Optimal scaling of discrete approximations to Langevin diffusions

Gareth O. Roberts†
University of Cambridge, UK

and Jeffrey S. Rosenthal
University of Toronto, Canada

[Received July 1995. Final revision April 1997]

Summary. We consider the optimal scaling problem for proposal distributions in Hastings–Metropolis algorithms derived from Langevin diffusions. We prove an asymptotic diffusion limit theorem and show that the relative efficiency of the algorithm can be characterized by its overall acceptance rate, independently of the target distribution. The asymptotically optimal acceptance rate is 0.574. We show that, as a function of dimension n, the complexity of the algorithm is $O(n^{1/3})$, which compares favourably with the $O(n)$ complexity of random walk Metropolis algorithms. We illustrate this comparison with some example simulations.

Keywords: Hastings–Metropolis algorithm; Langevin algorithm; Markov chain Monte Carlo method; Weak convergence

1. Introduction

This paper contains theoretical results related to the practical implementation of certain Metropolis–Hastings algorithms (Metropolis *et al.*, 1953; Hastings, 1970; Smith and Roberts, 1993) as used to explore probability distributions, e.g. in Bayesian statistics. Specifically, we consider issues related to discrete approximations to Langevin diffusions, as proposed in Grenander and Miller (1994), Phillips and Smith (1994) and Roberts and Tweedie (1996).

Hastings–Metropolis algorithms are now routinely used in many statistical applications (see for example Smith and Roberts (1993) and Besag and Green (1993)). The most commonly used algorithm of this type is the random walk algorithm, which is largely appealing because it is not problem specific and hence is easy to implement. However, as a result of this, it can frequently be slow to converge. Langevin algorithms use local problem-specific information and are therefore often almost as easy to implement.

In recent work of Roberts *et al.* (1997), the problem of optimal scaling of proposal variances for random walk Metropolis algorithms was considered. It was proved that, for Gaussian proposals and certain target distributions, the asymptotic acceptance probability should be tuned to be approximately 0.234 for optimal performance of the algorithm. Furthermore, it was shown that the proposal variance should scale as n^{-1} as the dimension $n \to \infty$. The paper thus provided a useful heuristic for running Metropolis algorithms

†*Address for correspondence*: Statistical Laboratory, University of Cambridge, 16 Mill Lane, Cambridge, CB2 1SB, UK.
E-mail: G.O.Roberts@statslab.cam.ac.uk

1369–7412/98/60255

efficiently. Although the result applied only asymptotically as $n \to \infty$, numerical studies (Gelman *et al.*, 1994) indicated that the result was a good approximation even in low dimensions. However, this result does not apply to more general Hastings algorithms. If the proposal density makes use of the structure of the target density, intuition suggests that a higher acceptance probability is likely to be preferred.

In this paper we carry out a similar study for a class of algorithms given by discrete approximations to Langevin diffusions. A Langevin diffusion for a multivariate probability density function π (with respect to Lebesgue measure) is a natural non-explosive diffusion which is reversible with respect to π. It makes use of the gradient of π to move more often in directions in which π is increasing. Thus, a discrete approximation to a Langevin diffusion should have an optimal acceptance probability which is larger than the 0.234 figure for random walk proposals.

Roberts and Tweedie (1996) demonstrated that Langevin algorithms can converge at subgeometric rates (at least for certain classes of target densities). However, it is also well known that Langevin algorithms can be significantly more efficient than their natural competitors, random walk Metropolis algorithms, particularly in high dimensional problems. In fact we shall be able to make direct comparisons of the efficiency of these algorithms asymptotically in dimension.

Our main results may be summarized as follows. For discrete approximations to Langevin diffusions for certain target distributions π, the optimal asymptotic scaling can be characterized as being that algorithm which has limiting acceptance probability in high dimensions (i.e. as $n \to \infty$) approximately 0.574. Furthermore, the proposal variance should scale as $n^{-1/3}$, and thus $O(n^{1/3})$ steps are required for the Langevin algorithm to converge. Therefore, Langevin algorithms are considerably more efficient than random-walk-based Metropolis methods which require $O(n)$ steps for the same class of target densities.

To understand this, we need to know by what criterion we are measuring optimality. Suppose that X is our Markov chain, and f is some function of interest, i.e. we wish to estimate $\pi(f) \equiv \mathbf{E}_\pi[f(X)]$. Assuming that a central limit theorem holds for X and f, a natural measure of efficiency is related to the variance of the ergodic estimate of $\pi(f)$:

$$e_f = \left[\lim_{t \to \infty}(t) \operatorname{var}\left\{ \sum_{i=1}^{t} f(X_i) \Big/ t \right\} \right]^{-1}.$$

Therefore e_f^{-1} is proportional to the number of iterations needed to achieve a particular accuracy for the ergodic estimate of $\pi(f)$. Unfortunately, the efficiency e_f varies with f, and furthermore there is no clear relationship between e_f and other natural measures of efficiency (perhaps because of the convergence rates of algorithms). However, for the diffusion limit, there is only one sensible measure of its efficiency, its speed measure. All other measures of efficiency are equivalent (up to a normalization constant), including those described above.

We note also that, whereas 0.574 is the optimal acceptance probability, the speed of the algorithm remains relatively high for acceptance probabilities between, say, 0.4 and 0.8. Practical implications such as these are considered in Section 3, and some simulations are described that show the asymptotic behaviour of the algorithms. In particular, a comparison between the Langevin algorithm and the random walk Metropolis algorithm, and a comparison with the asymptotically optimal algorithm is given in Fig. 1.

We prove our results formally only for target distributions of the form $\pi_n(\mathbf{x}) = \Pi_{i=1}^{n} f(x_i)$ corresponding to independent and identically distributed components. However, various generalizations are possible; see Section 4 and the similar discussion in Roberts *et al.* (1997), section 3. Furthermore, such optimal scaling results appear to be quite robust over changes in the model; see for example Gelman *et al.* (1994).

Similar algorithms have been studied in various contexts in the physics literature (Neal (1993), section 5.3, and Neal (1994)). Algorithms similar to discrete Langevin diffusions were proposed by Rossky *et al.* (1978). The idea that the proposal variance should scale as $n^{-1/3}$ is suggested in Kennedy and Pendleton (1991). Also, optimal acceptance probabilities are considered through simulations in Mountain and Thirumalai (1994).

