EXACT TRANSITION PROBABILITIES FOR THE INDEPENDENCE METROPOLIS SAMPLER

Richard L. Smith [1] , [2] **and Luke Tierney** [3]

May 6 1996.

Abstract. A recent result of Jun Liu's has shown how to compute explicitly the eigenvalues and eigenvectors for the Markov chain derived from a special case of the Hastings sampling algorithm, known as the indepdendence Metropolis sampler. In this note, we show how to extend the result to obtain exact n-step transition probabilities for any n. This is done first for a chain on a finite state space, and then extended to a general (discrete or continuous) state space. The paper concludes with some implications for diagnostic tests of convergence of Markov chain samplers.

1. Introduction

Suppose we want to generate a Monte Carlo sample from a discrete distribution $\pi_j,\ j = 1, ..., m$. The distribution is not easy to sample from directly, and it is possible that the functional form of π_j may be known only up to some unspecified normalizing constant, which will be inconvenient to compute directly if m is very large. One strategy is as follows: pick some other distribution $\{p_j,\ j = 1, ..., m\}$ which is easy to sample from, then define the Markov transition kernel

$$K_{ij} = \begin{cases} p_j \min\left(1, \frac{w_j}{w_i}\right), & j \neq i, \\ 1 - \sum_{k \neq i} K_{ik}, & j = i. \end{cases} \tag{1}$$

where $w_j = \frac{\pi_j}{p_j}$ is the weight associated with state j. It is easy to check that $\{K_{ij},\ 1 \leq i \leq m,\ 1 \leq j \leq m\}$ defines the transition kernel of a reversible Markov chain for which $\{\pi_j\}$ is a stationary distribution. If we can check that the chain is irreducible and aperiodic (a separate calculation, but usually straightforward) it then follows that the stationary distribution is unique and that the chain therefore converges to this distribution. The method is in fact a special case of the Hastings (1970) algorithm known as the independence Metropolis chain; see in particular Tierney (1994). Moreover, it readily extends to cases with infinite (countable or uncountable) state space.

[1] Address for correspondence: Statistical Laboratory, 16 Mill Lane, Cambridge CB2 1SB, U.K.

[2] Department of Statistics, University of North Carolina, Chapel Hill, N.C. 27599-3260, U.S.A.

[3] School of Statistics, University of Minnesota, Minneapolis 55455, U.S.A.

Recently Liu (1996) has shown that there exists an exact eigenanalysis of this system. Suppose the states are ordered so that $\{w_1 \geq w_2 \geq ... \geq w_m\}$. Then (1) simplifies to

$$K_{ij} = \begin{cases} p_j, & \text{if } j < i, \\ p_i + \lambda_i, & \text{if } j = i, \\ \frac{\pi_j}{w_i}, & \text{if } j > i, \end{cases} \tag{2}$$

where

$$\lambda_i = \sum_{k>i} \left(p_k - \frac{\pi_k}{w_i} \right), \quad 1 \leq i \leq m-1. \tag{3}$$

Define $\lambda_0 = 1,\ \lambda_m = 0$, and

$$\mathbf{v}_0 = (1, 1, 1, ..., 1)^T,$$
$$\mathbf{v}_k = (0, 0, ..., 0, S_{k+1}, -\pi_k, ..., -\pi_k)^T, \quad 1 \leq k \leq m-1,$$

where for $k > 0$ the first $k-1$ entries are 0, the k'th is $S_{k+1} \equiv \sum_{i=k+1}^{m} \pi_i$ and the remaining entries are all $-\pi_k$. Then Liu showed, as is easy to check directly, that the eigenvalues of K are $1 = \lambda_0 \geq \lambda_1 \geq ... \geq \lambda_{m-1}$ and that $\mathbf{v}_0, \mathbf{v}_1, ..., \mathbf{v}_{m-1}$ are the corresponding right eigenvectors.

In Section 2, we show how Liu's result may be extended to obtain exact n-step transition probabilities for the Markov chain with transition kernel K. In section 3 we extend the result to more general (mixed discrete/continuous) state spaces and provide a direct proof through integration by parts. Section 4 examines conditions for geometric convergence. The paper concludes in Section 5, with some discussion and an explicit example.

