Markov chain convergence: from finite to infinite

by

Jeffrey S. Rosenthal*

Department of Statistics, University of Toronto, Toronto, Ontario, Canada M5S 1A1

Phone: (416) 978-4594. Internet: `jeff@utstat.toronto.edu`

(November 1994; revised, September 1995.)

Summary. Bounds on convergence rates for Markov chains are a very widely-studied topic, motivated largely by applications to Markov chain Monte Carlo algorithms. For Markov chains on *finite* state spaces, previous authors have obtained a number of very useful bounds, including those which involve choices of paths. Unfortunately, many Markov chains which arise in practice are not finite. In this paper, we consider the extent to which bounds for finite Markov chains can be extended to infinite chains.

Our results take two forms. For countably-infinite state spaces $\mathcal{X}$, we consider the process of *enlargements* of Markov chains, namely considering Markov chains on finite state spaces $\mathcal{X}_1, \mathcal{X}_2, \ldots$ whose union is $\mathcal{X}$. Bounds for the Markov chains restricted to $\mathcal{X}_d$, if uniform in d, immediately imply bounds on $\mathcal{X}$. Results for finite Markov chains, involving choices of paths, can then be applied to countable chains. We develop these ideas and apply them to several examples of the Metropolis-Hastings algorithm on countable state spaces.

For uncountable state spaces, we consider the process of *refinements* of Markov chains. Namely, we break the original state space $\mathcal{X}$ into countable numbers of smaller and smaller pieces, and define a Markov chain on these finite pieces which approximates the original chain. Under certain continuity assumptions, bounds on the countable Markov chains, including those related to choices of paths, will imply bounds on the original chain. We develop these ideas and apply them to an example of an uncountable state space Metropolis algorithm.

Acknowledgements. I am grateful to Eric Belsley for several very helpful conversations, especially related to Proposition 1 herein. I thank the referee for very insightful comments.

* Partially supported by NSERC of Canada.

1. Introduction.

Quantitative geometric rates of convergence for Markov chains is now a widely studied topic, motivated in large part by applications to Markov chain Monte Carlo algorithms (see Gelfand and Smith, 1990; Smith and Roberts, 1993). On *finite* state spaces, much progress has recently been made, both in the form of general results (Diaconis, 1988; Sinclair and Jerrum, 1988; Jerrum and Sinclair, 1988; Diaconis and Stroock, 1991; Sinclair, 1992), and of results specifically related to Markov chain Monte Carlo (Hanlon, 1992; Frieze, Kannan, and Polson, 1994; Frigessi, Hwang, Sheu, and Di Stefano, 1993; Ingrassia, 1994; Liu, 1992; Belsley, 1993). On *infinite* state spaces, however, progress is much more limited (though for partial results see Lawler and Sokal, 1988; Amit and Grenander, 1991; Amit, 1991, 1993; Hwang, Hwang-Ma and Sheu, 1993; Meyn and Tweedie, 1993; Rosenthal, 1995a, 1995b, 1994; Baxter and Rosenthal, 1995; Roberts and Rosenthal, 1994).

In this paper we consider the extent to which previous bounds for finite chains (especially those involving choices of paths) can be extended to bounds for infinite chains. Our results fall into two categories. To study countably infinite chains, we consider *enlargements* of a sequence of related finite chains, and show that many of the finite results carry over to the countable chains. To study uncountable chains, we consider *refinements* of a sequence of related countable chains, and derive related quantitative bounds in this manner. Both techniques are illustrated through examples, all of which come from the Metropolis-Hastings algorithm (Metropolis et al., 1953; Hastings, 1970).

A review of results about finite Markov chains is given in Section 2. In Section 3 we discuss enlargements, and in Section 4 we discuss refinements. Three examples of enlargements, plus one example of a refinement, are given in Section 5.

2. Needed facts about finite chains.

Let $\mathcal{X}$ be a finite state space, and let $P(x,y)$ be an irreducible matrix of transition probabilities on $\mathcal{X}$. Assume P has a stationary distribution π, so that $\pi P = \pi$, and $\pi(x) > 0$ for all $x \in \mathcal{X}$. Let $\mathcal{M}$ be the set of all functions from $\mathcal{X}$ to $\mathbf{C}$, and let P act on $\mathcal{M}$ by $(fP)(y) = \sum_x f(x)P(x,y)$. Let an initial distribution be given by μ_0, regarded as an element of $\mathcal{M}$, so that $\mu_k = \mu_0 P^k$ is the distribution of the Markov chain after k

iterations. We are interested in bounds on the total variation distance

$$\|\mu_k - \pi\|_{\text{var}} \;=\; \sup_A |\mu_k(A) - \pi(A)| \;=\; \frac{1}{2}\sum_x |\mu_k(x) - \pi(x)|$$

between the distribution of the Markov chain after k iterations, and the stationary distribution π.

We introduce some notation (which shall also apply in the next section for countably infinite $\mathcal{X}$). Define an inner product on $\mathcal{M}$ by $< f, g >_{L^2(1/\pi)} = \sum_{x\in X} f(x)g(x)/\pi(x)$, and set $\|f\|_{L^2(1/\pi)} = \sqrt{< f, f >_{L^2(1/\pi)}}$. Finally, let $W = \{f \in \mathcal{M} \mid \sum_x f(x) = 0\}$, and set $\|P\big|_W\|_{L^2(1/\pi)} = \sup\{\|fP\|_{L^2(1/\pi)} \mid f \in W, \|f\|_{L^2(1/\pi)} = 1\}$.

Proposition 1. *We have*

$$\|\mu_k - \pi\|_{\text{var}} \;\leq\; \frac{1}{2}\|\mu_0 - \pi\|_{L^2(1/\pi)}\; \|P\big|_W\|^k_{L^2(1/\pi)}\,.$$

Proof. We have that

$$\begin{aligned}
\|\mu_k - \pi\|_{\text{var}} &= \frac{1}{2}\langle|\mu_k - \pi|,\; \pi\rangle_{L^2(1/\pi)} \\
&\leq \frac{1}{2}\|\mu_k - \pi\|_{L^2(1/\pi)} \\
&= \frac{1}{2}\|(\mu_0 - \pi)P^k\|_{L^2(1/\pi)} \\
&\leq \frac{1}{2}\|\mu_0 - \pi\|_{L^2(1/\pi)}\; \|P\big|_W\|^k_{L^2(1/\pi)}
\end{aligned}$$

as required. (We have used the Cauchy-Schwarz inequality and the definition of $\|P\big|_W\|_{L^2(1/\pi)}$, plus the observation that $(\mu_0 - \pi) \in W$.) ∎

Remarks.

1. The quantity $\|P\big|_W\|_{L^2(1/\pi)}$ is often referred to as the "second eigenvalue" of the Markov chain. For reversible Markov chains it is equal to the largest absolute value of any eigenvalue of P, excluding the eigenvalue 1 corresponding to the stationary distribution π.

2. It is easily computed that $\|\mu_0 - \pi\|_{L^2(1/\pi)} = \left(\sum_{x\in\mathcal{X}} \frac{\mu_0(x)^2}{\pi(x)}\right) - 1$; this may be helpful for computations.
3. If $\mu_0 = \delta_{x_0}$ is a point mass at x_0, then $\|\mu_0 - \pi\|^2_{L^2(1/\pi)} = \frac{1-\pi(x_0)}{\pi(x_0)}$. For such μ_0, with P reversible, this proposition reduces to Proposition 3 of Diaconis and Stroock (1991). The greater generality for μ_0 allowed here shall be especially important when we consider refinements in Section 4 below; there the individual probabilities $\pi(x_0)$ will all be approaching zero, so the bound of Diaconis and Stroock cannot be used directly.

In what follows we shall assume P is reversible with respect to π, meaning that $\pi(x)P(x,y) = \pi(y)P(y,x)$ for all $x,y \in \mathcal{X}$. Furthermore, for simplicity, we shall assume that P satisfies the following strong form of aperiodicity:

$$P(x,x) \;\geq\; a > 0, \qquad x \in \mathcal{X}\,.$$

This immediately implies that the eigenvalues of P are all real and are all at least $-1+2a$. Weaker conditions can be used instead to get lower bounds on the eigenvalues; see for example Proposition 2 of Diaconis and Stroock (1991). But such methods are not usually required, and for simplicity we do not consider them further here.