Our formal definitions are as follows. The reversible Langevin diffusion for the n-dimensional density π_n, with variance σ^2, is the diffusion process $\{\mathbf{\Lambda}_t\}$ which satisfies the stochastic differential equation

$$\mathrm{d}\mathbf{\Lambda}_t = \sigma\,\mathrm{d}\mathbf{B}_t + \frac{\sigma^2}{2}\nabla \log\{\pi_n(\mathbf{\Lambda}_t)\}\,\mathrm{d}t,$$

where $\mathbf{B}_t$ is standard n-dimensional Brownian motion. Thus, the natural discrete approximation can be written

$$\tilde{\mathbf{\Lambda}}_{t+1} = \tilde{\mathbf{\Lambda}}_t + \sigma_n \mathbf{Z}_{t+1} + \frac{\sigma_n^2}{2}\nabla \log\{\pi_n(\tilde{\mathbf{\Lambda}}_t)\}$$

where the random variables $\mathbf{Z}_t$ are distributed as independent standard normal, and where σ_n^2 is the chosen step variance. However, such discrete approximations can have vastly different asymptotic behaviours from the diffusion process that they attempt to approximate (Roberts and Tweedie, 1996). Specifically, these approximations can be transient, no matter how small the step variance σ_n^2. Therefore, to construct a sound algorithm based on the Langevin diffusion, it is not sufficient merely to approximate the diffusion itself. It is necessary to introduce a Metropolis accept or reject step (Metropolis *et al.*, 1953; Hastings, 1970) which ensures that π_n is a stationary distribution for the process.

The algorithm proceeds as follows. Given $\mathbf{X}_t$, we choose a proposal random variable $\mathbf{Y}_{t+1}$ by

$$\mathbf{Y}_{t+1} = \mathbf{X}_t + \sigma_n \mathbf{Z}_{t+1} + \frac{\sigma_n^2}{2}\nabla \log\{\pi_n(\mathbf{X}_t)\}$$

and then set $\mathbf{X}_{t+1} = \mathbf{Y}_{t+1}$ with probability

$$\alpha_n(\mathbf{X}_t,\ \mathbf{Y}_{t+1}) = \frac{\pi_n(\mathbf{Y}_{t+1})\, q_n(\mathbf{Y}_{t+1},\ \mathbf{X}_t)}{\pi_n(\mathbf{X}_t)\, q_n(\mathbf{X}_t,\ \mathbf{Y}_{t+1})} \wedge 1$$

where

$$q_n(\mathbf{x},\ \mathbf{y}) = \frac{1}{(2\pi\sigma_n^2)^{n/2}} \exp\left[\frac{-1}{2\sigma_n^2}\left\|\mathbf{y} - \mathbf{x} - \frac{\sigma_n^2}{2}\nabla \log\{\pi_n(\mathbf{x})\}\right\|_2^2\right] \equiv \prod_{i=1}^{n} q(x_i^n,\ y_i),$$

and $\|\cdot\|_2$ is the usual L^2-norm. Otherwise, with probability $1 - \alpha_n(\mathbf{X}_t,\ \mathbf{Y}_{t+1})$, we set $\mathbf{X}_{t+1} = \mathbf{X}_t$.

Thus the discrete algorithm has the desired stationary distribution π_n. However, the practical problem of determining the size of σ_n^2 remains. Specifically, a larger value of σ_n^2 corresponds to a larger proposal step size. This potentially allows for faster mixing, but only if the acceptance probabilities do not become unacceptably small. Such issues are the subject of the present paper.

2. Main results

We consider the Metropolis-adjusted discrete approximations $\{\mathbf{X}_t\}$ to the Langevin diffusion for π_n as above, with

$$\pi_n(\mathbf{x}) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \exp\{g(x_i)\} \tag{2.1}$$

a fixed probability distribution on $\mathbf{R}^n$. Throughout, we shall assume that $\mathbf{X}_0$ is distributed according to the stationary measure π. We further assume that g is a C^8-function with derivatives $g^{(i)}$ satisfying

$$|g(x)|, |g^{(i)}(x)| \leqslant M_0(x), \tag{2.2}$$

$1 \leqslant i \leqslant 8$, for some polynomial $M_0(\cdot)$, and that

$$\int_{\mathbf{R}} x^k f(x) \, \mathrm{d}x < \infty, \qquad k = 1, 2, 3, \ldots . \tag{2.3}$$

Finally, to apply standard stochastic differential equation results, we assume that g' is a Lipschitz function.

To compare these discrete approximations with limiting continuous time processes, it is convenient to define the discrete approximations as jump processes with exponential holding times. Specifically, we let $\{J_t\}$ be a Poisson process with rate $n^{1/3}$ and let $\mathbf{\Gamma}^n = \{\mathbf{\Gamma}^n_t\}_{t \geqslant 0}$ be the n-dimensional jump process defined by $\mathbf{\Gamma}^n_t = \mathbf{X}_{J_t}$ where we take $\sigma_n^2 = l^2 n^{-1/3}$ in the definitions from the previous section, with l an arbitrary positive constant. We assume throughout that $\{\mathbf{X}_t\}$ is non-explosive. We let

$$a_n(l) = \int\int \pi_n(\mathbf{x}) \, q_n(\mathbf{x}, \mathbf{y}) \, \alpha_n(\mathbf{x}, \mathbf{y}) \, \mathrm{d}\mathbf{x} \, \mathrm{d}\mathbf{y} \;=\; \mathbf{E}\left[\frac{\pi_n(\mathbf{Y}) \, q_n(\mathbf{Y}, \mathbf{X})}{\pi_n(\mathbf{X}) \, q_n(\mathbf{X}, \mathbf{Y})} \wedge 1\right]$$

be the π_n-average acceptance rate of the algorithm which generates $\mathbf{\Gamma}$.

The two main results in this paper are the following.

Theorem 1. We have

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

where $a(l) = 2\,\Phi(-Kl^3/2)$, with

$$\Phi(x) = \frac{1}{\surd(2\pi)} \int_{-\infty}^{x} \exp\left(-\frac{t^2}{2}\right) \mathrm{d}t$$

and

$$K = \sqrt{\mathbf{E}\left[\frac{5\,g'''(X)^2 - 3\,g''(X)^3}{48}\right]} > 0, \tag{2.4}$$

with the expectation taken over X having density $f = \exp(g)$.

Theorem 1 gives a formula for the asymptotic acceptance probability of the algorithm. We note that, in the Gaussian case where $g(x) = -x^2/2 + C$, this theorem reduces precisely to equation (2.3) of Kennedy and Pendleton (1991).

Theorem 2. Let $\{U^n\}_{t \geqslant 0}$ be the process corresponding to the first component of $\mathbf{\Gamma}^n$. Then, as $n \to \infty$, the process U^n converges weakly (in the Skorokhod topology) to the Langevin diffusion $\mathbf{U}$ defined by

$$\mathrm{d}U_t = h(l)^{1/2}\,\mathrm{d}B_t + \frac{1}{2}\,h(l)\frac{\mathrm{d}}{\mathrm{d}x}\,\log\{\pi_1(U_t)\}\,\mathrm{d}t,$$

where $h(l) = 2l^2\,\Phi(-Kl^3/2)$ is the speed of the limiting diffusion. Furthermore, $h(l)$ is maximized at the unique value of l for which $a(l) = 0.574$ (to three decimal places).

