The Annals of Applied Probability
1997, Vol. 7, No. 1, 110–120

WEAK CONVERGENCE AND OPTIMAL SCALING OF RANDOM WALK METROPOLIS ALGORITHMS[1]

BY G. O. ROBERTS, A. GELMAN AND W. R. GILKS

University of Cambridge, Columbia University and Institute of Public Health, Cambridge

This paper considers the problem of scaling the proposal distribution of a multidimensional random walk Metropolis algorithm in order to maximize the efficiency of the algorithm. The main result is a weak convergence result as the dimension of a sequence of target densities, n, converges to ∞. When the proposal variance is appropriately scaled according to n, the sequence of stochastic processes formed by the first component of each Markov chain converges to the appropriate limiting Langevin diffusion process.

The limiting diffusion approximation admits a straightforward efficiency maximization problem, and the resulting asymptotically optimal policy is related to the asymptotic acceptance rate of proposed moves for the algorithm. The asymptotically optimal acceptance rate is 0.234 under quite general conditions.

The main result is proved in the case where the target density has a symmetric product form. Extensions of the result are discussed.

1. Introduction. The random walk algorithm of Metropolis et al. (1953) is known to be an effective Markov chain Monte Carlo method for many diverse problems. However, its efficiency depends crucially on the scaling of the proposal density. If the variance of the proposal is too small, the Markov chain will converge slowly since all its increments will be small. Conversely, if the variance is too large, the Metropolis algorithm will reject too high a proportion of its proposed moves. A number of authors have suggested informal guidelines for scaling proposal to target variance ratios [e.g., Besag and Green (1993)] or monitoring accept/reject ratios [see, e.g., Besag, Green, Higdon and Mengersen (1995)]. However, although such rules of thumb often work well in practice, to date there have been no theoretical results to support them.

In this paper, we consider the asymptotic problem as the dimension of the state space, n, converges to infinity. By considering suitably regular sequences of canonical target densities and rescaling the proposal variance by a factor $(1/n)$, we obtain a weak convergence result for the sequence of algorithms restricted to a fixed finite set of components, C, to the appropriate

Received September 1995; revised October 1996.
[1]Supported by NSF Grants SBR-92-23637, DMS-94-04305 and DMS-94-57824.
AMS 1991 *subject classifications.* Primary 60F05; secondary 65U05.
Key words and phrases. Metropolis algorithm, weak convergence, optimal scaling, Markov chain Monte Carlo.

Langevin diffusion on $\mathbb{R}^c$. Finding the asymptotically optimal scaling is then a simple matter of optimizing the speed of the Langevin diffusion.

Specifically, let

$$\pi_n(\mathbf{x}^n) = \prod_{i=1}^{n} f(x_i^n) \tag{1.1}$$

be an n-dimensional product density with respect to Lebesgue measure. The random walk Metropolis algorithm with Gaussian proposal density,

$$q_n(\mathbf{x}^n, \mathbf{y}^n) = \frac{1}{(2\pi\sigma_n^2)^{n/2}} \exp\left\{\frac{-1}{2\sigma_n^2}|\mathbf{y}^n - \mathbf{x}^n|^2\right\},$$

produces a Markov chain $\mathbf{X}^n = \{\mathbf{X}_0^n, \mathbf{X}_1^n, \dots\}$, where $\mathbf{X}_m^n$ is chosen randomly as follows. We adopt the notation $\mathbf{x}^n$ for an n-vector with components $x_1^n, \dots, x_n^n$. Generate $\mathbf{Y}^n$ according to $q_n(\mathbf{X}_{m-1}^n, \cdot)$ and set $\mathbf{X}_m^n = \mathbf{Y}^n$ with probability

$$\alpha(\mathbf{X}_{m-1}^n, \mathbf{Y}^n) \equiv 1 \wedge \frac{\pi_n(\mathbf{Y}^n)}{\pi_n(\mathbf{X}_{m-1}^n)}.$$

Otherwise, we set $\mathbf{X}_m^n = \mathbf{X}_{m-1}^n$. Therefore, $\alpha(\cdot,\cdot)$ is known as the acceptance function. Produced in this way, it is easy to see that $\{\mathbf{X}_m^n\}$ is a Markov chain, reversible with respect to π_n, and is π_n-irreducible, aperiodic and hence ergodic [see, e.g., Roberts and Smith (1994) or Mengersen and Tweedie (1966)].