Under these assumptions, we can state general bounds of previous authors regarding $\|P\big|_W\|_{L^2(1/\pi)}$. Suppose, for each $x,y \in \mathcal{X}$ with $x \neq y$, we have chosen a *path* γ_{xy} from x to y consisting of a finite sequence of distinct directed "edges" $((v_0,v_1),(v_1,v_2),\ldots,(v_{L-1},v_L))$ with $v_0 = x$, $v_L = y$, and $P(v_i,v_{i+1}) > 0$ for each i. Then in terms of these paths, we have

(a) (Sinclair, 1992, Corollary 4)

$$\|P\big|_W\|_{L^2(1/\pi)} \leq \max(1-2a, 1-\frac{1}{8\eta^2})\,,$$

where $\eta = \sup_e Q(e)^{-1} \sum_{\gamma_{xy}\ni e} \pi(x)\pi(y)$, and if $e = (u,v)$ is a directed edge, then $Q(e) = \pi(u)P(u,v)$.

(b) (Sinclair, 1992, Theorem 5)

$$\|P\big|_W\|_{L^2(1/\pi)} \leq \max(1-2a, 1-\frac{1}{K})\,,$$

where $K = \sup_e Q(e)^{-1} \sum_{\gamma_{xy} \ni e} |\gamma_{xy}| \pi(x)\pi(y)$, and $|\gamma_{xy}|$ is the number of edges in γ_{xy}.

(c) (Diaconis and Stroock, 1991, Proposition 1)

$$\|P\big|_W\|_{L^2(1/\pi)} \leq \max(1-2a, 1-\frac{1}{\kappa}),$$

where $\kappa = \sup_e \sum_{\gamma_{xy} \ni e} |\gamma_{xy}|_Q \pi(x)\pi(y)$, and $|\gamma_{xy}|_Q = \sum_{e \in \gamma_{xy}} Q(e)^{-1}$.

Remarks.

1. In each of these bounds, the supremum is taken over all directed edges $e = (u, v)$ with $P(u, v) > 0$, and the sum is over points x and y such that (u, v) appears in γ_{xy}.
2. On a finite space $\mathcal{X}$, there are of course only a finite number of possible edges e, so the supremums above are actually maximums. However, we write the expressions as supremums so that the same formula will also apply in the next section.
3. If the collection of paths $\{\gamma_{xy}\}$ is itself symmetric, in the sense that for all $x \neq y$, γ_{yx} is simply the reversal of γ_{xy}, then clearly the direction of the edge e does not matter, so it suffices to take the supremums over edges pointed in only one direction. This shall be the case in all of the examples we consider.
4. These bounds remain true if the paths $\{\gamma_{xy}\}$ are chosen *randomly*, with η, κ, and K instead defined as supremums of *expected values* of the respective quantities (see Sinclair, 1992, Section 4). This fact shall be important in the final proof in Section 4 below.
5. In continuous time the situation is even simpler. Write $\overline{P}^t(x, y) = \sum_{n=0}^{\infty} e^{-t} \frac{t^n}{n!} P^n(x, y)$ for the corresponding continuous-time Markov operator (with mean-1 exponential holding times). Then if P is reversible with eigenvalues $\{\beta_i\}$, then the eigenvalues of $\overline{P}^t$ are $\{\sum_{n=0}^{\infty} e^{-t} \frac{t^n}{n!} \beta_i^n\} = \{e^{-t(1-\beta_i)}\}$ and hence are all positive. The bounds corresponding to the above are then

$$\|\overline{P}^t\big|_W\|_{L^2(1/\pi)} \leq e^{-t \max(\frac{1}{8\eta^2}, \frac{1}{K}, \frac{1}{\kappa})}.$$

In particular, the condition $P(x, x) \geq a > 0$ is no longer required.

3. Enlargements of Markov chains.

We suppose now that $\mathcal{X}$ is a countably infinite state space, and $P(x, y)$ is an irreducible Markov chain defined on $\mathcal{X}$ with initial distribution μ_0. We further assume that P is reversible with respect to a stationary distribution π on $\mathcal{X}$.

The idea of enlargements is as follows. We decompose $\mathcal{X}$ as $\mathcal{X} = \cup_d \mathcal{X}_d$ where each $\mathcal{X}_d \subseteq \mathcal{X}$ is finite, and $\mathcal{X}_1 \subseteq \mathcal{X}_2 \subseteq \ldots$. For d large enough so that $\pi(\mathcal{X}_d) > 0$ and $\mu_0(\mathcal{X}_d) > 0$, let π_d be the probability measure on $\mathcal{X}_d$ defined by $\pi_d(x) = \pi(x)/\pi(\mathcal{X}_d)$ for $x \in \mathcal{X}_d$, and similarly let $\mu_{0,d}(x) = \mu_0(x)/\mu_0(\mathcal{X}_d)$ for $x \in \mathcal{X}_d$. Further define $P_d(x, y)$ on $\mathcal{X}_d$ by $P_d(x, y) = P(x, y)$ for $x, y \in \mathcal{X}_d$, $x \neq y$, and $P_d(x, x) = 1 - \sum_{y \neq x} P_d(x, y) = P(x, x) + P(x, \mathcal{X}_d^C)$. Then clearly P_d is reversible with respect to π_d on $\mathcal{X}_d$.

Proposition 2. *Let $P(\cdot, \cdot)$ be an irreducible Markov chain on a countable state space $\mathcal{X}$, reversible with respect to $\pi(\cdot)$, and with initial distribution $\mu_0(\cdot)$. Let $\mathcal{X}_d$, π_d, $\mu_{0,d}$, and $P_d(\cdot, \cdot)$ be as above. Set $\mu_k = \mu_0 P^k$ and $\mu_{k,d} = \mu_{0,d} P_d^k$. Then for each fixed $x \in \mathcal{X}$ and $k \geq 0$, as $d \to \infty$ we have $\pi_d(x) \to \pi(x)$ and $\mu_{k,d}(x) \to \mu_k(x)$, and furthermore $\|\mu_{0,d} - \pi_d\|_{L^2(1/\pi_d)} \to \|\mu_0 - \pi\|_{L^2(1/\pi)}$ and $\|\mu_{k,d} - \pi_d\|_{\text{var}} \to \|\mu_k - \pi\|_{\text{var}}$.*

Proof. Since $\{\mathcal{X}_d\}$ are increasing, we have $\pi(\mathcal{X}_d) \to \pi(\mathcal{X}) = 1$. Hence for d large enough that $x \in \mathcal{X}_d$, we have $\pi_d(x) = \pi(x)/\pi(\mathcal{X}_d) \to \pi(x)$. For the second statement, write $\mu_k(x) = P(A_d) + P(B_d) + P(C_d)$, where A_d is the event that the path of the original Markov chain ends up at x (after k steps) without ever leaving $\mathcal{X}_d$ and without ever holding at a point, while B_d is the event that it ends up at x but does leave $\mathcal{X}_d$ at some point during the first k steps, and C_d is the event that it ends up at x without leaving $\mathcal{X}_d$ but with holding at least once. Now, as for $\mu_{k,d}(x)$, since $\mu_{0,d}(s) = \mu_0(s)/\mu_0(\mathcal{X}_d)$ for $s \in \mathcal{X}_d$, we have that $\mu_{k,d}(x) = P(A_d)/\mu_0(\mathcal{X}_d) + P(D_d)$, where D_d is the event that the chain corresponding to $P_d(\cdot, \cdot)$ ends up at x but holds at least once. Now, as $d \to \infty$, we have $\mu_0(\mathcal{X}_d) \to 1$, $P(D_d) \to P(C_d)$, and $P(B_d) \to 0$, so that $\mu_{k,d}(x) \to \mu_k(x)$.

For the statement about $L^2(1/\pi)$, we have that

$$\|\mu_{0,d} - \pi_d\|^2_{L^2(1/\pi_d)} = \sum_{x \in \mathcal{X}_d} \frac{\mu_{0,d}(x)^2}{\pi_d(x)} - 1 = \frac{\pi(\mathcal{X}_d)}{\mu_0(\mathcal{X}_d)^2} \sum_{x \in \mathcal{X}_d} \frac{\mu_0(x)^2}{\pi(x)} - 1$$

$$\to \sum_{x \in \mathcal{X}} \frac{\mu_0(x)^2}{\pi(x)} - 1 = \|\mu_0 - \pi\|^2_{L^2(1/\pi)} .$$