Theorem 2 may be interpreted as follows. For a given target density π_n as above, with n large, suppose that a Metropolis-adjusted discrete approximation to the Langevin diffusion for π_n is run with proposal steps of variance σ_n^2. Then, setting $l_n = \sigma_n n^{1/6}$, the theorem says that the speed (which is proportional to the mixing rate) of the process is approximately given by $h(l_n)$. Furthermore, the optimal value $\hat{l}_n$ of l_n which maximizes this speed is that for which the asymptotic acceptance probability $a_n(\hat{l}_n)$ is approximately 0.574. Hence σ_n^2 should be tuned to be approximately $\hat{l}_n^2\,n^{-1/3}$, which will make the acceptance probability approximately 0.574. If it is discovered that the acceptance rate is substantially smaller or substantially larger than 0.574, then the value of σ_n^2 should be modified accordingly.

3. Practical implications

In this section, we shall illustrate our results with a collection of simulation studies. The asymptotic results of Section 2 hold approximately in rather low dimensional problems as we shall observe. We shall also see the efficiency advantage of the Langevin algorithm over the Metropolis algorithm ($O(n^{2/3})$ as discussed earlier).

Measurements of efficiency in low dimensional Markov chains are not unique (see Gelman *et al.* (1994) for a discussion). Perhaps the most natural measure of efficiency would be the asymptotic variance of a relevant quantity of interest. Since we do not wish to specify such a quantity of interest, we have opted instead for a different (but asymptotically equivalent) criterion. We therefore measure the average squared jumping distance for the algorithm, which we call the *first-order efficiency*.

Definition 1. First-order efficiency of a multidimensional Markov chain $\mathbf{X}$ with first component $X^{(1)}$ say is defined by

$$\text{first-order efficiency} = \mathbf{E}[(X_{t+1}^{(1)} - X_t^{(1)})^2]$$

where $\mathbf{X}_t$ is assumed to be stationary.

We note that maximizing this quantity is equivalent to minimizing the lag 1 autocorrelations. Recall also that all sensible measurements of efficiency of the algorithm are asymptotically equivalent (up to a normalization constant) in the diffusion limit.

Fig. 1 shows estimated first-order efficiencies as a function of the acceptance rate (what we term the *efficiency curve*) for a product of standard normals. Dimensions 1, 5, 10 and 20 are given, and in each case two envelopes of curves are given: the upper envelope for each graph represents the first-order efficiency for the Langevin algorithm, and the lower envelope, the random walk Metropolis algorithm. The full curve represents the asymptotic efficiency curve normalized to have the same maximum value as the finite dimensional Langevin efficiency. Each point on the graph represents an estimate of first-order efficiency based on 2×10^5 iterations of the algorithm, and the roughness of the curves shows that some error in our estimates of efficiency still exist, although the error is not sufficiently large to obscure the

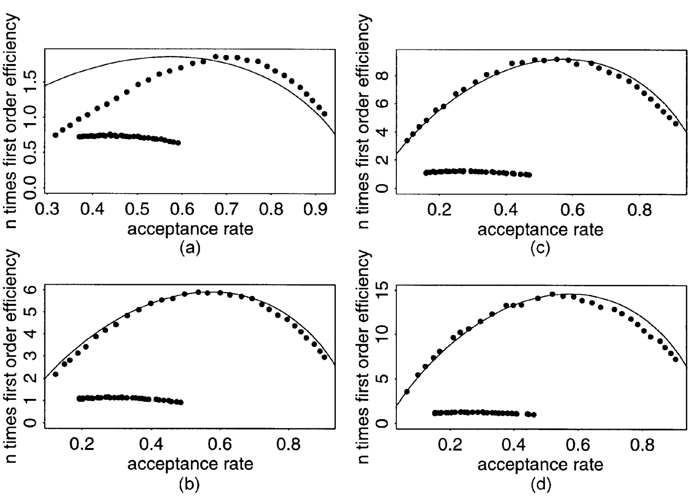

Fig. 1. First-order efficiency of Langevin and random walk Metropolis algorithms as a function of overall acceptance rates (efficiency curves) (———, asymptotic curve for the Langevin algorithm; the upper envelope of points shows the Langevin algorithm and the lower envelope shows the Metropolis algorithm): (a) $n = 1$; (b) $n = 5$; (c) $n = 10$; (d) $n = 20$

general picture. (Note that we plot first-order efficiency times dimension as this is the scale on which the Metropolis algorithm's efficiency is asymptotically constant in dimension.) The value of the first-order efficiency of the Langevin algorithm is therefore an absolute measure of efficiency *relative to* the Metropolis dynamics.

The efficiency gain for the Langevin method is immediately obvious and is increasing with dimension. In one dimension, the optimal acceptance rate is in excess of 0.7; however, in five or more dimensions the optimal acceptance rate is very close to the limiting value of 0.574. Furthermore, the whole efficiency curve is very well approximated by the asymptotic curve for dimensions greater than 5.

Where f is not symmetric, the asymptotic picture is not as clear in low dimensions. Fig. 2 considers the 10-dimensional case where $g = -x^2/2 - \exp(x^{-2}/2), x \geqslant 0$, and $g = -x^2/2$ for $x < 0$. In this case the efficiency curve is shifted towards lower acceptance rates, and in particular the optimal acceptance rate is less than 0.5. However, the asymptotically optimal acceptance rate (0.574) still gives a relative efficiency for this particular problem in excess of 0.95. Furthermore, additional simulations for the same density f, but in 1000 dimensions, show excellent agreement with the asymptotic efficiency curve.

Finally in this section, we give a word of caution. The main results in this paper require very smooth regularity conditions in g. It is tempting to suggest that these are technical conditions required for the proofs, but perhaps inconsequent to the actual asymptotic efficiency curve in practical situations. Although it is highly unlikely that the C^8-condition which we impose on g is *necessary* for the results in this paper, some smoothness conditions are necessary.

Consider, for instance, the following example. Suppose that $f(x) = \frac{2}{3}$ for $0 \leqslant x \leqslant 1$ and $f(x) = \frac{1}{3}$ for $-1 \leqslant x < 0$, with $f(x) = 0$ for $|x| > 1$. In this case, since $g'(x) = 0$ almost everywhere, the Langevin algorithm coincides exactly with the random walk Metropolis algorithm (from almost all starting values). Therefore, since the asymptotic relative efficiency curves for these algorithms are different, it is impossible that our results and those of Roberts

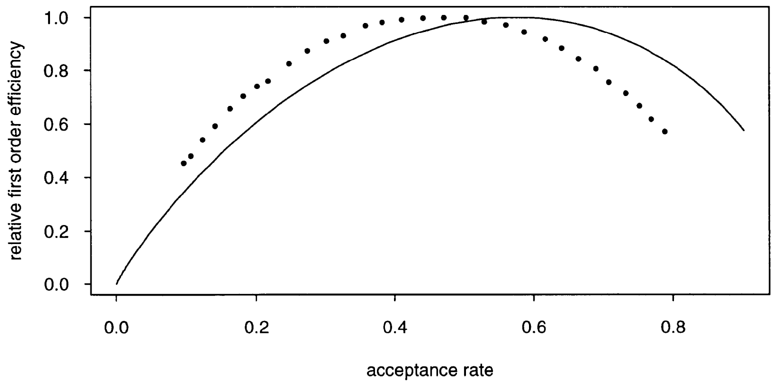

Fig. 2. Relative efficiency of a non-symmetric 10-dimensional density (●) and the asymptotic curve (———)

et al. (1997) both apply to this example; this shows that the discontinuity of f (and hence of g) is significant. Furthermore, additional simulations (not given here) involving continuous log-densities g with discontinuous derivatives suggest that the asymptotics are very different, and asymptotically optimal acceptance rates are likely to be less than 0.574.

