Carlo (MCMC) integration, $E_\pi g$ is approximated by the ergodic average $\bar{g}_n = n^{-1}\sum_{i=0}^{n-1} g(X^{(i)})$ on a partial realization of a Markov chain having $\pi$ as its stationary distribution. The use of $\bar{g}_n$ is usually justified through Birkhoff's ergodic theorem. Now (1) along with a moment condition on $g$ ensures the existence of a central limit theorem for the Monte Carlo error, i.e., there exists $0 < \sigma_g^2 < \infty$ such that as $n \to \infty$,

$$\sqrt{n}(\bar{g}_n - E_\pi g) \xrightarrow{d} \mathrm{N}(0, \sigma_g^2) \; .$$

The condition (1) also ensures that we can use the method of regenerative simulation to construct asymptotically valid MCMC standard errors. Regenerative simulation presents some complications not seen in block sampling algorithms, thus we develop a recipe for regeneration in variable-at-a-time implementations. Although we do not pursue these applications here, our algorithm for implementing regeneration also can be used to choose reasonable starting values (Hobert et al., 2006) or even as the basis of an adaptive component-wise MCMC sampler, perhaps following the basic recipes of Gilks et al. (1998) or Brockwell and Kadane (2005).

The remainder of this paper is organized as follows. In Section 2 we introduce variable-at-a-time implementations, consider convergence rates for two general algorithms, and develop conditions for the uniform convergence of the variable-at-a-time sampler with state-independent proposals. In Section 3 we introduce the method of regenerative simulation for estimating Monte Carlo standard errors, and demonstrate its implementation for variable-at-a-time samplers. In Section 4 we consider two examples, a toy problem in Bayesian inference and a logit-normal generalized linear mixed model (GLMM).

## 2 Convergence rates under variable-at-a-time updates

### 2.1 Basic Technique: Minorization

There are constructive techniques for verifying the existence of an appropriate $M$ and $t$ in (1); see Meyn and Tweedie (1993, Chapter 15). Following is a method for establishing uniform ergodicity. Suppose there exist a positive integer $n_0$, a number $\epsilon > 0$, a set $C \in \mathcal{B}$, and a

probability measure $Q$ on $\mathcal{B}$ such that

$$P^{n_0}(x, A) \geq \epsilon Q(A) \quad \text{for all} \quad x \in C, \ A \in \mathcal{B}. \tag{2}$$

We then say that a *minorization condition* holds on the set $C$, called a *small set*. If (2) holds with $C = \mathsf{X}$, then $\Phi$ is uniformly ergodic and $\|P^n(x, \cdot) - \pi(\cdot)\| \leq (1-\epsilon)^{\lfloor n/n_0 \rfloor}$ where $\lfloor \cdot \rfloor$ is the greatest integer function.

## 2.2 General variable-at-a-time updates

Suppose $\mathsf{X} = \mathsf{X}_1 \times \cdots \times \mathsf{X}_d$ with Borel $\sigma$-algebra $\mathcal{B}$. We allow each $\mathsf{X}_i \subseteq \mathbb{R}^{b_i}$ so that the total dimension is $b_1 + \cdots + b_d$. Under a variable-at-a-time, or *component-wise*, implementation of Metropolis-Hastings, the update of $X^{(k)} = x = (x_1, \ldots, x_d)$, where $x_i \in \mathsf{X}_i$, to $X^{(k+1)}$, is obtained as follows. Letting $i$ index the component of the chain due for update, a candidate $x_i^*$ is drawn from $p_i(x, \cdot)$, a proposal density on $\mathsf{X}_i$. Then, if $x^* = (x_1, \ldots, x_{i-1}, x_i^*, x_{i+1}, \ldots, x_d)$, the acceptance probability for the $i$th update is given by

$$\alpha_i(x, x_i^*) = \min\left\{1, \frac{\pi(x^*)}{\pi(x)} \frac{p_i(x^*, x_i)}{p_i(x, x_i^*)}\right\}.$$

Two ways to combine these component-wise updates involve *composition* and *simple mixing*. Before describing these we require more notation. We will continue to use a subscript to indicate the position of a vector component and a parenthetical superscript to indicate the step in a Markov chain, so $X^{(k)} = x = (x_1, \ldots, x_d)$, where $x_i \in \mathsf{X}_i$, denotes the current state of the chain. For each $i = 1, \ldots, d$ let $x_{[i]} = (x_1, \ldots, x_i)$ and $x^{[i]} = (x_i, \ldots, x_d)$; let $x_{[0]}$ and $x^{[d+1]}$ be null (vectors of dimension 0). Next, $f(u, v|w)$ denotes a Markov transition density (MTD) if it is a density in $v$ conditional on $u$ and $w$; the reader should think of a Markov chain moving from $u$ to $v$ possibly conditionally on some other variable $w$.

Composition of the component-wise updates corresponds to deterministically cycling through them one at a time. Thus the MTD for a move from $X^{(k)} = x$ to $X^{(k+1)} = y$ is

$$f_{comp}(x, y) = \prod_{i=1}^{d} f_i(x_i, y_i \mid y_{[i-1]}, x^{[i+1]}) \tag{3}$$

where $f_i(x_i, \cdot \mid y_{[i-1]}, x^{[i+1]})$ is the MTD for a Metropolis-Hastings algorithm on $\mathsf{X}_i$ having, given current state $x_i$, proposal density $p_i((y_{[i-1]}, x^{[i]}), \cdot)$, and acceptance probability

$$\alpha_i((y_{[i-1]}, x^{[i]}), y_i) = \min\left\{1, \frac{\pi(y_{[i]}, x^{[i+1]})}{\pi(y_{[i-1]}, x^{[i]})} \frac{p_i((y_{[i]}, x^{[i+1]}), x_i)}{p_i((y_{[i-1]}, x^{[i]}), y_i)}\right\}. \tag{4}$$

Our first result establishes general conditions for uniform ergodicity of the Markov chain determined by $f_{comp}$.

**Theorem 1.** *Suppose there exist positive constants $\varepsilon_i$ and positive functions $q_i$ on $\mathsf{X}_1 \times \cdots \times \mathsf{X}_i$ such that*

$$f_i(x_i, y_i | y_{[i-1]}, x^{[i+1]}) \geq \varepsilon_i q_i(y_{[i]})$$

*for all $x^{[i]}$ and $y_{[i]}$, for each $i = 1, \ldots d$. Then the Markov chain corresponding to $f_{comp}$ is uniformly ergodic. Moreover, if we let*

$$C = \int_{\mathsf{X}} q_1(x_1) q_2(x_1, x_2) \cdots q_d(x_1, \ldots, x_d) \mu(dx),$$

*then after $n$ iterations the total variation distance to stationarity is bounded above by $(1 - C\varepsilon_1 \cdots \varepsilon_d)^n$.*

*Proof.* We will establish (2) for $n_0 = 1$. Now

$$f_{comp}(x, y) = \prod_{i=1}^{d} f_i(x_i, y_i | y_{[i-1]}, x^{[i+1]}) \geq \prod_{i=1}^{d} \varepsilon_i q_i(y_{[i]}) \; .$$

Let $P_{comp}$ denote the one-step transition kernel corresponding to the MTD $f_{comp}$. Then

$$P_{comp}(x, A) \geq \varepsilon_1 \cdots \varepsilon_d \; C \; Q(A),$$

where

$$Q(A) = \frac{1}{C} \int_A q_1(x_{[1]}) q_2(x_{[2]}) \cdots q_d(x_{[d]}) \mu(dx) \; .$$

□

In simple mixing, each update consists of just one randomly selected component-wise update, where the component selection probabilities do not depend on the state of the chain. Let $r_i > 0$ and $\sum_{i=1}^{d} r_i = 1$, we can then write the MTD for the transition $X^{(k)} = x$ to $X^{(k+1)} = y$ as

$$f_{mix}(x, y) = \sum_{i=1}^{d} r_i f_i(x_i, y_i | x_{(-i)}) I(y_{(-i)} = x_{(-i)}) \tag{5}$$

where $x_{(-i)} = x \setminus x_i$ and similarly for $y_{(-i)}$.

For the remainder of this paper we will focus on composition. This turns out to be sufficient because, as the following theorem asserts, if a component-wise algorithm with updates combined by composition is uniformly ergodic, then the sampler is uniformly ergodic under simple mixing as well.

**Theorem 2.** *Suppose there exists $\epsilon > 0$ and a probability measure $Q$ such that $P_{comp}(x, A) \geq \epsilon Q(A)$ for all $x \in \mathsf{X}$ and $A \in \mathcal{B}$, and thus the composition sampler is uniformly ergodic. Then*

*the corresponding simple mixing algorithm (with the same component-wise proposal kernels) is uniformly ergodic as well. Moreover, if* $r_1, \ldots, r_d$ *denote the component selection probabilities, then after* $n \times d$ *iterations the total variation distance to stationarity is bounded above by* $(1 - \epsilon\ r_1 \cdots r_d)^n$.

*Proof.* Let $f^k_{mix}$ denote the $k$-step MTD of the mixing algorithm, given by (5) for $k = 1$. Now,

$$\begin{aligned} f^d_{mix}(x, y) &\geq \prod_{i=1}^{d} f_{mix}\left((y_{[i-1]}, x^{[i]}), (y_{[i]}, x^{[i+1]})\right) \\ &= \prod_{i=1}^{d} r_i f_i\left(x_i, y_i | (y_{[i-1]}, x^{[i+1]})\right) \\ &= \prod_{i=1}^{d} r_i\ f_{comp}(x, y) \end{aligned}$$

where the inequality follows from the fact that the right-hand side only accounts for the case that the updates occur in the order $1, 2, \ldots, d$; the first equality follows from the definition of $f_{mix}$; and the second equality follows from the definition of $f_{comp}$.

If we let $P_{mix}$ denote the corresponding Markov transition kernel, we have

$$P^d_{mix}(x, A) \geq \prod_{i=1}^{d} r_i\ P_{comp}(x, A) \geq \epsilon \prod_{i=1}^{d} r_i\ Q(A)\ .$$

□

*Remark* 1. Consider the upper bounds on the total variation distance in both theorems, specifically, $(1 - C\varepsilon_1 \cdots \varepsilon_d)^n$ for $f_{comp}$ and $(1 - \epsilon\ r_1 \cdots r_d)^n$ for $f^d_{mix}$. Notice that as $d \to \infty$ both of these bounds tend to 1. Also, in Theorem 2 the assumed minorization for $P_{comp}$ implies a total variation upper bound of $(1 - \epsilon)^n$. Of course, the upper bound for $P_{mix}$ is larger, that is, $(1 - \epsilon)^n < (1 - \epsilon\ r_1 \cdots r_d)^n$, but since these are just upper bounds on the total variation distance we caution against using them to compare the two Markov chains.

*Remark* 2. Roberts and Rosenthal (1997, Proposition 3.2) proved a result that has a similar flavor. Specifically, they showed that if the deterministically updated Gibbs sampler is uniformly ergodic, then so is the so-called random scan Gibbs sampler when the component-wise selection probabilities are equal. Note that equal selection probabilities $r_i \equiv 1/d$ achieves the tightest upper bound available from Theorem 2 by maximizing the product $r_1 \cdots r_d$.

### 2.3 Component-wise independence sampler

In a component-wise independence sampler (CWIS), the component-wise proposal kernels $p_i(x, \cdot)$ do not depend on the current state of the chain $x$. Again letting $i$ index the component of the chain due for update, a proposal $x_i^*$ is drawn from $p_i(\cdot)$, a density on $\mathsf{X}_i$. Then, if $x^* = (x_1, \ldots, x_{i-1}, x_i^*, x_{i+1}, \ldots, x_d)$, the acceptance probability for the $i$th update is given by

$$\alpha_i(x, x_i^*) = \min\left\{1, \frac{\pi(x^*)}{\pi(x)} \frac{p_i(x_i)}{p_i(x_i^*)}\right\}.$$

We are interested in establishing conditions under which the CWIS is uniformly ergodic. Clearly the conditions of Theorem 1 may be difficult to verify in practical applications. Further, as a CWIS update will likely be some combination of accepted and rejected component-wise proposals, the CWIS is not truly an independence sampler at all, and thus the results from Mengersen and Tweedie (1996) are not directly applicable. It is tempting to think that extending their work to component-wise updates will be straightforward. If we let $p(x) = \prod_{i=1}^d p_i(x_i)$, a density on $\mathsf{X}$, is the existence of $\epsilon > 0$ such that $p(x) \geq \epsilon\pi(x)$ a sufficient condition for uniform ergodicity of CWIS? It is not at all clear that it is. Attempts to generalize Mengersen and Tweedie's (1996) argument break down with the proliferation of cases to consider. In Theorem 3 we give a pair of conditions that together are sufficient for uniform ergodicity of CWIS.