We introduce the following conditions on f: we assume that f'/f is Lipschitz continuous and

$$\mathbb{E}_f\left[\left(\frac{f'(X)}{f(X)}\right)^8\right] \equiv M < \infty, \tag{A1}$$

$$\mathbb{E}_f\left[\left(\frac{f''(X)}{f(X)}\right)^4\right] < \infty. \tag{A2}$$

The main result of this paper is therefore that for each fixed one-dimensional component of $\{\mathbf{X}^n, n \geq 1\}$, the one-dimensional process converges weakly to the appropriate Langevin diffusion.

Let C^2 denote the space of real-valued functions with continuous second derivative. Let $\sigma_n^2 = l^2/(n-1)$, and for integers n, define $U_t^n = X_{[nt],1}^n$. In other words U_t^n consists of the first component of $\mathbf{X}_{[nt]}^n$.

Note that in the definition of σ_n^2, we use the divisor $n-1$. This could be replaced by n, which seems to be a more appropriate divisor for small n at least in the Gaussian case [see Gelman, Roberts and Gilks (1996)]. The asymptotic result is unaltered by the choice of divisor; however, our preferred choice here leads to a simpler proof.

Let us denote weak convergence of processes in the Skorokhod topology by $\Rightarrow$ [see, e.g., Ethier and Kurtz (1986)].

We can now state the result more precisely as follows.

THEOREM 1.1. *Suppose f is positive and C^2 and that* (A1) *and* (A2) *hold. Let $\mathbf{X}_0^\infty = (X_{0,1}^1\, X_{0,2}^2, \dots)$ be such that all of its components are distributed according to f and assume that $X_{0,j}^i = X_{0,j}^j$ for all $i \le j$. Then, as $n \to \infty$,*

$$U^n \Rightarrow U,$$

where U_0 is distributed according to f and U satisfies the Langevin SDE

$$dU_t = (h(l))^{1/2}\, dB_t + h(l) \frac{f'(U_t)}{2f(Ut)}\, dt \tag{1.2}$$

and

$$h(l) = 2l^2 \Phi\left(-\frac{l\sqrt{I}}{2}\right)$$

with Φ being the standard normal cumulative cdf and

$$I \equiv \mathbb{E}_f\left[\left(\frac{f'(X)}{f(X)}\right)^2\right].$$

Here $h(l)$ is sometimes called a *speed* measure for the diffusion process. We can write $U_t = V_{h(l)t}$, where V is the Langevin diffusion with speed measure unity:

$$dV_t = dB_t + \frac{f'(V_t)}{2f(V_t)}\, dt.$$

Therefore the "most efficient" asymptotic diffusion has the largest speed measure.

The result is illuminating for two reasons. First, since U^n is produced by speeding up time by a favor of n, the complexity of the algorithm is therefore n. Although complexity results exist for Markov chain Monte Carlo with finite state spaces [see, e.g., Frigessi and den Hollander (1993)], no such results are available in continuous state spaces. In Section 3 we will discuss the generalization of Theorem 1.1 to the case where the components are dependent, and the related ideas of phase transition.