4. Extensions

Our main results are proved for sequences of densities of the product form (2.1). It appears that the asymptotic efficiency curve (efficiency against acceptance rate) is robust to changes in correlation structure in several different extensions, although some of the technical issues addressed in the proofs of Appendix A are even more delicate in such extensions. Some of these types of extensions are discussed in Roberts *et al.* (1997). Here we shall briefly discuss one extension to illuminate why the efficiency curve remains invariant, even in the presence of correlation. No attempt at a formal proof will be made here. Consider a sequence of densities $\{\pi_n\}$ with

$$\pi_n(\mathbf{x}) = f_0(\mathbf{x}_0) \prod_{i=0}^{n} f(\mathbf{x}_i)$$

where we assume that the $\{\mathbf{x}_i\}$ are a-vectors for some $a > 1$, and $\mathbf{x}_0$ is a vector (not necessarily a dimensional).

Suppose that we use the usual scaling for the proposal variance, $\sigma_n^2 = l^2 n^{-1/3}$, and consider the process produced by just looking at the $\mathbf{x}_0$-co-ordinates. The formal asymptotics leading to theorems 1 and 2 still apply here (although with more involved technicalities), so that analogously the $\mathbf{x}_0$-process has a multidimensional diffusion limit with scaling:

$$h(l) = 2l^2\, \Phi(-K_{\text{mult}} l^3/2) \tag{4.1}$$

where K_{mult} is a multidimensional analogue of K appearing in equation (2.4), which incorporates information about correlation in π_n. However, analogously to theorem 1, the acceptance rate also satisfies

$$a(l) = 2\, \Phi(-K_{\text{mult}} l^3/2). \tag{4.2}$$

It remains to observe that the efficiency curve from equations (4.1) and (4.2) only varies with

K_{mult} via a multiplicative constant. Specifically, given a constant $0 < a_0 < 1$, and the scaling $l(a_0)$ which achieves acceptance rate a_0,

$$h\{l(a_0)\} = \frac{a[-2\,\Phi^{-1}\{a(l)/2\}]^{2/3}}{K_{\mathrm{mult}}^{2/3}}.$$

Thus, the relative efficiency curve remains unaltered.

We stress that the absolute efficiency of the algorithm varies considerably with many properties of the target density, though the relative efficiency as a function of acceptance rate remains unaltered. Similar arguments can be used to justify (at least heuristically) other extensions of the results of Section 2.

We have only considered the spherical proposal case here. For problems where components have very different scales, it will be sensible to allow differently scaled proposals in different dimensions. Some interesting problems are involved in choosing variance scaling componentwise, perhaps also in an attempt to allow some components to converge quicker than others. We do not pursue these problems here.

Finally, we note that extensions to weak convergence of multivariate components are also possible. For instance, with π_n of the form given in equation (2.1), it is possible (by very similar techniques to those used below) to show that the process described by the first $c > 1$ components of $\mathbf{\Gamma}^n$ converges weakly to an n-dimensional Markov process consisting of independent components, each of the type given by the $\mathbf{U}$ of theorem 2.

Acknowledgements

We are grateful to Radford Neal for bringing references from the physics literature to our attention. We also thank Richard Gibbens, Michael Miller and Richard Tweedie, for helpful discussions, and would like to thank the referees very much for their work. This work was supported in part by the Engineering and Physical Sciences Research Council of the UK and by the Natural Sciences and Engineering Research Council of Canada.

Appendix A: Theorem proofs

Let us define the generators of the discrete approximation process $\mathbf{\Gamma}^n$ and of the (first-component) Langevin diffusion process with speed $h(l)$, i.e.

$$G_n\, V(\mathbf{x}^n) = n^{1/3}\,\mathbf{E}\left[\{V(\mathbf{Y}) - V(\mathbf{x}^n)\}\left\{\frac{\pi_n(\mathbf{Y})\,q_n(\mathbf{Y},\mathbf{x}^n)}{\pi_n(\mathbf{x}^n)\,q_n(\mathbf{x}^n,\mathbf{Y})} \wedge 1\right\}\right],$$

where the expectation is taken over $\mathbf{Y} \sim q_n(\mathbf{x}^n, \cdot)$, and

$$G\,V(\mathbf{x}^n) = h(l)\{\tfrac{1}{2}\,V''(x_1) + \tfrac{1}{2}\,g'(x_1)\,V'(x_1)\}$$

(where $g(x_1) = \log\{f(x_1)\}$ as before).

To prove the weak convergence of the processes as in theorem 2 it suffices (Ethier and Kurtz (1986), chapter 4, corollary 8.7) to show that there are events $F_n^* \subseteq \mathbf{R}^n$ such that for all t

$$\mathbf{P}(\mathbf{\Gamma}_s^n \in F_n^* \text{ for all } 0 \leqslant s \leqslant t) \to 1 \tag{A.1}$$

and

$$\lim_{n\to\infty}\,\sup_{\mathbf{x}^n \in F_n^*} |G_n\,V(\mathbf{x}) - G\,V(\mathbf{x})| \;=\; 0$$

for all test functions V in the domain of a 'core' for the generator G, provided that this domain strongly separates points. In the present context, by the nature of G, we can restrict to functions V which depend only on the first co-ordinate x_1 of $\mathbf{x}^n$. Furthermore, we may restrict (Ethier and Kurtz (1986), chapter 8, theorem 2.1) to functions V which are in C_c^∞, i.e. which are infinitely differentiable with compact support.

The essence of the proof will be showing the uniform convergence of G_n to G, as $n \to \infty$ (and hence $\sigma_n^2 \to 0$), as above. This will involve careful Taylor series expansions with uniform bounds on remainder terms. It will also involve a quantitative version of the Lindeberg central limit theorem.

To proceed, we expand $G_n\,V(\mathbf{x})$ in a power series involving powers of $n^{-1/6}$.

