The Annals of Applied Probability
1998, Vol. 8, No. 1, 1–9

A NOTE ON METROPOLIS–HASTINGS KERNELS FOR GENERAL STATE SPACES

BY LUKE TIERNEY[1]

University of Minnesota

The Metropolis–Hastings algorithm is a method of constructing a reversible Markov transition kernel with a specified invariant distribution. This note describes necessary and sufficient conditions on the candidate generation kernel and the acceptance probability function for the resulting transition kernel and invariant distribution to satisfy the detailed balance conditions. A simple general formulation is used that covers a range of special cases treated separately in the literature. In addition, results on a useful partial ordering of finite state space reversible transition kernels are extended to general state spaces and used to compare the performance of two approaches to using mixtures in Metropolis–Hastings kernels.

1. Introduction. The Metropolis–Hastings algorithm [Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller (1953), Hastings (1970)] is a method of constructing a reversible regular Markov transition kernel $P(x, dy)$ on a measurable space $(E, \mathscr{E})$ with a specified invariant distribution π. The algorithm requires a transition kernel $Q(x, dy)$ and a measurable function $\alpha(x, y)$: $E \times E \to [0, 1]$. If the current state is x, then a candidate, or proposal, for the next state is generated from $Q(x, dy)$. An observed candidate of y is accepted with probability $\alpha(x, y)$. Otherwise, the candidate is rejected and the process remains at x. The resulting transition kernel is

$$P(x, dy) = Q(x, dy)\alpha(x, y) + \delta_x(dy) \int (1 - \alpha(x, u))Q(x, du), \tag{1}$$

where δ_x is point mass at x. Originally proposed for discrete state spaces, this algorithm has recently been applied extensively to more general state spaces, in particular as a method for examining posterior distributions in Bayesian inference [Besag and Green (1993), Smith and Roberts (1993), Tierney (1994), Green (1994)]. This note has two objectives. The first is to give a statement of necessary and sufficient conditions on $Q(x, dy)$ and $\alpha(x, y)$ for the algorithm to be reversible with respect to the target distribution π. These conditions unify a number of special cases treated separately in the literature. The second objective is to extend a result of Peskun (1973) on the ordering of asymptotic variances for finite state spaces to general state spaces. This result is then used to compare two approaches to using mixtures of kernels in Metropolis–Hastings samplers.

Received July 1995; revised May 1997.
[1]Research supported in part by NSF Grant DMS-93-03557.
AMS 1991 *subject classifications*. 60J05, 65C05, 62-04.
Key words and phrases. Markov chain Monte Carlo, Peskun's theorem, mixture kernels.

2. Reversibility. A Markov chain with initial distribution π and transition kernel P is reversible if and only if the detailed balance relation

$$\pi(dx)P(x, dy) = \pi(dy)P(y, dx) \tag{2}$$

is satisfied. The two sides of this identity are measures on $\mathscr{E} \otimes \mathscr{E}$, and detailed balance means these measures are identical. If detailed balance holds, then for any real-valued f,

$$\iint f(y)\pi(dx)P(x, dy) = \iint f(y)\pi(dy)P(y, dx) = \int f(y)\pi(dy)$$

and thus π is invariant for P. The Metropolis–Hastings kernel (1) satisfies (2) if and only if

$$\pi(dx)Q(x, dy)\alpha(x, y) = \pi(dy)Q(y, dx)\alpha(y, x), \tag{3}$$

that is, the diagonal component does not matter.

The following proposition gives a few useful facts about measures on product spaces.