Recall the notation previously defined. We use a subscript to indicate the position of a vector component and a parenthetical superscript to indicate the step in a Markov chain, so $X^{(k)} = x = (x_1, \ldots, x_d)$, where $x_i \in \mathsf{X}_i$, denotes the current state of the chain. For each $i = 1, \ldots, d$ let $x_{[i]} = (x_1, \ldots, x_i)$ and $x^{[i]} = (x_i, \ldots, x_d)$; let $x_{[0]}$ and $x^{[d+1]}$ be null (vectors of dimension 0). Now an explicit statement of the CWIS update rule for $X^{(k)} = x$ is

1. For each $i = 1, \ldots, d$,
   (a) Simulate $x_i^* \sim p_i(\cdot)$
   (b) Calculate
   $$\alpha_i((y_{[i-1]}, x^{[i]}), x_i^*) = \min\left\{1, \frac{\pi(y_{[i-1]}, x_i^*, x^{[i+1]})}{\pi(y_{[i-1]}, x_i, x^{[i+1]})} \frac{p_i(x_i)}{p_i(x_i^*)}\right\}$$
   (c) Set $y_i = \begin{cases} x_i^* & \text{with probability} \quad \alpha_i \\ x_i & \text{with probability} \quad 1 - \alpha_i \end{cases}$
2. Set $X^{(k+1)} = (y_1, \ldots, y_d)$.

**Theorem 3.** *Consider a CWIS on state space* $\mathsf{X} = \mathsf{X}_1 \times \cdots \times \mathsf{X}_d$ *with target density* $\pi$ *and proposal densities* $p_i$ *for* $i = 1, \ldots, d$. *Define* $p(x) := \prod_{i=1}^d p_i(x_i)$, *a density on* $\mathsf{X}$, *and suppose* $p(x) = 0$ *if and only if* $\pi(x) = 0$. *Further suppose there exists* $\delta > 0$ *such that* $p(x) \geq \delta\pi(x)$ *for all* $x \in \mathsf{X}$. *Finally, suppose there exists* $\varepsilon > 0$ *such that for any* $x, y \in \mathsf{X}$ *with* $\pi(x) > 0$ *and* $\pi(y) > 0$,

$$\pi(x)\pi(y) = \pi(x_{[i-1]}, x^{[i]})\pi(y_{[i-1]}, y^{[i]}) \geq \varepsilon\pi(x_{[i-1]}, y^{[i]})\pi(y_{[i-1]}, x^{[i]}) > 0 \tag{6}$$

*for each* $i = 1, \ldots, d-1$. *If we let* $P_{cwis}$ *denote the Markov transition kernel, then, for any* $x \in \mathsf{X}$ *and* $A \in \mathcal{B}(\mathsf{X})$,

$$P_{cwis}(x, A) \geq \delta\varepsilon^{\lfloor d/2 \rfloor}\pi(A)$$

*where* $\lfloor \cdot \rfloor$ *denotes the greatest integer function, and thus the Markov chain is uniformly ergodic. Moreover, the total variation distance to stationarity after* $n$ *iterations is bounded by* $(1 - \delta\varepsilon^{\lfloor d/2 \rfloor})^n$.

*Proof.* See Appendix A. □

*Remark* 3. As in Section 2.2 we see that the total variation upper bound for the CWIS approaches 1 as $d \to \infty$. Moreover, the upper bound for CWIS is larger than that of the MHIS which is $(1 - \delta)^n$ in the notation of the theorem. However, we will encounter an example in Section 4.2 where, despite the larger upper bound, CWIS is convincingly better than the MHIS.

The following two corollaries indicate settings under which (6) is easily verified.

**Corollary 1.** *Consider a CWIS on state space* $\mathsf{X} = \mathsf{X}_1 \times \cdots \times \mathsf{X}_d$ *with target density* $\pi$ *and proposal densities* $p_i$ *for* $i = 1, \ldots, d$. *Define* $p(x) := \prod_{i=1}^d p_i(x_i)$, *a density on* $\mathsf{X}$, *and suppose* $p(x) = 0$ *if and only if* $\pi(x) = 0$. *Further suppose there exists* $\delta > 0$ *such that* $p(x) \geq \delta\pi(x)$ *for all* $x \in \mathsf{X}$. *Finally, suppose there exist pairs of functions* $g_i$ *and* $h_i$ *on* $\mathsf{X}_i$ *for* $i = 1, \ldots, d$ *such that*

$$\prod_{i=1}^d g_i(x_i) \leq \pi(x) \leq \prod_{i=1}^d h_i(x_i) \tag{7}$$

*for any* $x \in \mathsf{X}$, *and* $\sup_{x_i \in \mathsf{X}_i} \{h_i(x_i)/g_i(x_i)\} < \infty$ *for each* $i = 1, \ldots, d$. *Then the Markov chain is uniformly ergodic.*

*Proof.* We need only show that (7) implies (6). Let $\rho_i = \inf_{x_i \in \mathsf{X}_i} \{g_i(x_i)/h_i(x_i)\}$ for $i = 1, \ldots, d$; by assumption each $\rho_i > 0$. Then, for any $x, y \in \mathsf{X}$ and each $i = 1, \ldots, d$ we have

$$\frac{\pi(x_{[i-1]}, x^{[i]})\pi(y_{[i-1]}, y^{[i]})}{\pi(x_{[i-1]}, y^{[i]})\pi(y_{[i-1]}, x^{[i]})} \;\geq\; \prod_{j=1}^{i-1} \frac{g_j(x_j)g_j(y_j)}{h_j(x_j)h_j(y_j)} \prod_{j=i}^{d} \frac{g_j(y_j)g_j(x_j)}{h_j(y_j)h_j(x_j)} \;\geq\; \prod_{j=1}^{d} \rho_j^2 \,.$$

Thus (6) holds with $\varepsilon = (\rho_1 \cdots \rho_d)^2$. □

*Remark* 4. An immediate consequence of Corollary 1 is that if the target density $\pi$ can be expressed as a product of $d$ densities (each $\pi_i$ a density on $\mathsf{X}_i$), that is, if the components of a random vector $X \sim \pi$ are mutually independent, a CWIS with $p(x) \geq \delta\pi(x)$ is uniformly ergodic. Further, since (6) holds for $\varepsilon = 1$ in this case, the upper bound on total variation distance to stationarity is $(1-\delta)^n$, the same bound as can be obtained for the MHIS.

**Corollary 2.** *Consider a CWIS on state space $\mathsf{X} = \mathsf{X}_1 \times \cdots \times \mathsf{X}_d$ with target density $\pi$ and proposal densities $p_i$ for $i = 1, \ldots, d$. Define $p(x) := \prod_{i=1}^d p_i(x_i)$, a density on $\mathsf{X}$, and suppose $p(x) = 0$ if and only if $\pi(x) = 0$. If there exist $0 < a \leq b < \infty$ and $c < \infty$ such that $a \leq \pi(x) \leq b$ and $p(x) \geq c$ for $\pi$-almost all $x$, then the chain is uniformly ergodic.*

*Proof.* The conditions of Theorem 3 hold with $\delta = c/b$ and $\varepsilon = (a/b)^2$. □

*Remark* 5. Although, in the interest of brevity, we have not stated it here, using Theorem 2 it is straightforward to see that the corresponding simple mixing algorithm with state-independent proposals is uniformly ergodic under the conditions of Theorem 3 and Corollaries 1 and 2.

### 2.4 Metropolised Gibbs

In a Gibbs sampler, the update of $X^{(k)} = x = (x_1, \ldots, x_d)$, where each $x_i \in \mathsf{X}_i$, to $X^{(k+1)} = y = (y_1, \ldots, y_d)$ is conducted by, for each $i = 1, \ldots, d$, taking $y_i$ to be a draw from the conditional density $\pi_i(\cdot|y_{[i-1]}, x^{[i+1]})$. If the means to simulate from some or all of the conditional densities $\pi_i(x_i|x_{(-i)})$ are unavailable, one might apply the Metropolis-Hastings idea to the component-wise updates. Letting $p_i((y_{[i-1]}, x^{[i]}), \cdot)$ denote the proposal density for the $i$th component, this "Metropolised" Gibbs algorithm proceeds by, for each $i = 1, \ldots, d$,

1. Simulate $x_i^* \sim p_i((y_{[i-1]}, x^{[i]}), \cdot)$
2. Calculate
$$\alpha_i = \min\left\{1, \frac{\pi_i(x_i^*|y_{[i-1]}, x^{[i+1]})}{\pi_i(x_i|y_{[i-1]}, x^{[i+1]})} \frac{p_i((y_{[i-1]}, x_i^*, x^{[i+1]}), x_i)}{p_i((y_{[i-1]}, x^{[i]}), x_i^*)}\right\} \tag{8}$$
3. Set
$$y_i = \begin{cases} x_i^* & \text{with probability} \quad \alpha_i \\ x_i & \text{with probability} \quad 1-\alpha_i \end{cases}$$

Of course, this approach is identical to the composition approach to variable-at-a-time updates of Section 2.2.

**Theorem 4.** *The Metropolised Gibbs sampler with Metropolis-Hastings proposal densities* $p_i((y_{[i-1]}, x^{[i]}), \cdot)$, $i = 1, \ldots, d$, *is equivalent to* $f_{comp}$ *with proposal densities* $p_i((y_{[i-1]}, x^{[i]}), \cdot)$.

*Proof.* Proposal densities are identical, so we need only show that acceptance probabilities are equivalent (which they must be, as both algorithms are constructed to have the same invariant distribution). If we let $x^* = (x_1, \ldots, x_{i-1}, x_i^*, x_{i+1}, \ldots, x_d)$, we have

$$\frac{\pi(x^*)}{\pi(x)} = \frac{\pi_{-i}(x_{(-i)})\pi_i(x_i^*|x_{(-i)})}{\pi_{-i}(x_{(-i)})\pi_i(x_i|x_{(-i)})} = \frac{\pi_i(x_i^*|x_{(-i)})}{\pi_i(x_i|x_{(-i)})}$$

and the equivalence of (4) and (8) follows immediately. □

# 3 Regenerative simulation

Here we consider a generalization to the minorization condition introduced in Section 2.1, and assume it holds for $n_0 = 1$. Specifically, we suppose there exists a function $s : \mathsf{X} \to [0, 1]$ with $E_\pi s > 0$ and a probability measure $Q$ on $\mathcal{B}$ such that

$$P(x, A) \geq s(x)Q(A) \quad \text{for all } x \in \mathsf{X} \text{ and } A \in \mathcal{B}\,. \tag{9}$$

The method of *regenerative simulation* (RS) requires simulation of the split chain $\left\{(X^{(n)}, \delta^{(n)})\right\}$, generated as follows. Given $X^{(k)} = x$, find $X^{(k+1)}$ and $\delta^{(k)}$ by

1. Simulate $X^{(k+1)} \sim P(x, \cdot)$
2. Simulate $\delta^{(k)}$, a Bernoulli random variable with success probability $r(X^{(k)}, X^{(k+1)})$ where
$$r(x, y) = \frac{s(x)Q(dy)}{P(x, dy)} \tag{10}$$
denotes the conditional probability of regeneration given a jump from $x$ to $y$ in the $\left\{X^{(n)}\right\}$ chain.

By construction, the sub-chain $\left\{X^{(n)}\right\}$ has Markov transition kernel given by $P$. Also, the set of $n$ for which $\delta^{(n-1)} = 1$, called *regeneration times*, represent times at which the chain "probabilistically restarts itself." An important application of regenerative simulation is Monte Carlo standard error estimation, as described in Appendix B; see any of Hobert et al. (2002); Jones et al. (2006); Jones and Hobert (2001); Mykland et al. (1995) for further details.