Lemma 1. Defining Z_i by

$$Y_i = x_i^n + \sigma_n Z_i + \frac{\sigma_n^2}{2}\, g'(x_i^n)$$

(so that Z_i is distributed as standard normal), and recalling that

$$q(x, y) = \frac{1}{\sqrt{(2\pi\sigma_n^2)}} \exp\left[-\frac{1}{2\sigma_n^2}\left\{y - x - \frac{\sigma_n^2}{2}\, g'(x)\right\}^2\right],$$

there is a sequence of sets $F_n \in \mathbf{R}^n$, with $\lim_{n\to\infty}\{n^{1/3}\,\pi_n(F_n^C)\} = 0$, such that

$$\log\left\{\frac{f(Y_i)\,q(Y_i, x_i^n)}{f(x_i^n)\,q(x_i^n, Y_i)}\right\} = C_3(x_i^n, Z_i)n^{-1/2} + C_4(x_i^n, Z_i)n^{-2/3} + C_5(x_i^n, Z_i)n^{-5/6} + C_6(x_i^n, Z_i)n^{-1} + C_7(x_i^n, Z_i, \sigma_n),$$

where

$$C_3(x_i^n, Z_i) = l^3\left\{-\frac{1}{4} Z_i\, g'(x_i^n)\, g''(x_i^n) - \frac{1}{12} Z_i^3\, g'''(x_i^n)\right\},$$

and where $C_4(x_i^n, Z_i)$, $C_5(x_i^n, Z_i)$ and $C_6(x_i^n, Z_i)$ are (also) polynomials in Z_i and the derivatives of g. Furthermore, if $\mathbf{E}_Z$ denotes expectation with $Z \sim N(0, 1)$, and $\mathbf{E}_X$ denotes expectation with X having density $f(\cdot)$, then

$$\mathbf{E}_X\mathbf{E}_Z[C_3(X, Z)] = \mathbf{E}_X\mathbf{E}_Z[C_4(X, Z)] = \mathbf{E}_X\mathbf{E}_Z[C_5(X, Z)] = 0,$$

whereas

$$\mathbf{E}_X\,\mathbf{E}_Z[C_3(X, Z)^2] = l^6K^2 = -2\mathbf{E}_X\,\mathbf{E}_Z[C_6(X, Z)] > 0.$$

In addition,

$$\lim_{n\to\infty}\sup_{\mathbf{x}^n\in F_n}\left[\mathbf{E}\left|\sum_{i=2}^n \log\left\{\frac{f(Y_i)\,q(Y_i, x_i^n)}{f(x_i^n)\,q(x_i^n, Y_i)}\right\} - \left\{n^{-1/2}\sum_{i=2}^n C_3(x_i^n, Z_i) - \frac{l^6K^2}{2}\right\}\right|\right] = 0.$$

Proof. The (Taylor series) expansion follows from straightforward (but messy) computation, done using the MATHEMATICA computation system (Wolfram, 1988). By inspection of the result, the coefficients are polynomials in Z_i and in g and its derivatives. The fact that

$$\mathbf{E}_X\,\mathbf{E}_Z[C_3(X, Z)] = \mathbf{E}_X\,\mathbf{E}_Z[C_5(X, Z)] = 0$$

is then immediate because these coefficients contain only terms involving odd powers of Z. The facts that $\mathbf{E}_X\,\mathbf{E}_Z[C_4(X, Z)] = 0$ and that

$$\mathbf{E}_X\{\mathbf{E}_Z[C_3(X, Z)^2] + 2\,\mathbf{E}_Z[C_6(X, Z)]\} = 0$$

follow from first replacing the even powers of Z by the appropriate moments of the standard normal distribution, and then finding (again using MATHEMATICA) explicit antiderivatives of $\exp\{g(x)\}C_4$

and $\exp\{g(x)\}(C_3^2 + 2C_6)$ respectively, which are of the form $\exp\{g(x)\}$ times a polynomial in derivatives of $g(x)$, and thus clearly approach 0 as $x \to \pm\infty$. (Note: the *existence* of $\mathbf{E}_X\,\mathbf{E}_Z[C_3(X, Z)^2]$, and hence of $\mathbf{E}_X\,\mathbf{E}_Z[C_6(X, Z)]$, follows since all moments of π_n exist.)

We now construct the sequence of sets F_n on which we uniformly control each of the four terms in the expansion (excluding just the C_3-term). The only thing to ensure is that $\lim_{n\to\infty}\{n^{1/3}\pi_n(F_n^C)\} = 0$. For $j = 4, 5, 6$, set $C_j(x) = \mathbf{E}_Z\,[C_j(x, Z)]$, and $V_j(x_i) = \operatorname{var}\{C_j(x_i, Z)\}$. Because of the polynomial restrictions on g, C_j and V_j are bounded by polynomials. Now,

$$\mathbf{E}\left[\left\{\sum_{i=2}^{n} C_j(x_i, Z_i) - \mathbf{E}_X[C_j(X)]\right\}^2\right] = \sum_{i=2}^{n} V_j(x_i) + \left[\sum_{i=2}^{n}\{C_j(x_i) - \mathbf{E}_X[C_j(X)]\}\right]^2.$$

So setting

$$F_{n,j} = \left\{x; \left|\sum_{i=2}^{n}\{C_j(x_i^n) - \mathbf{E}_X[C_j(X)]\}\right| < n^{5/8}\right\} \bigcap \left\{x; \left|\sum_{i=2}^{n}\{V_j(x_i^n) - \mathbf{E}_X[V_j(X)]\}\right| < n^{6/5}\right\}$$

it is easy to show by Markov's inequality applied to moments of the constrained functions that $n^{1/3}\,\pi_n(F_{n,j}^c) \to 0$ as $n \to \infty$, and, for $x \in F_{n,j}$,

$$\mathbf{E}\left[\left\{\sum_{i=2}^{n} C_j(x_i, Z_i) - \mathbf{E}_X[C_j(X)]\right\}^2\right] \leqslant O(n^{5/4}).$$

L^1-convergence of the fourth, fifth and sixth terms in the Taylor expansion is thus assured.

It remains to consider $C_7(x_i^n, Z_i, \sigma_n)$. However, by using the remainder formula of the Taylor series expansion, and again using inequality (2.2), we derive the bound

$$|C_7(x_i^n, Z_i, \sigma_n)| \leqslant n^{-7/6}\,p(x_i^n, w)$$

for some polynomial p with either $0 \leqslant w_i \leqslant Z_i$ or $Z_i \leqslant w_i \leqslant 0$. Now we can always take such a bounding polynomial to be of the form $A(1 + x_i^N)(1 + w^N)$ for sufficiently large A and for a sufficiently large even integer N. Since we can bound this polynomial in turn by $A(1 + x_i^N)(1 + Z_i^N)$ and since all polynomial moments of Z_i exist, it follows that we can write

$$\mathbf{E}_Z|C_7(x_i^n, Z, \sigma_n)| \leqslant n^{-7/6}\,p(x_i^n)$$

for a suitable polynomial $p(\cdot)$. Then, setting $u_7 = \mathbf{E}_X[p(X)]$, $v_7 = \operatorname{var}_X\{p(X)\}$ and

$$F_{n,7} = \left\{\mathbf{x}^n \in \mathbf{R}^n; \left|\frac{1}{n}\sum_{i=1}^{n} p(x_i^n) - u_7\right| < 1\right\},$$

Chebychev's inequality implies that

$$\pi_n(F_{n,7}^C) \leqslant v_7 n^{-1}.$$

Furthermore, for $\mathbf{x}^n \in F_{n,7}$,

$$\sum_{i=2}^{n} \mathbf{E}_Z|C_7(x_i^n, Z, \sigma_n)| \leqslant (u_7 + 1)n^{-1/6}.$$

We put $F_n = F_{n,4} \cap F_{n,5} \cap F_{n,6} \cap F_{n,7}$. On F_n, terms 4, 5, 6 and 7 converge uniformly on L^1, and it follows that the third term must also. Thus the last statement of the lemma also holds. □

The main point of this lemma is that, in the log-expansion of the proposal density components, the terms corresponding to $n^{-1/6}$ and $n^{-1/3}$ vanish, and the next three terms each have vanishing expectation. (This is in contrast with the situation for random walk Metropolis expansions, in which no terms cancel and only the first has vanishing expectation.)