Second, and perhaps more importantly in practice, Theorem 1.1 has the following corollary. First let

$$a_n(l) = \int\int \pi_n(\mathbf{x}^n)\, \alpha(\mathbf{x}^n, \mathbf{y}^n) q_n(\mathbf{x}^n, \mathbf{y}^n)\, d\mathbf{x}^n\, d\mathbf{y}^n$$

be the average acceptance rate of the random walk Metropolis algorithm in n dimensions, and let

$$a(l) = 2\Phi\left(-\frac{l\sqrt{I}}{2}\right)$$

COROLLARY 1.2.

(i) $$\lim_{n\to\infty} a_n(l) = a(l).$$

(ii) *$h(l)$ is maximized (to two decimal places) by*

$$l = \hat{l} = \frac{2.38}{\sqrt{I}}.$$

Also

$$a(\hat{l}) = 0.23$$

and $h(\hat{l}) = 1.3/I$.

This result gives rise to the useful heuristic for random walk Metropolis in practice:

Tune the proposal variance so that the average acceptance rate is roughly $1/4$.

The accompanying paper [Gelman, Roberts and Gilks (1996)] discusses the use of this heuristic in practice, and other related issues.

Note that the optimal value l is scaled, not by the standard deviation of the target density (as is often suggested), but by $1/\sqrt{I}$. However if f is Gaussian, it is easy to verify that I is exactly the reciprocal of the variance of f. In general, I is a measure of "roughness" of f—high values of I lead to $\hat{l}$ having to be small.

We only state Theorem 1.1 for univariate components, although implicit in our method of proof is the stronger statement that for integers $c > 1$, the process consisting of the first c components of $\mathbf{X}^n_{[nt]}$ converges to a collection of c independent processes each distributed according to (1.2).

2. Proof of Theorem 1.1. Define the (discrete time) generator of $\mathbf{X}^n$,

$$G_n V(\mathbf{x}^n) = n\mathbb{E}\left[(V(\mathbf{Y}^n) - V(\mathbf{x}^n))\left(1 \wedge \frac{\pi_n(\mathbf{Y}^n)}{\pi_n(\mathbf{x}^n)}\right)\right],$$

for any function V for which this definition makes sense.

The expectation here is taken with respect to the proposal distribution. Therefore, $\mathbf{Y}^n - \mathbf{x}^n \sim N(\mathbf{0}, l^2/(n-1)I_n)$. In the Skorokhod topology, it does not cause any problems to treat G_n as a continuous time generator (of a process with jumps at times of a Poisson process at rate n). We shall restrict attention to test functions V which are functions of the first component only.

Our proof of Theorem 1.1 will demonstrate uniform convergence of G_n to G, the generator of the limiting (one-dimensional) Langevin diffusion, for a suitably large class of real-valued functions V, where

$$GV(x) = h(l)\left[\frac{1}{2}V''(x) + \frac{1}{2}\frac{d}{dx}(\log f)(x)V'(x)\right].$$

Notice that G_n acts on functions of $\mathbb{R}^n$, whereas in the limit we are merely interested in functions of the first component, so that G generally just acts on functions of $\mathbb{R}$. This will involve a minor abuse of notation, but this will nevertheless add to the clarity of the sequel. Now, by Theorem 2.1 of Chapter 8 of Ethier and Kurtz (1986), since $(d/dx)\log f$ is Lipschitz, a core for the generator has domain C_c^∞ (infinitely differentiable functions on compact support). This will enable us to restrict attention to functions in C_c^∞.

Although the putative diffusion limit is Markov, the sequence of approximations $\{U^n, n > 1\}$ is not, although the approximations can be considered to be embedded in the sequence of Markov processes $\{\mathbf{Z}^n, n > 1\}$ with

$$\mathbf{Z}_t^n = \left(X_{[nt],1}^n, \dots, X_{[nt],n}^n\right),$$

so that U^n is the first component of $\mathbf{Z}^n$.