2. Discrete case

Suppose $\mathbf{e}_k$ is the unit vector with 1 in the k'th position ($1 \leq k \leq m$) and 0 everywhere else. As in the previous section, we define $S_k = \sum_{i \geq k} \pi_i$, with $S_1 = 1$.

Proposition 1. For $1 \leq k \leq m-1$,

$$\mathbf{e}_k = \pi_k \mathbf{v}_0 + \frac{1}{S_k} \mathbf{v}_k - \pi_k \sum_{j=1}^{k-1} \frac{\mathbf{v}_j}{S_j S_{j+1}}$$

while for $k = m$,

$$\mathbf{e}_m = \pi_k \mathbf{v}_0 - \pi_m \sum_{j=1}^{m-1} \frac{\mathbf{v}_j}{S_j S_{j+1}}.$$

Before proving this, we note the following simple identity:

Lemma 1. For $1 < i \leq m$,

$$1 + \sum_{j=1}^{i-1} \frac{\pi_j}{S_j S_{j+1}} = \frac{1}{S_i}. \tag{4}$$

Proof of Lemma 1. For $i = 2$, this follows at once from $S_2 + \pi_1 = S_1 = 1$. For $i > 2$, the result then follows by induction, using the identity

$$\frac{1}{S_{i+1}} - \frac{1}{S_i} = \frac{\pi_i}{S_i S_{i+1}}.$$

Proof of Proposition 1

Let

$$\mathbf{e}_k^* = \pi_k \mathbf{v}_0 + \frac{1}{S_k}\mathbf{v}_k - \pi_k \sum_{j=1}^{k-1} \frac{\mathbf{v}_j}{S_j S_{j+1}},$$

where, in the case $k = m$, we define $\mathbf{v}_m$ to be the zero vector.

For $i < k$, the i'th component of $\mathbf{e}_k^*$ is

$$\pi_k + \pi_k \sum_{j=1}^{i-1} \frac{\pi_j}{S_j S_{j+1}} - \frac{\pi_k S_{i+1}}{S_i S_{i+1}} = \pi_k \left(1 + \sum_{j=1}^{i-1} \frac{\pi_j}{S_j S_{j+1}} - \frac{1}{S_i} \right) = 0$$

using Lemma 1.

The k'th component of $\mathbf{e}_k^*$ is

$$\begin{aligned} \pi_k + \frac{S_{k+1}}{S_k} + \pi_k \sum_{j=1}^{k-1} \frac{\pi_j}{S_j S_{j+1}} &= \pi_k \left(1 + \sum_{j=1}^{k-1} \frac{\pi_j}{S_j S_{j+1}} \right) + \frac{S_{k+1}}{S_k} \\ &= \frac{\pi_k}{S_k} + \frac{S_{k+1}}{S_k} = \frac{S_k}{S_k} = 1. \end{aligned}$$

To make this argument valid in the case $k = m$, we need to define $S_{m+1} = 0$.

For $i > k$, the i'th component of $\mathbf{e}_k^*$ is

$$\pi_k - \frac{\pi_k}{S_k} + \pi_k \sum_{j=1}^{k-1} \frac{\pi_j}{S_j S_{j+1}} = \pi_k \left(1 - \frac{1}{S_k} + \sum_{j=1}^{k-1} \frac{\pi_j}{S_j S_{j+1}} \right) = 0.$$

Thus we have shown that $\mathbf{e}_k^* = \mathbf{e}_k$, and the proof is complete.

We now use Proposition 1 to derive exact n-step transition probabilities. Let $\mathbf{K}$ denote the matrix with entries (K_{ij}) and let $\mathbf{K}^n$ denote its n'th power, with entries $(K_{ij}^{(n)})$, so that $K_{ij}^{(n)}$ is the n'th-order transition probability from state i to state j.