Now consider a variable-at-a-time Metropolis-Hastings algorithm having transition density defined at (3), i.e.,

$$f_{comp}(x, y) = \prod_{i=1}^{d} f_i(x_i, y_i \mid y_{[i-1]}, x^{[i+1]})$$

where each $f_i(x_i, \cdot \mid y_{[i-1]}, x^{[i+1]})$ is an MTD with proposal density $p_i((y_{[i-1]}, x^{[i]}), \cdot\,)$. Assume that for $i = 1, \ldots, d$ there exist positive functions $s_i$ on $\mathsf{X}_i \times \cdots \times \mathsf{X}_d$ and $q_i$ on $\mathsf{X}_1 \times \cdots \times \mathsf{X}_i$ such that

$$p_i((y_{[i-1]}, x^{[i]}), y_i) \geq s_i(x^{[i]}) q_i(y_{[i]}) \,. \tag{11}$$

Next, assume that there are positive functions $w_i$ for $i = 1, \ldots, d$ such that

$$w_i(x) p_i(x, y_i) = w_i(y) p_i(y, x_i) \quad \text{for all} \;\; x, y \in \mathsf{X} \,. \tag{12}$$

*Remark* 6. For the CWIS note that (11) holds with $s_i(x^{[i]}) = 1$ and $q_i(y_{[i]}) = p_i(y_i)$ and (12) holds with $w_i(x) = p_i(x_i)$.

Set $r_i(x) = \pi(x)/w_i(x)$ and also suppose there exists a set of functions $g_{i1}, g_{i2}, h_{i1}, h_{i2}$ for $i = 1, \ldots, d$ such that for any $x$ with $\pi(x) > 0$ we have

$$0 \leq g_{i1}(x_{[i]}) g_{i2}(x^{[i+1]}) \leq r_i(x) \leq h_{i1}(x_{[i-1]}) h_{i2}(x^{[i]}) \text{ for each } i = 1, \ldots, d \tag{13}$$

and assume further that

$$E_\pi g_{i1} g_{i2} > 0 \text{ for each } i = 1, \ldots, d \,. \tag{14}$$

*Remark* 7. Consider the conditions (13) and (14), which are not as restrictive as they may at first appear. Indeed these conditions hold quite generally for bounded target densities. For example, consider the CWIS setting and take $w_i(x) = p_i(x_i)$ as above. For each $i = 1, \ldots, d-1$, select a point $\tilde{x}_{[i]} \in \mathsf{X}_1 \times \cdots \times \mathsf{X}_i$, and a set $D_i \subset \mathsf{X}_{i+1} \times \cdots \times \mathsf{X}_d$ such that $\inf\left\{\pi(x) : x^{[i+1]} \in D_i\right\} > 0$. Then take

$$g_{i1}(x_{[i]}) = \frac{1}{p_i(x_i)} \inf_{x^{[i+1]} \in D_i} \left[ \frac{\pi(x_{[i]}, x^{[i+1]})}{\pi(\tilde{x}_{[i]}, x^{[i+1]})} \right] ,$$

and

$$g_{i2}(x^{[i+1]}) = \pi(\tilde{x}_{[i]}, x^{[i+1]}) I_{D_i}(x^{[i+1]}) \,.$$

Assume $\sup \pi(x) < \infty$. We can take

$$h_{i1}(x_{[i-1]}) = \sup_{x^{[i]} \in X_i \times \cdots \times X_d} \pi(x_{[i-1]}, x^{[i]})$$

and $h_{i2}(x^{[i]}) = 1/p_i(x_i)$, and (13) holds. Finally, (14) holds by definition of the $D_i$.

We now show that (11)–(14) together are sufficient to guarantee a minorization. First, note that

$$f_{comp}(x, y) \geq \prod_{i=1}^{d} p_i((y_{[i-1]}, x^{[i]}), y_i) \alpha_i((y_{[i-1]}, x^{[i]}), y_i)$$

as the right-hand side only accounts for the case where every component-wise update proposal is accepted. Also,

$$\begin{aligned}
\alpha_i((y_{[i-1]}, x^{[i]}), y_i) &= \min\left\{1, \frac{\pi(y_{[i]}, x^{[i+1]})}{\pi(y_{[i-1]}, x^{[i]})} \frac{p_i((y_{[i]}, x^{[i+1]}), x_i)}{p_i((y_{[i-1]}, x^{[i]}), y_i)}\right\} \\
&= \min\left\{1, \frac{\pi(y_{[i]}, x^{[i+1]})}{w_i(y_{[i]}, x^{[i+1]})} \frac{w_i(y_{[i-1]}, x^{[i]})}{\pi(y_{[i-1]}, x^{[i]})}\right\} \\
&= \min\left\{1, \frac{r_i(y_{[i]}, x^{[i+1]})}{r_i(y_{[i-1]}, x^{[i]})}\right\} \\
&\geq \min\left\{1, \frac{g_{i1}(y_{[i]}) g_{i2}(x^{[i+1]})}{h_{i1}(y_{[i-1]}) h_{i2}(x^{[i]})}\right\} \\
&\geq \min\left\{1, \frac{c_i g_{i1}(y_{[i]})}{h_{i1}(y_{[i-1]})}\right\} \min\left\{1, \frac{g_{i2}(x^{[i+1]})}{c_i h_{i2}(x^{[i]})}\right\} ,
\end{aligned}$$

where the equality follows from (12), the first inequality follows from (13) and the second is due to the fact that for any nonnegative real numbers $a$ and $b$, $\min\{1, ab\} \geq \min\{1, a\} \min\{1, b\}$, and $c_i$ is any positive constant. Now (11) implies

$$f_{comp}(x, y) \geq \prod_{i=1}^{d} s_i(x^{[i]}) \min\left\{1, \frac{g_{i2}(x^{[i+1]})}{c_i h_{i2}(x^{[i]})}\right\} q_i(y_{[i]}) \min\left\{1, \frac{c_i g_{i1}(y_{[i]})}{h_{i1}(y_{[i-1]})}\right\} .$$

This implies a minorization condition (9) with

$$s(x) \propto \prod_{i=1}^{d} s_i(x^{[i]}) \min\left\{1, \frac{g_{i2}(x^{[i+1]})}{c_i h_{i2}(x^{[i]})}\right\} , \tag{15}$$

and probability density

$$q(y) \propto \prod_{i=1}^{d} q_i(y_{[i]}) \min\left\{1, \frac{c_i g_{i1}(y_{[i]})}{h_{i1}(y_{[i-1]})}\right\} . \tag{16}$$

Note that some care must be taken in the selection of the $s_i$ and the $g_{i1}, g_{i2}$ to ensure that $E_\pi s > 0$, as required for the minorization condition. We will not need the normalizing constant for $q$ since the probability of regeneration only depends on the product of $s$ and $q$. We now give the update rule for the component-wise split chain. Letting $X^{(k)} = x$ denote the current state of the chain, $X^{(k+1)}$ and $\delta^{(k)}$ are found by

1. Generate $X^{(k+1)} = y$ according to the component-wise update rule or $f_{comp}(x, \cdot)$

2. If even a single component-wise update proposal was rejected, set $\delta^{(k)} = 0$. If every component-wise update proposal was accepted, let $\delta^{(k)}$ be a Bernoulli random variable with success probability

$$r_A(x,y) = \prod_{i=1}^{d} \frac{s_i(x^{[i]}) \min\left\{1, \frac{g_{i2}(x^{[i+1]})}{c_i h_{i2}(x^{[i]})}\right\} q_i(y_{[i]}) \min\left\{1, \frac{c_i g_{i1}(y_{[i]})}{h_{i1}(y_{[i-1]})}\right\}}{p_i((y_{[i-1]}, x^{[i]}), y_i) \min\left\{1, \frac{\pi(y_{[i]}, x^{[i+1]})}{\pi(y_{[i-1]}, x^{[i]})} \frac{p_i((y_{[i]}, x^{[i+1]}), x_i)}{p_i((y_{[i-1]}, x^{[i]}), y_i)}\right\}} \tag{17}$$

We thus have the following theorem, which extends Mykland et al.'s (1995) Theorem 2 to component-wise updates, guaranteeing that the above update rule yields the correct regeneration probability (10).

**Theorem 5.** *A component-wise sampler with MTD $f_{comp}$ having target density $\pi$ satisfying* (11)–(14) *has a minorization condition $P(x,A) \geq s(x)Q(A)$ for all $x \in \mathsf{X}$ and $A \in \mathcal{B}(\mathsf{X})$. Further, in the component-wise split chain algorithm defined above, the probability of a regeneration occurring on a jump from $x$ to $y$ is given by*

$$\Pr(\delta^{(k)} = 1 | X^{(k)} = x, X^{(k+1)} = y) = \frac{s(x)Q(dy)}{P(x,dy)}.$$

*Proof.* The first assertion was proved above. Let $A_k$ denote the event that every single component-wise update proposal is accepted in the $k+1$ update. Since $\delta^{(k)}$ is zero if even a single update proposal is rejected, we have for any $x, y \in \mathsf{X}$,

$$1 = \Pr(A_k | \delta^{(k)} = 1, X^{(k)} = x, X^{(k+1)} = y) = \frac{\Pr(A_k, \delta^{(k)} | X^{(k)} = x, X^{(k+1)} = y)}{\Pr(\delta^{(k)} = 1 | X^{(k)} = x, X^{(k+1)} = y)}$$

and thus

$$\begin{aligned}
\Pr(\delta^{(k)} = 1 | X^{(k)} = x, X^{(k+1)} = y) &= \Pr(A_k, \delta^{(k)} = 1 | X^{(k)} = x, X^{(k+1)} = y) \\
&= \Pr(\delta^{(k)} = 1 | A_k, X^{(k)} = x, X^{(k+1)} = y) \\
&\quad \times \Pr(A_k | X^{(k)} = x, X^{(k+1)} = y) \\
&= r_A(x,y) \times \frac{\prod_{i=1}^{d} p_i((y_{[i-1]}, x^{[i]}), y_i) \alpha_i((y_{[i-1]}, x^{[i]}), y_i) \mu(dy)}{P(x,dy)} \\
&= \frac{s(x)q(y)\mu(dy)}{P(x,dy)} = \frac{s(x)Q(dy)}{P(x,dy)}
\end{aligned}$$

where $r_A$ is given by (17), and $q$ and $s$ are given by (16) and (15), respectively. □

*Remark* 8. Implementation of RS in the algorithm defined by MTD $f_{mix}$ is challenging. In particular, we were unsuccessful in establishing a minorization of the form (9). However, it is at least theoretically possible to implement RS using the minorization developed in the proof of Theorem 2. Of course, this is not appealing from a practical point of view.

# 4 Examples

## 4.1 A toy problem in Bayesian inference

Let $Y_1, \ldots, Y_m$ be i.i.d. Normal$(\mu, \theta)$, and let the prior on $(\mu, \theta)$ be given by the density $\pi_0(\mu, \theta) \propto I_A(\mu) I_B(\theta)/\sqrt{\theta}$, where $A$ and $B$ are Borel subsets of $\mathbb{R}$ and $\mathbb{R}^+$, respectively. Consider the problem of simulating an ergodic Markov chain with stationary distribution given by the posterior density

$$\pi(\mu, \theta) \propto \theta^{-\frac{m+1}{2}} e^{-\frac{1}{2\theta}\left[s^2 + m(\mu - \bar{y})^2\right]} I_A(\mu) I_B(\theta)$$

where $\bar{y} = \frac{1}{m}\sum_{i=1}^m y_i$ and $s^2 = \sum_{i=1}^m (y_i - \bar{y})^2$. Jones and Hobert (2001) showed that, provided $m \geq 5$, the Gibbs sampler is geometrically ergodic, even with $A = \mathbb{R}$ and $B = \mathbb{R}^+$. Here we propose a Metropolis-Hastings independence sampler that we show to be uniformly ergodic provided $A$ is bounded, and a component-wise independence sampler that we will show is uniformly ergodic provided $A$ is bounded and $B$ is bounded below away from zero. Then we show how to use RS in both samplers and conclude with an empirical comparison.

### 4.1.1 An independence sampler

If $(\mu^{(k)}, \theta^{(k)}) = (\mu, \theta) \in A \times B$ denotes the current state of the chain, the update $(\mu^{(k+1)}, \theta^{(k+1)})$ is found by

1. Simulate proposals $\mu^* \sim \mathrm{N}(\bar{y}, \frac{s^2}{m})$ on $A$ (a truncated normal distribution) and $\theta^* \sim \mathrm{IG}(\frac{m-1}{2}, \frac{s^2}{2})$ on $B$ (a truncated inverse gamma). Simulating the proposals is easily done by rejection sampling.

