Rates of Convergence of the Diaconis-Holmes-Neal Markov Chain Sampler *

Martin Hildebrand

December 18, 2000

Abstract

The Diaconis-Holmes-Neal Markov chain sampler is a modification of the Metropolis algorithm based on the nearest neighbor random walk. For certain stationary probabilities π and choices of a parameter θ, this paper proves that the Diaconis-Holmes-Neal Markov chain approaches its stationary distribution faster than the Markov chain of the corresponding Metropolis algorithm.

1 Introduction

The Metropolis algorithm has been of interest in Monte Carlo methods using Markov chain sampling. In a number of fields (such as statistical physics), there are probability distributions which are difficult to sample from directly (perhaps because the probabilities are known up to a difficult to compute normalizing constant) but for which a Markov chain which converges to the distribution can be constructed more readily. For more detailed descriptions of the applications, see the references in Diaconis, Holmes, and Neal [2] or in Neal [8]. The Metropolis algorithm introduced by Metropolis *et*

*AMS 2000 Subject Classifications. Primary 60J10; secondary 65C05. Keywords: Metropolis algorithm, sampling, Markov chains.

al. [6] provides examples of such Markov chains. When the Metropolis algorithm has nearest neighbor random walk as its underlying Markov chain, the algorithm often exhibits a diffusive behavior which can be slow. To overcome this diffusive behavior, Diaconis, Holmes, and Neal developed a related non-reversible Markov chain sampler which uses ratios of probabilities similar to the ratios used in the Metropolis algorithm. The Diaconis-Holmes-Neal algorithm involves a duplicate copy of the set of states and a direction on each copy of the set of states; while on a given copy, the algorithm only moves in that direction. In this paper, we shall consider show some examples where the Diaconis-Holmes-Neal Markov chain sampler goes faster than the corresponding Markov chain sampler from the Metropolis algorithm.

For the Diaconis-Holmes-Neal algorithm and the corresponding Metropolis algorithm, we shall consider a one-dimensional set of states $X = \{1, \ldots, n\}$. The corresponding Metropolis algorithm (with nearest neighbor random walk as its underlying Markov chain) is as follows.

Suppose we are at state j. Each step of the Markov chain defined by the Metropolis algorithm consists of the following two substeps.

1. Let $k = j + 1$ with probability $1/2$ (independent of any steps prior to arriving at state j); otherwise let $k = j - 1$.

2. If $k \notin X$, stay at j. Otherwise with probability $\min(1, \pi(k)/\pi(j))$,

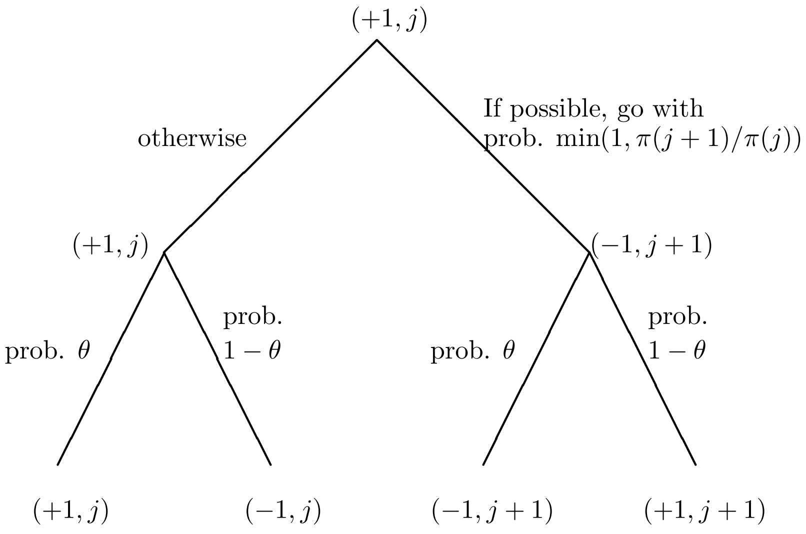

Figure 1: Transitions Starting at $(+1, j)$

go to k. Stay at j otherwise.

The stationary distribution of this Markov chain is π.

For the Diaconis-Holmes-Neal Markov chain sampler, we shall consider two copies of the state space. Define $\tilde{X}$ to be $\{(\epsilon, j) | j \in X \text{ and } \epsilon = \pm 1\}$; $\tilde{X}$ can be viewed as the two copies of X. Suppose π is such that $\pi(j) > 0$ for $j \in X$ and $\sum_{j=1}^{n} \pi(j) = 1$. Let $\tilde{\pi}((\epsilon, j)) = (1/2)\pi(j)$. The algorithm of Diaconis, Holmes, and Neal [2] for a one-dimensional space is as follows. Let $\theta \in (0, 1)$ be given; θ may depend on n. Suppose we start at a point (ϵ, j). Perform one step of a Markov chain by going through the following two substeps:

1. If $j + \epsilon \notin X$, stay at (ϵ, j). Otherwise go to $(-\epsilon, j + \epsilon)$ with probability $\min(1, \pi(j + \epsilon)/\pi(j))$ and stay at (ϵ, j) otherwise.

2. After substep 1, suppose we are at (ϵ', j'). With probability $1 - \theta$

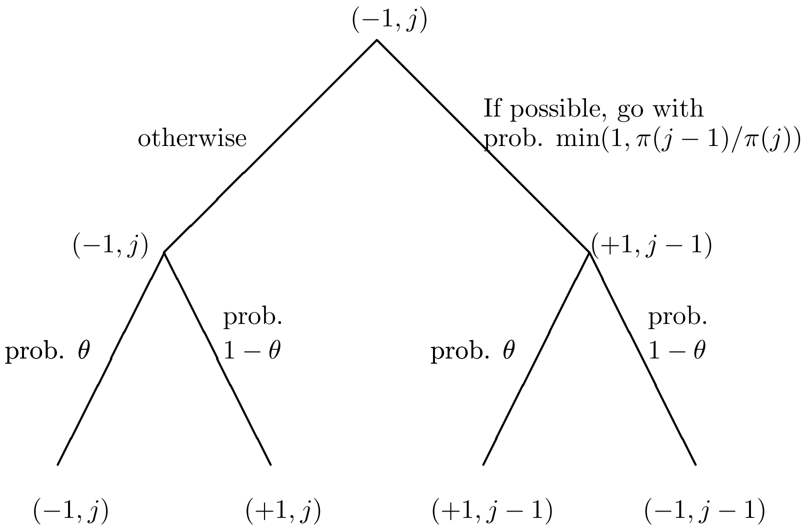

Figure 2: Transitions Starting at $(-1, j)$

go to $(-\epsilon', j')$. Otherwise stay at (ϵ', j').

Figure 1 illustrates these substeps if we start at $(+1, j)$, and Figure 2 illustrates these substeps if we start at $(-1, j)$. Note that one step of the Markov chain is from the top of the figure all the way to the bottom of the figure. Figure 3 illustrates the possible transitions for each step of the Markov chain; the thin lines represent transitions which occur only for certain π. Figure 4 illustrates the possible transitions if $\pi(j) = 1/n$ for all $j \in X$.

The stationary distribution for this Markov chain is $\tilde{\pi}$. Diaconis, Holmes, and Neal [2] note that both substeps in the transition leave the distribution $\tilde{\pi}$ invariant. The second substep leaves $\tilde{\pi}$ invariant since $\tilde{\pi}((+1, j)) = \tilde{\pi}((-1, j))$, and the first substep leaves $\tilde{\pi}$ invariant with reasoning following the reasoning of the stationary distribution

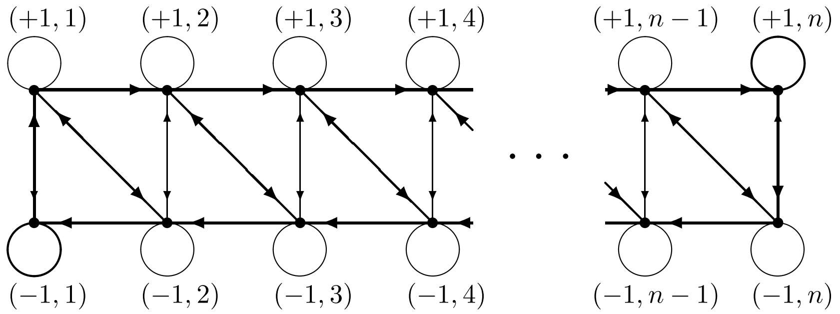

Figure 3: Possible Transitions in the Non-Reversible Markov Chain Sampler

of the corresponding Metropolis algorithm.

Note that if $\theta = 0.5$ and the Diaconis-Holmes-Neal algorithm starts at $(+1, j)$ or $(-1, j)$ with probability $1/2$ each, then the second coordinates of the points arising after each step of the Diaconis-Holmes-Neal Markov chain sampler will have the same probabilities as the points arising after each step of the Metropolis algorithm. On the other hand, if θ is small and we're in a region where π is relatively flat (i.e. $\pi(k)/\pi(j)$ is close to 1 if j and k are neighboring points), then the typical behavior of the Diaconis-Holmes-Neal algorithm will be to proceed in the same direction for a fairly large number of steps. On the positive copy of X within $\tilde{X}$, the algorithm would typically add 1 to the second coordinate; on the negative copy, the algorithm would typically subtract 1 from the second coordinate. This behavior contrasts with the diffusive behavior of the Metropolis algorithm. The Diaconis-Holmes-Neal algorithm on occasion will switch from the positive copy to the negative copy or vice versa.

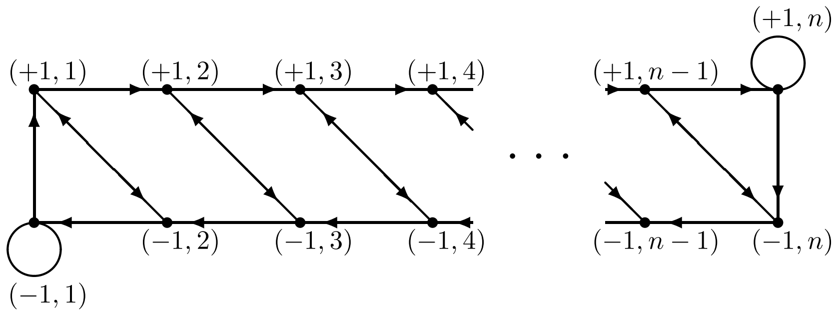

Figure 4: Transitions in the Non-Reversible Markov Chain Sampler if $\pi(j) = 1/n$

This switch may occur at $(+1,n)$, $(-1,1)$, where the second substep which occurs with probability θ occurs, or when $\pi(j+\epsilon)/\pi(j)$ is less than 1 and the first substep is such that the second coordinate does not change. The randomness in the Diaconis-Holmes-Neal algorithm for small values of θ arises in large part from the randomness of these switches as opposed to the diffusion in the Metropolis algorithm.