Define the sequence of sets $\{F_n \subseteq \mathbb{R}^n, n > 1\}$ by

$$F_n = \left\{|R_n(x_2, \dots, x_n) - I| < n^{-1/8}\right\} \cap \left\{|S_n(x_2, \dots, x_n) - I| < n^{-1/8}\right\},$$

where

$$R_n(x_2, \dots, x_n) = \frac{1}{n-1} \sum_{i=2}^{n} \left[(\log f(x_i))'\right]^2$$

and

$$S_n(x_2, \dots, x_n) = \frac{-1}{n-1} \sum_{i=2}^{n} \left[(\log f(x_i))''\right].$$

LEMMA 2.1. *For fixed t,*

$$\mathbb{P}\left[\mathbf{Z}_s^n \in F_n, 0 \le s \le t\right] \to 1 \quad \text{as } n \to \infty.$$

PROOF. Since $\mathbf{Z}_0^n \sim \pi_n$, $\mathbf{Z}_s^n \sim \pi_n, 0 \le s \le t$, since π is stationary. Therefore,

$$\mathbb{P}\left[\mathbf{Z}_s^n \notin F_n, \text{for some } 0 \le s \le t\right] \le tn\mathbb{P}_{\pi_n}[\mathbf{Z} \notin F_n].$$

Note that $\mathbb{E}[R_n(X_2, \dots, X_n)]$(according to π_n) $= I$, so that, by the weak law of large numbers, for all $\varepsilon > 0$,

$$\mathbb{P}_{\pi_n}\left[|R_n(\mathbf{Z}) - I| > \varepsilon\right] \to 0 \quad \text{as } n \to \infty.$$

Moreover, by Markov's inequality and (A1),

$$\begin{aligned}\mathbb{P}_{\pi_n}[\mathbf{Z} \notin F_n] &\le \mathbb{E}_{\pi_n}\left[(R_n(\mathbf{Z}) - I)^4\right] n^{1/2} \\ &\le \frac{3M}{(n-1)^{3/2}}.\end{aligned}$$

It follows that

$$\mathbb{P}\left[\mathbf{Z}_s^n \in \left\{|R_n - I| < n^{-1/8}\right\}, 0 \le s \le t\right] \to 1$$

as required. The proof that

$$\mathbb{P}\left[\mathbf{Z}_s^n \in \left\{|S_n(x_2,\dots,x_n) - I| < n^{-1/8}\right\}, 0 \le s \le t\right] \to 1$$

follows similarly using (A2). □

In the sequel, we shall use the following collection of preliminary results.

PROPOSITION 2.2. *The function $g(x) = 1 \wedge e^x$ is Lipschitz with coefficient 1. That is,*

$$|g(x) - g(y)| \le |x - y| \qquad \forall x, y \in \mathbb{R}$$

LEMMA 2.3. *Let*

$$W_n(= W_n(x_i)) = \sum_{i=2}^{n} \left[\frac{(\log f(x_i))''}{2}(Y_i - x_i)^2 + \frac{l^2}{2(n-1)}(\log f(x_i))'^2 \right],$$

where $Y_i \sim N(x_i, l^2/(n-1))$ independently for all $i = 2,\dots,n$. Then

$$\sup_{\mathbf{x}^n \in F_n} \mathbb{E}[|W_n|] \to 0 \quad \text{as } n \to \infty.$$

PROOF.

$$\begin{aligned}\mathbb{E}[|W_n|]^2 &\le \mathbb{E}\left[W_n^2\right] \\ &= \frac{1}{4(n-1)^2}\left(\sum_{i=2}^{n} (\log f(x_i))'' + \left((\log f(x_i))'\right)^2\right)^2 \\ &\quad + \frac{2}{4(n-1)^2}\sum_{i=2}^{n}\left((\log f(x_i))''\right)^2\end{aligned}$$

by direct calculation. However, for $\mathbf{x}^n \in F_n$,

$$\left|\sum_{i=1}^{n} \frac{(\log f(x_i))'' + \left((\log f(x_i))'\right)^2}{2(n-1)}\right| \le n^{-1/8}$$

and, since $(\log f)''$ is bounded, $\mathbb{E}[|W_n|]^2 \to 0$ uniformly for $\mathbf{x}^n \in F_n$. □

PROPOSITION 2.4. *If $A \sim N(\mu, \sigma^2)$, then*

$$\mathbb{E}[1 \wedge e^A] = \Phi\left(\frac{\mu}{\sigma}\right) + \exp(\mu + \sigma^2/2)\Phi\left(-\sigma - \frac{\mu}{\sigma}\right),$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function.