For the statement about variation distance, given $\epsilon > 0$, choose a finite subset $S \subseteq \mathcal{X}$ with $\pi(S) \geq 1 - \epsilon/4$ and $\mu_k(S) \geq 1 - \epsilon/4$. Then choose d_0 with $S \subseteq \mathcal{X}_{d_0}$, and with $|\mu_{k,d}(x) - \mu_k(x)| \leq \epsilon/4|S|$ and $|\pi_d(x) - \pi(x)| \leq \epsilon/4|S|$ for all $d \geq d_0$ and all $x \in S$. We then have, for $d \geq d_0$, that $\pi_d(S^C) \leq \epsilon/2$ and $\mu_{k,d}(S^C) \leq \epsilon/2$, so

$$2\|\mu_{k,d} - \pi_d\|_{\text{var}} = \sum_{x \in \mathcal{X}_d} |\mu_{k,d}(x) - \pi_d(x)|$$

$$\leq \epsilon/2 + \sum_{x \in S} |\mu_{k,d}(x) - \mu_k(x) + \mu_k(x) - \pi(x) + \pi(x) - \pi_d(x)|$$

$$\leq \epsilon/2 + \sum_{x \in S} (\epsilon/4|S| + \epsilon/4|S| + |\mu_k(x) - \pi(x)|)$$

$$\leq \epsilon + 2\|\mu_k - \pi\|_{\text{var}} .$$

It follows that $\limsup \|\mu_{k,d} - \pi_d\|_{\text{var}} \leq \|\mu_k - \pi\|_{\text{var}}$. Similarly $\liminf \|\mu_{k,d} - \pi_d\|_{\text{var}} \geq \|\mu_k - \pi\|_{\text{var}}$. The result follows. ∎

Combining the above two propositions, and letting $d \to \infty$ (along a subsequence if necessary), we obtain

Corollary 3. *Under the above assumptions, if* $\liminf_{d \to \infty} \|P_d\big|_W\|_{L^2(1/\pi_d)} \leq \beta$, *then*

$$\|\mu_k - \pi\|_{\text{var}} \leq \frac{1}{2} \|\mu_0 - \pi\|_{L^2(1/\pi)} \beta^k .$$

This corollary says that we can bound the distance to stationarity on the countably infinite chain $\mathcal{X}$ by any uniform bound on the sequence of finite chains $\{\mathcal{X}_d\}$. (A similar idea is used in Belsley, 1993, Theorem VI-4-2.)

To make use of this fact, we make the following definition. A set of paths $\{\gamma_{xy}\}$ on $\mathcal{X}$ is *unfolding* if there exists a sequence of finite subsets $\mathcal{X}_d$ of $\mathcal{X}$ with $\mathcal{X}_1 \subseteq \mathcal{X}_2 \subseteq \ldots$ and $\mathcal{X} = \cup_d \mathcal{X}_d$, such that for any $x, y \in \mathcal{X}_d$, the path γ_{xy} connecting x to y lies entirely inside $\mathcal{X}_d$. Not all collections of paths will be unfolding: for example, suppose $\mathcal{X}$ is the non-negative integers, and for each $x > y$, the path from x to y includes the point $x+1$. However, natural choices of paths will usually be unfolding. And for such unfolding paths, we can use the finite-chain bounds to obtain information about the infinite chain, as follows.

Theorem 4. *Let $P(x,y)$ be an irreducible Markov chain on a countably infinite state space $\mathcal{X}$, reversible with respect to a probability distribution π, and with $P(x,x) \geq a > 0$ for all $x \in \mathcal{X}$. Suppose that for each $x, y \in \mathcal{X}$ with $x \neq y$ we have chosen a path γ_{xy} from x to y, and suppose further that this collection of paths is unfolding as defined above. Then given an initial distribution μ_0 on $\mathcal{X}$, and setting $\mu_k = \mu_0 P^k$, we have*

$$\|\mu_k - \pi\|_{\text{var}} \leq \frac{1}{2} \|\mu_0 - \pi\|_{L^2(1/\pi)} \beta^k$$

where $\beta = \max(1-2a, \min(1-\frac{1}{8\eta^2}, 1-\frac{1}{K}, 1-\frac{1}{\kappa}))$, with η, K, and κ as defined in Section 2. (Note that these quantities now involve supremums over infinite numbers of edges and hence might be infinite, in which case we adopt the convention that $\frac{1}{\infty} = 0$.)

Proof. Let $\{\mathcal{X}_d\}$ be a nested sequence of subsets of $\mathcal{X}$ with respect to which the paths $\{\gamma_{xy}\}$ are unfolding. Then $\{\gamma_{xy}\}_{x,y \in \mathcal{X}_d}$ is a collection of paths on $\mathcal{X}_d$. The finite-chain bounds of the previous section, together with Proposition 1, immediately imply the analogous bounds for the finite chain P_d as above. The stated results for $\mathcal{X}$ follow by taking the limit $d \to \infty$ and using the previous corollary. ∎

Remarks.

1. As in the final remark of Section 2, in continuous time the situation is even simpler, and we obtain

$$\|\overline{P}^t - \pi\|_{\text{var}} \leq \frac{1}{2} \|\mu_0 - \pi\|_{L^2(1/\pi)} \, e^{-t \max(\frac{1}{8\eta^2}, \frac{1}{K}, \frac{1}{\kappa})},$$

with no requirement that $P(x,x) \geq a > 0$.

2. Our original goal was to generalize to infinite chains the elegant results of Ingrassia (1994) regarding bounds on finite versions of Metropolis-Hastings and Gibbs sampler

algorithms. Unfortunately, this appears not to be possible. For example, his results use quantities such as $b_\Gamma = \max_e \#\{\gamma_{xy} \,|\, e \in \gamma_{xy}\}$ and $d^* = \max_{x\in\mathcal{X}} \#\{y \,|\, P(x,y) > 0\}$, and it is easily seen that if both of these quantities are finite, then (since a point $x \in \mathcal{X}$ must be connected to *every* point $y \neq x$) we must have $|\mathcal{X}| \leq b_\Gamma \, d^* + 1 < \infty$. Hence, we have instead concentrated on generalizing the less specific results involving choices of paths.

4. Refinements of Markov chains.

In this section we consider extensions of the theory to uncountable state spaces. We assume throughout that $\mathcal{X}$ is an open subset of $\mathbf{R}^m$ with C^1 boundary. (More general spaces are also possible, but we will use differentiability properties so the generalizations are non-trivial.) We consider a Markov chain with initial distribution $\mu_0(\cdot)$, and transition probabilities $P(x,\cdot)$, reversible with respect to a stationary distribution $\pi(\cdot)$, and irreducible with respect to Lebesgue measure λ on $\mathbf{R}^m$.

We impose some regularity conditions. Call a subset of $\mathbf{R}^n$ *gentle* if it is contained in some finite union of C^1 hypersurfaces inside $\mathbf{R}^n$. (Intuitively, a gentle set is small and unimportant.) We assume that $\mu_0(\cdot)$ has density r (with respect to λ), and that $\pi(\cdot)$ has density h, such that $h > 0$ on $\mathcal{X}$, and such that r^2/h is a bounded function. We further assume that for each $x \in \mathcal{X}$, $P(x,\cdot)$ is of the form $P(x,dy) = a_x \delta_x(dy) + f_x(y)\lambda(dy)$. We assume that $a = \inf_x a_x > 0$, that each of $r(\cdot)$, $h(\cdot)$, $a_\cdot$, and $\int_A f_\cdot d\lambda$ are uniformly continuous functions off of some specific gentle subset of $\mathbf{R}^m$, and that $f_x(y)$ is a uniformly continuous function of $(x,y) \in \mathcal{X} \times \mathcal{X}$ off of some specific gentle subset of $\mathbf{R}^{2m}$. Reversibility then implies that $h(x)f_x(y) = h(y)f_y(x)$ except when (x,y) is in some specific gentle set in $\mathbf{R}^{2m}$.

Remark. Allowing discontinuities on certain gentle sets is only a very minor and unimportant weakening of our basic regularity conditions, and it is not really the main thrust of our result. We include it simply to allow for such probability distributions $P(x,\cdot)$ as, say, uniform distributions on nice subsets of $\mathcal{X}$.

To describe our result, we again choose a collection of paths. Specifically, for each $x \neq$

y, we let $\gamma_{xy} = \big((v_0, v_1), \ldots, (v_{L-1}, v_L)\big)$ be a sequence of edges, for some $L = |\gamma_{xy}| < \infty$, where each $v_i \in \mathcal{X}$, with $v_0 = x$, $v_L = y$, with $\{v_\ell\}_{0\le\ell\le L}$ distinct, and with $f_{v_i}(v_{i+1}) > 0$ for $0 \le i \le L-1$. We set $\gamma_{xy}(\ell) = v_\ell$ for $\ell \le |\gamma_{xy}|$. We assume that, for each ℓ, the subset of $\mathcal{X}\times\mathcal{X}$ on which $\gamma_{xy}(\ell)$ is defined has C^1 boundaries, and that on that subset $\gamma_{xy}(\ell)$ is a C^1 function of (x, y) except on some gentle subset of $\mathbf{R}^{2m}$. We further assume that $\{\gamma_{xy}\}$ is *unfolding* in the sense that there are bounded sets $S_1 \subseteq S_2 \subseteq \ldots$ with C^1 boundaries and with $\mathcal{X} = \cup_n S_n$, such that if $x, y \in S_j$, then $\gamma_{xy}(\ell) \in S_j$ for all $0 \le \ell \le |\gamma_{xy}|$. (If $\mathcal{X}$ is itself bounded then this condition is satisfied automatically.)