The Diaconis-Holmes-Neal algorithm for certain probabilities π and appropriate values for θ is considerably faster than the corresponding Metropolis algorithm. If $\pi(j) = 1/n$ for all $j \in X$ and $\theta = 1/n$, then Diaconis, Holmes, and Neal [2] showed that order n steps suffice to get close in variation distance to the uniform distribution for their algorithm. The corresponding Metropolis algorithm will be the nearest neighbor random walk and needs order n^2 steps to get close in variation distance to the uniform distribution. In this paper, we shall consider other expressions for θ while keeping $\pi(j) = 1/n$ for all $j \in X$. We shall find that if $\theta = c(n)$

with $\lim_{n\to\infty} c(n) = \infty$ and $\lim_{n\to\infty} c(n)/n = 0$, then the Diaconis-Holmes-Neal algorithm with this θ will be faster (for sufficiently large n) than the corresponding Metropolis algorithm. If $\theta = 1/n$ and $\pi(j) = z(2|(n/2) - j| + 1)$ where z is a normalizing constant, computer number-crunching suggests that order n^2 steps suffice for the Diaconis-Holmes-Neal algorithm to get close to uniform while techniques from Diaconis and Saloff-Coste [3] can be used to show that order $n^2 \log n$ steps are needed for the Metropolis algorithm to get close to uniform. (Diaconis, Holmes, and Neal [2].) We shall prove that the result suggested by the computer number-crunching is correct not only for this π but for a range of probabilities with one "valley" in the middle.

Recall that the variation distance between two probabilities P and Q on a set S is defined by

$$\|P - Q\| := \max_{A\subseteq S} |P(A) - Q(A)| = \frac{1}{2}\sum_{s\in S} |P(s) - Q(s)|.$$

In the following 5 theorems, X_m is the position of the Diaconis-Holmes-Neal algorithm after m steps, $P_m(j)$ is the probability $X_m = j$, and the value m will depend on n. Unless otherwise specified, X will remain $\{1, \ldots, n\}$.

Theorem 1 *Suppose* $\pi(j) = 1/n$ *for* $j = 1, \ldots, n$. *Suppose* $X_0 = a$ *with probability* 1 *for some* $a \in \tilde{X}$. *Suppose* $\theta = c(n)/n$ *where either* $c(n)$ *is a constant or* $\lim_{n\to\infty} c(n) = \infty$. *Assume there is a value* N_1

such that if $n > N_1$, then $c(n)/n \leq 1/2$. Let $\epsilon > 0$ be given. Then there exist positive values N, c_1, and c_2 (all possibly depending on the value ϵ and the function $c(n)$) such that if $n > N$ and $m > c_1 c(n) n$, then $\|P_m - U\| < \epsilon$ while if $n > N$ and $m < c_2 c(n) n$, then $\|P_m - U\| > 1 - \epsilon$.

Diaconis, Holmes, and Neal [2] proved this theorem in the case $\theta = 1/n$. The proof of the case where $c(n)$ is a constant is a minor modification of their argument while the case where $\lim_{n\to\infty} c(n) = \infty$ is a more substantial modification of their argument. Diaconis, Holmes, and Neal also considered convergence in the L^2 norm for this π. The optimal choice for θ in variation distance differs significantly from the optimal choice for θ the L^2 norm.

In Theorem 2, we shall consider probabilities π satisfying the following conditions:

1. $\pi(j)$ is a positive probability on $\{1, \ldots, n\}$.
2. There exists a value k such that $0.1n \leq k \leq 0.9n$ with $\pi(1) \geq \pi(2) \geq \ldots \geq \pi(k)$ and $\pi(k) \leq \pi(k+1) \leq \ldots \leq \pi(n)$.
3. $\pi(\lfloor 0.1n \rfloor)/\pi(1) \geq 0.5$ and $\pi(\lceil 0.9n \rceil)/\pi(n) \geq 0.5$.
4. $\pi(k)/\pi(1) \geq 0.5/n$ and $\pi(k)/\pi(n) \geq 0.5/n$.

Examples of probabilities satisfying 1 through 4 include $\pi(j) = z(2|((n+1)/2) - j| + 1)$ where z is a normalizing constant. This probability is shaped like a V. (See Figure 5.) The probability $\pi(j) = z(2|(n/2) - j| + 1)$ where z is a normalizing constant also

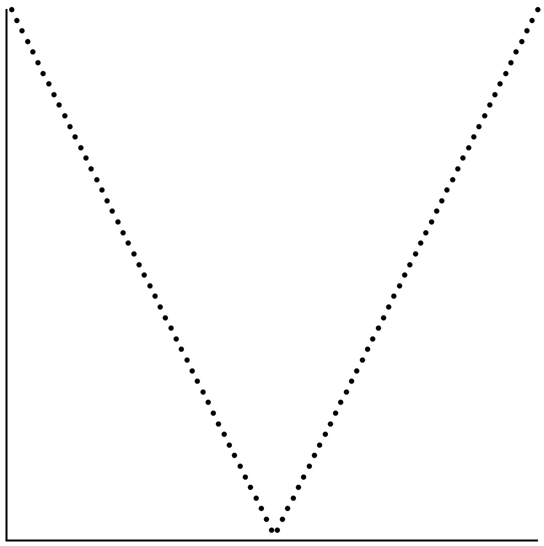

Figure 5: Graph of $\pi = z(2|((n+1)/2) - j| + 1)$

satisfies these conditions; this probability was considered by Diaconis, Holmes, and Neal [2] in their computer number-crunching. This number crunching suggested that order n^2 steps suffice to make the variation distance of their Markov chain sampler from the stationary distribution small. The following theorem verifies and generalizes this suggestion.

Theorem 2 *Let $\theta = 1/n$. Suppose π satisfies properties 1 through 4 above. Suppose $X_0 = (+1, 1)$. Let $\epsilon > 0$ be given. There exist positive values c and N (not depending on the choice of $\pi(n)$ but possibly depending on ϵ) such that if $n > N$ and $m > cn^2$, then $\|P_m - \tilde{\pi}\| < \epsilon$.*

The following theorem gives a lower bound for the Diaconis-

Holmes-Neal algorithm with the V-shaped probability.

Theorem 3 *Suppose $\pi = z(2|((n+1)/2) - j| + 1)$ for $j = 1, \ldots, n$ where z is a normalizing constant. Let $\theta = 1/n$ and $X_0 = (+1, 1)$. Let $\epsilon \in (0, 1/2)$ be given. Then for $n > 0$, there exists a value $C > 0$ (which may depend on ϵ) such that if $m < Cn^2$, then $\|P_m - \tilde{\pi}\| > \epsilon$.*

The lower bound can be generalized to some other probabilities satisfying conditions 1 through 4 in Theorem 2 but not to all such probabilities. We shall explore the generalizations with the proof of the theorem.

The following theorem deals with a W-shaped π and uses the state space $\{1, \ldots, 2n\}$ in place of $\{1, \ldots, n\}$.

Theorem 4 *Suppose $X = \{1, \ldots, 2n\}$. Let $\pi(j) = z(2|((n+1)/2) - j| + 1)$ if $1 \leq j \leq n$ and $\pi(j) = \pi(j - n)$ if $n + 1 \leq j \leq 2n$ where z is a normalizing constant. Let $\theta = 1/n$ and $X_0 = (+1, 1)$. Let $\epsilon > 0$ be given. There are positive values N and C (which may depend on ϵ) such that if $n > N$ and $m > Cn^2$, then $\|P_m - \tilde{\pi}\| < \epsilon$.*

The lower bound in Theorem 3 can be readily extended to the W-shaped π.

Theorem 5 *Suppose X, X_0, π, and θ are as in Theorem 4. Let $\epsilon \in (0, 3/4)$ be given. Then for $n > 0$, there exists a value $C > 0$ (which may depend on ϵ) such that if $m < Cn^2$, then $\|P_m - \tilde{\pi}\| > \epsilon$.*

For the W-shaped π, the algorithm at first concentrates on the leftmost quarter of the points while for the V-shaped π, the algorithm at first concentrates on the leftmost half of the points. This difference accounts for the larger set of possible values for ϵ in Theorem 5 as opposed to Theorem 3.

2 Lemma Used in Several of the Proofs

Suppose p_{ij} is the transition matrix of a connected Markov chain on $X := \{1, \ldots, N\}$. Let $p_{ij}^{(m)}$ be the probability of going from state i to state j in exactly m steps. Let $(p_1, \ldots, p_n)$ be the stationary distribution for this Markov chain. Define probabilities $Q_i^{(m)}$ and Q on X by $Q_i^{(m)}(j) = p_{ij}^{(m)}$ and $Q(j) = p_j$.

Lemma 1 *If there is a set $J \subseteq X$ with $|J| = N_1$ and*

$$\min_{1 \le i \le N,\ j \in J} p_{ij}^{(m)} = \delta > 0,$$

then $\|Q_i^{(\ell)} - Q\| \le (1 - N_1\delta)^{\lfloor \ell/m \rfloor}$.

This lemma follows immediately from Theorem 16.2.4 of Meyn and Tweedie [7] with

$$\nu_m(j) = \begin{cases} \delta & \text{if } j \in J \\ 0 & \text{if } j \notin J. \end{cases}$$

This lemma can also be proved with an adaptation of the argument on pages 173-174 of Doob [4].

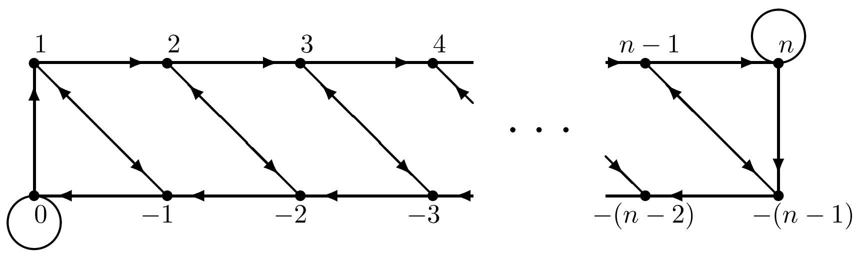

Figure 6: Transitions for a Random Process on $\mathbf{Z}/(2n)\mathbf{Z}$

3 Proof of Theorem 1

3.1 Conventions Used in the Proofs of the Upper Bound and Lower Bound

In the proof of Theorem 1, we shall consider the following Markov chain on $\mathbf{Z}/2n\mathbf{Z}$. With $c(n)$ as in Theorem 1, we shall define the transition probability

$$p_{ij} = \begin{cases} c(n)/n & \text{if } j = -i \\ 1 - c(n)/n & \text{if } j = i+1 \\ 0 & \text{otherwise.} \end{cases}$$

This Markov chain is a relabeling of the states in the Diaconis-Holmes-Neal Markov chain with $\pi(j)$ as in Theorem 1; see Figure 6 and compare it with Figure 4. The convergence properties of these two Markov chains will be identical, and we shall use the Markov chain on $\mathbf{Z}/2n\mathbf{Z}$.

We shall let $T_1, T_2, \ldots$ be the first, second, $\ldots$ time that we choose the option with probability θ in the second substep; each time we make such a choice, we shall say we have a flip. Suppose that $X_0 = a$. Suppose there are f flips in the first m steps. The following lemma relates X_m with $T_1, \ldots, T_f$.