2. Calculate the acceptance probability $\alpha = \min\left\{1, \frac{\pi(\mu^*, \theta^*)}{\pi(\mu, \theta)} \frac{p(\mu, \theta)}{p(\mu^*, \theta^*)}\right\}$, where $p(\mu, \theta)$ is the product of the two univariate proposal densities. The acceptance probability reduces to the minimum of 1 and

$$\exp\left\{-\frac{m}{2}\left[\left(\frac{1}{s^2} - \frac{1}{\theta}\right)(\mu - \bar{y})^2 - \left(\frac{1}{s^2} - \frac{1}{\theta^*}\right)(\mu^* - \bar{y})^2\right]\right\}$$

3. Set

$$(\mu^{(k+1)}, \theta^{(k+1)}) = \begin{cases} (\mu^*, \theta^*) & \text{with probability} \quad \alpha \\ (\mu, \theta) & \text{with probability} \quad 1 - \alpha \end{cases}$$

To prove that this algorithm is uniformly ergodic we will show the existence of $\epsilon > 0$ such that $p(\mu, \theta) > \epsilon \pi(\mu, \theta)$ for all $(\mu, \theta) \in A \times B$ . Suppose $A$ is bounded, and let $\tilde{\mu}$ satisfy

$|\tilde{\mu} - \bar{y}| = \sup_{\mu \in A} |\mu - \bar{y}| < \infty$. Now

$$\frac{p(\mu,\theta)}{\pi(\mu,\theta)} \propto \exp\left\{-\frac{m}{2}\left(\frac{1}{s^2} - \frac{1}{\theta}\right)(\mu - \bar{y})^2\right\}.$$

Thus the inequality holds for $\epsilon = k \exp\left\{-\frac{m}{2s^2}(\tilde{\mu} - \bar{y})^2\right\} > 0$ where $k$ is a ratio of normalizing constants.

To simulate the MHIS split chain we follow the recipe of Mykland et al. (1995). The conditional probability of regeneration on a jump from $(\mu', \theta')$ to $(\mu, \theta)$, given that the Metropolis-Hastings proposal was accepted, is

$$r_A(x,y) = \begin{cases} \min\left\{\frac{\pi(\mu',\theta')}{cp(\mu',\theta')}, \frac{\pi(\mu,\theta)}{cp(\mu',\theta')}\right\} & \pi(\mu',\theta') < cp(\mu',\theta'),\ \pi(\mu,\theta) < cp(\mu,\theta) \\ \min\left\{\frac{cp(\mu',\theta')}{\pi(\mu',\theta')}, \frac{cp(\mu,\theta)}{\pi(\mu,\theta)}\right\} & \pi(\mu',\theta') > cp(\mu',\theta'),\ \pi(\mu,\theta) > cp(\mu,\theta) \\ 1 & \text{otherwise .} \end{cases}$$

We take the constant $c > 0$ to be the median value of $\pi(\mu,\theta)/p(\mu,\theta)$ from a preliminary run.

### 4.1.2 A component-wise independence sampler

In a CWIS the proposals $\mu^*$ and $\theta^*$ are considered independently, allowing the possibility that one component of the chain might move while the other does not. Let $(\mu^{(k)}, \theta^{(k)}) = (\mu, \theta) \in A \times B$ denote the current state of the chain. The updated value $\mu^{(k+1)}$ is found by

1. Simulate $\mu^*$ from candidate density $p_1(\cdot) \sim \mathrm{N}(\bar{y}, \frac{s^2}{m})$ on $A$ (truncated normal distribution).

2. Calculate the acceptance probability $\alpha_1 = \min\left\{1, \frac{\pi(\mu^*,\theta)}{\pi(\mu,\theta)}\frac{p_1(\mu)}{p_1(\mu^*)}\right\}$, which reduces to the minimum of 1 and

$$\exp\left\{-\frac{m}{2}\left(\frac{1}{s^2} - \frac{1}{\theta}\right)\left[(\mu - \bar{y})^2 - (\mu^* - \bar{y})^2\right]\right\}$$

3. Set

$$\mu^{(k+1)} = \begin{cases} \mu^* & \text{with probability} \quad \alpha_1 \\ \mu & \text{with probability} \quad 1 - \alpha_1 \end{cases}$$

If we now let $(\mu^{(k+1)}, \theta^{(k)}) = (\mu, \theta) \in A \times B$, we find $\theta^{(k+1)}$ by

1. Simulate $\theta^*$ from candidate denstiy $p_2(\cdot) \sim \mathrm{IG}(\frac{m-1}{2}, \frac{s^2}{2})$ on $B$ (truncated inverse gamma distribution).

2. Calculate acceptance probability $\alpha_2 = \min\left\{1, \frac{\pi(\mu,\theta^*)}{\pi(\mu,\theta)}\frac{p_2(\theta)}{p_2(\theta^*)}\right\}$, which reduces to the minimum of 1 and
$$\exp\left\{-\frac{m}{2}\left(\frac{1}{\theta^*}-\frac{1}{\theta}\right)(\mu-\bar{y})^2\right\}$$

3. Set
$$\theta^{(k+1)} = \begin{cases} \theta^* & \text{with probability} \quad \alpha_2 \\ \theta & \text{with probability} \quad 1-\alpha_2 \end{cases}$$

To show the above CWIS is uniformly ergodic, we must find $\delta > 0$ such that $p_1(\mu)p_2(\theta) \geq \delta\pi(\mu,\theta)$, and $\varepsilon > 0$ such that (6) holds. Assuming $A$ is bounded, let $\tilde{\mu}$ satisfy $|\tilde{\mu}-\bar{y}| = \sup_{\mu\in A}|\mu-\bar{y}| < \infty$, and the former condition holds with $\delta = k\exp\left\{-\frac{m}{2s^2}(\tilde{\mu}-\bar{y})^2\right\} > 0$ by the argument of the previous subsection. Further,
$$\begin{aligned}\frac{\pi(\mu,\theta)\pi(\mu',\theta')}{\pi(\mu,\theta')\pi(\mu',\theta)} &= \exp\left\{-\frac{m}{2}\left(\frac{1}{\theta}-\frac{1}{\theta'}\right)\left[(\mu-\bar{y})^2-(\mu'-\bar{y})^2\right]\right\} \\ &\geq \exp\left\{-\frac{m}{2\min\{\theta,\theta'\}}\max\left\{(\mu-\bar{y})^2,(\mu'-\bar{y})^2\right\}\right\}\end{aligned}$$
If we also assume $B$ is bounded below away from zero, and denote the infimum of $B$ by $\theta_* > 0$, then (6) holds with $\varepsilon = \exp\left\{-\frac{m(\tilde{\mu}-\bar{y})^2}{2\theta_*}\right\}$. Uniform ergodicity follows from Theorem 3.

Now consider implementing RS. Let $r_1(\mu,\theta) \propto \pi(\mu,\theta)/p_1(\mu)$ and $r_2(\mu,\theta) \propto \pi(\mu,\theta)/p_2(\theta)$. For the CWIS split chain we seek pairs of functions $g_1, g_2$ and $h_1, h_2$ such that $r_1(\mu,\theta) \geq g_1(\mu)g_2(\theta)$ and $r_2(\mu,\theta) \leq h_1(\mu)h_2(\theta)$. Now
$$\begin{aligned}r_1(\mu,\theta) &= \exp\left\{-\frac{m+1}{2}\log\theta - \frac{s^2}{2\theta} - \frac{m}{2}\left(\frac{1}{\theta}-\frac{1}{s^2}\right)(\mu-\bar{y})^2\right\} \\ &\geq \exp\left\{-\frac{m}{2}\left(\frac{1}{\tilde{\theta}}-\frac{1}{s^2}\right)(\mu-\bar{y})^2\right\} \times \exp\left\{-\frac{m+1}{2}\log\theta - \frac{s^2}{2\theta}\right\} I(\theta > \tilde{\theta})\end{aligned}$$
for any choice of $\tilde{\theta} \in B$, suggesting a sensible choice of $g_1$ and $g_2$. Also
$$r_2(\mu,\theta) = \exp\left\{-\frac{m}{2\theta}(\mu-\bar{y})^2\right\}$$
and thus the desired inequality holds for $h_1(\mu) = 1$ and $h_2(\theta) = 1$.

The conditional probability of a regeneration on a jump from $(\mu',\theta')$ to $(\mu,\theta)$, conditional on both component-wise proposals being accepted, is given by
$$r_A((\mu',\theta'),(\mu,\theta)) = \frac{\min\left\{1, \frac{g_2(\theta')}{r_1(\mu',\theta')}\right\}\min\{1, g_1(\mu)\}}{\min\left\{1, \frac{r_1(\mu,\theta')}{r_1(\mu',\theta')}\right\}} \frac{\min\left\{1, \frac{1}{h_2(\theta')}\right\}\min\left\{1, \frac{r_2(\mu,\theta)}{h_1(\mu)}\right\}}{\min\left\{1, \frac{r_2(\mu,\theta)}{r_2(\mu,\theta')}\right\}}$$
with the components of this expression as defined above.

### 4.1.3 An empirical comparison

Let $m = 10$, $\bar{y} = 10.2$, and $s^2 = 6.5$, and suppose our goal is to estimate the posterior mean inverse coefficient of variation $E(\mu/\sqrt{\theta}|y)$.

We let $A = (0, 100)$ and $B = \mathbb{R}^+$ and implemented the MHIS of subsection 4.1.1, starting the chain at $(\mu^{(0)}, \theta^{(0)}) = (10, 1)$. Results are shown in Figure 1. Autocorrelation tails off to zero by lag 10 or so.

We let $A = (0, 100)$ and $B = (.01, \infty)$ and considered the CWIS of subsection 4.1.2, again starting the chain at $(\mu^{(0)}, \theta^{(0)}) = (10, 1)$. As Figure 2 demonstrates, the autocorrelation function for CWIS tails off more rapidly than did that of the MHIS.

To study the performance of the regenerative simulation approach to MCMC standard error estimation (see Appendix B) we conducted the following simulation study. For each sampler we produced 20,000 chains of a fixed length $n$. From each realized chain we simulated the associated split chain using the techniques described above, and computed a 95% confidence interval for the quantity of interest based on (24)–(26) with $g(\mu, \theta) = \mu/\sqrt{\theta}$. The "partial tours" at the beginning and end of each run, those draws that preceded the first regeneration and those that came subsequent to the last observed regeneration, were discarded. We note that discarding the partial tour at the end of the run introduces a bias, but it will be negligible in this simple example. We can compare MHIS to CWIS with respect to interval widths and realized coverage rates. We considered $n = 5000$ and $n = 1000$; results are summarized in Table 1.

The last two columns of Table 1 describe the frequency of regeneration for each sampler. The mean tour length for MHIS was 2.57, and that of CWIS was 5.22. Recall that CWIS permits a regeneration only if both component-wise proposals were accepted, thus it is not surprising that CWIS regenerates less frequently than MHIS.

While high frequency of regeneration is desirable, this is not the correct criterion on which to be comparing samplers. The middle columns in Table 1 report coverage rates and interval half-widths. While CWIS produces less frequent regeneration than MHIS, it also yields narrower confidence intervals, with no discernible sacrifice in coverage probability.

## 4.2 A logit-normal GLMM

In a generalized linear mixed model (GLMM) the distribution of the observable data is specified conditionally on an unobserved vector of random effects. Consider a GLMM in which, conditional on $U = u$, the observations $Y_{ij}$ are independently distributed as Bernoulli($p_{ij}$), where $\text{logit}(p_{ij}) = \beta x_{ij} + u_i$ for $j = 1, \ldots, m_i$ and $i = 1, \ldots, q$, where the $x_{ij}$ are covariates.