To deal with the possible discontinuities on gentle sets, we use the following notation. Given a function f which may have discontinuities or even be undefined at some points, we let $\lceil f\rceil(x)$ [resp. $\lfloor f\rfloor(x)$] be the limsup [resp. liminf] of f, i.e. the limit as $\epsilon \searrow 0$ of the supremum [resp. infemum] of the values of f (where defined) in an ϵ-ball around x. Thus

$$\lfloor f\rfloor_z(w) = \lim_{\epsilon\to 0^+} \inf_{\substack{\|z'-z\|<\epsilon \\ \|w'-w\|<\epsilon}} f_{z'}(w')\,,$$

etc. Where f is continuous, we of course have $\lceil f\rceil = \lfloor f\rfloor = f$.

Finally, for $\ell \le |\gamma_{xy}|$, we define

$$J_{xy}(\ell) \;=\; \det \begin{pmatrix} \frac{\partial}{\partial x_i}\left(\gamma_{xy}(\ell)\right)_j & \frac{\partial}{\partial x_i}\left(\gamma_{xy}(\ell+1)\right)_j \\ \frac{\partial}{\partial y_i}\left(\gamma_{xy}(\ell)\right)_j & \frac{\partial}{\partial y_i}\left(\gamma_{xy}(\ell+1)\right)_j \end{pmatrix}$$

to be the Jacobian of the mapping $(x, y) \mapsto (\gamma_{xy}(\ell), \gamma_{xy}(\ell+1))$. (If we are at an exceptional point where $\gamma_{xy}(\ell)$ or $\gamma_{xy}(\ell{+}1)$ is not C^1, then $J_{xy}(\ell)$ may not be defined, but $\lfloor J\rfloor_{xy}(\ell)$ and $\lceil J\rceil_{xy}(\ell)$ still will be.) We assume that $\lfloor J\rfloor_{xy}(\ell) > 0$ for all $x, y \in \mathcal{X}$ and all $0 \le \ell \le |\gamma_{xy}|$.

We can now state

Theorem 5. *Let $P(\cdot,\cdot)$ be a Markov chain on an open subset $\mathcal{X} \subseteq \mathbf{R}^m$, reversible with respect to $\pi(\cdot)$, with initial distribution $\mu_0(\cdot)$, and satisfying the regularity conditions as above. Let $\{\gamma_{xy}\}$ be a collection of paths for each $x \ne y$, unfolding and satisfying the regularity conditions as above. Then*

$$\|\mu_k - \pi\|_{\mathrm{var}} \;\le\; \frac{1}{2}\,\|\mu_0 - \pi\|_{L^2(1/\pi)}\,\beta^k$$

where

$$\|\mu_0 - \pi\|^2_{L^2(1/\pi)} = \int\limits_{\mathcal{X}} \frac{(r(x) - h(x))^2}{h(x)} \ \lambda(dx)\,,$$

and where $\beta = \max(1 - 2a, \min(1 - \frac{1}{8\overline{\eta}^2}, 1 - \frac{1}{\overline{K}}, 1 - \frac{1}{\overline{\kappa}}))$, *with*

$$\overline{\eta} = \sup_{z,w \in \mathcal{X}} \Big(h\rfloor(z) \lfloor f \rfloor_z(w) \Big)^{-1} \sum_{\substack{x,y,\ell \\ \gamma_{xy}(\ell)=z \\ \gamma_{xy}(\ell+1)=w}} \frac{\lceil h \rceil(x) \lceil h \rceil(y)}{\lfloor J \rfloor_{xy}(\ell)}\,;$$

$$\overline{K} = \sup_{z,w \in \mathcal{X}} \Big(\lfloor h \rfloor(z) \lfloor f \rfloor_z(w) \Big)^{-1} \sum_{\substack{x,y,\ell \\ \gamma_{xy}(\ell)=z \\ \gamma_{xy}(\ell+1)=w}} |\gamma_{xy}| \frac{\lceil h \rceil(x) \lceil h \rceil(y)}{\lfloor J \rfloor_{xy}(\ell)}\,;$$

$$\overline{\kappa} = \sup_{z,w \in \mathcal{X}} \sum_{\substack{x,y,\ell \\ \gamma_{xy}(\ell)=z \\ \gamma_{xy}(\ell+1)=w}} |\gamma_{xy}|_Q \frac{\lceil h \rceil(x) \lceil h \rceil(y)}{\lfloor J \rfloor_{xy}(\ell)}\,,$$

where

$$|\gamma_{xy}|_Q = \sum_{j=0}^{|\gamma_{xy}|-1} \left(\frac{\lfloor h \rfloor \left(\gamma_{xy}(j)\right) \lfloor f \rfloor_{\gamma_{xy}(j)} \left(\gamma_{xy}(j+1)\right)}{\lceil J \rceil_{xy}(j)} \right)^{-1}\,.$$

Remark. We emphasize that this theorem says nothing about the existence or properties of paths $\{\gamma_{xy}\}$ satisfying the stated regularity conditions; it merely provides a bound on $\|\mu_k - \pi\|_{\text{var}}$, *assuming* that such paths have been constructed. Furthermore, our regularity conditions can likely be improved upon; we have not made a serious effort to find the weakest conditions possible.

Proof. For each $d = 1, 2, 3, \ldots$, partition $\mathcal{X}$ into connected measurable subsets $\{B_{di}\}_{i \in I_d}$, where I_d is finite or countable, where B_{di} has diameter and Lebesgue-measure both less than $1/d$, and where furthermore there is a nested sequence of subsets $\{S_j\}$ as above, with respect to which $\{\gamma_{xy}\}$ is unfolding, such that for each j and d there are only a finite number of i with $B_{di} \cap S_j$ non-empty, and for each such i we have $B_{di} \subseteq S_j$.

In terms of such a partition, we define a new Markov chain by $\mathcal{X}_d = I_d$, $\mu_{0,d}(i) = \mu_0(B_{di}) = \int\limits_{B_{di}} r\, d\lambda$, $\pi_d(i) = \pi(B_{di}) = \int\limits_{B_{di}} h\, d\lambda$, and

$$P_d(i,j) \;=\; \mathbf{E}_\pi \left(P(x, B_{dj}) \,|\, x \in B_{di} \right) \;=\; \frac{\int\limits_{B_{di}} \int\limits_{B_{dj}} f_x(y) h(x) \lambda(dy) \lambda(dx)}{\pi_d(i)}\,.$$

Then it is easily verified that $P_d(\cdot,\cdot)$ is a Markov chain on $\mathcal{X}_d$ which is reversible with respect to $\pi_d(\cdot)$. We let $\mu_{k,d} = \mu_{0,d} P_d^k$ be the distribution of this Markov chain after k iterations.

We define paths $\{\gamma_{dij}\}$ on $\mathcal{X}_d$ *randomly* (see Remark 4 at the end of Section 2) as follows. First choose points $x_{di} \in B_{di}$ for each $i \in I_d$, chosen randomly according to normalized Lebesgue measure on B_{di}. Then, in notation as above, set $\gamma_{dij}(\ell) = c$ if $\gamma_{x_{di}x_{dj}}(\ell) \in B_{dc}$. Our assumptions imply that the random paths $\{\gamma_{dij}\}$ are unfolding, with probability 1, in the countable-$\mathcal{X}$ sense of the previous section.