Lemma 2 *If f is even, then*

$$\begin{aligned} X_m &= m + a + 2(T_1 - T_2) + 2(T_3 - T_4) \\ &\quad + \ldots + 2(T_{f-1} - T_f) \pmod{2n} \end{aligned}$$

and if f is odd, then

$$\begin{aligned} X_m &= m + 1 - a - 2T_1 + 2(T_2 - T_3) \\ &\quad +2(T_4 - T_5) + \ldots + 2(T_{f-1} - T_f) \pmod{2n} \end{aligned}$$

Proof: The proof can be done via an elementary induction argument. The result is clear if $m = 0$. The induction from m steps to $m + 1$ steps is clear if the $(m + 1)$-st step is not a flip. If the $(m + 1)$-st step is a flip, then note that $T_{f+1} = m + 1$. If f is even and at least 2 and the $(m + 1)$-st step is a flip, then

$$\begin{aligned} X_{m+1} &= -X_m \\ &= -m - a - 2T_1 + 2(T_2 - T_3) + \ldots + 2(T_{f-2} - T_{f-1}) + 2T_f \\ &= (m + 1) + 1 - a - 2T_1 + 2(T_2 - T_3) + \ldots + 2(T_f - T_{f+1}) \end{aligned}$$

Similar arguments apply if $f = 0$ or f is odd. $\diamondsuit$

3.2 Upper Bound Where $c(n)$ is Constant

The case where $c(n)$ is a constant is a minor modification of the ideas in Diaconis, Holmes, and Neal [2] when $\theta = 1/n$. Let $A_1(m)$ (respectively, $A_2(m)$) be the event that the number of flips in the first m steps is 1 (respectively, 2). Thus $P(A_1(m)) = m(c(n)/n)(1-(c(n)/n))^{m-1}$ and $P(A_2(m)) = \binom{m}{2}(c(n)/n)^2(1-(c(n)/n))^{m-2}$. If $m = 4n$, then both $P(A_1(m))$ and $P(A_2(m))$ are bounded below by positive constants c_1 and c_2 (respectively) for n large enough so that $1-(c(n)/n) > 0.5$.

If there is one flip (in the first m steps with $m = 4n$), then the probability that $T_1 = b \pmod{n}$ is $1/n$ for $b \in \{1, \dots n\}$. Thus if precisely one of $m + a$ and j is odd, then $P(X_m = j) \geq c_1/n$ for n large enough.

If there are two flips, then there are at least $2n$ pairs of T_1 and T_2 with $1 \leq T_1 < T_2 \leq 4n$ and $T_1 - T_2 = b \pmod{n}$ for $b \in \{1, \dots, n\}$. There are less than $8n^2$ pairs of T_1 and T_2 with $1 \leq T_1 < T_2 \leq 4n$, and each such pair is equally likely conditioned on precisely two flips occurring in the first $4n$ steps. Thus if either both $m + a$ and j are even or both $m + a$ and j are odd, then $P(X_m = j) \geq c_2/4n$ for n large enough.

This case of the upper bound follows from Lemma 1 with $J = X$.

3.3 Upper Bound Where $c(n)$ Is Not Constant

In this section, we consider the upper bound where $\lim_{n\to\infty} c(n) = \infty$. We shall use Lemma 2. Observe that T_1, $T_2 - T_1$, $T_3 - T_2$, $\dots$ are i.i.d. geometric random variables with parameter $c(n)/n$. The variance of each of these geometric random variables is $(n/c(n))^2(1-$

$(c(n)/n))$. Note that since $c(n)/n \leq 1/2$ for large enough n, the variance will be at least $(1/2)(n/c(n))^2$ for these n. Adding approximately $c(n)^2$ such variables gives us a random variable with variance approximately a multiple of n^2. Heuristically, this suggests that if there are likely to be at least $c(n)^2$ flips in the first m steps, then perhaps the expressions for X_m in Lemma 2 will be at least a positive constant times $1/(2n)$. If we can make this conclusion, then we will be able to use Lemma 1 to complete the proof of this upper bound.

We shall show the following lemmas. In these lemmas, we assume n is large enough so that $c(n)/n \leq 0.5$.

Lemma 3 *Let ℓ be a positive even integer which is at least $2\lfloor c(n)^2 \rfloor + 2$. Let S_2 be a set of $\lfloor c(n)^2 \rfloor$ positive even integers which are less than ℓ. Let $S = \{1, \ldots, \ell\} - S_2$. For each $i \in S$, define a_i so that the following three conditions all hold.*

A. If $i + 1 \in S$, then $a_i < a_{i+1}$.

B. If i is even and $i < \max S_2$, then either $a_{i+1} - a_{i-1} \notin [\lfloor n/c(n) \rfloor + 1, 2\lfloor n/c(n) \rfloor]$ or $a_i - a_{i-1} > \lfloor n/c(n) \rfloor$.

C. If $i + 1 \in S_2$, then $a_{i+2} - a_i \in [\lfloor n/c(n) \rfloor + 1, 2\lfloor n/c(n) \rfloor]$.

Then there exist values $c_1 > 0$ and $N > 0$ such that if $n > N$, then $P((T_1 - T_2) + (T_3 - T_4) + \ldots + (T_{\ell-1} - T_\ell) = k_0 \pmod{n} | T_i = a_i$ for all $i \in S$ and $T_i - T_{i-1} \leq \lfloor n/c(n) \rfloor$ for all $i \in S_2) \geq c_1/n$ for all $k_0 \in \{0, 1, \ldots, n-1\}$.

The set S_2 in Lemma 3 is meant to be the first $\lfloor c(n)^2 \rfloor$ positive even integers k such that $T_k - T_{k-1} \leq \lfloor n/c(n) \rfloor$ and $\lfloor n/c(n) \rfloor + 1 \leq T_{k+1} - T_{k-1} \leq 2\lfloor n/c(n) \rfloor$. The following lemma considers the probability that these integers will be determined in the first $\lfloor c_2 c(n) \log n \rfloor$ steps for some positive multiple c_2.

Lemma 4 *There exists a constant $c_2 > 0$ such that if $A(n)$ is the event that there are at least $\lfloor c(n)^2 \rfloor$ positive even integers k such that $T_k - T_{k-1} \leq \lfloor n/c(n) \rfloor$, $\lfloor n/c(n) \rfloor + 1 \leq T_{k+1} - T_{k-1} \leq 2\lfloor n/c(n) \rfloor$, and $T_{k+2} \leq c_2 c(n) n$, then $\lim_{n \to \infty} P(A(n)) = 1$.*

Lemma 5 *Let c_2 be the constant defined in Lemma 4. Let $E(n)$ be the event that the largest integer k with $T_k \leq c_2 c(n) n$ is even. Then $\lim_{n \to \infty} P(E(n)) = 1/2$.*

Observe that in Lemmas 4 and 5, the events $A(n)$ and $E(n)$ depend not only on the integer n but also on the function $c(n)$ and the choice of constant c_2. The limits presume that $c(n)$ is a fixed function and c_2 is a fixed constant.

Observe that Lemmas 3, 4, and 5 imply that there are constants $c_2 > 0$ and $c > 0$ such that $p_{ij}^{(m)} \geq c/(2n)$ for all $i, j \in \mathbf{Z}/2n\mathbf{Z}$ where $m := m(n) := \lfloor c_2 c(n) n \rfloor$. This conclusion is what we need to prove the upper bound.

Proof of Lemma 3: Observe that $T_2 - T_1$, $T_3 - T_2$, $T_4 - T_3$, $\ldots$ (with no restrictions on the number of flips) are i.i.d. geometric random variables with mean $n/c(n)$. If $i < k_1 < j$, observe that

$$P(T_2 = k_1 | T_1 = i, T_3 = j) = \frac{1}{j - i + 1}$$

In other words, given values for T_1 and T_3, the value T_2 will be chosen uniformly on the integers between T_1 and T_3. Also observe that if $i < k_1 \leq k_2 < j$, then

$$P(T_2 = k_1 | T_1 = i, T_3 = j, T_2 \leq k_2) = \frac{1}{k_2 - i}$$

Thus to prove this lemma, we simply need to consider what happens to the sum of $\lfloor c(n)^2 \rfloor$ i.i.d. random variables which are uniform on $\{1, 2, \ldots, \lfloor n/c(n) \rfloor\}$. Let μ_n be the mean and σ_n^2 be the variance of one such variable.

For each n, let $X_{n,m}$ be independent random variables which are uniform on

$$\left\{ \frac{1 - \mu_n}{\sigma_n c(n)}, \frac{2 - \mu_n}{\sigma_n c(n)}, \ldots, \frac{\lfloor n/c(n) \rfloor - \mu_n}{\sigma_n c(n)} \right\}$$

for $m \in \{1, \ldots, \lfloor c(n)^2 \rfloor\}$. Note that $\lim_{n\to\infty} \sum_{m=1}^{\lfloor c(n)^2 \rfloor} EX_{n,m}^2 = 1$ and that for each $\epsilon > 0$, we get $\lim_{n\to\infty} \sum_{m=1}^{\lfloor c(n)^2 \rfloor} E(X_{n,m}^2; |X_{n,m}| > \epsilon) = 0$ where $E(Y; A) = E(Y 1_A)$ and 1_A is the indicator function of A. Thus we can use the Lindeberg-Feller Theorem (as in, for instance, Durrett [5], pp. 116-117, or Chung [1], p. 187) to prove that $S_n := \sum_{m=1}^{\lfloor c(n)^2 \rfloor} X_{n,m}$ converges in distribution to the standard normal random variable.

Since $n/c(n) \geq 2$, we may conclude that there exist positive constants c_3 and c_4 such that

$$c_3 \leq \frac{\sigma_n}{n/c(n)} \leq c_4.$$

Thus there exists a constant $c_1 > 0$ and N such that if $n > N$, we get $P(n < \sigma_n c(n) S_n < 2n) > c_1$. Since $\sigma_n c(n) S_n$ is the sum of unimodal symmetric integer-valued random variables with mean 0, it is also a unimodal symmetric integer-valued random variable with mean 0. Thus if $n > N$, then $P(\sigma_n c(n) S_n = n) > c_1/n$ and $P(\sigma_n c(n) S_n = k_0) > c_1/n$ for $k_0 \in \{0, 1, \ldots, n-1\}$.