PROPOSITION 1. *Let $\mu(dx, dy)$ be a sigma-finite measure on the product space $(E \times E, \mathscr{E} \otimes \mathscr{E})$ and let $\mu^T(dx, dy) = \mu(dy, dx)$. Then there exists a symmetric set $R \in \mathscr{E} \otimes \mathscr{E}$ such that μ and μ^T are mutually absolutely continuous on R and mutually singular on the complement of R, R^c. The set R is unique up to sets that are null for both μ and μ^T. Let μ_R and μ_R^T be the restrictions of μ and μ^T to R. Then there exists a version of the density*

$$r(x, y) = \frac{\mu_R(dx, dy)}{\mu_R^T(dx, dy)}$$

such that $0 < r(x, y) < \infty$ and $r(x, y) = 1/r(y, x)$ for all $x, y \in E$.

PROOF. Let $\nu(dx, dy) = \mu(dx, dy) + \mu^T(dx, dy) = \mu(dx, dy) + \mu(dy, dx)$. Then ν is symmetric and both μ and μ^T are absolutely continuous with respect to ν. Let $h(x, y)$ be a density of μ with respect to ν. Then $\mu^T(dx, dy) = h(y, x)\nu(dy, dx) = h(y, x)\nu(dx, dy)$ and thus $h(y, x)$ is a density of μ^T with respect to ν. Let $R = \{(x, y): h(x, y) > 0 \text{ and } h(y, x) > 0\}$. Then R is symmetric, the restrictions of μ and μ^T to R are mutually absolutely continuous with $r(x, y) = h(x, y)/h(y, x)$ on R, and on R^c the measures μ and μ^T are mutually singular. The function $r(x, y)$ can be set to one on R^c. If R^* is any other set with the specified properties, then μ and μ^T must be mutually absolutely continuous as well as mutually singular on $R \setminus R^*$ and on $R^* \setminus R$, which means these sets must be null sets for both μ and μ^T. $\square$

For a given proposal generation kernel Q, let $\mu(dx, dy) = \pi(dx)Q(x, dy)$. The set R for this measure μ can be viewed as consisting of those state pairs (x, y) for which transitions from x to y and from y to x are both possible in the Markov chain with initial distribution π and transition kernel Q. The function $r(x, y)$ measures the relative rate of these transitions. The detailed balance condition (3) can be written in terms of μ as

$$\mu(dx, dy)\alpha(x, y) = \mu^T(dx, dy)\alpha(y, x). \tag{4}$$

Examining this identity separately for the sets R and R^c yields the following result.

THEOREM 2. *A Metropolis–Hastings transition kernel satisfies the detailed balance condition* (4) *if and only if the following two conditions hold.*

(i) *The function α is μ-almost everywhere zero on R^c.*
(ii) *The function α satisfies $\alpha(x, y)r(x, y) = \alpha(y, x)$ μ-almost everywhere on R.*

PROOF. Let $\eta(dx, dy) = \mu(dx, dy)\alpha(x, y)$ and $\eta^T(dx, dy) = \mu^T(dx, dy) \cdot \alpha(y, x)$. If α is μ-almost everywhere zero on R^c, then $\eta(R^c) = 0$. Since R^c is symmetric, this implies that $\eta^T(R^c) = 0$ also, and hence detailed balance holds on R^c. Conversely, since the measures η and η^T are mutually singular on R^c, detailed balance implies that they must satisfy $\eta(R^c) = \eta^T(R^c) = 0$, which implies that α is μ-almost everywhere zero on R^c. On R, the measures μ and μ^T are equivalent with $d\mu/d\mu^T = r(x, y)$. So (4) holds on R if and only if $\alpha(x, y)r(x, y) = \alpha(y, x)$ holds μ^T-almost everywhere, which holds if and only if $\alpha(x, y)r(x, y) = \alpha(y, x)$ holds μ-almost everywhere. □

The set R may be empty; if it is, then $P(x, dy) = \delta_x(dy)$.