Defining $\mathbf{v}_m = \mathbf{0}$ and $\lambda_m = 0$, we have from Proposition 1, for $1 \le k \le m$,

$$\mathbf{e}_k = \pi_k \mathbf{v}_0 + \frac{1}{S_k}\mathbf{v}_k - \pi_k \sum_{j=1}^{k-1} \frac{\mathbf{v}_j}{S_j S_{j+1}}$$

and so

$$\mathbf{K}^n \mathbf{e}_k = \pi_k \mathbf{v}_0 + \frac{\lambda_k^n}{S_k}\mathbf{v}_k - \pi_k \sum_{j=1}^{k-1} \frac{\lambda_j^n \mathbf{v}_j}{S_j S_{j+1}}.$$

Hence if v_{ji} denotes the i'th component of $\mathbf{v}_j$,

$$K_{ik}^{(n)} = \pi_k + \frac{\lambda_k^n}{S_k} v_{ki} - \pi_k \sum_{j=1}^{k-1} \frac{\lambda_j^n v_{ji}}{S_j S_{j+1}}.$$

Noting that

$$v_{ji} = \begin{cases} 0 & \text{if } i < j, \\ S_{j+1} & \text{if } i = j, \\ -\pi_j & \text{if } i > j, \end{cases}$$

we have shown:

Proposition 2.

$$K_{ik}^{(n)} = \begin{cases} \pi_k \left(1 + \sum_{j=1}^{i-1} \frac{\lambda_j^n \pi_j}{S_j S_{j+1}} - \frac{\lambda_i^n}{S_i}\right) & \text{for } i < k, \\ \pi_k \left(1 + \sum_{j=1}^{k-1} \frac{\lambda_j^n \pi_j}{S_j S_{j+1}} - \frac{\lambda_k^n}{S_k}\right) + \lambda_k^n & \text{for } i = k, \\ \pi_k \left(1 + \sum_{j=1}^{k-1} \frac{\lambda_j^n \pi_j}{S_j S_{j+1}} - \frac{\lambda_k^n}{S_k}\right) & \text{for } i > k. \end{cases}$$

3. The general case

The result of Proposition 2 may be extended to general distributions on a measurable space $(E, \mathcal{E})$. Suppose we wish to simulate from a probability measure Π on $(E, \mathcal{E})$. We assume that we are able to sample from a distribution P. We must assume that Π is absolutely continuous with respect to P, and that we can evaluate a density $w(x) = \Pi(dx)/P(dx)$. The independence Metropolis-Hastings sampler is then as follows: Given a current state $X_n = x$, choose a new state y from the probability measure P, then accept the new state and set $X_{n+1} = y$ with probability $\min\{1, w(y)/w(x)\}$. Otherwise reject the new state and set $X_{n+1} = x$. When both $w(x)$ and $w(y)$ are zero, we reject the new step and remain at x. As usual, both the drawing from P and the accept/reject decision are independent of each other and of all past random events.

If the second part of the Metropolis-Hastings step results in the process rejecting the proposed move, then we say that a rejection has occurred. This does not include the case when the random drawing from P selects the same value x, which is automatically accepted. The distinction is important whenever x is an atom of P. The probability that a rejection occurs, given that the current state is x, is $\lambda(w(x))$, where

$$\begin{aligned}\lambda(w) &= \int_{\{y:\ w(y)\le w\}} \left\{1 - \frac{w(y)}{w}\right\} P(dy) \\ &= \int_{\{y:\ w(y)\le w\}} \left\{P(dy) - \frac{\Pi(dy)}{w}\right\}\end{aligned}$$

for $w > 0$ and $\lambda(0) = P(\{y : w(y) = 0\})$.

For a more restricted set-up than we are considering here, Liu (1996), generalizing his result for discrete chains, showed that the spectrum of the transition operator consists of all values of $\lambda(w(x))$, $x \in E$, together with 1. If w^* is the essential supremum of $w(x)$ (essential with respect to Π or P), then the essential supremum of $\lambda(w)$ is $1 - 1/w^*$. Thus if $w^* < \infty$ then there is a "spectral gap" of size $1/w^*$, and the n-th order transition probabilities of the chain differ from the stationary measure Π by a term of order at most $(1 - 1/w^*)^n$. Thus we have uniform geometric convergence whenever $w^* < \infty$. Our subsequent development will allow these results to be re-derived as a consequence of general formulae for the n-step transition probabilities.