If U_n is the sum of $\lfloor c(n)^2 \rfloor$ i.i.d. random variables which are uniform on $\{1, 2, \ldots, \lfloor n/c(n) \rfloor\}$, then U_n and $\sigma_n c(n) S_n + n\mu_n$ have the same distribution. Thus $P(U_n = k_0 \pmod{n}) > c_1/n$ for $k_0 \in \{0, 1, \ldots, n-1\}$. Lemma 3 follows. $\diamondsuit$

Proof of Lemma 4: Note that there is a constant $c_5 > 0$ such that $P(T_k - T_{k-1} \leq \lfloor n/c(n) \rfloor$ and $\lfloor n/c(n) \rfloor + 1 \leq T_{k+1} - T_{k-1} \leq 2\lfloor n/c(n) \rfloor) > c_5$ for any positive even integer k and every positive integer n with $c(n)/n \leq 1/2$. Choose a value $c_6 > 1/c_5$. Observe that since $c(n) \to \infty$ as $n \to \infty$, in the first $\lfloor c_6 c(n)^2 \rfloor$ positive even integers, the probability that there are at least $c(n)^2$ positive even integers k with $T_k - T_{k-1} \leq \lfloor n/c(n) \rfloor$ and $\lfloor n/c(n) \rfloor + 1 \leq T_{k+1} - T_{k-1} \leq 2\lfloor n/c(n) \rfloor$ approaches 1 as n approaches infinity. Furthermore there exists a constant $c_2 > 0$ such that $\lim_{n\to\infty} P(T_{2\lfloor c_6 c(n)^2 \rfloor + 2} \leq c_2 c(n) n) = 1$; to see this, note that for each n, the random variable $T_{2\lfloor c_6 c(n)^2 \rfloor + 2}$ is the sum of $2\lfloor c_6 c(n)^2 \rfloor + 2$ independent random variables each with mean $n/c(n)$ and variance between $(1/2)(n/c(n))^2$ and $(n/c(n))^2$ if $c(n)/n \leq 1/2$, that $\lim_{n\to\infty} E(T_{2\lfloor c_6 c(n)^2\rfloor+2})/(c(n)n)$ is finite, and that $\lim_{n\to\infty} \left(Var(T_{2\lfloor c_6 c(n)^2\rfloor+2})\right)^{1/2} / E(T_{2\lfloor c_6 c(n)^2\rfloor+2}) = 0$. Thus the constant c_2 is the desired constant. $\diamondsuit$

Intuitively, one should hope that Lemma 5 is true since the integer k can be viewed as the value of a binomial random variable with parameters $\lfloor c_2 c(n)n\rfloor$ and $c(n)/n$ and since for large n, the probability that that binomial random variable will be a given integer k will differ little from the probability that that random variable will be $k+1$. The proof below expands upon this idea.

Proof of Lemma 5: Let Y_n be a binomial random variable with parameters $\lfloor c_2 c(n)n\rfloor$ and $c(n)/n$. Observe that for each n, $P(Y_n = k)$ is a unimodal function of k and that $\lim_{n\to\infty} \max_k P(Y_n = k) = 0$. Suppose $k_{\max}$ is a value which maximizes $P(Y_n = k)$. Let's consider the case where $k_{\max}$ is even. Then since

$$\begin{aligned}
P(Y_n = k_{\max}) &\geq P(Y_n = k_{\max} - 1) \\
P(Y_n = k_{\max} - 2) &\geq P(Y_n = k_{\max} - 3) \\
&\cdots \\
P(Y_n = k_{\max}) &\geq P(Y_n = k_{\max} + 1) \\
P(Y_n = k_{\max} + 2) &\geq P(Y_n = k_{\max} + 3) \\
&\cdots
\end{aligned}$$

we may conclude that $P(Y_n = k_{\max}) + P(Y_n \text{ is even}) \geq P(Y_n \text{ is odd})$.

Also since

$$\begin{aligned}
P(Y_n = k_{\max} - 1) &\geq P(Y_n = k_{\max} - 2) \\
P(Y_n = k_{\max} - 3) &\geq P(Y_n = k_{\max} - 4) \\
&\ldots \\
P(Y_n = k_{\max} + 1) &\geq P(Y_n = k_{\max} + 2) \\
P(Y_n = k_{\max} + 3) &\geq P(Y_n = k_{\max} + 4) \\
&\ldots
\end{aligned}$$

we may conclude $P(Y_n \text{ is odd}) \geq P(Y_n \text{ is even}) - P(Y_n = k_{\max})$. A similar argument applies if $k_{\max}$ is odd. Since $\lim_{n\to\infty} P(Y_n = k_{\max}) = 0$ and $\lim_{n\to\infty} P(Y_n \text{ is odd}) + P(Y_n \text{ is even}) = 1$, we get $\lim_{n\to\infty} P(Y_n \text{ is even}) = 1/2$. The lemma follows. $\diamondsuit$

3.4 Lower Bound

First let's consider the case where $c(n)$ is a constant. If m is no more than a small enough multiple of n, then the probability of having a flip in the first m steps is no more than $\epsilon/2$ and hence $\|P_m - U\| > (1-\epsilon/2) - 1/(2n) > 1 - \epsilon$ for large enough n. Thus we have the lower bound in this case.

Otherwise $c(n) \to \infty$ as $n \to \infty$. Let f be the number of flips after m steps; note that m will depend on n. Furthermore for a given m and n, the number f is a random variable with mean $mc(n)/n$ and variance $(mc(n)/n)(1 - c(n)/n)$ and that $T_2 - T_1$, $T_3 - T_2$, $\ldots$ are i.i.d. geometric random variables with mean $n/c(n)$ and variance $(1-c(n)/n)(n/c(n))^2$.

Let $\ell = mc(n)/n$. The sum of ℓ i.i.d. geometric random variables with the same mean and variance as $T_2 - T_1$ has variance $(mc(n)/n)(1-c(n)/n)(n/c(n))^2 = m(n/c(n))(1-c(n)/n)$.

Lemma 6 *There exists a constant $c_1 > 0$ such that if $m < c_1 c(n) n$, then $P(\max(|(T_1-T_2)-E(T_1-T_2)|, |(T_1-T_2)+(T_3-T_4)-E(T_1-T_2)-E(T_3-T_4)|, \ldots, |(T_1-T_2)+\ldots+(T_{2\ell-1}-T_{2\ell})-E(T_1-T_2)-\ldots-E(T_{2\ell-1}-T_{2\ell})|) > (\epsilon/32)n) < \epsilon/8$.*

Proof: Use Kolmogorov's maximal inequality (see, for example, pp. 62-63 of Durrett [5]), and choose c_1 small enough so that the variance of the sum of ℓ i.i.d. geometric random variables is a small enough multiple of n. ♢

Proposition 1

$$P\left(\left|f - \frac{mc(n)}{n}\right| > \beta\left(m\frac{c(n)}{n}\left(1-\frac{c(n)}{n}\right)\right)^{1/2}\right) < \frac{1}{\beta^2}$$

Proof: Use Chebyshev's inequality. ♢

Now let's consider the case where f is even. Observe $m + a + 2(T_1-T_2)+\ldots+2(T_{f-1}-T_f) = a+2(T_1-T_2+n/c(n))+\ldots+2(T_{f-1}-T_f+n/c(n))-(f/2)2(n/c(n))+m$. By Proposition 1, the probability that $|-fn/c(n)+m| > (\epsilon/16)n$ can be made to be under $\epsilon/8$ by choosing $c_1 > 0$ to be small enough. Furthermore, by Lemma 6, we get $P(|2(T_1 - T_2 + n/c(n)) + \ldots + 2(T_{f-1} - T_f + n/c(n))| > (\epsilon/16)n) < (\epsilon/8) + P(f > 2\ell)$. Note that $P(f > 2\ell) \to 0$ as $n \to \infty$ since f has mean ℓ and variance $o(\ell^2)$. Thus, if f is even, then $P(|X_m - a| > (\epsilon/8)n) < (\epsilon/8) + (\epsilon/8) + o(1) = (\epsilon/4) + o(1)$.

A similar argument applies for odd f, but we also need to use the fact that $P(|2(T_1 - n/c(n))| > (1/f(n))n) = o(1)$ where $f(n) \to \infty$ as $n \to \infty$ and $\lim_{n\to\infty} f(n)/c(n) = 0$.

Thus

$$\begin{aligned} \|P_m - U\| &> \left(1 - \frac{2(\epsilon/4)n}{2n}\right) - \frac{\epsilon}{2} - o(1) \\ &= 1 - \frac{\epsilon}{4} - \frac{\epsilon}{2} - o(1) \\ &> 1 - \epsilon \end{aligned}$$

for sufficiently large n. This completes the proof of Theorem 1.

4 Proof of Theorem 2

4.1 Overview

In this section, we will give a proof of Theorem 2 and assume that π satisfies the conditions required by this theorem. Some of the ideas in this proof appear in computations for the case where $\pi(j) = z(2|(n+1)/2 - j| + 1)$ where z is a normalizing constant and $\theta = 1/n$. Figures 7 and 8 give computations of the probability of being at various points for this θ and π after m steps for various m and $n = 100$. The computations assume the starting point is $(+1, 1)$. We see that after one step, we are most likely to be at $(+1, 2)$, and after two steps, we are mostly likely to be at $(+1, 3)$. The probability of being at $(+1, j+1)$ after j steps will get considerably smaller for j around $n/2$, but this probability will be much larger than the probability of being at $(+1, j)$, $(+1, j-1)$, or other nearby points. Even after $(5/4)n$ steps, there still is a sharp local maximum of probabilities around $(-1, (3/4)n)$. Direct computation will show that this maximum value is at least some positive constant times $1/n$. In the more general case for π considered in Theorem 2, analogous results appear for the first $\lfloor 1.1n \rfloor$ steps and are used in Lemma 9 below. Furthermore, the probability of returning to $(+1, 1)$ on step j appears to be at least a positive constant times $1/n$ if $j \geq 2$ and j is less than a multiple of n. We will use the probability of returning to $(+1, 1)$ after j steps in our argument.

Throughout this section, we shall discuss the probability that the Markov chain follows paths such as $(+1, 1) \to (+1, 2) \to \ldots \to (+1, j)$. The probabilities will be conditioned on starting at the first point, and the Markov chain is the one in Diaconis-Holmes-Neal sampler. Proofs of lemmas to prove this theorem will involve probabilities of such paths and of concatenations of such paths.

4.2 Lemmas Used to Prove Theorem 2

To prove Theorem 2, we shall show the following lemma.

```
m=  1 (+1,1) 0.00020  (+1,2) 0.97020  (+1,3) 0.00000  (+1,4) 0.00000
 (+1,24) 0.00000 (+1,25) 0.00000 (+1,26) 0.00000

m=  2 (+1,1) 0.01970  (+1,2) 0.00039  (+1,3) 0.94090  (+1,4) 0.00000
 (+1,24) 0.00000 (+1,25) 0.00000 (+1,26) 0.00000

m=  3 (+1,1) 0.01000  (+1,2) 0.01921  (+1,3) 0.00058  (+1,4) 0.91208
 (+1,24) 0.00000 (+1,25) 0.00000 (+1,26) 0.00000

m= 23 (+1,1) 0.01034  (+1,2) 0.02023  (+1,3) 0.01008  (+1,4) 0.01910
 (+1,24) 0.42855 (+1,25) 0.00000 (+1,26) 0.00000

m= 24 (+1,1) 0.02086  (+1,2) 0.01025  (+1,3) 0.01971  (+1,4) 0.00998
 (+1,24) 0.00276 (+1,25) 0.40855 (+1,26) 0.00000

m= 25 (+1,1) 0.01041  (+1,2) 0.02034  (+1,3) 0.01014  (+1,4) 0.01920
 (+1,24) 0.00941 (+1,25) 0.00279 (+1,26) 0.38891
```

Figure 7: Computations for $\pi(j) = z(2|(n+1)/2-j|+1)$, $n = 100$, and $\theta = 1/n$

Lemma 7 *There exist constants N, $c_1 > 0$, and $c_2 > 0$ such that if $n > N$, then $p_{ij}^{(c_1 n)} > c_2/n^2$ for all $i \in \tilde{X}$ and all $j \in J := \{(+1,\ell)|\ell \le 0.1n\} \cup \{(-1,\ell)|\ell \ge 0.9n\}$.*

Lemmas 7 and 1 imply Theorem 2.

To prove Lemma 7, we shall use two base points ($(+1,1)$ and $(-1,n)$). The following lemmas will imply Lemma 7. In these lemmas, we assume that n is larger than some integer N; the constants will be valid for all such n.