Our previous theorem (for countable chains) thus implies bounds on the Markov chain $P_d(\cdot,\cdot)$ on $\mathcal{X}_d$, in terms of its corresponding quantities η_d, K_d, and κ_d. The current theorem will thus follow from the following lemma:

Lemma 6. *Under the above conditions, and assuming* $\|\mu_0 - \pi\|_{L^2(1/\pi)} < \infty$, *we have*

$$\lim_{d\to\infty} \|\mu_{0,d} - \pi_d\|_{L^2(1/\pi_d)} \;=\; \|\mu_0 - \pi\|_{L^2(1/\pi)}\,;$$

$$\limsup_{d\to\infty} \|\mu_{k,d} - \pi_d\|_{\mathrm{var}} \;=\; \|\mu_k - \pi\|_{\mathrm{var}}\,;$$

and furthermore

$$\limsup_{d\to\infty} \eta_d \le \overline{\eta}\,; \quad \limsup_{d\to\infty} K_d \le \overline{K}\,; \quad \limsup_{d\to\infty} \kappa_d \le \overline{\kappa}\,.$$

Proof. For the statement about $L^2(1/\pi)$, we have

$$1 + \|\mu_{0,d} - \pi_d\|_{L^2(1/\pi_d)} \;=\; \sum_{i\in I_d} \frac{\mu_{0,d}(i)^2}{\pi_d(i)} \;=\; \sum_{i\in I_d} \frac{\left(\int\limits_{B_{di}} r\,d\lambda\right)^2}{\int\limits_{B_{di}} h\,d\lambda}\,.$$

By continuity, off of gentle sets, we can find $x_{di}^*, y_{di}^* \in B_{di}$ with $\int\limits_{B_{di}} r\,d\lambda = r(x_{di}^*)\lambda(B_{di})$ and $\int\limits_{B_{di}} h\,d\lambda = h(y_{di}^*)\lambda(B_{di})$. Hence,

$$1 + \|\mu_{0,d} - \pi_d\|_{L^2(1/\pi_d)} \;=\; \sum_{i\in I_d} \frac{r(x_{di}^*)^2}{h(y_{di}^*)}\,\lambda(B_{di})\,.$$

This is essentially a Riemann sum for r^2/h, except that we may have $x^*_{di} \neq y^*_{di}$. But since r^2/h is bounded and uniformly continuous and integrable (since $\|\mu_0 - \pi\|_{L^2(1/\pi)} < \infty$), it still follows (cf. Spivak, 1980, p. 263) that as $d \to \infty$, the sum will converge to

$$\int \frac{h^2}{r}\, d\lambda \;=\; 1 + \|\mu_0 - \pi\|_{L^2(1/\pi)}\,,$$

as required.

For the statement about variation distance, fix $\epsilon > 0$, and choose a bounded subset $S \subseteq \mathcal{X}$ with $\pi(S^C) < \epsilon$, and with probability less than ϵ that the continuous chain escapes from S in the first k steps. (Note that if $\mathcal{X}$ is bounded there is no need to consider S at all.) Assume for notational ease that $\lambda(S) \geq 1$. Then choose d_1 sufficiently large that for $d \geq d_1$, there is probability less than 2ϵ that the chain on $\mathcal{X}_d$ will escape from S, and furthermore probability less than ϵ that the chain on $\mathcal{X}_d$ will in the first k steps move from point i to point j where B_{dj} intersects S^C or a point of discontinuity of $f_{x_i}(\cdot)$. These conditions ensure that for $d \geq d_1$, the limitations of the set S and the discontinuities of the $f_x(\cdot)$ will only affect probabilities by $O(\epsilon)$ and hence can (and shall) be ignored.

Furthermore since the values a_x are uniformly continuous, it follows that as $d \to \infty$ the chains on $\mathcal{X}_d$ will hold with probabilities approaching the correct values for the original chain on $\mathcal{X}$. Thus, holding probabilities also can (and shall) be ignored in the calculation which follows.

Now, by uniform continuity, choose d_2 such that for $d \geq d_2$, the values of each of $f_x(\cdot)$, $r(\cdot)$, and $r_k(\cdot) - h(\cdot)$ vary by less than $\epsilon/\lambda(S)^{k+1}$ on each subset B_{dj}. Set $d_0 = \max(d_1, d_2)$.

Then for $d \geq d_0$, it follows that for any choices of $j_0, j_1, \ldots, j_{k-1}$, we will have $\mu_{0,d}(j_0) P_d(j_0, j_1) \ldots P(j_{k-2}, j_{k-1}) P(j_{k-1}, i)$ within $\epsilon\, \lambda(B_{dj_2}) \ldots \lambda(B_{dj_{k-1}}) \lambda(B_{di}) / \lambda(S)^{k+1}$ of the probability that the continuous chain goes from B_{dj_0} to B_{dj_1} to $\ldots$ to $B_{dj_{k-1}}$ to B_{di} in its first k steps. Thus, summing over those $j_0, \ldots, j_{k-1}$ for which $B_{dj_i} \subset S$, we see that $\mu_{k,d}(i)$ will be within $\epsilon/\lambda(S)$ of $\mu_k(B_{di})$.

We conclude that $|\mu_{k,d}(i) - \pi_d(i)|$ will be within $O(\epsilon/\lambda(S))$ of $|\mu_k(B_{di}) - \pi(B_{di})| = \big|\int_{B_{di}} (r_k - h) d\lambda\big|$ (where r_k is the density of μ_k with respect to λ). Summing over i and dividing by 2, we see that $\|\mu_{k,d} - \pi_d\|_{\text{var}}$ will be within $O(\epsilon)$ of

$$\frac{1}{2} \sum_{i \in I_d} \Big| \int_{B_{di}} (r_k - h) d\lambda \Big| \, .$$

Now, if $r_k - h$ does not change sign on B_{di}, then $\big| \int_{B_{di}} (r_k - h) d\lambda \big| = \int_{B_{di}} |r_k - h| d\lambda$. Furthermore, since $d \geq d_2$, the contribution to the above sum made by those i for which $r_k - h$ *does* change sign on B_{di} (and hence satisfies $|r_k - h| < \epsilon / \lambda(S)^{k+1}$ there), will be less than $\epsilon / \lambda(S)^k \leq \epsilon$. Hence, $\|\mu_{k,d} - \pi_d\|_{\text{var}}$ will be within $O(\epsilon)$ of

$$\frac{1}{2} \sum_{i \in I_d} \int_{B_{di}} |r_k - h| d\lambda \;=\; \frac{1}{2} \int_{\mathcal{X}} |r_k - h| d\lambda \;=\; \|\mu_k - \pi\|_{\text{var}} \, .$$

The statement about variation distance follows.

For the statement about η_d, recall that our paths are now *random,* so we must bound the *expected values* of quantities like $\sum_{\gamma_{xy} \ni e} \pi(x)\pi(y)$. To proceed, consider first the case where there are no gentle sets of discontinuity or non-differentiability. Consider an edge $e = (i, j)$ of $\mathcal{X}_d$. If a path γ_{dab} has $\gamma_{dab}(\ell) = i$ and $\gamma_{dab}(\ell + 1) = j$, then the corresponding points x_{da} and x_{db} must satisfy $\gamma_{x_{da}, x_{db}}(\ell) \in B_{di}$ and $\gamma_{x_{da}, x_{db}}(\ell + 1) \in B_{dj}$. That is, we must have $(x_{da}, x_{db}) \in g^{-1}(B_{di} \times B_{dj})$, where g_ℓ is the function on $\mathcal{X} \times \mathcal{X}$ taking (x, y) to $(\gamma_{xy}(\ell), \gamma_{xy}(\ell + 1))$ (with $g_\ell(x, y)$ undefined if $|\gamma_{xy}| < \ell + 1$). Hence taking expected values (with respect to the random choice of paths), and recalling that $\pi(\cdot)$ has density h, we obtain (writing $\mathbf{I}(\cdot)$ for the indicator function of an event) that

$$\mathbf{E}\Bigg(\sum_{\substack{a,b \\ \gamma_{dab}(\ell) = i \\ \gamma_{dab}(\ell+1) = j}} \pi_d(a)\pi_d(b) \Bigg) \;=\; \mathbf{E}\Bigg(\sum_{a,b} \pi_d(a)\pi_d(b)\, \mathbf{I}\Big(\gamma_{dab}(\ell) = i, \;\gamma_{dab}(\ell+1) = j \Big) \Bigg)$$

$$= \; \sum_{a,b} \pi_d(a)\pi_d(b)\, \mathbf{P}\Big(\gamma_{dab}(\ell) = i, \;\gamma_{dab}(\ell+1) = j \Big)$$

$$= \; \sum_{a,b} \pi_d(a)\pi_d(b) \frac{\lambda\left(g^{-1}(B_{di} \times B_{dj}) \cap (B_{da} \times B_{db}) \right)}{\lambda(B_{da} \times B_{db})}$$

$$= \; \int_{g^{-1}(B_{di} \times B_{dj})} \Theta(x, y) \lambda(dx) \lambda(dy) \, ,$$

where Θ is the piecewise-constant function defined by

$$\Theta(x,y) \;=\; \frac{\pi_d(a)\pi_d(b)}{\lambda(B_{da}\times B_{db})} \;=\; \frac{\int_{B_{da}\times B_{db}} h(x')h(y')\lambda(dx')\lambda(dy')}{\lambda(B_{da}\times B_{db})}\,, \qquad \text{for } (x,y)\in B_{da}\times B_{db}\,.$$

Now, as $d\to\infty$, the diameters of the sets $\{B_{dk}\}$ approach 0. Thus, uniform continuity implies (writing $\approx$ to mean that the ratio of the two quantities uniformly approaches 1 as $d\to\infty$) that $\Theta(x,y)\approx h(x)h(y)$. The sum then becomes a Riemann sum, whence we obtain that

$$\mathbf{E}\Bigg(\sum_{\substack{a,b\\ \gamma_{dab}(\ell)=i\\ \gamma_{dab}(\ell+1)=j}} \pi_d(a)\pi_d(b)\Bigg) \;\approx\; \int\limits_{g_\ell^{-1}(B_{di}\times B_{dj})} h(x)h(y)\lambda(dx)\lambda(dy)\,.$$