Define $\tilde{\Pi}(w) = \Pi(\{y : w(y) \le w\})$ and $\tilde{P}(w) = P(\{y : w(y) \le w\})$ for all real values w, and let $\tilde{\Pi}(dw)$ and $\tilde{P}(dw)$ denote the corresponding probability measures on the Borel subsets of the real line. Let $\tilde{\Pi}(w-)$ denote the left hand limit of $\tilde{\Pi}(x)$ at w. The function $\lambda(w)$ can then be written as

$$\lambda(w) = \int_{v \le w} \left(1 - \frac{v}{w}\right) \tilde{P}(dv) = \tilde{P}(w) - \frac{\tilde{\Pi}(w)}{w}. \tag{5}$$

Note that $\lambda(w) = 1 - \frac{1}{w}$ when $w > w^*$.

By Fubini's theorem, we also have

$$\begin{aligned}
\lambda(w) &= \tilde{P}(0) + \int_{(0,w]} \left(1 - \frac{v}{w}\right) \tilde{P}(dv) \\
&= \tilde{P}(0) + \int_{(0,w]} \left(\frac{1}{v} - \frac{1}{w}\right) \tilde{\Pi}(dv) \\
&= \tilde{P}(0) + \int_{(0,w]} \int_v^w \frac{1}{u^2} du \tilde{\Pi}(dv) \\
&= \tilde{P}(0) + \int_0^w \int_{(0,u]} \frac{1}{u^2} \tilde{\Pi}(dv) du \\
&= \tilde{P}(0) + \int_0^w \frac{\tilde{\Pi}(u)}{u^2} du
\end{aligned} \tag{6}$$

for $w \geq 0$. In particular, $\lambda(w)$ is differentiable in the ordinary sense with derivative $\tilde{\Pi}(w)/w^2$ whenever $w > 0$ and w is a continuity point of $\tilde{\Pi}$.

Define $T_n(w) = 0$ when $\tilde{\Pi}(w) = 0$ and

$$T_n(w) = 1 + \int_{v>w} \frac{\lambda^n(v)}{\tilde{\Pi}(v)\tilde{\Pi}(v-)} d\tilde{\Pi}(v) - \frac{\lambda^n(w)}{\tilde{\Pi}(w)} \tag{7}$$

when $\tilde{\Pi}(w) > 0$. Then Proposition 2 suggests the following general result:

Theorem 1. The n-step transition kernel for the independence Metropolis-Hastings chain is given by

$$p_n(x, dy) = T_n(w(x) \vee w(y))\Pi(dy) + \lambda^n(w(x))\delta_x(dy)$$

where $\delta_x(dy)$ denotes point mass at x and $x \vee y = \max\{x, y\}$.

To prove this result we first develop an alternate representation for $T_n(w)$. We begin with a preliminary result:

Lemma 2. For any $a \leq b$ with $\tilde{\Pi}(a) > 0$,

$$\int_{(a,b]} \frac{\tilde{\Pi}(dv)}{\tilde{\Pi}(v)\tilde{\Pi}(v-)} = \frac{1}{\tilde{\Pi}(a)} - \frac{1}{\tilde{\Pi}(b)}.$$

Proof of Lemma 2. For each integer $n \geq 1$ let $x_0 = a \leq x_1 \leq \cdots \leq x_n = b$ be a partition of the interval $(a, b]$ and assume that $\Delta_n = \max\{x_i - x_{i-1} : i = 1, \ldots n\} \to 0$ as $n \to \infty$. Then for each n

$$\frac{1}{\tilde{\Pi}(a)} - \frac{1}{\tilde{\Pi}(b)} = \sum_{i=1}^n \left[\frac{1}{\tilde{\Pi}(x_{i-1})} - \frac{1}{\tilde{\Pi}(x_i)}\right] = \sum_{i=1}^n \frac{\tilde{\Pi}(x_i) - \tilde{\Pi}(x_{i-1})}{\tilde{\Pi}(x_i)\tilde{\Pi}(x_{i-1})} = \int_{(a,b]} h_n(v)\tilde{\Pi}(dv)$$