The standard Metropolis–Hastings rejection probability $\alpha_{MH}(x, y)$ can be written as

$$\alpha_{MH}(x, y) = \begin{cases} \min\{1, r(y, x)\}, & \text{if } (x, y) \in R, \\ 0, & \text{if } (x, y) \notin R. \end{cases}$$

Condition (i) holds by construction. For $(x, y) \in R$

$$\begin{aligned} \alpha_{MH}(x, y)r(x, y) &= \min\{r(x, y), r(y, x)r(x, y)\} \\ &= \min\{r(x, y), 1\} = \alpha_{MH}(x, y) \end{aligned}$$

since $r(x, y) = 1/r(y, x)$ on R, and thus (ii) is satisfied as well.

This general formulation covers a number of special cases that are usually treated separately in the literature.

1. *Common dominating measure* [e.g., Tierney (1994), Section 2.3]. Suppose there is a measure ν such that $\pi(dx) = \pi(x)\nu(dx)$ and $Q(x, dy) = q(x, y)\nu(dy)$. Then

$$R = \{(x, y)\colon \pi(x)q(x, y) > 0 \quad \text{and} \quad \pi(y)q(y, x) > 0\}$$

and

$$r(x, y) = \frac{\pi(x)q(x, y)}{\pi(y)q(y, x)}.$$

Detailed balance holds if and only if $\pi(x)q(x, y)\alpha(x, y) = 0$ for $\nu \times \nu$-almost all $(x, y) \notin R$ and $\alpha(x, y)r(x, y) = \alpha(y, x)$ for $\nu \times \nu$-almost all $(x, y) \in R$.

2. *Deterministic proposals* [e.g., Tierney (1994), Section 2.3.4]. Suppose T is a one-to-one transformation from E onto E such that $T^{-1} = T$, and let $Q(x, y) = \delta_{T(x)}(dy)$. Thus when the current state is x then the proposal is $T(x)$. Define the measure π' by $\pi'(A) = \pi(T^{-1}(A))$ for all $A \in \mathscr{E}$, and let $\nu(dx) = \pi(dx) + \pi'(dx)$. Let $h(x)$ be a density for $\pi(dx)$ with respect to $\nu(dx)$. Then $h(T(x))$ is a density for $\pi'(dx)$ with respect to ν. Let $A = \{x \in E: h(x) > 0 \text{ and } h(T(x)) > 0\}$. Then $R = \{(x, y): x \in A \text{ and } y = T(x)\}$ and $r(x, y) = h(x)/h(T(x))$ on R. Detailed balance holds if and only if $\alpha(x, T(x)) = 0$ for π-almost all $x \notin A$ and

$$\alpha(x, T(x))\frac{h(x)}{h(T(x))} = \alpha(T(x), x)$$

for π-almost all $x \in A$.

3. *Green's dimension-changing kernel.* Green (1995) applies the Metropolis–Hastings algorithm to Bayesian model selection. His formulation leads to a state space in which elements have two components, an index from a discrete set $\mathscr{I}$ and a value from a set E_i that depends on the index. The index i might be an integer and E_i might be $\mathbb{R}^i$. Thus $E = \{(i, x): i \in \mathscr{I} \text{ and } x \in E_i\}$ and $\mathscr{E}$ is the sigma algebra generated by the sets $\{(i, x): x \in A\}$ for $i \in \mathscr{I}$ and $A \in \mathscr{E}_i$, where $\mathscr{E}_i$ is a sigma algebra on E_i. Write $\pi_i(A)$ for the probability of observing an index of i and a value in $A \in \mathscr{E}_i$, and define $Q_{ij}(x, dy)$ and $\alpha_{ij}(x, y)$ analogously. For each $i, j \in \mathscr{I}$, let $\mu_{ij}(dx, dy) = \pi_i(dx)Q_{ij}(x, dy)$ and $\nu_{ij}(dx, dy) = \pi_i(dx)Q_{ij}(x, dy) + \pi_j(dy)Q_{ji}(y, dx)$ be measures on $\mathscr{E}_i \otimes \mathscr{E}_j$, let $h_{ij} = d\mu_{ij}/d\nu_{ij}$ and set

$$R_{ij} = \{(x, y): x \in E_i,\ y \in E_j,\ h_{ij}(x, y) > 0, \text{ and } h_{ji}(y, x) > 0\}.$$