Armed with these preliminary results, we are now in a position to state two uniform convergence results which play major roles in the proof of Theorem 1.1

LEMMA 2.5. *For $V \in C_c^\infty$,*

$$\limsup_{n\to\infty} \sup_{x_1 \in \mathbb{R}} n\left|\mathbb{E}[V(Y_1) - V(x_1)]\right| < \infty,$$

where as usual, Y_1 is distributed $N(x_1, \sigma_n^2)$ and the expectation is taken with respect to this normal random variable.

PROOF.

$$V(Y_1) - V(x_1) = V'(x_1)(Y_1 - x_1) + \tfrac{1}{2}V''(Z_1)(Y_1 - x_1)^2$$

for some $Z_1 \in (x_1, Y_1)$ or (Y_1, x_1). Therefore,

$$n\mathbb{E}[V(Y_1) - V(x_1)] \le \frac{1}{2}K\frac{n}{n-1}l^2,$$

where K is an upper bound for V''. □

LEMMA 2.6. *Suppose $V \in C_c^\infty$ is a function of the first component of $\mathbf{Z}^n$. Then*

$$\sup_{\mathbf{x}^n \in F_n} |G_n V(\mathbf{x}^n) - GV(x_1)| \to 0 \quad \text{as } d \to \infty.$$

PROOF. Decomposing $\mathbf{Y}^n$ (the proposal) into $(Y_1, \mathbf{Y}^{n-})$,

$$G_n V(\mathbf{x}^n) = n\mathbb{E}_{Y_1}\left[(V(\mathbf{Y}^n) - V(\mathbf{x}^n))\mathbb{E}_{\mathbf{Y}^{n-}}\left[\left(1 \wedge \frac{\pi_n(\mathbf{Y}^n)}{\pi_n(\mathbf{x}^n)}\right)\right]\right],$$

we begin by concentrating on the inner expectation, which we will call E $[= E(Y_1)]$. Write

$$E = \mathbb{E}\left[1 \wedge \exp\left\{\varepsilon(Y_1) + \sum_{i=2}^{n} (\log f(Y_i) - \log f(x_i))\right\}\right]$$

$$\left[\text{where we write } \varepsilon(Y_1) = \log\left(\frac{(f(Y_1))}{(f(x_1))}\right)\right]$$

$$= \mathbb{E}\left[1 \wedge \exp\left\{\varepsilon(Y_1) + \sum_{i=2}^{n}\left[(\log f(x_i))'(Y_i - x_i) + \frac{1}{2}(\log f(x_i))''(Y_i - x_i)^2 + \frac{1}{6}(\log f(Z_i))'''(Y_i - x_i)^3\right]\right\}\right]$$

for some $Z_i \in (x_i, Y_i)$ *or* (Y_i, x_i).

Therefore, by Proposition 2.2, we can write

$$\left|E - \mathbb{E}\left[1 \wedge \exp\left\{\varepsilon(Y_1) + \sum_{i=2}^{n}\left[(\log f(x_i))'(Y_i - x_i) - \frac{l^2}{2(n-1)}((\log f(x_i))')^2\right]\right\}\right]\right|$$

$$\le \mathbb{E}[|W_n|] + \sup_{z \in \mathbb{R}} |\log f(z)'''| \frac{1}{6(n-1)^{1/2}} \frac{4l^3}{(2\pi)^{1/2}},$$