We also have that

$$\pi_d(i)P_d(i,j) \;=\; \int\limits_{B_{di}\times B_{dj}} h(x)\mathbf{E}_\pi(f_w(y)\,|\,w\in B_{di})\lambda(dx)\lambda(dy)\,.$$

Now, as $d\to\infty$, continuity implies that $\mathbf{E}_\pi(f_w(y)\,|\,w\in B_{di})\approx f_x(y)$ for any choice of $x=x(d)\in B_{di}$, whence

$$\pi_d(i)P_d(i,j) \;\approx\; \int\limits_{B_{di}\times B_{dj}} h(x)f_x(y)\lambda(dx)\lambda(dy)\,.$$

Finally, standard multi-variable calculus says that if $U_d\searrow\{z\}$ and $V_d\searrow\{w\}$, then

$$\frac{\int_{g_\ell^{-1}(U_d\times V_d)} h(x)h(y)\lambda(dx)\lambda(dy)}{\int_{U_d\times V_d} h(x)f_x(y)\lambda(dx)\lambda(dy)} \;\to\; \Big(h(z)f_z(w)\Big)^{-1}\sum_{\substack{x,y\\ g(x,y)=(z,w)}} \frac{h(x)h(y)}{J_{xy}(\ell)}\,.$$

We thus conclude that, as $d\to\infty$, quantities of the form

$$\mathbf{E}\Bigg(\Big(\pi_d(i)P_d(i,j)\Big)^{-1}\sum_{\substack{a,b\\ \gamma_{dab}(\ell)=i\\ \gamma_{dab}(\ell+1)=j}} \pi_d(a)\pi_d(b)\Bigg)$$

will be uniformly arbitrarily close to an expression of the form

$$\Big(h(z)f_z(w)\Big)^{-1}\sum_{\substack{x,y\\ \gamma_{xy}(\ell)=z\\ \gamma_{xy}(\ell+1)=w}} \frac{h(x)h(y)}{J_{xy}(\ell)}\,,$$

for an appropriate choice of $(z,w) \in \mathcal{X} \times \mathcal{X}$.

It follows that for fixed ℓ,

$$\limsup_{d\to\infty} \sup_{i,j\in I_d} \mathbf{E}\Bigg(\Big(\pi_d(i)P_d(i,j)\Big)^{-1} \sum_{\substack{a,b \\ \gamma_{dab}(\ell)=i \\ \gamma_{dab}(\ell+1)=j}} \pi_d(a)\pi_d(b)\Bigg)$$

$$\leq \sup_{z,w\in\mathcal{X}} \Big(h(z)f_z(w)\Big)^{-1} \sum_{\substack{x,y \\ \gamma_{xy}(\ell)=z \\ \gamma_{xy}(\ell+1)=w}} \frac{h(x)h(y)}{J_{xy}(\ell)} \,.$$

Summing over ℓ, the statement about η_d follows for this case.

To take account of possible discontinuities on gentle sets, we simply replace each computed quantity by its "worst case" values (thus preserving the inequality). This amounts to using the $\lceil\cdot\rceil$ operation in the numerators, and the $\lfloor\cdot\rfloor$ operation in the denominators, as done in the statement of the theorem.

The statements about κ_d and K_d are entirely similar. This completes the proof of the lemma, and hence also the proof of the theorem. ∎

Remarks.

1. The regularity conditions and proof above may seem rather technical. The essence, however, is that for well-behaved Markov chains $P(\cdot,\cdot)$ on uncountable sets $\mathcal{X} \subseteq \mathbf{R}^m$, bounds involving choices of paths can be used analogously to their use for finite chains.
2. Again, in continuous time the situation is even simpler, and we obtain

$$\|\overline{P}^t - \pi\|_{\text{var}} \;\leq\; \frac{1}{2}\,\|\mu_0 - \pi\|_{L^2(1/\pi)}\; e^{-t\,\max(\frac{1}{8\overline{\eta}^2},\frac{1}{\overline{K}},\frac{1}{\overline{\kappa}})}\,,$$

with no requirement that $P(x,x) \geq a > 0$.

5. Examples.

In this section we apply Theorems 4 and 5 to several examples. We note that all of the examples are versions of the Metropolis-Hastings algorithm (Metropolis et al, 1953; Hastings, 1970) with appropriate proposal distributions.

5.1. A geometric birth-death chain.

We suppose that $\mathcal{X} = \{0, 1, 2, \ldots\}$, and that for some numbers $a, b, c > 0$ with $a + b + c = 1$ and $b > c$, we have for all $x \geq 1$, $P(x, x) = a$, $P(x, x-1) = b$, $P(x, x+1) = c$, while for $x = 0$ we have $P(0, 0) = a + b$ and $P(0, 1) = c$. Such a chain is reversible with respect to the stationary distribution given by $\pi(x) = C(c/b)^x$ with $C = 1 - (c/b)$.

These and much more general birth-death chains have been studied in great detail by Belsley (1993, Chapter VI), using sophisticated ideas related to orthogonal polynomials. We here apply the ideas of this paper, by choosing paths and bounding the quantity η. (It appears that the quantity κ is always infinite for this example.)

We define unfolding paths γ_{xy} in the obvious way, namely that for $x < y$, $\gamma_{xy} = ((x, x+1), (x+1, x+2), \ldots, (y-1, y))$, with γ_{yx} defined symmetrically. Such paths are obviously unfolding, with respect to $\mathcal{X}_d = \{0, 1, 2, \ldots, d\}$. Then if $e = (z, z+1)$, then

$$Q(e)^{-1} \sum_{\gamma_{xy} \ni e} \pi(x)\pi(y) \;=\; \frac{1}{C\left(\frac{c}{b}\right)^z c} \sum_{x=0}^{z} \sum_{y=z+1}^{\infty} \pi(x)\pi(y)$$

$$= \frac{1}{Cc^{z+1}/b^z} \left(1 - (c/b)^{z+1}\right) (c/b)^{z+1} = \frac{1}{b-c} \left(1 - (c/b)^{z+1}\right) < \frac{1}{b-c} \,.$$

It follows that $\eta \leq \frac{1}{b-c}$, so that by Theorem 4,

$$\|\mu_k - \pi\|_{\text{var}} \;\leq\; \frac{1}{2} \, \|\mu_0 - \pi\|_{L^2(1/\pi)} \left(1 - \min(2a, \frac{1}{8(b-c)^2})\right)^k \,.$$

In particular, if $\mu_0 = \delta_{x_0}$ is a point mass, then

$$\|\mu_k - \pi\|_{\text{var}} \;\leq\; \frac{1}{2} \left(\frac{1 - \pi(x_0)}{\pi(x_0)}\right) \left(1 - \min(2a, \frac{1}{8(b-c)^2})\right)^k \,.$$

5.2. An infinite star.

Choose positive weights $\{w_i\}_{i=1}^{\infty}$ with $\sum_i w_i = \frac{1}{2}$. Then define a Markov chain on $\mathcal{X} = \{0, 1, 2, \ldots\}$ by $P(x, x) = \frac{1}{2}$ for all x, and for $i \geq 1$, $P(0, i) = w_i$ and $P(i, 0) = \frac{1}{2}$. This Markov chain is reversible with respect to the stationarity distribution given by $\pi(0) = \frac{1}{2}$ and $\pi(i) = w_i$ for $i \geq 1$. It may be pictured as an infinite "star", with 0 in the center and all the positive integers connected to 0 around the sides. (A finite version of this example, with equal weights, is discussed in Diaconis and Stroock, 1991, p. 49.)