where

$$h_n(v) = \frac{1}{\tilde{\Pi}(x_i)\tilde{\Pi}(x_{i-1})}$$

for $x_{i-1} < v \leq x_i$. Since $\Delta_n \to 0$,

$$h_n(v) \to \frac{1}{\tilde{\Pi}(v)\tilde{\Pi}(v-)}$$

for any $v \in (a, b]$ as $n \to \infty$. Since h_n is nonnegative and $\int_{(a,b]} h_n(v)\tilde{\Pi}(dv)$ is finite and constant in n, an argument analogous to the proof of Scheffé's theorem shows that

$$\int_{(a,b]} h_n(v)\tilde{\Pi}(dv) \to \int_{(a,b]} \frac{\tilde{\Pi}(dv)}{\tilde{\Pi}(v)\tilde{\Pi}(v-)},$$

which completes the proof of Lemma 2.

Using Lemma 2 we can now rewrite $T_n(w)$:

Lemma 3. For any w with $\tilde{\Pi}(w) > 0$,

$$T_n(w) = \int_w^\infty \frac{n\lambda^{n-1}(v)}{v^2} dv.$$

Proof of Lemma 3. Using the integration by parts formula for Lebesgue-Stieltjes integrals (e.g. Hewitt and Stromberg, 1975, Theorem 21.67) together with (5), (6) and the result of Lemma 2,

$$\begin{aligned}
T_n(w) &= 1 - \frac{\lambda^n(w)}{\tilde{\Pi}(w)} + \int_{v>w} \lambda^n(v) d\left(-\frac{1}{\tilde{\Pi}(v)}\right) \\
&= 1 - \frac{\lambda^n(w)}{\tilde{\Pi}(w)} + \left[-\frac{\lambda^n(v)}{\tilde{\Pi}(v)}\right]_w^\infty + \int_w^\infty \frac{1}{\tilde{\Pi}(v)} \frac{d}{dv}(\lambda^n(v)) dv \\
&= \int_w^\infty \frac{1}{\tilde{\Pi}(v)} \frac{n\lambda^{n-1}(v)\tilde{\Pi}(v)}{v^2} dv \\
&= \int_w^\infty \frac{n\lambda^{n-1}(v)}{v^2} dv,
\end{aligned}$$

which completes the proof of Lemma 3.

The representation of Lemma 3 shows that $\tilde{T}_1(w) = 1/w$ when $\tilde{\Pi}(w) > 0$, and thus proves Theorem 1 for $n = 1$. To prove the theorem for $n > 1$ we need to show that the formula for $p_n(x, dy)$ satisfies the Chapman-Kolmogorov equation

$$p_{n+1}(x, dy) = \int p_n(z, dy) p_1(x, dz).$$

Since

$$\begin{aligned}
&\int \left[T_n(w(z)\vee w(y))\Pi(dy)+\lambda^n(w(z))\delta_z(dy)\right]p_1(x,dz)\\
&\quad=\int \left[T_n(w(z)\vee w(y))\Pi(dy)+\lambda^n(w(z))\delta_z(dy)\right]\left[\frac{1}{w(x)\vee w(z)}\Pi(dz)+\lambda(w(x))\delta_x(dz)\right]\\
&\quad=\left[\int \frac{1}{w(x)\vee w}T_n(w\vee w(y))\tilde{\Pi}(dw)+\frac{\lambda^n(w(y))}{w(x)\vee w(y)}+\lambda(w(x))T_n(w(x)\vee w(y))\right]\Pi(dy)\\
&\qquad+\lambda^{n+1}(w(x))\delta_x(dy),
\end{aligned}$$

it is enough to show that

$$T_{n+1}(u\vee v)=\int \frac{1}{u\vee w}T_n(w\vee v)\tilde{\Pi}(dw)+\frac{\lambda^n(v)}{u\vee v}+\lambda(u)T_n(u\vee v) \tag{8}$$

for $n\geq 1$ and all v with $\tilde{\Pi}(v)>0$.