Lemma 8 *There is a constant $c_3 > 0$ such that for any $i \in \tilde{X}$, the Diaconis-Holmes-Neal sampler starting at i either will reach the base point $(+1,1)$ in the first $2n$ steps with probability at least c_3 or will reach the base point $(-1,n)$ in the first $2n$ steps with probability*

```
m= 49 (+1,50) 0.01222  (+1,51) 0.00000  (+1,52) 0.00000
(+1,98) 0.00000  (+1,99) 0.00000  (+1,100) 0.00000
(-1,76) 0.00000  (-1,75) 0.00000  (-1,74) 0.00000

m= 50 (+1,50) 0.00043  (+1,51) 0.01210  (+1,52) 0.00000
(+1,98) 0.00000  (+1,99) 0.00000  (+1,100) 0.00000
(-1,76) 0.00000  (-1,75) 0.00000  (-1,74) 0.00000

m= 51 (+1,50) 0.00030  (+1,51) 0.00042  (+1,52) 0.01198
(+1,98) 0.00000  (+1,99) 0.00000  (+1,100) 0.00000
(-1,76) 0.00000  (-1,75) 0.00000  (-1,74) 0.00000

m= 97 (+1,50) 0.00038  (+1,51) 0.00030  (+1,52) 0.00038
(+1,98) 0.00754  (+1,99) 0.00000  (+1,100) 0.00000
(-1,76) 0.00008  (-1,75) 0.00004  (-1,74) 0.00008

m= 98 (+1,50) 0.00031  (+1,51) 0.00038  (+1,52) 0.00031
(+1,98) 0.00052  (+1,99) 0.00747  (+1,100) 0.00000
(-1,76) 0.00004  (-1,75) 0.00008  (-1,74) 0.00005

m= 99 (+1,50) 0.00038  (+1,51) 0.00030  (+1,52) 0.00038
(+1,98) 0.00024  (+1,99) 0.00052  (+1,100) 0.00739
(-1,76) 0.00008  (-1,75) 0.00005  (-1,74) 0.00008

m=124 (+1,50) 0.00032  (+1,51) 0.00041  (+1,52) 0.00033
(+1,98) 0.00054  (+1,99) 0.00058  (+1,100) 0.00056
(-1,76) 0.00306  (-1,75) 0.00009  (-1,74) 0.00006

m=125 (+1,50) 0.00042  (+1,51) 0.00032  (+1,52) 0.00042
(+1,98) 0.00058  (+1,99) 0.00056  (+1,100) 0.00059
(-1,76) 0.00032  (-1,75) 0.00291  (-1,74) 0.00009

m=126 (+1,50) 0.00033  (+1,51) 0.00041  (+1,52) 0.00033
(+1,98) 0.00055  (+1,99) 0.00060  (+1,100) 0.00057
(-1,76) 0.00023  (-1,75) 0.00031  (-1,74) 0.00277
```

Figure 8: Computations for $\pi(j) = z(2|(n+1)/2 - j| + 1)$, $n = 100$, and $\theta = 1/n$

at least c_3.

Lemma 9 *There exists a constant* $c_4 > 0$ *such that for each* $j \in J$ *and each base point, the Diaconis-Holmes-Neal sampler started at the base point will reach* j *in the first* $2n$ *steps with probability at least* c_4/n.

Lemma 10 *There exists a constant* $c_5 > 0$ *such that for any* $m \in [n, 5n]$ *and each base point, the probability that the Diaconis-Holmes-Neal sampler started at the base point will again be at that base point after precisely* m *steps is at least* c_5/n.

To prove that Lemmas 8, 9, and 10 prove Lemma 7, observe that for each $i \in \tilde{X}$ and each $j \in J$, there exists a base point $B.P. \in \{(+1,1),(-1,n)\}$ such that

$$\begin{aligned}
& \sum_{i_1=0}^{2n}\sum_{i_2=0}^{2n} P(X_{i_1} = B.P. \text{ and } X_k \neq B.P. \text{ for } k < i_1 | X_0 = i) \\
& \times P(X_{i_2} = j \text{ and } X_k \neq B.P. \text{ if } 1 \leq k \leq i_2 | X_0 = B.P.) \\
& \times P(X_{5n-i_1-i_2} = B.P. | X_0 = B.P.) \\
\geq \quad & c_3 \frac{c_4}{n} \frac{c_5}{n}
\end{aligned}$$

Note that the double sum is a lower bound for $p_{ij}^{(5n)}$ since for each i_1 and i_2, the product gives the probability that the Markov chain which starts at i will first reach $B.P.$ on step i_1, will be at $B.P.$ on step $5n - i_2$ for the last time in the first $5n$ steps, and will reach j on step $5n$. This shows that Lemma 7 follows from Lemmas 8, 9, and 10.

4.3 Proof of Lemma 8

Suppose k is as in condition 2 of the conditions for Theorem 2. Let $k_l = \min\{j : j \leq k \text{ and } \pi(j) < 2\pi(k)\}$ and $k_r = \max\{j : j \geq k$ and $\pi(j) < 2\pi(k)\}$. Note that k_l and k_r mark the points furthest to the left of k and furthest to the right of k, respectively, where the stationary probability is under twice its minimum value.

Case 1: Here $i = (+1, j)$ for some $j < k_l$.

If the Diaconis-Holmes-Neal sampler starts at i and always has the "usual" second substeps (i.e. the ones with probability $1-\theta$), it will likely go to the negative copy before reaching $(+1, k)$. Once it reaches the negative copy, the Markov chain will eventually go to $(-1, 1)$ and $(+1, 1)$. Thus in this case, the base point $(+1, 1)$ will be examined.

Given j' with $j \leq j' < k$, the probability that the Markov chain goes along the path

$$(+1, j) \to (+1, j{+}1) \to \ldots \to (+1, j') \to (-1, j') \to (-1, 1) \to (+1, 1)$$

is

$$\begin{aligned}
& \frac{\pi(j+1)}{\pi(j)}(1-\theta)\ldots\frac{\pi(j')}{\pi(j'-1)}(1-\theta) \\
& \times \left(1 - \frac{\pi(j'+1)}{\pi(j')}\right)(1-\theta)(1-\theta)^{j'} \\
= \; & \frac{\pi(j')}{\pi(j)}\left(1 - \frac{\pi(j'+1)}{\pi(j')}\right)(1-\theta)^{j'+j'-j+1} \\
\geq \; & \frac{\pi(j')}{\pi(j)}\left(1 - \frac{\pi(j'+1)}{\pi(j')}\right)(1-\theta)^{2n}
\end{aligned}$$

Observe that

$$\begin{aligned} & \sum_{j'=j}^{k-1} \frac{\pi(j')}{\pi(j)} \left(1 - \frac{\pi(j'+1)}{\pi(j')}\right) (1-\theta)^{2n} \\ = \quad & (1-\theta)^{2n} \sum_{j'=j}^{k-1} \left(\frac{\pi(j')}{\pi(j)} - \frac{\pi(j'+1)}{\pi(j)}\right) \\ = \quad & (1-\theta)^{2n} \left(1 - \frac{\pi(k)}{\pi(j)}\right) \end{aligned}$$

Observe that $1 - (\pi(k)/\pi(j)) \geq 1/2$ since $\pi(j) \geq 2\pi(k)$. Since $\lim_{n\to\infty}(1-1/n)^{2n} = e^{-2} > 0$, this case is complete.

Case 2: Here $i = (+1, j)$ where $j \geq k_l$.

In this case, if the Diaconis-Holmes-Neal sampler starts at i and the "usual" second substeps occur on the first $2n$ steps, then either j is at least k and the Markov chain will go directly to $(+1, n)$ and then $(-1, n)$ or the Markov chain likely will go to $(+1, k)$ (where $\tilde{\pi}$ has its minimum) and then go directly to $(+1, n)$ and $(-1, n)$. So in this case, the base point will be $(-1, n)$. In fact, the probability that the Markov chain follows the path

$$(+1, j) \to (+1, j+1) \to \ldots \to (+1, n-1) \to (+1, n) \to (-1, n)$$

is $(1-\theta)^{n-j+1}$ if $j \geq k$ and is

$$\frac{\pi(j+1)}{\pi(j)} \cdots \frac{\pi(k)}{\pi(k-1)}(1-\theta)^{n-j+1} = \frac{\pi(k)}{\pi(j)}(1-\theta)^{n-j+1}$$

if $j < k$. In both possibilities, this probability is at least $0.5(1-1/n)^{2n}$, and this case is completed by the fact that $\lim_{n\to\infty}(1-1/n)^{2n} = e^{-2} > 0$.

Case 3: Here $i = (-1, j)$ for some $j > k_r$. In this case, the base point will be $(-1, n)$, and the argument is symmetric to that of case 1.

Case 4: Here $i = (-1, j)$ for some $j \leq k_r$. In this case, the base point will be $(+1, 1)$, and the argument is symmetric to that of case 2.

These 4 cases cover all possibilities for i, and the proof of Lemma 8 is finished.

4.4 Proof of Lemma 9

We shall consider the base point $(+1, 1)$; the argument for the base point $(-1, n)$ is similar.

If $j = (+1, \ell)$ with $\ell \leq 0.1n$, then the probability that the Markov chain follows the path

$$(+1, 1) \to (+1, 2) \to \ldots \to (+1, \ell)$$

is

$$\frac{\pi(2)}{\pi(1)} \frac{\pi(3)}{\pi(2)} \cdots \frac{\pi(\ell)}{\pi(\ell-1)} (1-\theta)^{\ell-1} \geq \frac{\pi(\ell)}{\pi(1)} (1-\theta)^{2n} \geq (1-\theta)^{2n} 0.5.$$

If $j = (-1, \ell)$ with $\ell \geq 0.9n$, then the probability that the Markov chain follows the path

$$(+1, 1) \to (+1, 2) \to \ldots \to (+1, n) \to (-1, n) \to \ldots \to (-1, \ell)$$

is

$$(1-\theta)^{2n-\ell} \frac{\pi(2)}{\pi(1)} \frac{\pi(3)}{\pi(2)} \cdots \frac{\pi(k)}{\pi(k-1)} \cdot \frac{\pi(n-1)}{\pi(n)} \cdots \frac{\pi(\ell)}{\pi(\ell+1)}$$

$$\begin{aligned} &= (1-\theta)^{2n-\ell}\frac{\pi(k)}{\pi(1)}\frac{\pi(\ell)}{\pi(n)} \\ &\geq (1-\theta)^{2n}\frac{0.5}{n}0.5 \end{aligned}$$

For either kind of j, this probability is at least $(1-1/n)^{2n}0.25/n$ and the lemma follows since $\lim_{n\to\infty}(1-1/n)^{2n} = e^{-2} > 0$.

4.5 Proof of Lemma 10

In this lemma, we shall consider the probability of returning to the base point after a certain number of steps. We shall consider the base point $(+1, 1)$. The proof for the base point $(-1, n)$ is similar.

We shall define $FR(m)$ to be the event that $X_m = (+1, 1)$ and $X_\ell \neq (+1, 1)$ if $0 < \ell < m$. We may view $FR(m)$ is the event the Markov chain first returns to $(+1, 1)$ on step m; this event will be used in the lemmas below.