We define paths in the obvious way, namely for $i, j \geq 1$ with $i \neq j$, set $\gamma_{ij} = ((i, 0), (0, j))$, while $\gamma_{0i} = (0, i)$ and $\gamma_{i0} = (i, 0)$. Hence $|\gamma_{xy}| \leq 2$ for all $x \neq y$. Then if $e = (i, 0)$, then

$$Q(e)^{-1} \sum_{\gamma_{xy} \ni e} |\gamma_{xy}| \pi(x) \pi(y) \leq 4(w_i)^{-1} \pi(i) \sum_{y \neq i} \pi(y) = 4(1 - w_i) \leq 4\,.$$

It follows that $K \leq 4$. Furthermore, we may take $a = \frac{1}{2}$ to get $P(x, x) \geq a$ for all x. Hence $\max(1 - 2a, 1 - \frac{1}{K}) \leq 3/4$, so that by Theorem 4,

$$\|\mu_k - \pi\|_{\text{var}} \;\leq\; \frac{1}{2}\, \|\mu_0 - \pi\|_{L^2(1/\pi)} \, \left(\frac{3}{4}\right)^k \,.$$

5.3. A two-dimensional Metropolis walk.

We here let $\mathcal{X} = \{0, 1, 2, \ldots\} \times \{0, 1, 2, \ldots\}$. For some fixed $0 < \rho < 1$, for all $i, j \geq 1$, we set $P((i, j), (i + 1, j)) = P((i, j), (i, j + 1)) = \rho/4$, and $P((i, j), (i - 1, j)) = P((i, j), (i, j - 1)) = 1/4$ with $P((i, j), (i, j)) = (1 - \rho)/2$. We set the boundary conditions in the obvious way by adding holding probability, so that $P((0, j), (0, j+1)) = P((i, 0), (i+1, 0)) = \rho/4$, $P((0, j), (1, j)) = P((i, 0), (i, 1)) = \rho/4$, $P((0, j), (0, j - 1)) = P((i, 0), (i - 1, 0)) = 1/4$, $P((0, j), (0, j)) = P((i, 0), (i, 0)) = (3 - 2\rho)/4$, and finally $P((0, 0), (1, 0)) = P((0, 0), (0, 1)) = \rho/4$, $P((0, 0), (0, 0)) = (1 - \rho)/2$.

This Markov chain is simply the Metropolized version of two-dimensional simple symmetric random walk, reversible with respect to $\pi((i, j)) = C\,\rho^{i+j}$ where $C = (1 - \rho)^2$. We again proceed by choosing paths and bounding η.

We choose paths as follows. If $i_1 \leq i_2$ and $j_1 \leq j_2$, with $(i_1, j_1) \neq (i_2, j_2)$, then we set

$$\gamma_{(i_1, j_1), (i_2, j_2)} = \big(((i_1, j_1), (i_1 + 1, j_1)), ((i_1 + 1, j_1), (i_1 + 2, j_1)), \ldots, ((i_2 - 1, j_1), (i_2, j_1)),$$

$$((i_2,j_1),(i_2,j_1+1)),\ldots,((i_2,j_2-1),(i_2,j_2))\big)\,,$$

while if $i_1 \le i_2$ but $j_1 > j_2$, we set

$$\gamma_{(i_1,j_1),(i_2,j_2)} = \big(((i_1,j_1),(i_1+1,j_1)),((i_1+1,j_1),(i_1+2,j_1)),\ldots,((i_2-1,j_1),(i_2,j_1)),$$

$$((i_2,j_1),(i_2,j_1-1)),\ldots,((i_2,j_2+1),(i_2,j_2))\big)\,.$$

For $i_1 > i_2$, we define $\gamma_{(i_1,j_1),(i_2,j_2)}$ to be the reversal of $\gamma_{(i_2,j_2),(i_1,j_1)}$. To summarize, then, $\gamma_{(i_1,j_1),(i_2,j_2)}$ is simply the path which adjusts each coordinate, one step at a time, adjusting the first coordinate first for $i_1 \le i_2$ and second for $i_1 > i_2$.

Now, if $e = ((i,j),(i+1,j))$, then

$$Q(e)^{-1}\sum_{\gamma xy \ni e}\pi(x)\pi(y) = \frac{1}{C\rho^{i+j+1}/4}\sum_{0\le i_1\le i}\pi((i_1,j))\sum_{\substack{i+1\le i_2<\infty\\ 0\le j_2<\infty}}\pi((i_1,j))\pi((i_2,j_2))$$

$$= \frac{4}{C\rho^{i+j+1}}\sum_{0\le i_1\le i}C\rho^{i_1+j}\sum_{i+1\le i_2<\infty}C\rho^{i_2}\sum_{0\le j_2<\infty}\rho^{j_2}$$

$$= \frac{4C}{\rho^{i+j+1}}\left(\frac{\rho^j-\rho^{i+1+j}}{1-\rho}\right)\left(\frac{\rho^{i+1}}{1-\rho}\right)\left(\frac{1}{1-\rho}\right)$$

$$< \frac{4C}{(1-\rho)^3} = \frac{4}{1-\rho}\,.$$

Similarly, if $e = ((i,j),(i,j+1))$, then

$$Q(e)^{-1}\sum_{\gamma xy \ni e}\pi(x)\pi(y) = \frac{1}{C\rho^{i+j+1}/4}\sum_{\substack{0\le i_1\le i\\ i+1\le i_2<\infty\\ 0\le j_0\le j}}\left(\pi((i_1,j_0))\pi((i_2,j)) + \pi((i_1,j))\pi((i_2,j_0))\right)$$

$$= \frac{4}{C\rho^{i+j+1}}\sum_{\substack{0\le i_1\le i\\ i+1\le i_2<\infty\\ 0\le j_0\le j}}C^2\left(\rho^{i_1+j_0}\rho^{i_2+j} + \rho^{i_1+j}\rho^{i_2+j_0}\right)$$

$$= \frac{8C}{\rho^{i+1}}\left(\frac{1-\rho^{i+1}}{1-\rho}\right)\left(\frac{\rho^{i+1}}{1-\rho}\right)\left(\frac{1-\rho^{i+1}}{1-\rho}\right)$$

$$< \frac{8C}{(1-\rho)^3} = \frac{8}{1-\rho}\,.$$

It follows that $\eta \leq \frac{8}{1-\rho}$, so that $1-\frac{1}{8\eta^2} \leq 1-\frac{(1-\rho)^2}{8^3}$. Furthermore we have $P(x,x) \geq a$ for all $x \in \mathcal{X}$, where $a = \frac{1-\rho}{2}$, so that $1-2a = \rho < 1 - \frac{(1-\rho)^2}{8^3}$. Theorem 4 thus implies that

$$\|\mu_k - \pi\|_{\text{var}} \;\leq\; \frac{1}{2}\,\|\mu_0 - \pi\|_{L^2(1/\pi)} \left(1 - \frac{(1-\rho)^2}{8^3}\right)^k .$$

5.4. A Metropolis chain for a truncated normal distribution.

We let $\mathcal{X} = \mathbf{R}$ be the one-dimensional real line, and consider a Metropolis algorithm with stationary distribution $\pi(\cdot)$ given by the standard normal distribution $N(0,1)$ truncated to $[-M, M]$, having density $h(y) = K\,e^{-y^2/2}$ for $|y| \leq M$ (with respect to Lebesgue measure λ), where $K^{-1} = \int\limits_{-M}^{M} e^{-y^2/2} dy$. We consider proposal distributions $Q(x,\cdot)$ given by the normal distributions $N(x,1)$ having densities $g_x(y) = \frac{1}{\sqrt{2\pi}} e^{-(y-x)^2/2}$. Assume that $M \geq 1$ so that $K \leq \left(\frac{3}{2}\right)\frac{1}{\sqrt{2\pi}} < 1$.

In the language of the Metropolis algorithm, the "acceptance probabilities" are thus given by $\alpha_{xy} = \min(1, \frac{e^{-y^2/2}}{e^{-x^2/2}})$ for $|y| \leq M$, and by $\alpha_{xy} = 0$ for $y > |M|$, and the Markov chain transitions are then given by

$$P(x, dy) \;=\; a_x\,\delta_x(dy) \;+\; \alpha_{xy} g_x(y)\lambda(dy) \;=\; a_x\,\delta_x(dy) \;+\; f_x(y)\lambda(dy)\,,$$

where $f_x(y) = \frac{1}{\sqrt{2\pi}} e^{-\max((y-x)^2,\,2y^2-2yx)/2}$ for $|y| \leq M$ and 0 otherwise, and $a_x = 1 - \int\limits_{-M}^{M} f_x(y)\lambda(dy)$.

We choose paths "linearly" as follows. Given $x, y \in \mathcal{X}$ with $x \neq y$, set $L_{xy} = \max(1, \lfloor |y-x| \rfloor)$ (i.e. the greatest integer not exceeding $\max(1, |y-x|)$), set $A_{xy} = \frac{y-x}{L_{xy}}$, and set $\gamma_{xy}(\ell) = x + \ell A_{xy}$ for $0 \leq \ell \leq L_{xy} = |\gamma_{xy}|$. (Note that $0 < A_{xy} < 2$ for all $x, y \in \mathcal{X}$.)

It is easily verified that all of the regularity and unfolding assumptions of Theorem 5 are then satisfied. It is further computed that

$$J_{xy}(\ell) \;=\; \det \begin{pmatrix} 1 - \frac{\ell}{L_{xy}} & \frac{\ell}{L_{xy}} \\ 1 - \frac{\ell+1}{L_{xy}} & \frac{\ell+1}{L_{xy}} \end{pmatrix} \;=\; 1/L_{xy}\,.$$

We now proceed to compute $\overline{\eta}$.