Let the random effects $U_1, \ldots, U_q$ be i.i.d. Normal$(0, \sigma^2)$. McCulloch (1997) discusses three Monte Carlo algorithms for finding maximum likelihood estimators of the unknown parameter $\theta = (\beta, \sigma^2)$. A common feature is that all three require simulation from the same target distribution, namely the conditional distribution of the random effects given the data. We thus wish to simulate an ergodic Markov chain with stationary density

$$\pi(u|y;\theta) \propto \exp\left\{\sum_{i=1}^{q}\left[u_i y_{i+} - \sum_{j=1}^{m_i}\log\left(1+e^{\beta x_{ij}+u_i}\right) - \frac{u_i^2}{2\sigma^2}\right]\right\} \tag{18}$$

where $y_{i+} = \sum_{j=1}^{m_i} y_{ij}$ for $i = 1, \ldots, q$. McCulloch's (1997) Monte Carlo EM algorithm requires a Monte Carlo approximation to the so-called $Q$-function

$$Q(\theta;\tilde{\theta}) = \int l_c(\theta; y, u)\pi(u|y;\tilde{\theta})du$$

where

$$l_c(\theta; y, u) = \sum_{i=1}^{q}\sum_{j=1}^{m_i}\left[y_{ij}(\beta x_{ij} + u_i) - \log\left(1+e^{\beta x_{ij}+u_i}\right)\right] - \frac{q}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{q}u_i^2$$

denotes the "complete-data log-likelihood," what the log-likelihood would be if the random effects were observable.

In the following we propose three samplers having target density $\pi(u|y;\theta)$. We begin with a Metropolis-Hastings random walk sampler which we show to be geometrically ergodic; next we consider an independence sampler which we show is uniformly ergodic; and finally we consider a component-wise independence sampler which is uniformly ergodic. For the latter we discuss the simulation of the split chain. Further, we include an empirical comparison of the three samplers, and report the performance of regenerative simulation in the CWIS. We conclude with a comment on the use of RS in the Metropolis random walk.

### 4.2.1 Metropolis random walk

A Metropolis random walk with normally distributed jump proposals, that is,

$$p(u, u^*) \propto \exp\left\{-\frac{1}{2\tau^2}|u^* - u|^2\right\}, \tag{19}$$

where $|\cdot|$ denotes the standard Euclidean norm, is geometrically ergodic, as we prove here using Theorem 4.3 of Jarner and Hansen (2000). We must show that

$$\lim_{|u|\to\infty}\frac{u^T\nabla\log\pi(u)}{|u|} = -\infty$$

and that

$$\limsup_{|u|\to\infty} \frac{u^T \nabla \log \pi(u)}{|u|\ |\nabla \log \pi(u)|} < 0 \ .$$

Now $\frac{\partial}{\partial u_i} \log \pi(u) = y_{i+} - p_{i+} - \frac{u_i}{\sigma^2}$, where $p_{i+} = \sum_{j=1}^{m_i} p_{ij}$, for $i = 1, \ldots, q$, and thus

$$\begin{aligned}
\lim_{|u|\to\infty} \frac{u^T \nabla \log \pi(u)}{|u|} &= \lim_{|u|\to\infty} \frac{\sum_{i=1}^q u_i(y_{i+} - p_{i+}) - \frac{1}{\sigma^2}\sum_{i=1}^q u_i^2}{\left(\sum_{i=1}^q u_i^2\right)^{1/2}} \\
&= -\frac{1}{\sigma^2} \lim_{|u|\to\infty} \frac{\sum_{i=1}^q u_i^2}{\left(\sum_{i=1}^q u_i^2\right)^{1/2}} = -\frac{1}{\sigma^2} \lim_{|u|\to\infty} |u| = -\infty \ .
\end{aligned}$$

Next,

$$\begin{aligned}
\limsup_{|u|\to\infty} \frac{u^T \nabla \log \pi(u)}{|u|\ |\nabla \log \pi(u)|} &= \limsup_{|u|\to\infty} \frac{\sum_{i=1}^q u_i(y_{i+} - p_{i+}) - \frac{1}{\sigma^2}\sum_{i=1}^q u_i^2}{\left(\sum_{i=1}^q u_i^2\right)^{1/2} \left(\sum_{i=1}^q \left(y_{i+} - p_{i+} - \frac{u_i}{\sigma^2}\right)^2\right)^{1/2}} \\
&= \limsup_{|u|\to\infty} \frac{-\frac{1}{\sigma^2}\sum_{i=1}^q u_i^2}{\left(\sum_{i=1}^q u_i^2\right)^{1/2} \left(\frac{1}{\sigma^4}\sum_{i=1}^q u_i^2\right)^{1/2}} \\
&= \frac{-1/\sigma^2}{-1/\sigma^2} \limsup_{|u|\to\infty} \frac{|u|^2}{|u|^2} = -1 \ .
\end{aligned}$$

### 4.2.2 An independence sampler

Consider a Metropolis-Hastings independence sampler with proposals drawn from the random effects' marginal distribution, $p(u)$. The acceptance probability for a proposed jump from $u$ to $u^*$ reduces to the minimum of 1 and $r(u^*)/r(u)$, where $r(u) \propto \pi(u)/p(u)$, the ratio of the target density to the proposal density. In the logit-normal example we can take

$$r(u) = \exp\left\{ \sum_{i=1}^q \left[ u_i y_{i+} - \sum_{j=1}^{m_i} \log\left(1 + e^{\beta x_{ij} + u_i}\right) \right] \right\}. \tag{20}$$

Recall that an MHIS is uniformly ergodic if and only if there exists $\epsilon > 0$ such that $p(u) \geq \epsilon\pi(u)$ for all $u$; equivalently, an MHIS is uniformly ergodic if and only if $r(u)$ is bounded. It is easily verified that $r(u) \leq \exp\left\{-\beta \sum_i \sum_j x_{ij} y_{ij}\right\}$ for all $u \in \mathbb{R}^q$, and thus our MHIS is uniformly ergodic.

### 4.2.3 A component-wise independence sampler

In McCulloch's (1997) implementation of the Monte Carlo EM and Monte Carlo Newton-Raphson algorithms, simulation from the target distribution is accomplished by a component-wise independence sampler with (univariate) proposals drawn from the marginal distribution

of the random effects. Specifically, for the $i$th component-wise update, we generate a proposal $u_i^* \sim \text{Normal}(0, \sigma^2)$; the acceptance probability is again equal to the minimum of 1 and $r(u^*)/r(u)$, where $u^* = (u_1, \ldots, u_{i-1}, u_i^*, u_{i+1}, \ldots, u_q)$ and $r$ is given by (20). The CWIS is also uniformly ergodic, an immediate consequence of our Corollary 1 and the fact that the random effects are conditionally independent given the data.

Regenerative simulation is straightforward in this problem because (13) holds with equality on both sides, again as a result of component-wise independence in the target distribution. Define $r_i(u_i) = \exp\left\{u_i y_{i+} - \sum_j \log(1 + e^{\beta x_{ij} + u_i})\right\}$ for $i = 1, \ldots, q$, so that $r(u) = \prod_{i=1}^q r_i(u_i)$. The conditional probability of regeneration on a jump from $u'$ to $u$, conditional on every component-wise proposal being accepted, is

$$r_A(u', u) = \prod_{i=1}^{q} \frac{\min\left\{\dfrac{c_i}{r_i(u_i')}, 1\right\} \min\left\{\dfrac{r_i(u_i)}{c_i}, 1\right\}}{\min\left\{\dfrac{r_i(u_i)}{r_i(u_i')}, 1\right\}}$$

or

$$r_A(u', u) = \prod_{i=1}^{q} \begin{cases} \min\left\{\frac{r_i(u_i')}{c_i}, \frac{r_i(u_i)}{c_i}\right\} & r_i(u_i') < c_i, r_i(u_i) < c_i \\ \min\left\{\frac{c_i}{r_i(u_i')}, \frac{c_i}{r_i(u_i)}\right\} & r_i(u_i') > c_i, r_i(u_i) > c_i \\ 1 & \text{otherwise .} \end{cases}$$

With the goal of maximizing the frequency of regeneration, it makes sense to take the $c_i$ to be the median values of $r_i(u_i)$ from a preliminary run.

#### 4.2.4 Simulation study

Suppose $x_{ij} = j/15$ for each $j = 1, \ldots, m_i \equiv 15$, for each $i = 1, \ldots, q = 10$. We consider a data set generated by Booth and Hobert (1999, Table 2), assuming $\theta = (\beta, \sigma^2) = (5.0, 0.5)$. Suppose we want to evaluate $Q(\theta; \tilde{\theta})$ for $\theta = \tilde{\theta} = (4.0, 1.5)$. We can take as a Markov chain Monte Carlo approximation the ergodic average of the chain $\left\{l_c(\theta; y, u^{(k)})\right\}$ where $\left\{u^{(k)} : k = 1, 2, \ldots\right\}$ is an ergodic Markov chain with stationary density $\pi(u|y; \tilde{\theta})$ as defined by (18).

We ran each of the three Metropolis-Hastings algorithms discussed above, in each case simulating a chain of length $n = 10^6$ and taking as our initial distribution $U^{(0)} \sim \text{N}_q(0, \sigma^2 I)$. For the Metropolis random walk we drew our jump proposals from a $\text{N}_q(0, \tau^2 I)$, with $\tau^2 = \sigma^2/10$ (this setting determined by trial and error, in order to minimize the autocorrelation in the resulting chain, and yielding an acceptance rate of 39.18%). A partial trace plot (the second 1000 updates) and the sample autocorrelation function are shown in Figure 3. Analogous plots for the MHIS and CWIS follow, in Figures 4 and 5, respectively.

Our most striking result is the dreadful performance of the MHIS. It is worth recalling that, as we proved above, Figure 4 depicts a uniformly ergodic Markov chain! Thus this example nicely illustrates the perils of over-reliance on asymptotic properties of a sampler, which provide little guarantee of favorable performance in finite-sample implementations. The Metropolis random walk (geometrically ergodic) mixes much faster than the MHIS, but still shows significant autocorrelation. The CWIS is, by a very wide margin, the best of the three samplers considered.

To study the performance of regenerative simulation in this setting we propose the following experiment. Using the CWIS algorithm defined above, run a chain for a fixed number of regenerations $R$, and calculate a 95% confidence interval for $Q(\theta;\tilde{\theta})$ based on the RS approach to MCMC standard error estimation (see Appendix B; specifically, we use (24)–(26) with $g(u) = l_c(\theta; y, u)$). Repeat a large number of times, noting coverage rates and interval widths.

We considered $R = 50$, $R = 25$, and $R = 10$, and generated 1000 chains for each. Results are summarized in Table 2. The right-most columns summarize the observed chain lengths (note that this experiment has the opposite design of the previous example, in that the number of regenerations is fixed and the total chain length is random). The overall average tour length is $\bar{N} = 10{,}732$, a result of the fact that a regeneration in CWIS requires that *every* component-wise proposal be accepted. Because the tours are so long, one might expect regenerative simulation to yield accurate results for even modest values of $R$. For $R = 50$ and $R = 25$ the attained coverage rate is in fact reasonably close to the nominal confidence level. For $R = 10$ the method undercovers.

### 4.2.5 Regeneration in Metropolis random walk

The reader might wonder why we did not conduct a similar study for the Metropolis random walk sampler. As we will show, it is not entirely clear how to implement regenerative simulation in high-dimensional Metropolis random walks; in particular, Mykland et al.'s (1995) approach is often not viable, as we demonstrate here.

The RS recipe of Mykland et al. (1995) requires functions $s$ and $q$ such that $p(u, u^*) \geq s(u)q(u^*)$ for all $u, u^*$. The conditional probability of regeneration on a jump from $U^{(k)} = u'$ to $U^{(k+1)} = u$, given that the Metropolis jump proposal was accepted, is given by

$$r_A(u', u) = \frac{s(u')q(u)}{p(u', u)} \times \begin{cases} \min\left\{\frac{\pi(u')}{c}, \frac{\pi(u)}{c}\right\} & \pi(u') < c,\ \pi(u) < c \\ \min\left\{\frac{c}{\pi(u')}, \frac{c}{\pi(u)}\right\} & \pi(u') > c,\ \pi(u) > c \\ 1 & \text{otherwise .} \end{cases}$$

To find appropriate functions $s$ and $q$ Mykland et al. (1995) suggest choosing a point $\tilde{u}$ and a set $D$ and taking $s(u) = \inf_{u^* \in D} [p(u, u^*)/p(\tilde{u}, u^*)]$ and $q(u^*) = p(\tilde{u}, u^*) I_D(u^*)$. Intuitively, it makes sense to take $\tilde{u}$ to be some *a priori* estimate of the "center" of the target distribution, perhaps the ergodic average from a preliminary run. We might then take $D$ to be a hypercube centered at $\tilde{u}$, that is, $D = \{u : |u_i - \tilde{u}_i| \leq b_i \;\; i = 1, \ldots, d\}$ for some $b_1, \ldots, b_d > 0$. For the Metropolis random walk $p(u, u^*)$ is given by (19). Then

$$\frac{s(u')q(u)}{p(u', u)} = \exp\left\{-\frac{1}{\tau^2}\sum_{i=1}^{q}(u_i' - \tilde{u}_i)[u_i - \tilde{u}_i + b_i \, \mathrm{sgn}(u_i' - \tilde{u}_i)]\right\} \prod_{i=1}^{q} I(|u_i - \tilde{u}_i| < b_i) \qquad (21)$$

where $\mathrm{sgn}(z) = z/|z|$ for $z \neq 0$.