Suppose $u,v\geq 0$ and $\tilde{\Pi}(v)>0$. Using Lemma 3 and Fubini's theorem the integral on the right hand side of (8) can be written as

$$\begin{aligned}
\int \frac{1}{u\vee w}T_n(w\vee v)\tilde{\Pi}(dw)&=\int\int_{w\vee v}^{\infty}\frac{1}{u\vee w}\frac{n\lambda^{n-1}(z)}{z^2}dz\tilde{\Pi}(dw)\\
&=\int_v^{\infty}\int_{[0,z]}\frac{1}{u\vee w}\tilde{\Pi}(dw)\frac{n\lambda^{n-1}(z)}{z^2}dz\\
&=\frac{1}{u}\int_v^{u\vee v}\frac{n\lambda^{n-1}(z)}{z^2}\tilde{\Pi}(z)dz\\
&\quad+\int_{u\vee v}^{\infty}\left[\frac{\tilde{\Pi}(u)}{u}+\tilde{P}(z)-\tilde{P}(u)\right]\frac{n\lambda^{n-1}(z)}{z^2}dz.
\end{aligned} \tag{9}$$

¿From (6) we have

$$\int_v^{u\vee v}\frac{n\lambda^{n-1}(z)}{z^2}\tilde{\Pi}(z)dz=\lambda^n(u\vee v)-\lambda^n(v), \tag{10}$$

and (5) implies that

$$\begin{aligned}
&\int_{u\vee v}^{\infty}\left[\frac{\tilde{\Pi}(u)}{u}+\tilde{P}(z)-\tilde{P}(u)\right]\frac{n\lambda^{n-1}(z)}{z^2}dz\\
&\quad=\int_{u\vee v}^{\infty}\frac{n\lambda^{n-1}(z)}{z^2}\tilde{P}(z)dz-\lambda(u)\int_{u\vee v}^{\infty}\frac{n\lambda^{n-1}(z)}{z^2}dz\\
&\quad=\int_{u\vee v}^{\infty}\frac{n\lambda^{n-1}(z)}{z^2}\tilde{P}(z)dz-\lambda(u)T_n(u\vee v).
\end{aligned} \tag{11}$$

Finally, using (5) and integration by parts the integral on the right hand side of (11) becomes

$$\begin{aligned}
\int_{u\vee v}^{\infty} \frac{n\lambda^{n-1}(z)}{z^2}\tilde{P}(z)dz &= \int_{u\vee v}^{\infty} \frac{n\lambda^{n}(z)}{z^2}dz + \int_{u\vee v}^{\infty} \frac{n\lambda^{n-1}(z)}{z^2}\frac{\tilde{\Pi}(z)}{z}dz \\
&= \frac{n}{n+1}T_{n+1}(u\vee v) + \int_{u\vee v}^{\infty} \frac{1}{z}\frac{d}{dz}\left(\lambda^n(z)\right)dz \\
&= \frac{n}{n+1}T_{n+1}(u\vee v) + \left[\frac{\lambda^n(z)}{z}\right]_{u\vee v}^{\infty} + \int_{u\vee v}^{\infty} \frac{\lambda^n(z)}{z^2}dz \qquad (12) \\
&= \frac{n}{n+1}T_{n+1}(u\vee v) - \frac{\lambda^n(u\vee v)}{u\vee v} + \frac{1}{n+1}T_{n+1}(u\vee v) \\
&= T_{n+1}(u\vee v) - \frac{\lambda^n(u\vee v)}{u\vee v}
\end{aligned}$$