To continue, we define $\tilde{G}_n$ by

$$\tilde{G}_n\,V(\mathbf{x}^n) = n^{1/3}\mathbf{E}\left[\{V(\mathbf{Y}) - V(\mathbf{x}^n)\}\prod_{i=2}^{n}\frac{f(y_i)\exp[-(1/2\sigma_n^2)\{x_i^n - y_i - (\sigma_n^2/2)\,g'(y_i)\}^2]}{f(x_i^n)\exp[-(1/2\sigma_n^2)\{y_i - x_i^n - (\sigma_n^2/2)\,g'(x_i^n)\}^2]} \wedge 1\right].$$

In fact, $\tilde{G}_n$ is like G_n except that the product omits the factor corresponding to $i = 1$. The following theorem shows that this omission is unimportant.

Theorem 3. There are sets $S_n \subseteq \mathbf{R}^n$ with $n^{1/3}\,\pi_n(S_n^C) \to 0$ such that, for any $V \in C_c^\infty$,

$$\lim_{n\to\infty} \sup_{\mathbf{x}^n \in S_n} \big| G_n\, V(\mathbf{x}^n) - \tilde{G}_n\, V(\mathbf{x}^n) \big| \;=\; 0.$$

Moreover,

$$\lim_{n\to\infty} \sup_{\mathbf{x}^n \in S_n} \left(\mathbf{E} \left| \frac{\pi_n(\mathbf{Y})\, q_n(\mathbf{Y}, \mathbf{x}^n)}{\pi_n(\mathbf{x}^n)\, q_n(\mathbf{x}^n, \mathbf{Y})} \wedge 1 - \prod_{i=2}^{n} \frac{f(y_i) \exp[-(1/2\sigma_n^2)\{x_i^n - y_i - (\sigma_n^2/2)\, g'(y_i)\}^2]}{f(x_i^n) \exp[-(1/2\sigma_n^2)\{y_i - x_i^n - (\sigma_n^2/2)\, g'(x_i^n)\}^2]} \wedge 1 \right| \right) = 0$$

Proof. Since the function $x \mapsto \exp(x) \wedge 1$ has Lipschitz constant 1, and since $Y_1 = x_1^n + \sigma_n Z + \frac{1}{2}\sigma_n^2\, g'(x_1^n)$, where $Z \sim N(0, 1)$, it follows that

$$\big| G_n\, V(\mathbf{x}^n) - \tilde{G}_n\, V(\mathbf{x}^n) \big| \leqslant n^{1/3}\, \mathbf{E}[|V(\mathbf{Y}) - V(\mathbf{x}^n)||R(x_1^n, Z, \sigma_n)|]$$

where

$$R(x, z, \sigma) = \frac{1}{2} \left(z^2 - \left[z + \frac{\sigma}{2}\, g'(x) + \frac{\sigma}{2}\, g' \left\{ x + \frac{\sigma^2}{2}\, g'(x) + \sigma z \right\} \right]^2 \right) + g \left\{ x + \frac{\sigma^2}{2}\, g'(x) + \sigma z \right\} - g(x).$$

By a first-order Taylor series expansion in σ, as in the argument for the proof of lemma 1, with the integral form of the remainder, we obtain

$$|R(x, z, \sigma_n)| \leqslant M_1(x)\, M_2(z) n^{-1/3} \tag{A.2}$$

for suitable positive polynomials $M_1(\cdot)$ and $M_2(\cdot)$.

Finally, since $V \in C_c^\infty$, there is a constant K_1 such that

$$|V(\mathbf{Y}) - V(\mathbf{x}^n)| \leqslant K_1 |Y_1 - x_1^n|.$$

Furthermore, since

$$|Y_1 - x_1^n| \leqslant \sigma_n |Z_1| + \frac{\sigma_n^2}{2} |g'(x_1^n)|,$$

and recalling that g' is assumed to be Lipschitz, we can write $g'(x) \leqslant K_2(1 + |x|)$ for a constant $K_2 \geqslant 1$. It follows that

$$n^{1/3} |V(\mathbf{Y}) - V(\mathbf{x}^n)||R(x_1^n, Z, \sigma_n)| \leqslant K_1 K_2 (1 + |x|)\, M_1(x) \left\{ \sigma_n |Z|\, M_2(Z) + \frac{\sigma_n^2}{2}\, M_2(Z) \right\}. \tag{A.3}$$

We now set S_n to be the set on which $M_1(x_1^n) \leqslant n^{1/12}$. By inequality (A.3), and recalling that $Z \sim N(0, 1)$ so that $M_2(Z)$ and $|Z|\, M_2(Z)$ are integrable, and using Markov's inequality, we see that

$$\pi_n(S_n^C) \;=\; \pi_n\big\{M_1(x_1^n)^5 \geqslant n^{5/12}\big\} \leqslant n^{-5/12}\, \mathbf{E}\big[M_1(x_1^n)^5\big].$$

The first result follows. The proof of the second is virtually identical, except that the estimation of $V(\mathbf{Y}) - V(\mathbf{x}^n)$ is not necessary. □

Lemma 2. There are sets $T_n \subseteq \mathbf{R}^n$ with $n^{1/3}\,\pi_n(T_n^C) \to 0$ such that, for any $V \in C_c^\infty$,

$$\lim_{n\to\infty} \sup_{\mathbf{x}^n \in T_n} \left| n^{1/3}\, \mathbf{E}\big[V(Y_1) - V(x_1^n)\big] \;-\; \frac{l^2}{2} \{V''(x_1^n) + g'(x_1^n)\, V'(x_1^n)\} \right| = 0.$$

Proof. We write

$$Y_1 = x_1^n + \sigma_n Z + \tfrac{1}{2}\sigma_n^2\, g'(x_1^n)$$

and

$$W(x, z, \sigma) = V\left\{x + \sigma z + \frac{\sigma^2}{2}\, g'(x)\right\}.$$

A second-order Taylor series expansion with respect to σ_n then gives that

$$\mathbf{E}[V(Y_1) - V(x_1^n)] = \frac{\sigma_n^2}{2}\{V''(x_1^n) + g'(x_1^n)\, V'(x_1^n)\} + \mathbf{E}\left[\int_0^{\sigma_n} \frac{\partial^3 V}{\partial \sigma^3}(x_1^n, z, \epsilon)\, \frac{(\sigma_n - \epsilon)^2}{2}\, \mathrm{d}\epsilon\right].$$