where W_n is as defined in Lemma 2.3. Also by that result,

$$\sup_{\mathbf{x}^n \in F_n} \left| E - \mathbb{E}\left[1 \wedge \exp\left\{\varepsilon(Y_1) + \sum_{i=2}^{n} \left[(\log f(x_i))'(Y_i - x_i) - \frac{l^2}{2(n-1)}((\log f(x_i))')^2\right]\right\}\right]\right|$$

$$= \varphi(n) \quad (\text{say}) \to 0 \text{ as } n \to \infty.$$

However,

$$\varepsilon(Y_1) + \sum_{i=2}^{n} \left[(\log f(x_i))'(Y_i - x_i) - \frac{l^2}{2(n-1)}((\log f(x_i))')^2\right]$$

is distributed $N(\varepsilon(Y_1) - l^2 R_n/2, l^2 R_n)$, so that, by Proposition 2.4,

$$\begin{aligned}
&\mathbb{E}\left[1 \wedge \exp\left\{\varepsilon(Y_1) + \sum_{i=2}^{n} \left[(\log f(x_i))'(Y_i - x_i) - \frac{l^2}{2(n-1)}((\log f(x_i))')^2\right]\right\}\right] \\
&\quad = \Phi\left(R_n^{-1/2}\left(l^{-1}\varepsilon(Y_1) - \frac{lR_n}{2}\right)\right) \\
&\quad\quad + \exp(\varepsilon(Y_1))\Phi\left(-\frac{lR_n^{1/2}}{2} - \varepsilon(Y_1)R_n^{-1/2}l^{-1}\right) \\
&\quad \triangleq M(\varepsilon) \quad \text{say}.
\end{aligned}$$

Therefore we can write

$$\sup_{\mathbf{x}^n \in F_n} \left| G_n V - n\mathbb{E}\left[(V(Y_1) - V(x_1))M\left(\log\left(\frac{f(Y_1)}{f(x_1)}\right)\right)\right]\right|$$

$$\le \varphi(n) n\mathbb{E}[|V(Y_1) - V(x_1)|] \to 0 \quad \text{as } n \to \infty.$$

It therefore remains to consider the term

$$n\mathbb{E}\left[(V(Y_1) - V(x_1))M\left(\log\frac{f(Y_1)}{f(x_1)}\right)\right].$$

Now a Taylor series expansion of the integrand about x_1 gives

$$\begin{aligned}
&(V(Y_1) - V(x_1))M\left(\log\frac{f(Y_1)}{f(x_1)}\right) \\
&\quad = \left(V'(x_1)(Y_1 - x_1) + \frac{1}{2}V''(x_1)(Y_1 - x_1)^2 + \frac{V'''(Z_1)}{6}(Y_1 - x_1)^3\right) \\
&\quad\quad \times \left[M(0) + (Y_1 - x_1)M'(0)(\log f(x_i))' + \frac{1}{2}(Y_1 - x_1)^2 T(x_1, W_1)\right],
\end{aligned}$$

where

$$T(x_1, W_1) = M''\left(\log\frac{f(W_1)}{f(x_1)}\right)\left((\log f(W_1))'\right)^2 + (\log f(W_1))'' M'\left(\log\frac{f(W_1)}{f(x_1)}\right)$$

and where $Z_1, W_1 \in [x_1, Y_1]$ or $[Y_1, x_1]$. However, since V has compact support, S say, there exists $K < \infty$ such that $|\log f)^{(i)}(x)|$, $V^{(i)}(x)| \le K$ for $x \in S$, $i = 1,2,3$, and it is easy to check that M' and M'' are bounded (again by K say). Now