Given an edge $e = (z, w)$, assume without loss of generality (by symmetry around 0) that $z > |w|$. If $z - w < 1$, then the only path using the edge e is the length-one path γ_{zw}. Furthermore $J_{zw}(0) = 1$, and we obtain

$$\sum_{\substack{x,y,\ell \\ \gamma_{xy}(\ell)=z \\ \gamma_{xy}(\ell+1)=w}} \frac{h(x)h(y)}{J_{xy}(\ell)} \;=\; h(w)/f_z(w) \;=\; K\, e^{-w^2/2} / \frac{1}{\sqrt{2\pi}} e^{-\max((w-z)^2,\, 2w^2-2wz)/2}$$

$$= \sqrt{2\pi}\, K\, e^{(\max(z^2,w^2)-2wz)/2} \;\leq\; \frac{3}{2}\sqrt{e}\,.$$

If $1 \leq z - w < 2$, then the only way the edge $e = (z, w)$ can be in the path γ_{xy} is if for some non-negative integers m and n, we have $x = z + m(z - w)$ and $y = w - n(z - w)$. Hence

$$\sum_{\substack{x,y,\ell \\ \gamma_{xy}(\ell)=z \\ \gamma_{xy}(\ell+1)=w}} \frac{h(x)h(y)}{J_{xy}(\ell)} \;\leq\; \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} K^2\,(z-w)(m+n+1)\, e^{-(z+m(z-w))^2/2} e^{-(w-n(z-w))^2/2}$$

$$\leq (z-w)\left(\sum_{m=0}^{\infty}(m+1)\, e^{-(z+m(z-w))^2/2}\right)\left(\sum_{n=0}^{\infty}(n+1)\, e^{-(w-n(z-w))^2/2}\right).$$

Since $z \geq 0$, the first of these two sums is easily bounded by

$$\sum_{m=0}^{\infty}(m+1)\, e^{-(z^2+m(z-w)^2)/2} \;=\; \frac{e^{-z^2/2}}{(1-e^{-(z-w)^2/2})^2}\,.$$

The second sum is more difficult, since if w is large and positive then it is possible that $(w - n(z - w))$ will be very small (or even 0) for a large value of n, in which case the sum could still be very large. However, by making the change of variables $j = n - \lfloor \frac{w}{z-w} \rfloor$ and allowing j to range over all integers, the sum can be bounded by

$$\frac{w}{z-w} + 1 + 2\,\frac{\frac{w}{z-w}+2}{\left(1-e^{-(z-w)^2/2}\right)^2}\,.$$

We conclude that

$$(h(z)f_z(w))^{-1} \sum_{\substack{x,y,\ell \\ \gamma_{xy}(\ell)=z \\ \gamma_{xy}(\ell+1)=w}} \frac{h(x)h(y)}{J_{xy}(\ell)}$$

$$\leq \left(\frac{1}{\sqrt{2\pi}} K e^{-z^2/2} e^{-(w-z)^2/2}\right)^{-1} \frac{1}{2\pi}(z-w)e^{-z^2/2}\,\frac{3\frac{w}{z-w}+5}{(1-e^{-1/2})^4} \;\leq\; \frac{6e^2M}{(1-e^{-1/2})^4}\,,$$

where we have used that $K\sqrt{2\pi} \geq 2/3$, that $z - w < 2$, and that $w \leq M$.

We conclude that $\overline{\eta} \leq \frac{6e^2 M}{(1-e^{-1/2})^4}$. Furthermore, it is easily verified that the holding probabilities satisfy

$$a_x \geq \frac{1}{2} - \frac{1}{2\pi} \int_x^\infty e^{-(2y^2-2yx)/2} dy = \frac{1}{2} - \frac{1}{2\pi} \int_0^\infty e^{-z^2-zx} dz \geq \frac{1}{2} - \frac{1}{2\sqrt{2}}\,.$$

Hence, Theorem 5 implies that

$$\|\mu_k - \pi\|_{\text{var}} \;\leq\; \frac{1}{2}\,\|\mu_0 - \pi\|_{L^2(1/\pi)} \,\left(1 - \frac{(1-e^{-1/2})^8}{8(6e^2M)^2}\right)^k\,.$$

Remark. This example successfully bounds $\overline{\eta}$ for an uncountable chain. However, it is easily seen that as $M \to \infty$, we have $\overline{\eta} \to \infty$. (Indeed, consider the single term $m = 0, n = \lfloor M \rfloor - 1$ in the above sum for $\overline{\eta}$, with $z = M$ and $w = L(1 - \frac{1}{\lfloor M \rfloor})$.) Thus, in this example at least, Theorem 5 does not help to bound convergence as the state space becomes unbounded.

REFERENCES

Y. Amit (1991), On the rates of convergence of stochastic relaxation for Gaussian and Non-Gaussian distributions. J. Multivariate Analysis **38**, 89–99.

Y. Amit (1993), Convergence properties of the Gibbs sampler for perturbations of Gaussians. Tech. Rep. **352**, Department of Statistics, University of Chicago.

Y. Amit and U. Grenander (1991), Comparing sweep strategies for stochastic relaxation. J. Multivariate Analysis **37**, No. **2**, 197–222.

J.R. Baxter and J.S. Rosenthal (1995), Rates of convergence for everywhere-positive Markov chains. Stat. Prob. Lett. **22**, 333–338.

E.D. Belsley (1993), Rates of convergence of Markov chains related to association schemes. Ph.D. dissertation, Dept. of Mathematics, Harvard University.

P. Diaconis (1988), Group representations in probability and statistics. IMS Lecture Series **11**, Institute of Mathematical Statistics, Hayward, California.

P. Diaconis and D. Stroock (1991), Geometric bounds for eigenvalues of Markov chains. Ann. Appl. Prob. **1**, 36–61.

A. Frieze, R. Kannan, and N.G. Polson (1994), Sampling from log-concave distributions. Ann. Appl. Prob. **4**, 812–837.

A. Frigessi, C.-R. Hwang, S.J. Sheu, and P. Di Stefano (1993), Convergence rates of the Gibbs sampler, the Metropolis algorithm, and other single-site updating dynamics. J. Roy. Stat. Soc. Ser. B **55**, 205–220.

A.E. Gelfand and A.F.M. Smith (1990), Sampling based approaches to calculating marginal densities. J. Amer. Stat. Assoc. **85**, 398–409.

P. Hanlon (1992), A Markov chain on the symmetric group and Jack symmetric functions. Discr. Math. **99**, 123–140.

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

C.-R. Hwang, S.-Y. Hwang-Ma, and S.-J. Sheu (1993), Accelerating Gaussian diffusions. Ann. Appl. Prob. **3**, 897–913.

S. Ingrassia (1994), On the rate of convergence of the Metropolis algorithm and Gibbs sampler by geometric bounds. Ann. Appl. Prob. **4**, 347–389.

M. Jerrum and A. Sinclair (1988), Approximating the permanent. SIAM J. Comput. **18**, 1149–1178.

G.F. Lawler and A.D. Sokal (1988), Bounds on the L^2 spectrum for Markov chains and Markov processes: A generalization of Cheeger's inequality. Trans. Amer. Math. Soc. **309**, 557–580.

J. Liu (1992), Eigen analysis for a Metropolis sampling scheme with comparisons to rejection sampling and importance resampling. Research Rep. **R–427**, Dept. of Statistics, Harvard University.

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

S.P. Meyn and R.L. Tweedie (1993), Computable bounds for convergence rates of Markov chains. Tech. Rep., Dept. of Statistics, Colorado State University.

G.O. Roberts and J.S. Rosenthal (1994), Shift-coupling and convergence rates of ergodic averages. Preprint.

J.S. Rosenthal (1994), Convergence of Gibbs sampler for a model related to James-

Stein estimators. Stat. and Comput., to appear.

J.S. Rosenthal (1995a), Rates of convergence for Gibbs sampler for variance components models. Ann. Stat. **23**, 740–761.

J.S. Rosenthal (1995b), Minorization conditions and convergence rates for Markov chain Monte Carlo. J. Amer. Stat. Assoc. **90**, 558-566.

A. Sinclair (1992), Improved bounds for mixing rates of Markov chains and multicommodity flow. Combinatorics, Prob., Comput. **1**, 351–370.

A. Sinclair and M. Jerrum (1989), Approximate counting, uniform generation and rapidly mixing Markov chains. Inform. Comput. **82**, 93–133.

A.F.M. Smith and G.O. Roberts (1993), Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods (with discussion). J. Roy. Stat. Soc. Ser. B **55**, 3–24.

M. Spivak (1980), Calculus, 2nd ed. Publish or Perish, Inc. Wilmington, Delaware.