Therefore, by inequality (2.2), and an argument for the remainder term similar to that used in the proof of lemma 1, there is a polynomial $M_3(\cdot)$ such that the remainder term is less than $n^{-1/2}\, M_3(x_1^n)$. Letting T_n be the set on which $M_3(x_1^n) \leqslant n^{1/12}$, the result follows by Markov's inequality as in the previous lemma. □

To proceed, we make some further definitions. Let $a(x) = -\frac{1}{4}\, g'(x)\, g''(x)$ and $b(x) = -\frac{1}{12} g'''(x)$, so that with $C_3(x, z)$ as in lemma 1 we have

$$C_3(x, z) = l^3\{a(x)z + b(x)z^3\}.$$

Set

$$Q_n(\mathbf{x}^n; \cdot) = \mathcal{L}\left\{n^{-1/2} \sum_{i=2}^{n} C_3(x_i^n, Z_i)\right\}$$

and let

$$\phi_n(\mathbf{x}^n; t) = \int \exp(\mathrm{i}tw)\, Q_n(\mathrm{d}w)$$

be the corresponding characteristic function. Finally, let

$$\phi(t) = \exp(-t^2 K^2/2)$$

be the characteristic function of the distribution $N(0, K^2)$, with K as in theorem 1.

Lemma 3. There is a sequence of sets $H_n \subseteq \mathbf{R}^n$ such that

(a)

$$\lim_{n\to\infty} \{n^{1/3}\, \pi_n(H_n^c\} = 0,$$

(b) for all $t \in \mathbf{R}$,

$$\lim_{n\to\infty} \sup_{\mathbf{x}^n \in H_n} |\phi_n(\mathbf{x}^n; t) - \phi(t)| = 0,$$

(c) For all bounded continuous functions r,

$$\lim_{n\to\infty} \sup_{\mathbf{x}^n \in H_n} \left| \int_{\mathbf{R}} Q_n(\mathbf{x}^n, \mathrm{d}y)\, r(y) - \frac{1}{\surd(2\pi l^6 K^2)} \int_{\mathbf{R}} r(y) \exp\left(-\frac{y^2}{l^6 K^2}\right) \mathrm{d}y \right| = 0$$

where K is as in theorem 1,

(d)

$$\lim_{n\to\infty} \sup_{\mathbf{x}^n \in H_n} \left| \mathbf{E}_Z\left[1 \wedge \exp\left\{n^{-1/2} \sum_{i=2}^{n} C_3(x_i^n, Z_i) - \frac{l^6 K^2}{2}\right\}\right] - 2\,\Phi\left(-\frac{l^3 K}{2}\right) \right| = 0.$$

Proof. We define H_n as a region on which certain functionals have average value that is close to their mean. Specifically, we let H_n be the set of $\mathbf{x}^n \in \mathbf{R}^n$ such that

$$\left| \frac{1}{n} \sum_{i=2}^{n} h(x_i^n) - \int h(x) f(x)\, \mathrm{d}x \right| \leqslant n^{-1/4}$$

and

$$|h(x_i^n)| \leqslant n^{3/4}, \qquad 1 \leqslant i \leqslant n, \tag{A.4}$$

for each of the functionals $h(x) = a(x)^2$, $b(x)^2$, $a(x)\,b(x)$, $a(x)^4$, $b(x)^4$, $a(x)^3\,b(x)$, $a(x)^2\,b(x)^2$, $a(x)\,b(x)^3$.

Statement (a) now follows from Chebychev's inequality together with inequalities (2.2) and (2.3).

Assuming statement (b) for the moment, statement (c) follows by the continuity theorem for characteristic functions (applied to an arbitrary sequence of $\{\mathbf{x}^n; n = 1, 2, \ldots\} \in H_1 \times H_2 \times \ldots$).

Statement (d) then follows since, if $R \sim N(-\alpha, 2\alpha)$, then $\mathbf{E}[1 \wedge \exp(R)] = 2\,\Phi\{-\surd(\alpha/2)\}$ (see Roberts *et al.* (1997), proposition 2.5), and furthermore $w \mapsto 1 \wedge \exp(w)$ is a bounded functional.

It remains to prove statement (b). Our proof is a quantitative modification of the standard proof of the Lindeberg central limit theorem (see Durrett (1991), pages 98–99).

Taking $\{\mathbf{x}^n\}$ to be a fixed sequence in $H_1 \times H_2 \times \ldots$, we set $W_i = C_3(x_i^n, Z_i)$, set

$$v(x_i^n) = \mathrm{var}_Z(W_i) = l^6\{a(x_i^n)^2 + 6\,a(x_i^n)\,b(x_i^n) + 15\,b(x_i^n)^2\},$$

and decompose $\phi_n(\mathbf{x}^n, t) = \Pi_{i=2}^n\,\theta_i^n(x_i^n, t)$ as a product of characteristic functions of $n^{-1/2}W_i$. Note that by inequality (A.4), for any $t \in \mathbf{R}$, we have $(t^2/2n)\,v(x_i^n) \leqslant 1$ for sufficiently large n. Hence, using equation (3.6) on p. 85 of Durrett (1991), for any $\epsilon > 0$, we have

$$\begin{aligned}\left|\theta_i^n(x_i^n, t) - \left\{1 - \frac{t^2}{2n}\,v(x_i^n)\right\}\right| &\leqslant \mathbf{E}_Z\left[\frac{|t|^3}{n^{3/2}}\frac{|W_i|^3}{3!} \wedge \frac{2t^2}{n}\frac{|W_i|^2}{2!}\right] \\ &\leqslant \mathbf{E}_Z\left[\frac{|t|^3}{n^{3/2}3!}|W_i|^3; |W_i| \leqslant n^{1/2}\epsilon\right] + \frac{t^2}{n}\mathbf{E}_Z[|W_i|^2; |W_i| > n^{1/2}\epsilon] \\ &\leqslant \frac{\epsilon|t|^3}{6n}\mathbf{E}_Z[|W_i|^2] + \frac{t^2}{\epsilon^2 n^2}\,\mathbf{E}_Z[|W_i|^4].\end{aligned}$$

Hence, since $\mathbf{x}^n \in H_n$, and using lemma (4.3) on p. 94 of Durrett (1991),

$$\begin{aligned}\left|\phi_n(\mathbf{x}^n; t) - \prod_{i=2}^n\left\{1 - \frac{t^2}{2n}\,v(x_i^n)\right\}\right| &\leqslant \sum_{i=2}^n\left(\frac{\epsilon|t|^3}{6n}\,\mathbf{E}_Z[|W_i|^2] + \frac{t^2}{\epsilon^2 n^2}\,\mathbf{E}_Z[|W_i|^4]\right) \\ &\leqslant \frac{K^2 + 22n^{-1/4}}{6}\,l^6\epsilon|t|^3 + \frac{t^2}{\epsilon^2\,n}\big(\xi + 14868 l^{12} n^{-1/4}\big).\end{aligned}$$

where $\xi = \mathbf{E}_X\,\mathbf{E}_Z[|W_i|^4]$.