$$M(0) = 2M'(0) = 2\Phi\left(-\frac{lR_n^{1/2}}{2}\right)$$

so that

$$\begin{aligned}\mathbb{E}&\left[n(V(Y_1) - V(x_1))M\left(\log\frac{f(Y_1)}{f(x_1)}\right)\right]\\ &= 2n\Phi\left(-\frac{R_n^{1/2}l}{2}\right)\left[\left(\frac{1}{2}V''(x_1) + \frac{1}{2}\log f(x_i)\right)'V'(x_1)\right]\mathbb{E}\left[(Y_1 - x_1)^2\right]\\ &\quad + \mathbb{E}[B(x_1, Y_1, n)],\end{aligned}$$

where

$$\mathbb{E}[|B(x_1, Y_1, n)|] \le a_1(K)n\mathbb{E}\left[|Y_1 - x|^3\right] + a_2(K)n\mathbb{E}\left[|Y_1 - x_1|^4\right] + a_3(K)n\mathbb{E}\left[|Y_1 - x_1|^5\right]$$

and a_1, a_2 and a_3 are polynomials in K. Therefore, $\mathbb{E}[|B(x_1, Y_1, n)|]$ is uniformly $0(n^{-1/2})$ and so

$$\sup_{\mathbf{x}^n \in F_n} |G_n V(\mathbf{x}) - GV(\mathbf{x})| \to 0 \quad \text{as } n \to \infty. \qquad \square$$

PROOF OF THEOREM 1.1. From Lemma 2.6, we have uniform convergence for vectors contained in a set of limiting probability 1. This essentially proves the result by Theorem 8.7 of Chapter 4 of Ethier and Kurtz (1986). There remains one further technical point: we need that C_c^∞ separates points [see Ethier and Kurtz (1986), page 113]. However, this is easily checked. $\square$

The proof of Corollary 1.2 follows directly from the proof of Theorem 1.1.

3. Extensions. Theorem 1.1 assumes that π_n has the product form given in (1.1). In this section we will discuss generalizations of this result.

Assume in this section that $\{\pi_n, n \ge 1\}$ is a sequence of densities satisfying the projective consistency requirement

$$\int \pi_{n+1}(\mathbf{x}^n, x_{n+1})\, dx_{n+1} = \pi_n(\mathbf{x}^n).$$

Crucial to the asymptotic arguments of Section 2 is the fact that the limiting value of $G_n V(\mathbf{x}^n)$ only depends on $\mathbf{x}^n$ through x_1. Therefore the first component of the process is asymptotically Markov. The following condition is therefore an essential condition in any generalization of Theorem 1.1

(E1) The tail σ-algebra, $\mathscr{T} = \bigcap_{n \geq 1} \sigma\{X_n,\ n \geq 1\}$ is π-trivial (where π is the appropriate limiting measure of the π_n's).

In a thermodynamic context, (E1) is a phase transition condition. The product form for π_n given in (1.1) is essentially the infinite temperature case.

It is difficult to formulate (E1) into a general result giving sufficient conditions for Theorem 1.1 to hold for a larger class of densities than those satisfying the form of (1.1). However, there are a number of interesting examples where extensions are possible. We briefly sketch three directions for extensions, although we do not give any formal proofs.

1. Suppose that π_n has the form

$$\pi_n(\mathbf{x}^n) = \prod_{i=1}^{n} f_i(\mathbf{x}_i).$$

We allow the functions to be different; however, in order for any sensible limit to be possible, an extra law of large numbers condition on these functions is necessary to ensure that analogy of Lemma 2.1 holds. Under such a condition, the proof of Theorem 1.1 can easily be generalized, and weak convergence is obtained to the Langevin diffusion in (1.2). We omit details of this. The most interesting consequence of this result is that although the form of $h(l)$ in this example will turn out to be more complicated than that appearing in (1.2), relative efficiency as a function of acceptance rate is unaltered, so that the "optimal efficiency" is again achieved at an acceptance rate of 0.234. The robustness of this result is the most useful practical implication of this paper. The product form density appearing here serves only to preclude the possibility of nontrivial tail σ-algebras. However the robustness of the relationship between acceptance rates and efficiency is likely to hold far more generally, where the tail σ-algebra is trivial.