First we shall consider a lemma for an odd number of steps.

Lemma 11 *Suppose $m \geq 3$, $m \leq 0.2n - 1$, and m is odd. Then there is a constant $c_1 > 0$ such that $P(FR(m)|X_0 = (+1, 1)) \geq c_1/n$*

Proof: Suppose $m = 2j - 1$ and $2 \leq j \leq 0.1n$. Then the probability that the Markov chain follows the path

$$(+1, 1) \to (+1, 2) \to \ldots \to (+1, j-1) \to$$
$$(-1, j) \to (-1, j-1) \to \ldots \to (-1, 1) \to (+1, 1)$$

is

$$\begin{aligned}
& \frac{\pi(2)}{\pi(1)}(1-\theta)\ldots\frac{\pi(j-1)}{\pi(j-2)}(1-\theta)\cdot\frac{\pi(j)}{\pi(j-1)}\theta\cdot(1-\theta)^j \\
= \; & \frac{\pi(j)}{\pi(1)}\theta(1-\theta)^{2j-2} \\
\geq \; & \frac{0.5}{n}(1-1/n)^{0.2n}
\end{aligned}$$

The lemma follows since $\lim_{n\to\infty}(1-1/n)^{0.2n} = e^{-0.2} > 0$. $\diamondsuit$

The next lemma will consider returning to $(+1,1)$ after an even number of steps.

Lemma 12 *Suppose* $n \geq 100$. *Suppose* m *is even and* $0.08n \leq m \leq 0.2n$. *Then there is a constant* $c_2 > 0$ *such that* $P(X_m = (+1,1)|X_0 = (+1,1)) \geq c_2/n$.

Proof: Observe that the number of odd integers in $[3, m/2]$ is at least $\lceil((m/2)-3)/2\rceil \geq m/4 - 1.5 \geq 0.02n - 0.015n = 0.005n$. Thus

$$\begin{aligned}
& P(X_m = (+1,1)|X_0 = (+1,1)) \\
\geq \; & \sum_{j:3\leq j\leq m/2 \text{ and } j \text{ is odd}} P(FR(j)|X_0 = (+1,1)) \\
& \times P(X_{m-j} = (+1,1)|X_0 = (+1,1)) \\
\geq \; & 0.005n\frac{c_1}{n}\frac{c_1}{n}
\end{aligned}$$

where c_1 is as in the proof of the previous lemma. Let $c_2 = 0.005c_1^2$. $\diamondsuit$

The next lemma will be proved by an induction argument. Note that the expression $c_2(j)$ in this lemma may approach 0 as j approaches ∞.

Lemma 13 *Suppose $j \geq 2$ and $n \geq 100$. There is a constant $c_2(j) > 0$ such that $P(X_m = (+1,1)|X_0 = (+1,1)) \geq c_2(j)/n$ whenever $m \in [0.04jn, 0.04(j+1)n)$.*

Proof: By using the previous two lemmas, we can find a value for $c_2(2)$ and $c_2(3)$; in particular, we can let $c_2(2) = c_2(3) = \min(c_1, c_2)$.

Suppose the lemma holds for values j and $j+1$. We shall show that it also holds for the value $j+2$. Suppose $m \in [0.04(j+2)n, 0.04(j+3)n)$. Then

$$
\begin{aligned}
& P(X_m = (+1,1)|X_0 = (+1,1)) && (1)\\
= & \sum_{i \text{ odd and } 0.04n \leq i < 0.08n} P(FR(i)|X_0 = (+1,1)) && (2)\\
& \times P(X_{m-i} = (+1,1)|X_0 = (+1,1))
\end{aligned}
$$

Observe that $m - i \in [0.04jn, 0.04(j+2)n)$ since $i \in [0.04n, 0.08n)$ and $m \in [0.04(j+2)n, 0.04(j+3)n)$. Also observe that the number of odd integers in $[0.04n, 0.08n)$ is at least $0.04n/2 - 1 \geq 0.02n - 0.01n = 0.01n$. Thus the expression in (1) is greater than or equal to $0.01n(c_1/n)\min(c_2(j)/n, c_2(j+1)/n)$, and we can let $c_2(j+2) = 0.01c_1 \min(c_2(j), c_2(j+1))$. So the lemma is complete.

$\diamondsuit$

Let $c_5 = \min(c_2(25), c_2(26), \ldots, c_2(125))$; Lemma 10 follows from Lemma 13.

4.6 Changes to Conditions

The techniques in the proof of Theorem 2 can be readily modified to handle certain changes in the conditions. For instance, if the lower bounds $0.5/n$ in condition 4 are replaced with $0.5/\sqrt{n}$, then the expression c_4/n in Lemma 9 can be replaced with $c_4/\sqrt{n}$, Lemmas 8 and 10 are unchanged, and the n^2 term in Theorem 2 can be improved to $n^{3/2}$.

5 Proof of Theorems 3 and 5

5.1 Overview

If $X_0 = (+1, 1)$, the behavior of the Diaconis-Holmes-Neal Markov chain will typically progress with the first coordinate $+1$ while increasing the second coordinate by one at each step. The low probabilities at the bottom of the V (or W) make it unlikely that the second coordinate exceeds $n/2$ before the first coordinate becomes -1. When the first coordinate is -1 with the second coordinate less than $n/2$, the Markov chain usually will subtract 1 from the second coordinate for awhile, but the Markov chain almost surely will return to a point with the first coordinate $+1$. Usually the second coordinate will be small enough when the first coordinate returns to $+1$ so that the Markov chain will again be unlikely to progress directly to a point with second coordinate at least $n/2$; however, the rare case where the second coordinate is slightly under $n/2$ when the first coordinate returns to $+1$ will need to be carefully considered.

5.2 Trials

To prove Theorem 3, we shall perform "trials" (which are not independent since they depend on the starting point) as follows. We start at the point $(+1, j)$ where $1 \leq j \leq 0.4n$ and perform steps of the Diaconis-Holmes-Neal Markov chain sampler (with π and θ as in Theorem 3) until the first of the following two conditions hold:
1. The Markov chain reaches a point $(+1, k)$ or $(-1, k)$ with $k > (n+1)/2$. If this happens, call the trial a "success".
2. The Markov chain reaches a point $(+1, k)$ or $(-1, k)$ with $k \leq 0.4n$ after at least $0.05n$ steps. If this happens, call the trial a "failure".

Note that almost surely, the trial will end in a "success" or a "failure". Furthermore, after a "failure", the Markov chain almost surely will either be at a starting point for another trial or will go to a starting point for another trial without the second coordinate exceeding $(n+1)/2$ in the period between trials.

By Lemma 14 below, the probability of obtaining a success in no more than $0.5(0.5-\epsilon)n/c$ trials (where $c > 0$ is the constant in that lemma) is no more than $0.5(0.5-\epsilon)$. Note that all trials last at least $0.05n$ steps. These two facts give us the conclusion that $\|P_m - \tilde{\pi}\| \geq ((n-1)/(2n)) - 0.5(0.5-\epsilon)$ if $m < (0.5(0.5-\epsilon)n/c)(0.05n)$. Observe that if $\epsilon \in (0, 0.5)$, then $((n-1)/(2n)) - 0.5(0.5-\epsilon) > \epsilon$ for n larger than some value N_ϵ. Thus Theorem 3 holds for $n > N_\epsilon$. If $n > 0$, then $\|P_0 - \tilde{\pi}\| \geq 0.5 > \epsilon$. So we can make $C = C_\epsilon > 0$ such that $C_\epsilon N_\epsilon^2 < 1$, and the theorem holds for the other values of $n > 0$.

Lemma 14 *For each trial, regardless of the starting point, the probability of success will be no more than c/n for some constant $c > 0$ and sufficiently large n.*

Proof: First note that the probability of the trial going $(+1, j) \to (+1, j+1) \to \ldots \to (+1, \lceil n/2 \rceil) \to (\pm 1, \lceil n/2 \rceil + 1)$ is under c_1/n for some constant $c_1 > 0$ and sufficiently large n. This will be a success unless $j + \lceil 0.05n \rceil \leq 0.4n$ (in which case the trial is doomed to fail).

There are three kinds of interruptions from the pattern of adding 1 to the second coordinate and leaving the first coordinate as $+1$. The first kind is to go $(+1, k) \to (-1, k)$ for some $k \leq \lceil n/2 \rceil$. Then the Markov chain typically will go $(-1, k) \to (-1, k-1) \to \ldots$, but the Markov chain may return the first coordinate to $+1$ before the trial is a failure. The second kind is to go $(+1, k) \to (-1, k+1)$ for some $k < \lceil n/2 \rceil$. (This move with $k = \lceil n/2 \rceil$ gives a success.) The Markov chain will then typically go $(-1, k+1) \to (-1, k) \to (-1, k-1) \to \ldots$, but again the Markov chain may return the first coordinate to $+1$ before the trial is a failure. The third kind is to go $(+1, k) \to (+1, k)$ for some $k \leq \lceil n/2 \rceil$.

In the following sections, we shall show the probability that the trial will have a success with exactly r interruptions will be bounded above by the rth term of geometric series which for large enough n has multiple under some constant less than 1. This will prove the lemma.

5.3 Exactly One Interruption of the First Kind

Let us consider the case where there is exactly one interruption and the interruption is of the first kind. If there is a success here, then the Markov chain will go $(+1, j) \to \ldots \to (+1, k_1) \to (-1, k_1) \to \ldots \to (-1, k_2) \to (+1, k_2 - 1) \to \ldots \to (+1, \lceil n/2 \rceil) \to (\pm 1, \lceil n/2 \rceil + 1)$ with $0.35n \le k_2 \le k_1 \le (n+1)/2$. (If $k_2 < 0.35n$, then the trial would automatically fail.) The probability of obtaining success with this kind of interruption is under

$$\sum_{0.35n \le k_2 \le k_1 \le \lceil n/2 \rceil} \frac{\pi(k_1)}{\pi(j)} \left(1 - \frac{\pi(k_1+1)}{\pi(k_1)}\right) \frac{1}{n} \frac{\pi(\lceil n/2 \rceil)}{\pi(k_2 - 1)} \qquad (3)$$

Note that $\pi(k_1)/\pi(j)$ deals with the steps from $(+1, j)$ to $(+1, k_1)$, $1 - \pi(k_1+1)/\pi(k_1)$ deals with the step from $(+1, k_1)$ to $(-1, k_1)$, $1/n$ deals with the step from $(-1, k_2)$ to $(+1, k_2 - 1)$, and $\pi(\lceil n/2 \rceil)/\pi(k_2 - 1)$ deals with the steps from $(+1, k_2 - 1)$ to $(+1, \lceil n/2 \rceil)$. (By a term dealing with steps, we mean that the term gives an upper bound on the probability of taking those particular steps.)