Given $\delta > 0$, we choose ϵ sufficiently small to make the first term less than $\delta/2$, and then choose n sufficiently large to make the second term less than $\delta/2$, to obtain that

$$\left|\phi_n(\mathbf{x}^n; t) - \prod_{i=2}^n\left\{1 - \frac{t^2}{2n}v(x_i^n)\right\}\right| < \delta.$$

However,

$$\begin{aligned}\left|\prod_{i=2}^n\left\{1 - \frac{t^2}{2n}\,v(x_i^n)\right\} - \exp\left(-t^2 l^6\frac{K^2}{2}\right)\right| &\leqslant \left|\exp\left(-t^2 l^6\frac{K^2}{2}\right) - \exp\left\{-t^2\sum_{i=2}^n\frac{v(x_i^n)}{2n}\right\}\right| \\ &\quad + \left|\prod_{i=2}^n\left\{1 - \frac{t^2}{2n}\,v(x_i^n)\right\} - \prod_{i=2}^n\exp\left\{-t^2\frac{v(x_i^n)}{2n}\right\}\right|.\end{aligned}$$

Now, the first term goes to 0 uniformly for $\mathbf{x}^n \in H_n$. Also, by lemma 4.3 on p. 94 of Durrett (1991), the second term is bounded above by $\Sigma_{i=2}^n t^4 v^2(x_i^n)/4n^2$, which goes to 0 uniformly for $\{\mathbf{x}^n\}$ sequences (such that $\mathbf{x}^n \in H_n$) since the individual terms of $v^2(x_i^n)$ converge uniformly to their respective limits. The result follows. □

Proof of theorem 1. Recalling that

$$a_n(l) = \mathbf{E}\left[\frac{\pi_n(\mathbf{y})\,q_n(\mathbf{y}, \mathbf{x})}{\pi_n(\mathbf{x})\,q_n(\mathbf{x}, \mathbf{y})} \wedge 1\right],$$

theorem 1 follows directly from the last statement in lemma 1, the second statement in theorem 3 and from part (d) of lemma 3. □

Proof of theorem 2. We take $F_n^* = H_n \cap S_n \cap T_n \cap F_n$. Then

$$\lim_{n\to\infty} \{n^{1/3}\, \pi_n(F_n^{*C})\} = 0,$$

and

$$\lim_{n\to\infty} \{\mathbf{P}(\Gamma_t^n \in F_n^*,\ 0 \leqslant t \leqslant T)\} = 1$$

for any fixed T. Also, from lemma 1, theorem 3 and lemmas 2 and 3, it follows that

$$\lim_{n\to\infty} \sup_{\mathbf{x}^n \in F_n^*} |G_n\, V(\mathbf{x}^n) - G\, V(\mathbf{x}^n)| \ = \ 0$$

for all $V \in C_c^\infty$ which depend only on the first co-ordinate. Therefore, as discussed at the beginning of this section, using corollary 8.7 of chapter 4 of Ethier and Kurtz (1986), the weak convergence in theorem 2 follows.

Finally, to prove the statement about maximizing $h(l)$, we note that this problem amounts to finding the value $\hat{l}$ of l which maximizes the function $2l^2\, \Phi(-Kl^3/2)$, and then evaluating $\hat{a} = a(\hat{l}) = 2\, \Phi(-K\hat{l}^3/2)$. Making the substitution $u = Kl^3/2$ shows that this is the same as finding the value $\hat{u}$ of u which maximizes $2^{5/3} K^{-2/3} u^{2/3}\, \Phi(-u)$, and then evaluating $\hat{a} = 2\, \Phi(-\hat{u})$. It follows that the value of $\hat{u}$, and hence also the value of $\hat{a}$, does not depend on the value of K (provided that $K > 0$), so it suffices to take $K = 2$. For $K = 2$ we find (again using MATHEMATICA) that $\hat{l} \doteq 0.82515$, so that $\hat{a} \doteq 0.57424$. This completes the proof of theorem 2. □

References

Besag, J. and Green, P. J. (1993) Spatial statistics and Bayesian computation (with discussion). *J. R. Statist. Soc.* B, **55**, 25–37, 53–102.

Durrett, R. (1991) *Probability: Theory and Examples*. Pacific Grove: Wadsworth and Brooks.

Ethier, S. N. and Kurtz, T. G. (1986) Markov processes, characterization and convergence. New York: Wiley.

Gelman, A., Roberts, G. O. and Gilks, W. R. (1994) Efficient Metropolis jumping rules. *Research Report 94-10*. Statistical Laboratory, University of Cambridge, Cambridge.

Grenander, U. and Miller, M. I. (1994) Representations of knowledge in complex systems (with discussion). *J. R. Statist. Soc.* B, **56**, 549–603.

Hastings, W. K. (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**, 97–109.

Kennedy, A. D. and Pendleton, B. (1991) Acceptances and autocorrelations in hybrid Monte Carlo. *Nucl. Phys.* B, suppl., **20**, 118–121.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. and Teller, E. (1953) Equations of state calculations by fast computing machines. *J. Chem. Phys.*, **21**, 1087–1091.

Mountain, R. D. and Thirumalai, D. (1994) Quantitative measure of efficiency of Monte Carlo simulations. *Physica* A, **210**, 453–460.

Neal, R. M. (1993) Probabilistic inference using Markov chain Monte Carlo methods. *Technical Report CRG-TR-93-1*. Department of Computer Science, University of Toronto, Toronto.

——— (1994) An improved acceptance procedure for the hybrid Monte Carlo algorithm. *J. Computnl Phys.*, **111**, 194–203.

Phillips, D. B. and Smith, A. F. M. (1994) Bayesian model comparison via jump diffusions. *Technical Report 94-20*. Imperial College of Science, Technology and Medicine, London.

Roberts, G. O., Gelman, A. and Gilks, W. R. (1997) Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Probab.*, **7**, 110–120.

Roberts, G. O. and Tweedie, R. L. (1996) Exponential convergence of Langevin diffusions and their discrete approximations. *Bernoulli*, **2**, 341–363.

Rossky, P. J., Doll, J. D. and Friedman, H. L. (1978) Brownian dynamics as smart Monte Carlo simulation. *J. Chem. Phys.*, **69**, 4628–4633.

Smith, A. F. M. and Roberts, G. O. (1993) Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods (with discussion). *J. R. Statist. Soc.* B, **55**, 3–23, 53–102.

Wolfram, S. (1988) *Mathematica: a System for Doing Mathematics by Computer*. New York: Addison-Wesley.