2. Suppose π_n has the Markov form

$$\pi_n(\mathbf{x}^n) = f_1(x_1) \prod_{i=1}^{n-1} P(x_i, x_{i+1}),$$

where P is the transition kernal of an ergodic Markov chain (which therefore has a trivial trail σ-algebra). In this case a generalization of Theorem 1.1 is possible. Here the weak limiting process needs to be an infinite-dimensional diffusion to preserve the Markov property, and the details of the analogous results to Lemmas 2.1, 2.5 and 2.6 need to be more involved, requiring a rate of convergence condition (such as geometric ergodicity for P).

3. Suppose π_n has an exchangeable form. Suppose $X_1, X_2, \ldots$ is a random sequence distributed according to π_n. Then by de Finetti's theorem, for θ measurable in $\mathscr{T}$, $X_1|\theta, X_2|\theta, \ldots$ are conditionally iid. Therefore we would expect the limiting Langevin diffusion to be the conditioned diffusion

$$dX_{t,1} = h(l)^{1/2}\, dB_t + \frac{1}{2}h(l)\frac{d(\log k)}{dx}(X_{t,1})\, dt,$$

where k is the conditional density (assuming that this exists) of X_1 given θ, where θ is determined by the initial behavior of the initial sequence $\mathbf{X}_0$. Recall that the initial value for the Markov chain is a initial sequence from the stationary distribution of the limiting process on $\mathbb{R}^n$. In this example, the limiting probability measure is not even defined. This gives therefore a kind of asymptotic reducibility, and so the Metropolis algorithm in this case is not $O(n)$.

Acknowledgments. The authors would like to thank Yuval Peres, who helped to prove a special case of Theorem 1.1 to one of us. We also acknowledge useful conversations with Jeff Rosenthal.

REFERENCES

BESAG, J. and GREEN, P. J. (1993). Spatial statistics and Bayesian computation. *J. Roy. Statist. Soc. Ser. B* **55** 25–38.

BESAG, J., GREEN, P. J., HIGDON, D. and MENGERSEN, K. (1995). Bayesian computation and stochastic systems. *Statist. Sci.* **10** 3–66.

ETHIER, S. N. and KURTZ, T. G. (1986). *Markov Processes, Characterization and Convergence*. Wiley, New York.

FRIGESSI, A. and DEN HOLLANDER, F. (1994). A dynamical phase transition in a caricature of a spin glass. *J. Statist. Phys.* **75** 585–625.

GELMAN, A., ROBERTS, G. O. and GILKS, W. R. (1996). Efficient Metropolis jumping rules. *Bayesian Statistics 5* 599–608. Clarendon, Oxford.

MENGERSEN, K. L. and TWEEDIE, R. L. (1996). Rates of convergence of the Hastings and Metropolis algrithms. *Ann. Statist.* **24** 101–121.

METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H. and TELLER, E. (1953). Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21** 1087–1092.

ROBERTS, G. O. and SMITH, A. F. M. (1994). Simple conditions for the convergence of the Gibbs sampler and Hastings–Metropolis algorithms. *Stochastic Process. Appl.* **49** 207–216.

G. O. ROBERTS
STATISTICAL LABORATORY
UNIVERSITY OF CAMBRIDGE
CAMBRIDGE, CB2 1SB
UNITED KINGDOM
E-MAIL: g.o.roberts@statslab.cam.ac.uk

A. GELMAN
DEPARTMENT OF STATISTICS
COLUMBIA UNIVERSITY
NEW YORK, NEW YORK 10027
E-MAIL: gelman@stat.columbia.edu

W. R. GILKS
MRC BIOSTATISTICS UNIT
INSTITUTE OF PUBLIC HEALTH
CAMBRIDGE, CB2 2SR
UNITED KINGDOM
E-MAIL: wally.gilks@mrc-bsu.cam.ac.uk