We ran a chain of length $n = 10^6$ and observed 391,804 accepted proposals. We take the $\tilde{u}_i$ to be the mean values from a preliminary run. The choice of the $b_i$ entails a Goldilocks problem: we want to observe $|u_i - \tilde{u}_i| < b_i$ with high frequency, but (21) depends on an infimum taken over the set of all such $u$. We considered six different values of the $b_i$ (six different multiples of the standard deviations from a preliminary run); results are summarized below. The middle column indicates the proportion of the 391,804 accepted proposals for which each $|u_i - \tilde{u}_i| < b_i$, i.e., the proportion of the time that (21) was non-zero. The right-most column indicates the average non-zero value of (21).

| $b_i$ | $\% \{|u_i - \tilde{u}_i| < b_i\}$ | (21) |
|---|---|---|
| 0.3 $\mathrm{SD}_i$ | 0.00 | |
| 0.5 $\mathrm{SD}_i$ | $5.10 \times 10^{-5}$ | 0.00107 |
| 1.0 $\mathrm{SD}_i$ | 0.0193 | $2.62 \times 10^{-06}$ |
| 1.5 $\mathrm{SD}_i$ | 0.2269 | $6.77 \times 10^{-09}$ |
| 2.0 $\mathrm{SD}_i$ | 0.6113 | $7.88 \times 10^{-11}$ |
| 2.5 $\mathrm{SD}_i$ | 0.8668 | $1.79 \times 10^{-12}$ |

There is no acceptable trade-off available here, and thus regeneration is not viable in this problem. We note two caveats on our negative conclusion: (i) that we only considered $D$ to be a hypercube centered at $\tilde{u}$, and (ii) that we considered only the approach recommended by Mykland et al. (1995). We are not aware of any other literature on this problem. It is not our contention that regeneration is impossible in high-dimensional Metropolis random walks. We merely hope to convey that it is a challenging proposition, and suggest it as a possible subject for future research.

# A Proof of Theorem 3

First note that $\varepsilon \leq 1$ as can be seen by taking $x = y$. Now, for $x, y \in \mathsf{X}$, define

$$\beta(x,y) = \prod_{i=1}^{d} p_i(y_i)\alpha_i((y_{[i-1]}, x^{[i]}), y_i).$$

We will show that there exists $\rho > 0$ such that $\beta(x,y) \geq \rho\pi(y)$ for all $x, y \in \mathsf{X}$ and thus, since $P(x, dy) \geq \beta(x,y)dy$, that the chain is uniformly ergodic.

For any $x$ and $y$ we can partition the index set $\{1, \ldots, d\}$ into $I_1$ and $I_0$ defined by

$$I_1(x,y) = \left\{ i : \alpha_i((y_{[i-1]}, x^{[i]}), y_i) = 1 \right\},$$

$$I_0(x,y) = \left\{ i : \alpha_i((y_{[i-1]}, x^{[i]}), y_i) < 1 \right\};$$

and write

$$\beta(x,y) = \prod_{i=1}^{d} p_i(y_i)\alpha_i((y_{[i-1]}, x^{[i+1]}), y_i) = \prod_{i \in I_1} p_i(y_i) \prod_{i \in I_0} p_i(x_i) \frac{\pi(y_{[i-1]}, y_i, x^{[i+1]})}{\pi(y_{[i-1]}, x_i, x^{[i+1]})}. \tag{22}$$

Now fix $x$ and $y$ in $\mathsf{X}$. There exists an even integer $k$ such that

$$\begin{aligned} I_0(x,y) &= \{d_0 + 1, \ldots, d_1, d_2 + 1, \ldots, d_3, d_4 + 1, \ldots, d_5, \ldots\ldots, d_k + 1, \ldots, d_{k+1}\} \\ I_1(x,y) &= \{d_1 + 1, \ldots, d_2, d_3 + 1, \ldots, d_4, d_5 + 1, \ldots, d_6, \ldots, d_{k-1} + 1, \ldots, d_k\} \end{aligned} \tag{23}$$

where $0 = d_0 \leq d_1 < d_2 < d_3 < \ldots < d_k \leq d_{k+1} = d$.

This representation of $I_0$ and $I_1$ appears to make sense only for the case where $\alpha_1$ and $\alpha_d$ are both less than 1, but in fact can be shown to accommodate the general case by allowing the first or last batch of $I_0$ to be null. We will explain further here, but first we must explain how $k$ and the $d_j$ are determined.

For each $j = 1, \ldots, k-1$, $d_j + 1$ is the index of a switch between $\alpha = 1$ and $\alpha < 1$: If $j$ is even the switch is from $\alpha_{d_j} = 1$ to $\alpha_{d_j+1} < 1$ and if $j$ is odd then $\alpha_{d_j} < 1$ and $\alpha_{d_j+1} = 1$. If $\alpha_1 < 1$ and $\alpha_d < 1$, then $k$ is the total number of such switches, which must be even. If only one of $\alpha_1$ and $\alpha_d$ is less than 1, there must be $k-1$ such switches, and if $\alpha_1 = \alpha_d = 1$ then the number of switches is $k-2$. We can define $d_1$ and $d_k$ a bit more precisely as

$$d_1 = \begin{cases} 0 & \text{if all } \alpha_i < 1 \\ \min\{i : \alpha_i = 1\} - 1 & \text{otherwise} \end{cases}$$

and

$$d_k = \begin{cases} d_0 = 0 & \text{if all } \alpha_i < 1 \\ \max\{i : \alpha_i = 1\} & \text{otherwise} \end{cases}$$

Thus the representation in (23) is completely general, since either or both of $d_1 = 0$ (in which case the set $\{d_0 + 1, \ldots, d_1\}$ is null) and $d_k = d$ (in which case $\{d_k + 1, \ldots, d_{k+1}\}$ is null) are allowed. A null contribution to $I_0$ is easily recognized: If $d_1 = 0$, as is the case where $\alpha_1 = 1$, the first batch of indices in $I_0$ is $\{d_0 + 1, \ldots, d_1\} = \{1, 0\}$, which should be considered null; if $d_k = d$, as is so when $\alpha_d = 1$, the last batch of indices in $I_0$ is $\{d_k + 1, \ldots, d_k\} = \{d + 1, d\}$, which should also be considered null.

In the special case where all $\alpha_i$ are less than 1, then $k = 0$ and $d_1 = d$; if all $\alpha_i$ are equal to 1, then $k = 2$, $d_1 = 0$ and $d_2 = d$.

We now find $\rho > 0$ such that $\beta(x, y) \geq \rho\pi(y)$ for any $x, y \in \mathsf{X}$. First, suppose $x$ and $y$ are such that all $\alpha_i$ are less than 1. Then $I_1 = \emptyset$, and

$$\beta(x, y) = p(x)\frac{\pi(y)}{\pi(x)} \geq \delta\pi(y)$$

thus taking care of that special case. Now suppose that at least one $\alpha_i = 1$. We will need a new notation which we introduce here. Given $0 = d_0 \leq d_1 < d_2 < \ldots < d_k \leq d_{k+1} = d$, we partition the vector $z \in \mathsf{X} = \mathsf{X}_1 \times \cdots \times \mathsf{X}_d$ as $z = (z_0, z_1, z_2, \ldots, z_k)$ where $z_j = (z_{d_j+1}, \ldots, z_{d_{j+1}})$. Thus we are changing our use of a subscript from indicator of a single component to indicator of "batch." Define $z_{[j]} = (z_0, z_1, \ldots, z_j)$ and $z^{[j]} = (z_j, z_{j+1}, \ldots, z_k)$ for $j = 0, 1, \ldots, k$. Let $z_{[-1]}$ and $z^{[k+1]}$ be null. By (22),

$$\begin{aligned}
\beta(x, y) &= p(x_0, y_1, x_2, y_3, x_4, \ldots, y_{k-1}, x_k) \prod_{j \in \{0,2,\ldots,k\}} \frac{\pi(y_{[j-1]}, y_j, x^{[j+1]})}{\pi(y_{[j-1]}, x_j, x^{[j+1]})} \\
&\geq \delta\pi(x_0, y_1, x_2, \ldots, y_{k-1}, x_k) \prod_{j \in \{0,2,\ldots,k\}} \frac{\pi(y_{[j-1]}, y_j, x^{[j+1]})}{\pi(y_{[j-1]}, x_j, x^{[j+1]})} \\
&= \delta\pi(x_0, y_1, x_2, \ldots, y_{k-1}, x_k) \frac{\pi(y_0, x^{[1]})}{\pi(x_0, x^{[1]})} \prod_{j \in \{2,4,\ldots,k\}} \frac{\pi(y_{[j-1]}, y_j, x^{[j+1]})}{\pi(y_{[j-1]}, x_j, x^{[j+1]})} \\
&\geq \delta\varepsilon\pi(y_{[1]}, x_2, y_3, x_4, \ldots, y_{k-1}, x_k) \prod_{j \in \{4,\ldots,k\}} \frac{\pi(y_{[j-1]}, y_j, x^{[j+1]})}{\pi(y_{[j-1]}, x_j, x^{[j+1]})} \\
&\geq \delta\varepsilon^2\pi(y_{[3]}, x_4, y_5, x_6, \ldots, y_{k-1}, x_k) \prod_{j \in \{4,\ldots,k\}} \frac{\pi(y_{[j-1]}, y_j, x^{[j+1]})}{\pi(y_{[j-1]}, x_j, x^{[j+1]})} \\
&\vdots \\
&\geq \delta\varepsilon^{k/2}\pi(y_{[k-1]}, x_k) \frac{\pi(y_{[k-1]}, y_k)}{\pi(y_{[k-1]}, x_k)} \\
&= \delta\varepsilon^{k/2}\pi(y)
\end{aligned}$$

where the first inequality follows from the assumption that $p(x) \geq \delta\pi(x)$ and the rest follow from (6). We have seen that the number of switches is either $k$, $k-1$, or $k-2$; thus it must be that $k \leq d+1$. But if the number of switches is $k-2$, that means the first batch of $I_0$ is null and we have one less $\varepsilon$ on the right hand side of the final inequality above. Thus $\beta(x,y) \geq \rho\pi(y)$ holds with $\rho = \delta\varepsilon^{\lfloor d/2 \rfloor}$, and thus the chain is uniformly ergodic.

# B MCMC standard error estimation via regenerative simulation

Let $\left\{X^{(0)}, X^{(1)}, X^{(2)}, \ldots\right\}$ be a Harris ergodic Markov chain on $(\mathsf{X}, \mathcal{B})$ with stationary distribution $\pi$ and let $g : \mathsf{X} \to \mathbb{R}^1$. Further let $\left\{(X^{(n)}, \delta^{(n)}) : n = 0, 1, 2, \ldots\right\}$ be an associated split chain with $\delta^{(0)} = 1$. Define $0 = \tau_0 < \tau_1 < \tau_2 < \ldots$ by $\tau_r = \min\left\{k > \tau_{r-1} : \delta^{(k)} = 1\right\}$ for $r = 1, 2, \ldots$; thus $\tau_r$ indicates the $r$th regeneration time. We call the path taken by the chain between regeneration times $\tau_{r-1}$ and $\tau_r$ the $r$th *tour*; let $N_r = \tau_r - \tau_{r-1}$ and $S_r = \sum_{k=\tau_{r-1}+1}^{\tau_r} g(X^{(k)})$ denote the tour length and tour sum, respectively, for $r = 1, 2, \ldots$. It follows from the split chain construction that the pairs $(N_r, S_r)$ are i.i.d., since each is based on a different tour.