Note that

$$(3) \quad = \quad \frac{\pi(\lceil n/2 \rceil)}{\pi(j)} \sum_{0.35n \le k_2 \le k_1 \le \lceil n/2 \rceil} \frac{1}{n} \frac{\pi(k_1)}{\pi(k_2 - 1)} \left(1 - \frac{\pi(k_1+1)}{\pi(k_1)}\right)$$

$$\begin{aligned}
&\leq \frac{c_1}{n} \sum_{0.35n \leq k_2 \leq k_1 \leq \lceil n/2 \rceil} \frac{1}{n} \frac{2z}{\pi(k_2 - 1)} \\
&\leq \frac{c_1}{n} \sum_{0.35n \leq k_2 \leq (n+1)/2} \frac{1}{n} \left(\frac{n+1}{2} - k_2 + 1 \right) \frac{2}{2\left(\frac{n+1}{2} - k_2\right) + 3} \\
&\leq \frac{c_1}{n} \left(0.15 + \frac{2}{n} \right)
\end{aligned}$$

5.4 Exactly One Interruption of the Second Kind

If a successful trial has exactly one interruption and this interruption is of the second kind, then the path will be $(+1, j) \to \ldots \to (+1, k_1) \to (-1, k_1 + 1) \to \ldots \to (-1, k_2 + 1) \to (+1, k_2) \to \ldots \to (+1, \lceil n/2 \rceil) \to (\pm 1, \lceil n/2 \rceil + 1)$ with $0.35n \leq k_2 \leq k_1 < \lceil n/2 \rceil$. Thus the probability of obtaining success in this case is under

$$\begin{aligned}
\sum_{0.35n \leq k_2 \leq k_1 \leq \lceil n/2 \rceil} \frac{c_1}{n} \frac{1}{n} \frac{1}{n} \frac{2\left(\frac{n+1}{2} - k_1\right) + 1}{2\left(\frac{n+1}{2} - k_2\right) + 1}
&\leq \frac{c_1}{n} (0.15n + 1)^2 \frac{1}{n^2} \\
&= \frac{c_1}{n} \left(0.15 + \frac{1}{n} \right)^2
\end{aligned}$$

5.5 Exactly One Interruption of the Third Kind

If a successful trial has exactly one interruption and this interruption is of the third kind, then the path will be $(+1, j) \to \ldots \to (+1, k) \to (+1, k+1) \to \ldots \to (+1, \lceil n/2 \rceil) \to (\pm 1, \lceil n/2 \rceil + 1)$ with $k \leq \lceil n/2 \rceil$. The probability of success here is under

$$\sum_{k=1}^{\lceil n/2 \rceil} \frac{c_1}{n} \frac{1}{n} \frac{2}{2\left(\frac{n+1}{2} - k\right) + 1}$$

$$\leq \frac{c_1}{n} \sum_{k=1}^{\lceil n/2 \rceil} \frac{1}{n} \left(\frac{2}{2(k-1)+1} \right)$$

Let

$$h(n) = \sum_{k=1}^{\lceil n/2 \rceil} \frac{2}{2(k-1)+1}$$

Observe that $h(n)/n \to 0$ as $n \to \infty$.

5.6 More Than One Interruption

Now let us consider the case where there are exactly two interruptions in a successful trial. First let us consider the probability of having both interruptions being of the first kind. This probability is under

$$\sum \frac{\pi(k_1)}{\pi(j)} \left(1 - \frac{\pi(k_1+1)}{\pi(k_1)}\right) \frac{1}{n} \frac{\pi(k_3)}{\pi(k_2-1)} \left(1 - \frac{\pi(k_3+1)}{\pi(k_3)}\right) \frac{1}{n} \frac{\pi(\lceil n/2 \rceil)}{\pi(k_4-1)} \tag{4}$$

where the sum is over k_1, k_2, k_3, k_4 such that $0.35n \leq k_2 \leq k_1 \leq \lceil n/2 \rceil$ and $0.35n \leq k_4 \leq k_3 \leq \lceil n/2 \rceil$ with $k_3 \geq k_2$. For each k_1, k_2, k_3, k_4, the corresponding path is $(+1, j) \to \ldots \to (+1, k_1) \to (-1, k_1) \to \ldots \to (-1, k_2) \to (+1, k_2-1) \to \ldots (+1, k_3) \to (-1, k_3) \to \ldots \to (-1, k_4) \to (+1, k_4 - 1) \to \ldots (+1, \lceil n/2 \rceil) \to (\pm 1, \lceil n/2 \rceil + 1)$. Observe that (4) is under

$$\begin{aligned} & \frac{\pi(\lceil n/2 \rceil)}{\pi(j)} \sum_{k_1,k_2:0.35n \leq k_2 \leq k_1 \leq \lceil n/2 \rceil} \frac{1}{n} \frac{\pi(k_1)}{\pi(k_2-1)} \left(1 - \frac{\pi(k_1+1)}{\pi(k_1)}\right) \\ & \times \sum_{k_3,k_4:0.35n \leq k_4 \leq k_3 \leq \lceil n/2 \rceil} \frac{1}{n} \frac{\pi(k_3)}{\pi(k_4-1)} \left(1 - \frac{\pi(k_3+1)}{\pi(k_3)}\right) \\ \leq \ & \frac{c_1}{n} \left(0.15 + \frac{2}{n}\right) \left(0.15 + \frac{2}{n}\right) \end{aligned}$$

Similar arguments apply for other kinds of interruptions, and we may conclude that the probability of obtaining a successful trial with exactly two interruptions is under

$$\frac{c_1}{n}\left(\left(0.15+\frac{2}{n}\right)+\left(0.15+\frac{1}{n}\right)^2+\frac{h(n)}{n}\right)^2$$

Similar arguments show that the probability of obtaining success with exactly r interruptions is under

$$\frac{c_1}{n}\left(\left(0.15+\frac{2}{n}\right)+\left(0.15+\frac{1}{n}\right)^2+\frac{h(n)}{n}\right)^r$$

Thus the probability of obtaining success is thus under

$$\sum_{r=0}^{\infty}\frac{c_1}{n}\left(\left(0.15+\frac{2}{n}\right)+\left(0.15+\frac{1}{n}\right)^2+\frac{h(n)}{n}\right)^r$$

Observe that

$$\left(\left(0.15+\frac{2}{n}\right)+\left(0.15+\frac{1}{n}\right)^2+\frac{h(n)}{n}\right)<0.5$$

for large enough n, and thus Lemma 14 and Theorem 3 are proved. $\diamondsuit$

5.7 Proof of Theorem 5

The proof of Theorem 5 is virtually identical to the proof of Theorem 3. One defines trials identically, and Lemma 14 holds again. The probability of success in no more than $0.5(0.75-\epsilon)n/c$ trials is no more than $0.5(0.75-\epsilon)$ where c is the constant in Lemma 14. Again, all trials last at least $0.05n$ steps. So $\|P_m-\tilde{\pi}\| \geq ((3n-1)/(4n)) - 0.5(0.75-\epsilon)$ if $m < 0.5(0.75-\epsilon)(n/c)(0.05n)$. If $\epsilon \in$

$(0, 0.75)$, then $((3n-1)/(4n)) - 0.5(0.75-\epsilon) > \epsilon$ for n larger than some value N_ϵ. Thus Theorem 5 holds for $n > N_\epsilon$ and can be extended to $n > 0$ just as Theorem 3 was extended. $\diamondsuit$

6 Proof of Theorem 4

6.1 Overview

The proof of Theorem 4 will share some characteristics with the proof of Theorem 2, but the different probability π will require some changes in the details. In particular, we will find a lower bound on $p_{ij}^{(cn)}$ for some constant $c > 0$, all $i \in \tilde{X}$, and all j in a subset of $\tilde{X}$. This lower bound will be based on probabilities of paths which are concatenations of paths which start at i and end at a base point, paths which start and end at the same base point, and paths which start at the base point and end at j. For this proof, the base points will be $(+1, 1)$, $(+1, n)$, and $(-1, 2n)$.

We will use the same convention about the Markov chain following paths as described in the overview of the proof of Lemma 2.

6.2 Lemmas Used to Prove This Theorem

The following lemma, along with Lemma 1, implies Theorem 4.

Lemma 15 *There exist constants N, $c_1 > 0$, and $c_2 > 0$ such that if $n > N$, then $p_{ij}^{(c_1 n)} > c_2/n^2$ for all $i \in \tilde{X}$ and all $j \in J :=$ $\{(+1, \ell) | 0.75n < \ell < 1.25\ell\} \cup \{(-1, \ell) | 0.75n < \ell < 1.25n\}$.*

The following three lemmas (with the base points in the previous subsection) imply Lemma 15 just as Lemmas 8, 9, and 10 imply Lemma 7. These lemmas apply for integers n larger than some positive integer N.

Lemma 16 *There is a constant $c_3 > 0$ such that for each $i \in \tilde{X}$, there is a base point (possibly depending on i) such that the Diaconis-Holmes-Neal sampler starting at i will reach this base point in the first $2n$ steps with probability at least c_3.*

Lemma 17 *There is a constant $c_4 > 0$ such that for each $j \in J$ and each base point, the probability that the Diaconis-Holmes-Neal sampler started at the base point will reach j in the first $2n$ steps with probability at least c_4/n.*

Lemma 18 *There exists a constant $c_5 > 0$ such that for any $m \in [n, 5n]$ and each base point, the probability that the Diaconis-Holmes-Neal sampler starting at the base point will again be at that base point after m steps is at least $c_5 > 0$.*

6.3 Proof of Lemma 16

Throughout this proof, we shall assume n is at least 6.

Observe that π is minimized at $k_1 := \lfloor (n+1)/2 \rfloor$ and $k_2 := n+k_1$. Let $k_{l,1} := \min\{j : j \leq k_1 \text{ and } \pi(j) < 2\pi(k_1)\}$, let $k_{l,2} := k_{l,1} + n$, let $k_{u,1} := \max\{j : j \leq n, j \geq k_1, \text{ and } \pi(j) < 2\pi(k_1)\}$, and let $k_{u,2} := k_{u,1} + n$.

Cases 1-4: Proofs for the cases $i = (+1, j)$ for some $j < k_{l,1}$ (Case 1), $i = (+1, j)$ for some $j \geq k_{l,2}$ (Case 2), $i = (-1, j)$ for some $j \leq k_{r,1}$ (Case 3), and $i = (-1, j)$ for some $j > k_{r,2}$ (Case 4) are analogous to those of corresponding cases in the proof of Lemma 8. In Cases 1 and 3, the base point is $(+1, 1)$, and in Cases 2 and 4, the base point is $(-1, 2n)$.

The other points will have base point $(+1, n)$ and will require more detail. Here the Diaconis-Holmes-Neal sampler will be likely to move around in the middle section and pass $(+1, n)$ in doing so.

Case 5: Here $i = (+1, j)$ for some j with $k_{l,1} \leq j \leq n$. The probability that the Markov chain follows the path $(+1, j) \to \ldots \to (+1, n)$ is $a(1 - \theta)^{n-j} \geq (1/2)(1 - 1/n)^n$ where

$$a := \begin{cases} 1 & \text{if } j \geq k_1 \\ \pi(k_1)/\pi(j) & \text{if } j < k_1 \end{cases}$$

Note that $a \geq 1/2$ and that this probability is at least a positive constant.

Case 6: Here $i = (+1, j)$ for some j with $n < j < k_{l,2}$. The probability that there exists a value $j' \leq k_2$ such that the steps starting at i go $(+1, j) \to \ldots \to (+1, j') \to (-1, j')$ is at least $(1/2)(1 - \theta)^{n/2}$ which is at least a positive constant. (To prove this, use the ideas of Case 1 of the proof of Lemma 8.)