Thus $(x, y) \in R_{ij}$ if and only if $(y, x) \in R_{ji}$. Then

$$R = \{(i, x, j, y): i, j \in \mathscr{I},\ (x, y) \in R_{ij}\},$$

and $r_{ij}(x, y) = h_{ij}(x, y)/h_{ji}(y, x)$. A Metropolis–Hastings kernel is reversible if and only if for each $i, j \in \mathscr{I}$ we have $\alpha_{ij}(x, y) = 0$ for μ_{ij}-almost all $(x, y) \notin R_{ij}$ and

$$\alpha_{ij}(x, y)r_{ij}(x, y) = \alpha_{ji}(y, x)$$

for μ_{ij}-almost all $(x, y) \in R_{ij}$.

3. A partial ordering.

Peskun (1973) introduces a useful partial ordering on transition kernels that can be called off-diagonal domination. If P_1 and P_2 are transition kernels with invariant distribution π, then P_1 dominates P_2 off the diagonal, $P_1 \succeq P_2$, if for π-almost all $x \in E$ we have $P_1(x, A \setminus \{x\}) \geq P_2(x, A \setminus \{x\})$ for all $A \in \mathscr{E}$.

LEMMA 3. *If P_1 and P_2 have invariant distribution π and $P_1 \succeq P_2$, then $P_2 - P_1$ is a positive operator on $L^2(\pi)$. That is, $\langle (P_2 - P_1)f, f\rangle = \iint f(x)f(y)(P_2(x,dy) - P_1(x,dy))\pi(dx) \geq 0$ for all $f \in L^2(\pi)$.*

PROOF. Let $H(dx,dy) = \pi(dx)(\delta_x(dy) - P_2(x,dy) + P_1(x,dy))$. Then $H(A) \geq 0$ for all $A \in \mathscr{E}\otimes\mathscr{E}$, and $H(E\times E) = 1$. That is, H is a probability on $\mathscr{E}\otimes\mathscr{E}$. Furthermore, $H(E\times A) = H(A\times E) = \pi(A)$; that is, the marginal distributions of H are both equal to π. So for $f \in L_2(\pi)$,

$$\begin{aligned}
&\iint f(x)f(y)(P_2(x,dy) - P_1(x,dy))\pi(dx)\\
&\quad= \iint f(x)f(y)(\pi(dx)\delta_x(dy) - H(dx,dy))\\
&\quad= \int f(x)^2\pi(dx) - \iint f(x)f(y)H(dx,dy)\\
&\quad= \tfrac{1}{2}\left(\int f(x)^2\pi(dx) + \int f(y)^2\pi(dy) - 2\iint f(x)f(y)H(dx,dy)\right)\\
&\quad= \tfrac{1}{2}\iint (f(x)-f(y))^2 H(dx,dy)
\end{aligned}$$

and the final right-hand side is nonnegative. □

This result shows that the lag one autocorrelations of a stationary Markov chain with a transition kernel P_1 are at most as large as for a chain with kernel P_2 when $P_1 \succeq P_2$. Higher order correlations need not be ordered, but if the kernels are reversible then the asymptotic variances of sample path averages for a chain with kernel P_1 are no larger than for a chain with kernel P_2 when $P_1 \succeq P_2$:

THEOREM 4. *Let P_1 and P_2 be reversible transition kernels with invariant distribution π and suppose $f \in L_0^2(\pi) = \{g \in L^2(\pi)\colon \int g\,d\pi = 0\}$. Let*

$$v(f,H) = \lim_{n\to\infty}\frac{1}{n}\operatorname{Var}_H\left(\sum_{i=1}^n f(X_i)\right)$$

where $X_0, X_1, \ldots$ is a Markov chain with initial distribution π and reversible transition kernel H. If $P_1 \succeq P_2$, then $v(f,P_1) \leq v(f,P_2)$.