Suppose we run the split chain for a total of $R$ regenerations (and hence observe $R$ tours). A natural estimator of $E_\pi g$ is

$$\hat{E}_\pi g = \bar{g}_{R,RS} := \frac{\bar{S}_R}{\bar{N}_R} = \frac{\sum_{r=1}^R S_r}{\sum_{r=1}^R N_r} = \frac{1}{\tau_R}\sum_{k=1}^{\tau_R} g(X^{(k)}) = \bar{g}_{\tau_R}. \tag{24}$$

Assume $E(N_r) < \infty$ and $E(|S_r|) < \infty$. Then, as $R \to \infty$, $\bar{g}_{R,RS} = \bar{g}_{\tau_R} \to E_\pi g$ by the strong law of large numbers. Further, if $E(N_r^2) < \infty$ and $E(S_r^2) < \infty$, then $\sqrt{R}(\bar{g}_{\tau_R} - E_\pi g) \xrightarrow{d} \mathrm{N}(0, \xi_g^2)$ as $R \to \infty$ by the central limit theorem. Hobert et al. (2002) proved that

$$\hat{\xi}_g^2 := \frac{1}{R\bar{N}^2}\sum_{r=1}^R (S_r - N_r\bar{g}_{\tau_R})^2 \tag{25}$$

is a consistent estimator of $\xi_g^2$. Finally, an approximate $100(1-\alpha)\%$ confidence interval for $E_\pi g$ is given by

$$\bar{g}_{\tau_R} \pm z^{\alpha/2}\frac{\hat{\xi}_g}{\sqrt{R}} \tag{26}$$

where $\Pr(Z < z^p) = 1 - p$ if $Z$ has the standard normal distribution.

## References


Bédard, M. and Rosenthal, J. S. (2008). Optimal scaling of Metropolis algorithms: Heading toward general target distributions. *The Canadian Journal of Statistics* (to appear).

Booth, J. G. and Hobert, J. P. (1999). Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. *Journal of the Royal Statistical Society,* Series B, 61:265–285.

Brockwell, A. E. and Kadane, J. B. (2005). Identification of regeneration times in MCMC simulation, with application to adaptive schemes. *Journal of Computational and Graphical Statistics*, 14:436–458.

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

Christensen, O. F., Moller, J., and Waagepetersen, R. P. (2001). Geometric ergodicity of Metropolis-Hastings algorithms for conditional simulation in generalized linear mixed models. *Methodology and Computing in Applied Probability*, 3:309–327.

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 Jones, G. L. (2008). Batch means and spectral variance estimators in Markov chain Monte Carlo. Technical report, University of Minnesota, Department of Statistics.

Fort, G., Moulines, E., Roberts, G. O., and Rosenthal, J. S. (2003). On the geometric ergodicity of hybrid samplers. *Journal of Applied Probability*, 40:123–146.

Gilks, W. R., Roberts, G. O., and Sahu, S. K. (1998). Adaptive Markov chain Monte Carlo through regeneration. *Journal of the American Statistical Association*, 93:1045–1054.

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.

Hobert, J. P., Jones, G. L., and Robert, C. P. (2006). Using a Markov chain to construct a tractable approximation of an intractable probability distribution. *Scandinavian Journal of Statistics*, 33:37–51.

Jarner, S. F. and Hansen, E. (2000). Geometric ergodicity of Metropolis algorithms. *Stochastic Processes and Their Applications*, 85:341–361.

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.

McCulloch, C. E. (1997). Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association*, 92:162–170.

Mengersen, K. and Tweedie, R. L. (1996). Rates of convergence of the Hastings and Metropolis algorithms. *The Annals of Statistics*, 24:101–121.

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*, 90:233–241.

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. (1998). Two convergence properties of hybrid samplers. *The Annals of Applied Probability*, 8:397–407.

Roberts, G. O. and Rosenthal, J. S. (2006). Harris recurrence of Metropolis-within-Gibbs and trans-dimensional Markov chains. *The Annals of Applied Probability*, 16:2123–2139.

Roberts, G. O. and Sahu, S. K. (1997). Updating schemes, correlation structure, blocking and parametrization for the Gibbs sampler. *Journal of the Royal Statistical Society,* Series B, 59:291–317.

Roberts, G. O. and Tweedie, R. L. (1996). Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika*, 83:95–110.

Rosenthal, J. S. (2008). Optimal proposal distributions and adaptive MCMC. In Brooks, S., Gelman, A., Meng, X.-L., and Jones, G. L., editors, *Handbook of Markov chain Monte Carlo* (to appear). Chapman & Hall, Boca Raton.

Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion). *The Annals of Statistics*, 22:1701–1762.

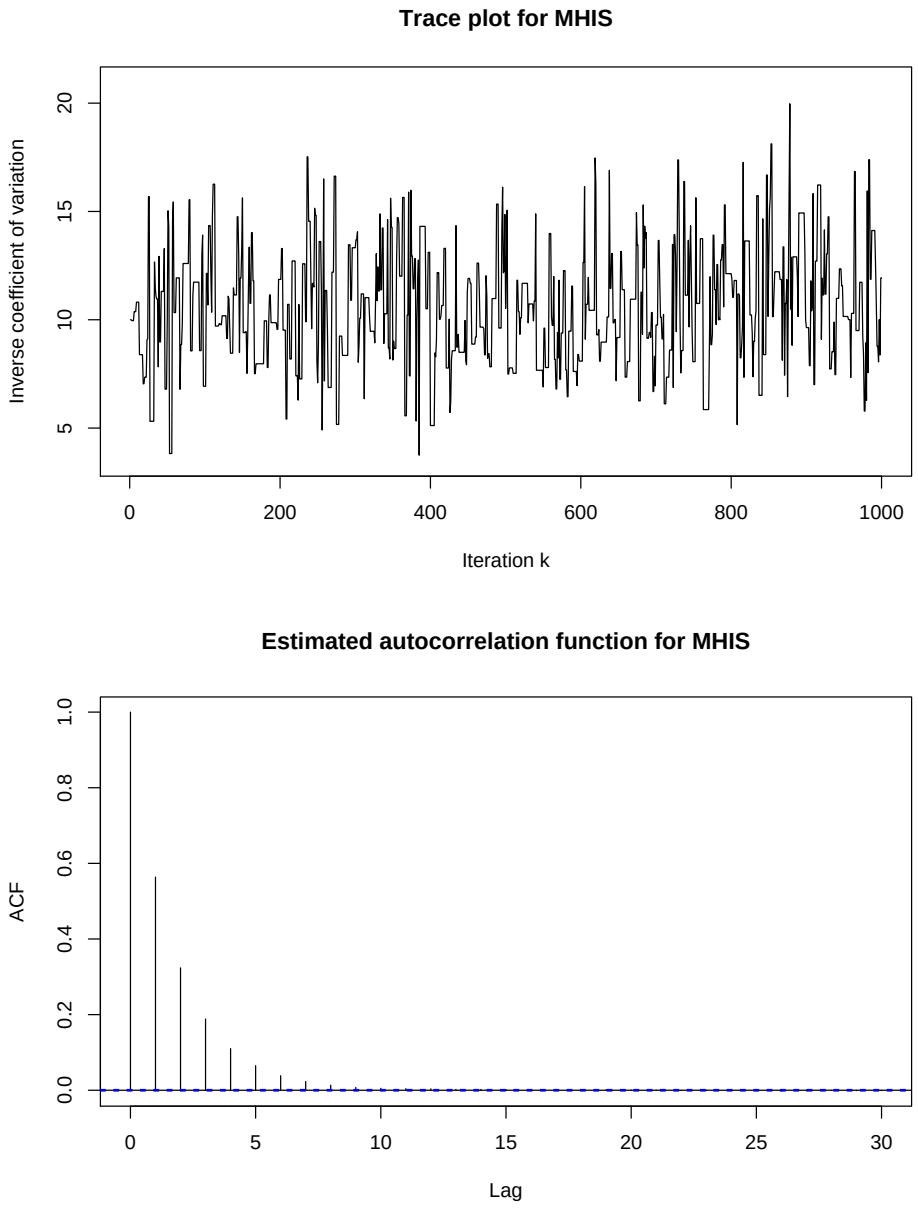


Figure 1: *Trace plot and estimated autocorrelation function of* $\left\{\mu^{(k)}/\sqrt{\theta^{(k)}}\right\}$ *for the MHIS in the example of Section 4.1.*

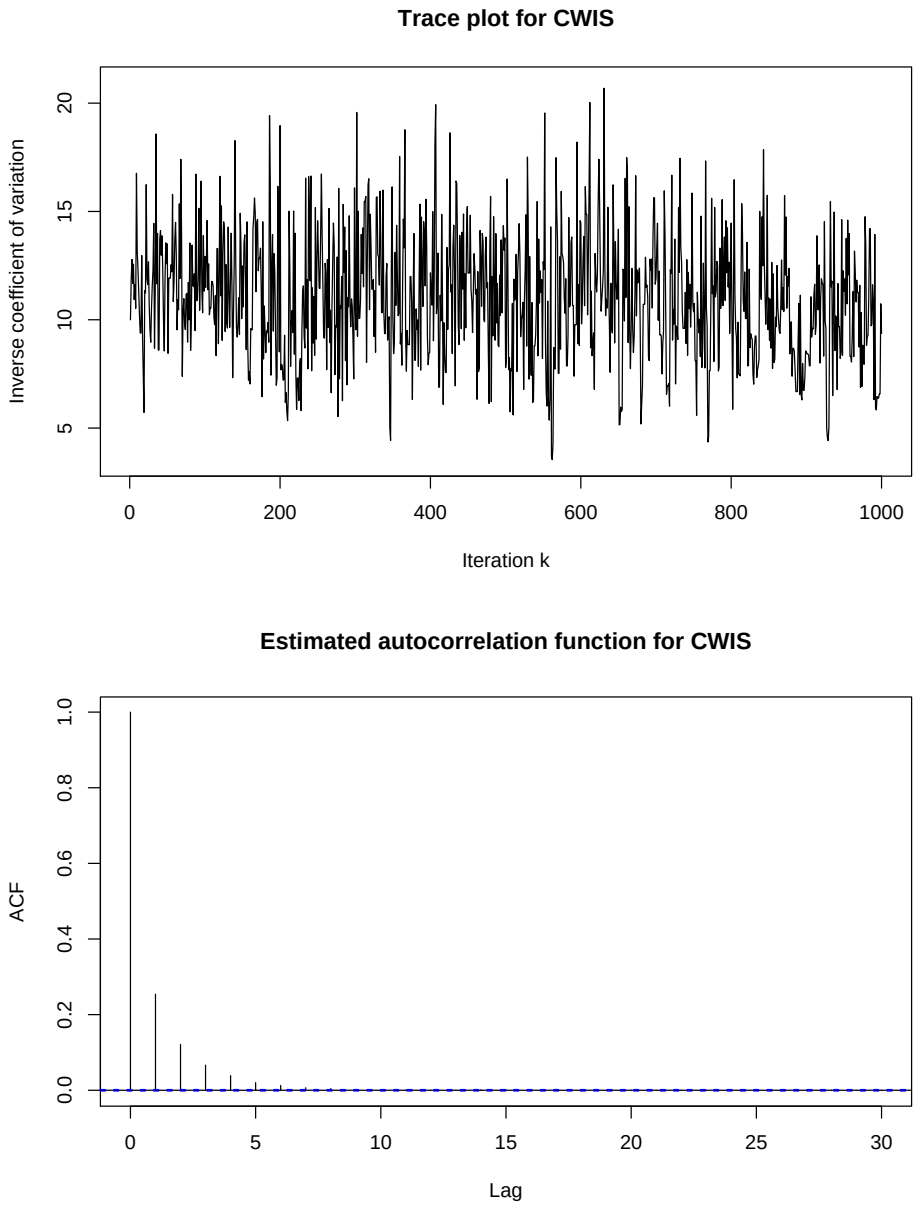


Figure 2: *Trace plot and estimated autocorrelation function of* $\left\{\mu^{(k)}/\sqrt{\theta^{(k)}}\right\}$ *for the CWIS in the example of Section 4.1.*

| | | Half-width | | Coverage | | Number of tours | |
|---|---|---|---|---|---|---|---|
| Sampler | Chain length | Mean | Std Dev | Rate | Std Error | Mean | Std Dev |
| MHIS | $n = 5000$ | 0.1494 | 0.0093 | 0.9495 | 0.0015 | 1944.47 | 34.65 |
| CWIS | | 0.1113 | 0.0069 | 0.9494 | 0.0016 | 958.01 | 27.31 |
| MHIS | $n = 1000$ | 0.3321 | 0.0293 | 0.9455 | 0.0016 | 388.03 | 15.54 |
| CWIS | | 0.2472 | 0.0278 | 0.9441 | 0.0016 | 190.88 | 12.20 |

Table 1: Half-widths and coverage rates of nominal 95% confidence intervals for $E(\mu/\sqrt{\theta}|y)$ in the example of Section 4.1. Based on 20,000 replications.

| | Coverage | | Half-width | | Total chain length | |
|---|---|---|---|---|---|---|
| Regenerations | Rate | Std Error | Mean | Std Dev | Mean | Std Dev |
| $R = 50$ | 0.928 | 0.0082 | 0.01172 | 0.00189 | 533,668 | 72,893 |
| $R = 25$ | 0.920 | 0.0086 | 0.01600 | 0.00355 | 267,392 | 52,479 |
| $R = 10$ | 0.862 | 0.0109 | 0.02377 | 0.00836 | 107,316 | 35,258 |

Table 2: Coverage rates and interval half-widths of nominal 95% confidence intervals for $Q(\theta;\tilde{\theta})$ in the logit-normal example of Section 4.2. Based on 1000 replications.