Combining (9), (10), (11), and (12) the integral on the right hand side of (8) is

$$\int \frac{1}{u\vee w}T_n(w\vee v)\tilde{\Pi}(dw) = \frac{\lambda^n(u\vee v)-\lambda^n(v)}{u} - \frac{\lambda^n(u\vee v)}{u\vee v} - \lambda(u)T_n(u\vee v) + T_{n+1}(u\vee v),$$

and thus the right hand side of (8) is

$$\begin{aligned}
&\frac{\lambda^n(v)}{u\vee v} + \frac{\lambda^n(u\vee v)-\lambda^n(v)}{u} - \frac{\lambda^n(u\vee v)}{u\vee v} + T_{n+1}(u\vee v) \\
&\quad = \left[\lambda^n(u\vee v)-\lambda^n(v)\right]\left[\frac{1}{u} - \frac{1}{u\vee v}\right] + T_{n+1}(u\vee v) \\
&\quad = T_{n+1}(u\vee v)
\end{aligned}$$

This proves (8) and completes the proof of Theorem 1.

4. Geometric rates of convergence

The essential supremum (with respect to P or Π) of the function $\lambda(w)$ is achieved when $w = w^*$ and is then $r = 1 - \frac{1}{w^*}$. It is clear from the form of (7) that $|1 - T_n(w)| \le Kr^n$ for some finite K, and all w. Hence for any Π-measurable set A,

$$\left|\int_{y\in A} p_n(x,dy) - \Pi(A)\right| = O(r^n) \quad \text{as } n\to\infty. \qquad (13).$$

Moreover, from the fact that the chain starting at x remains at x for at least n steps with probability $\lambda(w(x))^n$, this also serves as a lower bound on the rate of convergence, for any x.

Therefore, the chain has a geometric rate of convergence, with rate r, whenever $r < 1$. This occurs if and only if $w^* < \infty$, and then $r = 1 - \frac{1}{w^*}$. The converse result shows that this cannot be improved, and in particular, if $w^* = \infty$, the chain is not geometrically ergodic.

This result matches a number of others which have been obtained in recent years. Tierney (1994, Corollary 4) noted that when $w^* < \infty$, then the chain is geometrically ergodic with convergence rate r satisfying $r \leq 1 - \frac{1}{w^*}$. Mengersen and Tweedie (1996) showed that when $w^* = \infty$ the chain is not geometrically ergodic. Liu (1996) also showed $r \leq 1 - \frac{1}{w^*}$ directly by a simple coupling argument.

There are also some related results which apply to more general chains than the independence Metropolis sampler. For example, Roberts and Tweedie (1996) showed that for any algorithm with invariant measure Π not concentrated at a single point, for which the rejection probability $P(x, \{x\})$ is a measurable function of x,

$$\text{ess sup}_{\Pi} P(x, \{x\}) = 1$$

implies that the Markov chain is not geometriclaly ergodic.

Finally, we note that the independence sampler is *stochastically monotone* with respect to the function $w(\cdot)$: if x and x' are two starting states with $w(x) < w(x')$, then it is possible to construct two coupled chains X_n and X'_n, for $n \geq 0$ with $X_0 = x$, $X'_0 = x'$, such that $w(X_n) \leq w(X'_n)$ for all $n \geq 1$. A general theory for rates of convergence in stochastically monotone chains has been developed by Lund and Tweedie (1996).

5. An example: exponential distribution

Although the result of Theorem 1 is an explicit formula, it seems unlikely that there are many cases where the integrals can be explicitly evaluated and so expressed in terms of simple algebraic formulae. We present here one instance, albeit artificial, where this is possible.

Suppose $E = [0, \infty)$, μ is Lebesgue measure, and $\pi(x) = e^{-x}$, $x > 0$, $p(x) = \theta e^{-\theta x}$, $x > 0$. We assume $0 < \theta < 1$ to guarantee that $p(\cdot)$ has heavier tail density than $\pi(\cdot)$. Then $w(x) = \theta^{-1} e^{-(1-\theta)x}$ achieves its maximum $w^* = \theta^{-1}$ at $x = 0$, and the transition density is

$$p(x, y) = \begin{cases} \theta e^{-\theta y}, & y < x, \\ \theta e^{-y+(1-\theta)x}, & y > x. \end{cases}$$

It is easily checked that $\lambda(x) = (1 - \theta) e^{-\theta x}$, and

$$T_n(x) = 1 + (1-\theta)^n \frac{e^{x - n\theta x} - 1}{1 - n\theta} - (1-\theta)^n e^{x - n\theta x}.$$

Theorem 1 indicates that the transition probabilities are of the following form. Suppose $X_0 = x$. Then with probability $\lambda(x)^n$, the process does not move, i.e. $X_n = x$. Otherwise the process moves to a new point y say, and the transition density from x to y is given by

$$p_n(x,y) = \begin{cases} \pi(y)T_n(y), & y < x, \\ \pi(y)T_n(x), & y > x. \end{cases} \tag{14}$$

It is possible, though tedious, to verify (14) directly.