PROOF. The proof is based on the approach of Kipnis and Varadhan (1986). If H is a reversible transition kernel with invariant distribution π, then H represents a self-adjoint operator on $L_0^2(\pi)$ with spectral radius bounded by one. By the spectral decomposition theorem for self-adjoint operators, for each $f \in L_0^2(\pi)$ there exists a finite positive measure $e_{f,H}$ on the real line with support contained in the interval $[-1,1]$ such that $\langle f, H^n f\rangle = \int x^n e_{f,H}(dx)$ for all integers $n \geq 0$. Thus

$$\frac{1}{n}\operatorname{Var}_H\left(\sum_{i=1}^n f(X_i)\right) = \int 1 + 2\sum_{i=1}^n \frac{n-i}{n}x^i e_{f,H}(dx) \to \int \frac{1+x}{1-x}e_{f,H}(dx).$$

The limit is guaranteed to exit, but it may be infinite. Kipnis and Varadhan (1986) show that a central limit theorem holds when the limit is finite. Now define $v_\lambda(f, H) = \langle f, (I-\lambda H)^{-1}(I+\lambda H)f\rangle$ for $0 \le \lambda < 1$. This is well defined and finite since $I - \lambda H$ as an operator on $L_0^2(\pi)$ has a bounded inverse for $\lambda \in [0, 1)$. Then

$$v_\lambda(f, H) = \int \frac{1+\lambda x}{1-\lambda x} e_{f,H}(dx) \to v(f, H)$$

as $\lambda \to 1$, whether $v(f, H)$ is finite or infinite. Now suppose that P_1 and P_2 are reversible transition kernels with invariant distribution π and $P_1 \succeq P_2$. Let $H_\beta = P_1 + \beta(P_2 - P_1)$ for $0 \le \beta \le 1$ and $h_\lambda(\beta) = v_\lambda(f, H_\beta)$. Let $A_\lambda(\beta) = (I-\lambda H_\beta)^{-1}(I+\lambda H_\beta)$, and let $A'_\lambda(\beta) = \lim_{h\downarrow 0}(1/h)(A_\lambda(\beta+h) - A_\lambda(\beta))$ be the right-hand derivative of $A_\lambda(\beta)$ at $\beta \in [0, 1)$. Since $(I-\lambda H_\beta)A_\lambda(\beta) = I+\lambda H_\beta$, we have

$$(I-\lambda H_\beta)A'_\lambda - \lambda(P_2 - P_1)A_\lambda(\beta) = \lambda(P_2 - P_1)$$

and thus

$$\begin{aligned} A'_\lambda(\beta) &= \lambda(I-\lambda H_\beta)^{-1}(P_2 - P_1)(I + A_\lambda(\beta)) \\ &= 2\lambda(I-\lambda H_\beta)^{-1}(P_2 - P_1)(I-\lambda H_\beta)^{-1}. \end{aligned}$$

So the right-hand derivatives of h_λ are given by

$$\begin{aligned} h'_\lambda(\beta) &= \langle f, 2\lambda(I-\lambda H_\beta)^{-1}(P_2 - P_1)(I-\lambda H_\beta)^{-1}f\rangle \\ &= 2\lambda\langle (I-\lambda H_\beta)^{-1}f, (P_2 - P_1)(I-\lambda H_\beta)^{-1}f\rangle \\ &\ge 0 \end{aligned}$$

for $\beta \in [0, 1)$ since $P_2 - P_1$ is a positive operator. So h_λ is a nondecreasing function, and therefore

$$v_\lambda(f, P_2) = h_\lambda(1) \ge h_\lambda(0) = v_\lambda(f, P_1).$$

Taking limits as $\lambda \to 1$ then shows that $v(f, P_2) \ge v(f, P_1)$. □

This theorem generalizes the finite state space result of Peskun's theorem 2.1.1 to general state spaces.