The probability that there exists a value j'' with $k_1 + 1 \leq j'' \leq n$ such that the Markov chain follows the path $(-1, j') \to \ldots \to (-1, j'') \to (+1, j'')$ is at least $(1/2)(1 - \theta)^n$ which is at least a positive constant. (To prove this, again use the ideas of Case 1 of the proof of Lemma 8 and use the fact that $1-(\pi(k_1+1)/\pi(n)) \geq 1/2$ for large enough n.)

Like the previous case of this lemma, the probability that the Markov chain follows the path $(+1, j'') \to \ldots \to (+1, n)$ is at least a positive constant.

Concatenating these paths gives this case.

Case 7: Here $i = (-1, j)$ for some j with $k_{r,1} < j \leq k_{r,2}$. The probability that there exists a value $j' \geq k_1$ so that the Markov chain follows the path $(-1, j) \to \ldots \to (-1, j') \to (+1, j')$ is at least $a(1/2)(1-\theta)^{(3/2)n}$ where

$$a := \begin{cases} 1 & \text{if } j \leq k_2 \\ \pi(k_2)/\pi(j) & \text{if } j > k_2 \end{cases}$$

(Again, to prove this, use the ideas of Case 1 of the proof of Lemma 8.) Note that $a \geq 1/2$ and that this probability is at least a positive constant.

By Case 5 of this proof, the probability that the Markov chain follows the path $(+1, j') \to \ldots \to (+1, n)$ is at least a positive constant.

Concatenating these paths gives this case.

This case completes the proof of this lemma. $\diamondsuit$

6.4 Proof of Lemma 17

We shall again break the argument into various cases.

The first case is where the base point is $(+1, 1)$ and $j = (+1, \ell)$ with $(3/4)n < \ell < (5/4)n$. We'll let k_1 and k_2 be as in the proof of Lemma 16.

The probability that the Markov chain follows the path $(+1, 1) \rightarrow \ldots \rightarrow (+1, \ell)$ is

$$\frac{\pi(k_1)}{\pi(1)}(1-\theta)^{\ell-1}$$

if $\ell \leq n$ and is

$$\frac{\pi(k_1)}{\pi(1)} \frac{\pi(\ell)}{\pi(n+1)}(1-\theta)^{\ell-1}$$

if $\ell > n$. Note that $\pi(k_1)/\pi(1) \geq 1/n$ and $\pi(\ell)/\pi(n+1) \geq 1/2$. This case follows since $\theta = 1/n$.

The case where the base point is $(-1, 2n)$ and $j = (-1, \ell)$ with $(3/4)n < \ell < (5/4)n$ is similar.

Now let us consider the case where the base point is $(+1, 1)$ and $j = (-1, \ell)$ with $(3/4)n < \ell < (5/4)n$. Let us consider, for j' with $\max(n+1, \ell) \leq j' \leq (5/4)n + (1/8)n$, the probability that the Markov chain follows the path $(+1, 1) \rightarrow \ldots \rightarrow (+1, j') \rightarrow (-1, j') \rightarrow (-1, \ell)$. Note that each such path has length no more than $2n$. This probability is

$$\begin{aligned} & \frac{\pi(k_1)}{\pi(1)} \frac{\pi(j')}{\pi(n+1)} \left(1 - \frac{\pi(j'+1)}{\pi(j')}\right)(1-\theta)^{(j'-1)+1+(j'-\ell)} \\ \geq\ & \frac{1}{n}\frac{1}{4}\frac{2}{n}(1-\theta)^{2n} \end{aligned}$$

The probability of having such a path (given a start at $(+1, 1)$) is $\lfloor n/8 \rfloor (1/n)(1/4)(2/n)(1-1/n)^{2n}$ since there are at least $\lfloor n/8 \rfloor$ such paths. Thus this case is completed.

The proof for the case where the base point is $(-1, 2n)$ and $j = (+1, \ell)$ with $(3/4)n < \ell < (5/4)n$ is similar.

We also need to consider what happens if the base point is $(+1, n)$. In this case, the probability of going from $(+1, n)$ to a given $j \in J$ in no more than $2n$ steps will be at least a positive constant (for large enough n).

First suppose $j = (+1, \ell)$ with $n \leq \ell < (5/4)n$, then the probability of the Markov chain following the path $(+1, n) \to (+1, n+1) \to \ldots \to (+1, \ell)$ is $(\pi(\ell)/\pi(n))(1-\theta)^{\ell-n}$, which for all $n \geq 2$ is at least a positive constant.

In the case with $j = (-1, \ell)$ with $(3/4)n < \ell < (5/4)n$, let us consider the probability the Markov chain follows the path $(+1, n) \to (+1, n+1) \to \ldots \to (+1, j') \to (-1, j') \to \ldots \to (-1, \ell)$ for each j' with $\ell \leq j' \leq (5/4)n + (1/8)n$. Each such path has probability at least $(1/4)(2/n)(1/2)(1-\theta)^n$, and there are at least $\lfloor n/8 \rfloor$ such paths. Thus, for $n \geq 8$, the probability that one of these paths will be followed is at least a positive constant.

In the case with $j = (+1, \ell)$ with $(3/4)n < \ell < n$, we shall consider paths of the form $(+1, n) \to (+1, n+1) \to \ldots \to (+1, j') \to (-1, j') \to \ldots \to (-1, j'') \to (+1, j'') \to \ldots \to (+1, \ell)$ with $(3/4)n - (1/8)n \leq j'' \leq \ell < n \leq j' < (5/4)n$. There are at least $\lfloor n/4 \rfloor \lfloor n/8 \rfloor$ such paths, and $(1/2)(2/n)(1/4)(2/n)(1-\theta)^{2n}$ is a lower bound for the probability of each such path. So for $n \geq 8$, the probability of having such a path is at least a positive constant.

This completes the proof of the lemma. $\diamondsuit$

6.5 Proof of Lemma 18

The proofs for the base points $(+1, 1)$ and $(-1, 2n)$ are analogous to the proofs for the base points $(+1, 1)$ and $(-1, n)$ in Lemma 10.

For the base point $(+1, n)$, let us consider paths $(+1, n) \to \ldots \to (+1, j') \to (-1, j' + 1) \to \ldots (-1, j'') \to (+1, j'') \to \ldots \to (+1, n)$ with $(3/4)n < j'' \leq n \leq j' < (5/4)n$.

In this section, we define $FR(m)$ to be the event that $X_m = (+1, n)$ and $X_k \neq (+1, 1)$ for $1 \leq k < m$.

The following lemma will consider relatively small odd values for m. Its proof uses the paths under consideration.

Lemma 19 *Suppose* $0.04n \leq m \leq 0.2n - 1$ *and* m *is odd. Then there is a constant* $c_1 > 0$ *such that* $P(FR(m)|X_0 = (+1, n)) \geq c_1/n$ *is* $n \geq 75$.

Proof: For such an m, there will be at least one such path; we require that $m \geq 3$. For $m = 3$, there is one such path. For $m = 5$, there are two such paths. In general, there will be $(m - 1)/2$ such paths. The probability of each path is at least a positive constant divided by n^2. Since $m \geq 0.04n \geq 3$, the conclusion follows. $\diamondsuit$

By considering paths of the form $(+1, n) \to (+1, n + 1) \to \ldots \to$

$(+1, j') \to (-1, j') \to \ldots (-1, j'') \to (+1, j'') \to \ldots \to (+1, n)$ with $(3/4)n < j'' \leq n < j' < (5/4)n$, we get the following lemma.

Lemma 20 *Suppose $n \geq 50$. Suppose m is even and $0.08n \leq m \leq 0.2n$. Then there is a constant $c_2 > 0$ such that $P(X_m = (+1, n)|X_0 = (+1, n)) \geq c_2/n$.*

Proof: The proof is similar to the proof of Lemma 19, but we require $m \geq 4$. The number of such paths is $(m-2)/2$, and each path occurs with probability at least a positive constant divided by n^2. ♢

The following lemma will complete the proof of Lemma 18

Lemma 21 *Suppose $j \geq 2$ and $n \geq 75$. There is a constant $c_2(j) > 0$ such that $P(X_m = (+1, n)|X_0 = (+1, n)) \geq c_2(j)/n$ whenever $m \in [0.04jn, 0.04(j+1)n)$.*

Proof: The proof is analogous to the proof of Lemma 13, and we shall not repeat the details. ♢

Lemma 18 follows from Lemma 21 just as Lemma 10 follows from Lemma 13. ♢

7 Questions for Further Study

One may want to explore the Diaconis-Holmes-Neal algorithm for a broader range of stationary probabilities π. One set of probabilities which might be worth studying involves having a fixed number of minima albeit without necessarily having the symmetries of the W. It still may be appropriate to pick base points between the minima, but different values for the maxima or different values for the minima need to be carefully considered. Another set of probabilities which might be worth studying would have the number of minima growing with n. For example, π might have minima around $0.5\sqrt{n}$, $1.5\sqrt{n}$, $2.5\sqrt{n}$, etc. and maxima around 1, $\sqrt{n}$, $2\sqrt{n}$, etc. If the maximal values of π are about twice the minimal values and $\theta = 1/n$, then perhaps around order $n^{3/2}$ steps will suffice. The position of the Markov chain will stay between two minima for roughly order $\sqrt{n}$ steps before going to the left or the right. A more rigorous argument might verify the conjecture that order $n^{3/2}$ steps suffices here.

Another question one might consider is expanding to more than one dimension. By considering more copies of the state space, one can create various generalizations of the Diaconis-Holmes-Neal algorithm. These generalizations may also sometimes overcome the effects of diffusion in the corresponding Metropolis algorithm.

8 Acknowledgments

The author thanks Persi Diaconis for discussing the Diaconis-Holmes-Neal algorithm and suggesting questions related to the algorithm. The author also thanks Susan Holmes for suggesting some further questions related to the algorithm and Laurent Saloff-Coste for asking about a W-shaped probability. The author also thanks Ben Jamison, Karin Reinhold, Carlos Rodriguez, and an anonymous reader for some comments on earlier drafts of the paper.

References

[1] Chung, K.L. (1968). *A Course in Probability Theory.* New York: Harcourt, Brace, and World.

[2] Diaconis, P., Holmes, S., and Neal, R. (2000). *Analysis of a non-reversible Markov chain sampler*, *Ann. Appl. Prob.* **10**, 726-752.

[3] Diaconis, P., and Saloff-Coste, L. (1995). What do we know about the Metropolis algorithm? *Proceeding of the Twenty-Seventh Annual ACM Symposium on the Theory of Computing.* New York: Association for Computing Machinery.

[4] Doob, J.L. (1953). *Stochastic Processes.* New York: John Wiley and Sons.

[5] Durrett, R. (1996). *Probability Theory and Examples, Second Edition.* Pacific Grove, CA: Wadsworth & Brooks/Cole.

[6] Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E. (1953) Equations of state calculations by fast computing machines, *J. Chem. Phys.* **21**, 1087-1092.

[7] Meyn, S.P., and Tweedie, R.L. (1993) *Markov Chains and Stochastic Stability.* New York: Springer-Verlag.

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

Department of Mathematics and Statistics

State University of New York

University at Albany

Albany, NY 12222

e-mail: `martinhi@math.albany.edu`