Now let us consider an application. From (14), we deduce

$$\Pr\{X_n > y | X_0 = x\} = \begin{cases} e^{-y} - \frac{(1-\theta)^n}{1-n\theta}(e^{-y} - e^{-n\theta y}), & y < x, \\ e^{-y}\left\{1 + (1-\theta)^n \frac{e^{x-n\theta x}-1}{1-n\theta} - (1-\theta)^n e^{x-n\theta x}\right\}, & y \geq x. \end{cases} \tag{15}$$

We can now see what this implies in terms of the rate of convergence to the limit e^{-y}. Indeed, from (15) we see that, provided $x > 0$,

$$\Pr\{X_n > y | X_0 = x\} = \left\{1 + \frac{(1-\theta)^n}{n\theta}\right\} e^{-y} + o\left\{\frac{(1-\theta)^n}{n}\right\}. \tag{16}$$

The case $x = 0$ is different because in this case the dominant term in the error is $(1-\theta)^n$, i.e. $\lambda(0)^n$.

Garren (1994) and Garren and Smith (1995) have proposed a statistical technique for estimating the convergence rate of a Markov chain, based on approximations of the form

$$\Pr\{X_n \in E | X_0 = x\} = \rho + a_2 \lambda_2^n + o(\lambda_2^n) \tag{17}$$

where E is some subset of the state space, λ_2 is the second largest eigenvalue of the transition matrix, and a_2 is a constant depending on both x and E. An approximation of this form can be rigorously justified in the case that the transition operator is Hilbert-Schmidt, a condition which is valid for many instances of the Gibbs sampler, but not, in general, for Metropolis-Hastings chains. The question therefore arises as to what form of approximation is appropriate in cases where the operator is not Hilbert-Schmidt.

For this particular example, (16) shows that the convergence rate is of $O(\lambda_2^n/n)$, where $\lambda_2 = 1-\theta$. More generally, our result shows how the n-th order transition probabilities may be expressed as integrals of $\lambda^n(x)$ and this suggests some form of Laplace approximation to obtain the convergence rate. Thus one reason for being interested in having an explicit formula for the n-step transition probabilities is that we now have a class of examples for which the exact convergence rate may be calculated analytically.

Acknowledgements. We are grateful to Jun Liu for a prepublication copy of his paper. Comments by Richard Tweedie and Gareth Roberts are also gratefully acknowledged.

RLS's research was partially supported by NSF grant DMS-9205112, and LT's by NSF grant DMS-9303557.

References

Garren, S.T. (1994), Parametric and nonparametric modeling of data using sampling algorithms. PhD Thesis, Department of Statistics, University of North Carolina.

Garren, S.T. and Smith, R.L. (1995), Estimating the second largest eigenvalue of a Markov transition matrix. Preprint; submitted for publication.

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

Hewitt, E. and Stromberg, K. R. (1975), *Real and Abstract Analysis*, New York, N. Y.: Springer-Verlag.

Liu, J.S. (1996), Metropolized independent sampling with comparisons to rejection sampling and importance sampling. To appear, *Statistics and Computing*.

Lund, R.B. and Tweedie, R. (1996), Geometric convergence rates for pathwise ordered Markov chains. To appear, *Mathematics of Operations Research*.

Mengersen, K.L. and Tweedie, R.L. (1996), Rates of convergence of the Hastings and Metropolis algorithms. *Ann. Statist.* **24**, to appear.

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

Tierney, L. (1994), Markov chains for exploring posterior distributions. *Annals of Statistics* **22**, 1701-1728.