A similar analysis of the second derivative of h_λ shows that this function is convex. Thus

$$v_\lambda(f, P_2) \ge \langle g_\lambda, (P_2 - P_1)g_\lambda\rangle + v_\lambda(f, P_1),$$

where $g_\lambda = (I-\lambda P_1)^{-1}f$ for all $\lambda \in [0, 1)$. If there exists a function g such that $(I - P_1)g = f$ and $\langle g_\lambda, (P_2 - P_1)g_\lambda\rangle \to \iint g(x)(P_2(x, dy) - P_1(x, dy))g(y)\pi(dx)$, then the variance inequality is strict; that is, $v(f, P_2) > v(f, P_1)$, if $v(f, P_1)$ is finite and if the functions $P_1 g$ and $P_2 g$ are not equal π-almost everywhere. A sufficient condition for the existence of such a g is that the spectral radius of P_1 as an operator on $L_0^2(\pi)$ be strictly less than one; weaker conditions are possible.

When the inequality is strict, the finite sample variances for sufficiently large samples will be ordered as well. One might hope for the finite sample variances to be ordered for all n, but this is not true in general. As a simple counter example, consider the irreducible, doubly stochastic, symmetric transition matrices

$$P_1 = \begin{bmatrix} 0.0 & 0.2 & 0.8 & 0.0 \\ 0.2 & 0.0 & 0.0 & 0.8 \\ 0.8 & 0.0 & 0.2 & 0.0 \\ 0.0 & 0.8 & 0.0 & 0.2 \end{bmatrix}, \qquad P_2 = \begin{bmatrix} 0.1 & 0.1 & 0.8 & 0.0 \\ 0.1 & 0.1 & 0.0 & 0.8 \\ 0.8 & 0.0 & 0.2 & 0.0 \\ 0.0 & 0.8 & 0.0 & 0.2 \end{bmatrix}$$

and the function $f^T = (1, -1, -3, 3)$. Since P_1 and P_2 are doubly stochastic, they have the uniform distribution as their invariant distribution. Furthermore, $P_1 \succeq P_2$, but

$$\mathrm{Var}_{P_1}(f(X_0) + f(X_1) + f(X_2)) = \tfrac{3}{4} f^T f + f^T P_1 f + \tfrac{1}{2} f^T P_1^2 f = 15.4$$

and

$$\mathrm{Var}_{P_2}(f(X_0) + f(X_1) + f(X_2)) = \tfrac{3}{4} f^T f + f^T P_2 f + \tfrac{1}{2} f^T P_2^2 f = 14.8.$$

So the sum has smaller variance under P_2 than under P_1, and thus $P_1 \succeq P_2$ does not imply $\mathrm{Var}_{P_1}(\sum_{i=1}^n f(X_i)) \le \mathrm{Var}_{P_2}(\sum_{i=1}^n f(X_i))$ for all n.

As Peskun points out, for a given proposal generation kernel Q using the acceptance probability α_{MH} produces a transition kernel that is maximal with respect to off-diagonal domination among all Metropolis–Hastings kernels with the same proposal kernel. This follows from the observation that

$$\alpha(x, y) = \alpha(y, x) r(y, x) \le \min\{1, r(y, x)\} = \alpha_{MH}(x, y)$$

for any acceptance probability $\alpha(x, y)$ that produces a reversible transition kernel. The Metropolis–Hastings kernel for a given Q that uses the acceptance probability $\alpha_{MH}(x, y)$ will be called the maximal Metropolis–Hastings kernel for Q. In terms of minimizing the asymptotic variance of sample path averages, $\alpha_{MH}(x, y)$ is the optimal acceptance probability function.