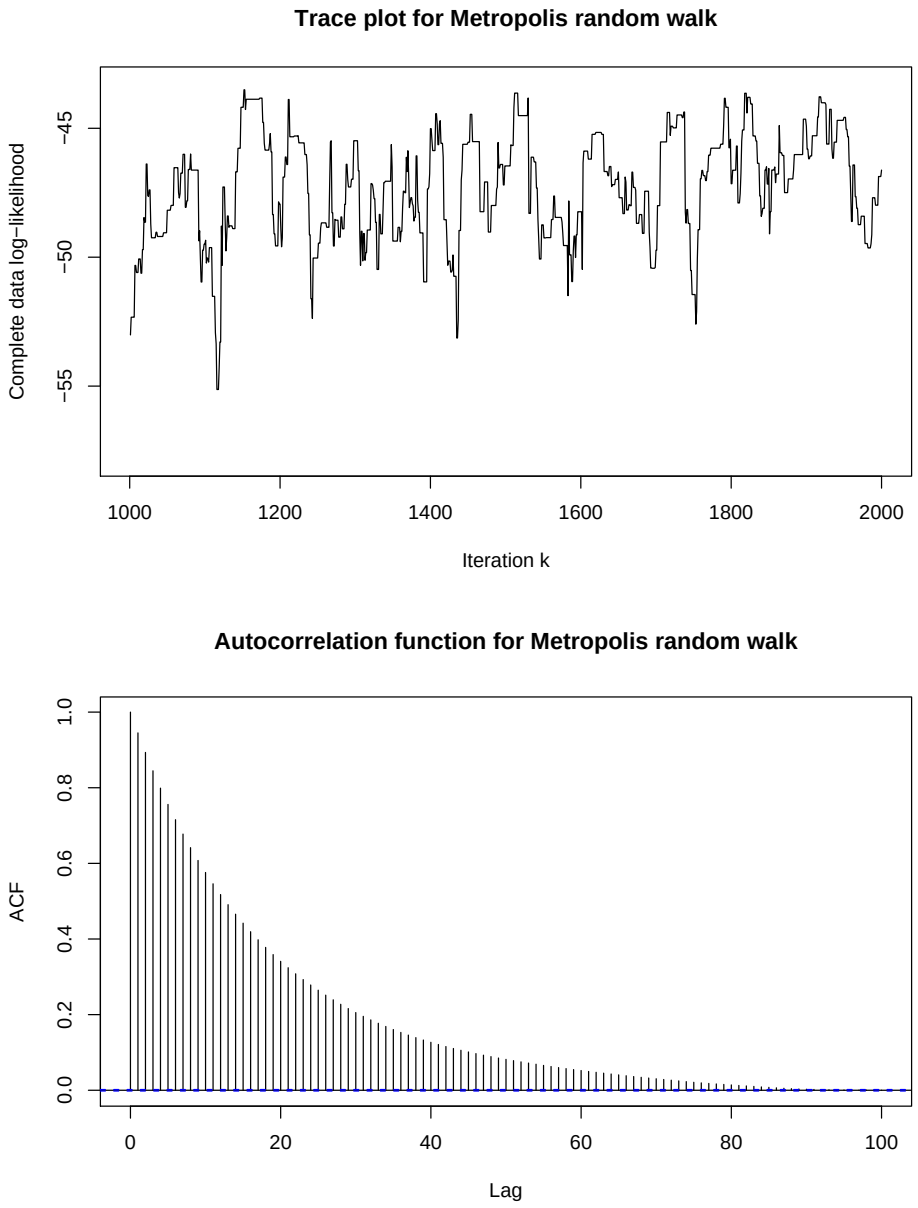


Figure 3: *Partial trace plot and autocorrelation function for the Markov chain* $\{l_c(\theta; y, u^{(k)})\}$ *generated by the Metropolis random walk sampler in the logit-normal example of Section 4.2.*

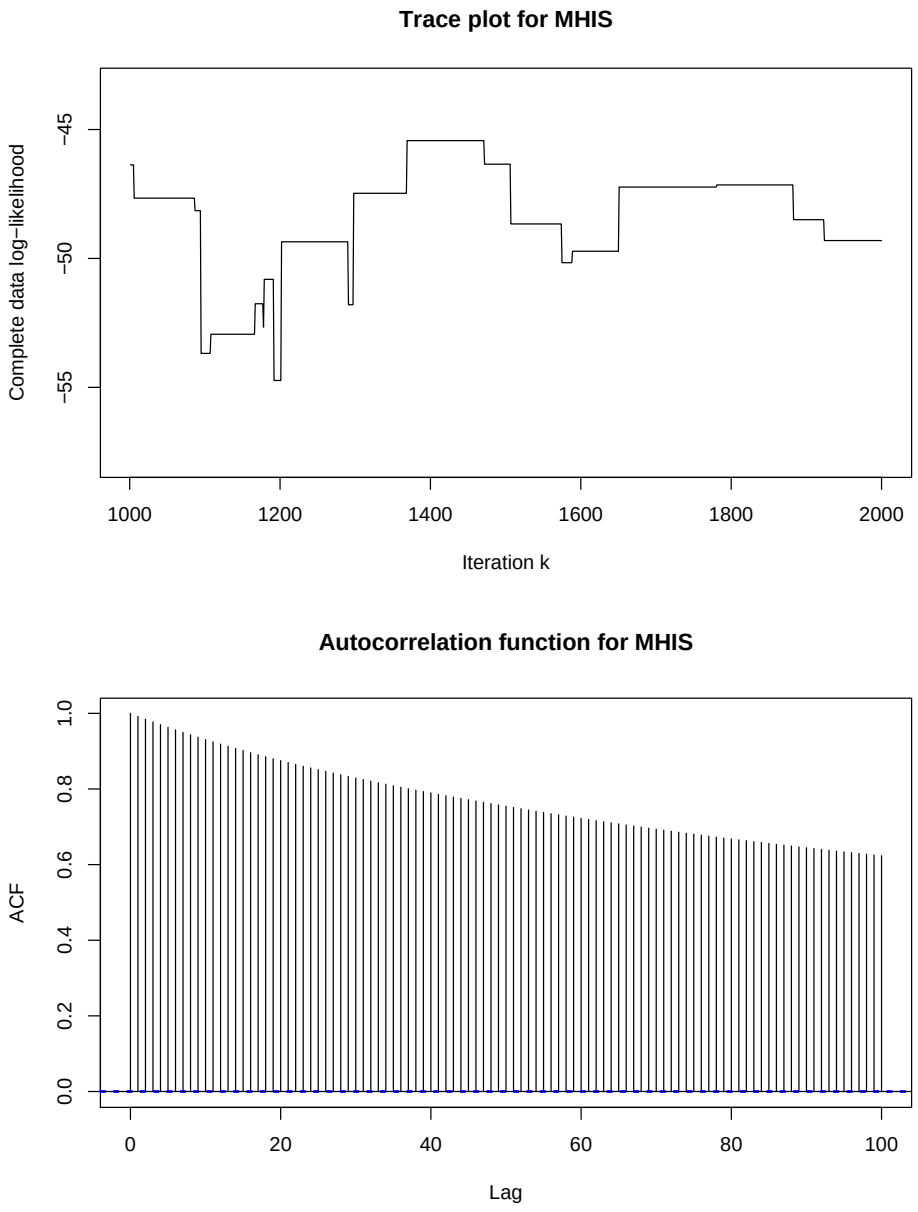


Figure 4: *Partial trace plot and autocorrelation function for the Markov chain* $\{l_c(\theta; y, u^{(k)})\}$ *generated by the Metropolis-Hastings independence sampler in the logit-normal example of Section 4.2.*

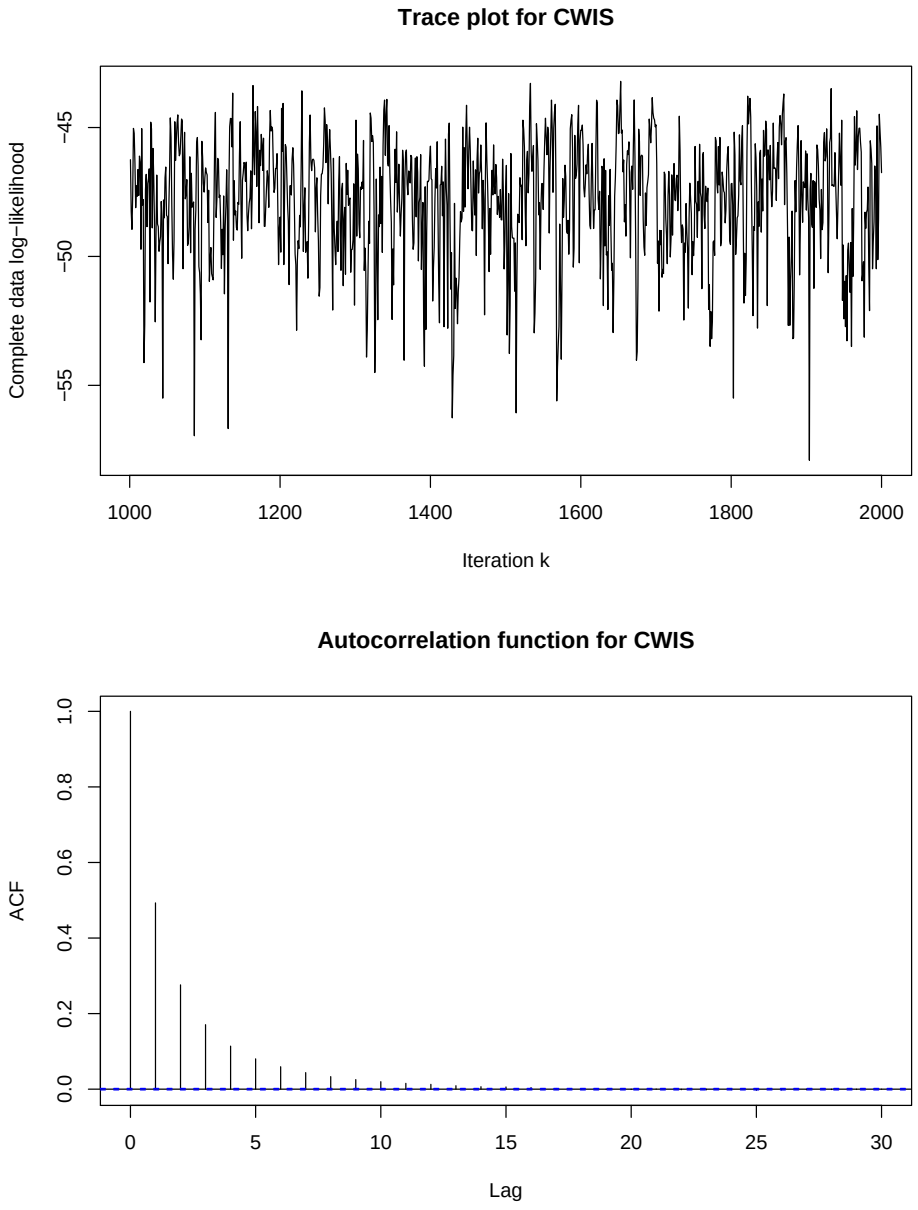


Figure 5: *Partial trace plot and autocorrelation function for the Markov chain* $\{l_c(\theta; y, u^{(k)})\}$ *generated by the component-wise independence sampler in the logit-normal example of Section 4.2.*