4. Mixing Metropolis–Hastings kernels. It is often useful to build up a sampler from simpler component samplers. Given proposal kernels Q_i we can combine them in a mixture in two ways: we can use a mixture of the Metropolis–Hastings kernels based on each proposal kernel, or we can form a mixture proposal kernel and use it to form a single Metropolis–Hasting kernel. When using the maximal acceptance probability, the second approach is preferable in terms of asymptotic variances of sample path averages.

PROPOSITION 5. *Let Q_i be a sequence of proposal kernels and let $\beta_i \ge 0$ with $\sum \beta_i = 1$ be a set of probabilities. Let P_i be the maximal Metropolis–Hastings kernels based on proposal kernels Q_i and let P be the maximal Metropolis–Hastings kernel based on the proposal kernel $Q = \sum \beta_i Q_i$. Then $P \succeq \sum \beta_i P_i$.*

PROOF. Suppose, without loss of generality, that there is a common symmetric dominating measure ν for all the measures $\mu_i(dx, dy) = \pi(dx)Q_i(x, dy)$. Let $h_i(x, y)$ be the density of μ_i with respect to ν and let $h(x, y) = \sum \beta_i h_i(x, y)$. Then

$$\begin{aligned}\pi(dx)Q(x, dy)\alpha_{MH}(x, y) &= h(x, y)\min\left\{\frac{h(y, x)}{h(x, y)}, 1\right\}\nu(dx, dy)\\ &= \min\{h(y, x), h(x, y)\}\nu(dx, dy)\\ &= \min\left\{\sum \beta_i h_i(y, x), \sum \beta_i h_i(x, y)\right\}\nu(dx, dy)\\ &\geq \sum \beta_i \min\{h_i(y, x), h_i(x, y)\}\nu(dx, dy)\\ &= \sum \beta_i \pi(dx)Q_i\alpha_{MH}^{(i)}(x, y),\end{aligned}$$

where $\alpha_{MH}^{(i)}(x, y)$ is the maximal acceptance probability function for the proposal kernel Q_i. □

This result must of course be treated with caution. To compute the acceptance probability for the mixture kernel $Q = \sum \beta_i Q_i$, one usually has to compute the transition densities for all the Q_i. In contrast, when using the mixture of Metropolis–Hastings kernels $\sum \beta_i P_i$, one only needs to compute the transition density for the proposal kernel that was actually used. Thus even though the number of iterations needed to achieve a particular level of accuracy is lower when using a mixture proposal kernel, the cost of each iteration may be higher, and in terms of CPU time it may be better to use a longer run of the cheaper chain than a shorter run of the more costly one.

While preparing the revision it came to my attention that a version of Proposition 5 for discrete chains was given in the rejoinder of Besag, Green, Higdon and Mengersen (1995). A similar version for discrete state spaces was also given in the appendix of Tierney (1991).

REFERENCES

BESAG, J. and GREEN, P. J. (1993). Spatial statistics and Bayesian computation (with discussion). *J. Roy. Statist. Soc. Ser. B* **55** 25–37.

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

GREEN, P. J. (1995). Reversible jump MCMC computation and Bayesian model determination. *Biometrika* **82** 711–732.

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

KIPNIS, C. and VARADHAN, S. R. S. (1986). Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Comm. Math. Phys.* **104** 1–19.

METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H. and TELLER, E. (1953). Equations of state calculations by fast computing machines. *J. Chemical Physics* **21** 1087–1091.

PESKUN, P. H. (1973). Optimum Monte Carlo sampling using Markov chains. *Biometrika* **60** 607–612.

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. Roy. Statist. Soc. Ser. B* **55** 3–24.

TIERNEY, L. (1991). Markov chains for exploring posterior distributions. Technical Report 560, School of Statistics, Univ. Minnesota.

TIERNEY, L. (1994). Markov chains for exploring posterior distributions. *Ann. Statist.* **22** 1701–1786.

SCHOOL OF STATISTICS
UNIVERSITY OF MINNESOTA
MINNEAPOLIS, MINNESOTA 55455
E-MAIL: luke@stat.umn.edu