Rates of Convergence for Lamplighter Processes

Olle Häggström* and Johan Jonasson
Chalmers University of Technology

April 7, 1997

Abstract

Consider a graph, G, for which the vertices can have two modes, 0 or 1. Suppose that a particle moves around on G according to a discrete time Markov chain with the following rules. With (strictly positive) probabilities p_m, p_c and p_r it moves to a randomly chosen neighbour, changes the mode of the vertex it is at or just stands still respectively. We call such a random process a (p_m, p_c, p_r)-lamplighter process on G. Assume that the process starts with the particle in a fixed position and with all vertices having mode 0. The convergence rate to stationarity in terms of the total variation norm is studied for the special cases when $G = \mathcal{K}_N$, the complete graph with N vertices, and when $G = \mathbf{Z} \bmod N$. In the former case we prove that as $N \to \infty$, $\frac{2p_c+p_m}{4p_cp_m} N \log N$ is a threshold for the convergence rate. In the latter case we show that the convergence rate is asymptotically determined by the cover time C^N in that the total variation norm after aN^2 steps is given by $P(C^N > aN^2)$. The limit of this probability can in turn be calculated by considering a Brownian motion with two absorbing barriers. In particular this means that there is no threshold for this case.

1 Introduction

Studies of quantative convergence rates for Markov chains is a steadily expanding area of modern probability. One reason, which should not be underestimated, for the activity in this area is that these problems are often easily described and raise people's curiosity. Card shuffling is one example of this. Assume that a deck of cards is initially arranged in some known order. We start to mix the deck with some shuffling technique. How many shuffles are required to properly mix the deck? This question has been studied and answered for a variety of different shuffling techniques by different authors. The most famous result of this type is found in Bayer and Diaconis [5] where it is shown that about seven ordinary riffle shuffles are sufficient for a deck of 52 cards to become reasonably well-mixed.

Another reason, perhaps more important from a practical point of view, is the application to Markov chain Monte Carlo algorithms, where a Markov chain is defined in such a way that its distribution converges to some probability distribution of interest. It is obviously very important to know how long time is required to

*Research supported by a grant from the Swedish Natural Science Research Council.

obtain satisfactory convergence. However, bounds which are at the same time rigorous and good enough to be of practical use are often very difficult to find. As a consequence of this, much of today's Markov chain Monte Carlo practice lacks rigorous theoretical justification.

There of course exist general results on rates of convergence for Markov chains, but in all but the very simplest situations these are much too crude to be of any use in practice. Therefore, one is forced to study special cases (or, at best, classes of special cases). During such studies many useful techniques have been developed, such as coupling, strong stationary times, eigenvalue analysis, etc. We refer to Rosenthal [16] who gives a survey of such methods. These methods have in common that they have all arisen as tools to solve particular problems but have later proved to apply to many different situations. Likewise, for every special case studied there is a hope that the methods used will prove useful in other situations as well. It is our hope that our studies of convergence rates for lamplighter processes will, apart from satisfying our own curiosity, shed light on other problems.

We now describe our setup. Imagine a finite connected graph, G, equipped with a lamp at each vertex. Imagine further a person doing a simple random walk on this graph, but apart from walking he can choose either to push the button or to rest, both with strictly positive probability. The random process obtained is an irreducible and aperiodic Markov chain on the state space $G \times \{0,1\}^G$ (we identify, with some abuse of notation, G with its set of vertices) of possible positions for the person and possible modes for the lamps. It is easy to see that the unique stationary distribution for the process consists of having each lamp independently on or off with probability $\frac{1}{2}$ each, and independently of this the position of the person is distributed according to the stationary distribution of simple random walk on G (i.e. the probability of standing at a vertex g is proportional to the number of edges incident to g). The principal question of this paper is the following:

> *If we start with the person at a fixed vertex and all lamps off, how long does it take to come close to the stationary distribution?*

By "close" we mean close with respect to total variation norm (see Section 2). The restriction to starting configurations with all lamps off is for simplicity of description only; a moment's thought reveals the convergence rates are the same for arbitrary initial states of the lamps. In the sequel we will be more formal and instead of talking about lamps we shall say that the vertices can have two modes, 0 or 1, and the person will be reduced to a particle. If the particle moves, changes the mode of the vertex it is at or rests with probabilities p_m, p_c and p_r respectively, we will call the process a *(p_m, p_c, p_r)-lamplighter process on G.*

The question of the convergence rate will be answered for two cases, namely for $G = \mathcal{K}_N$, i.e. for the complete graph on N vertices, and for $G = \mathbf{Z} \bmod N$, i.e. for the case where G consists of N vertices arranged in a circle with edges between adjacent vertices. In both cases the answer will be given in terms of the asymptotics as $N \to \infty$. We shall see that the answers are quite different in the two cases; for the complete graph a threshold phenomenon is exhibited, while for the circle the convergence to equilibrium is much smoother. This result seems to be closely related to the fact that the cover time of simple random walk on G (i.e. the time taken until the random walk has visited every vertex, see e.g. [1] and the references therein) exhibits a threshold phenomenon for $\mathcal{K}_N$ but not for $\mathbf{Z} \bmod N$. It would be very interesting if one could come up with some reasonably sharp result which for general G relates the convergence rate for its lamplighter process to the cover time for simple random walk on G.

We have chosen to work exclusively in a discrete time setting. All our results and methods have straightforward analogues for continuous time. For the case $G = \mathcal{K}_N$ the case of continuous time is in fact even a bit easier since the complications of Section 3.2 do not arise.

Lamplighters and related processes have been studied in a few previous papers. The lamplighter on $\mathbf{Z}$ was studied by Kaimanovich and Vershik [11]. It can be viewed as a random walk on a group, and some of its interest comes from the fact that it lies "between" random walk on $\mathbf{Z}^d$ and random walk on a Caley tree, in that on one hand it moves away from the starting point at sublinear speed while on the other hand the group has exponential growth. Variants of this model have been studied in [14] to demonstrate counterintuitive behaviour for biased random walks, and in [9] as a prototype for a random walker interacting mutually with its environment.

In the next section we give some necessary preliminaries. Section 3 treats the lamplighter on $\mathcal{K}_N$ while Section 4 is devoted to the lamplighter on $\mathbf{Z}$ mod N.

2 Preliminaries

In the study of finite state Markov chains, the common way to measure the distance between two probability measures is in terms of the *total variation norm.*

DEFINITION. Let F be a finite set and let P and Q be probability measures on F. The *total variation norm* of $P - Q$ is given by

$$||P - Q|| = \sup_{A \subseteq F} (P(A) - Q(A)) = \frac{1}{2} \sum_{x \in F} |P(x) - Q(x)|.$$

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $\{X_n\}$ be an irreducible and aperiodic Markov chain with finite state space, S, defined on this probability space. Then the distribution of X_n converges to the unique stationary distribution, π. Assume that the Markov chain starts in some fixed state. When studying the rate of the convergence to stationarity we observe the behavior of $||P(X_n \in \cdot) - \pi||$ and say that we are close to stationarity if this is close to 0 and far from stationarity if it is close to 1. (Note that the total variation norm is a quite conservative measure of how close you are to stationarity. Assume for instance that $\{X_n\}$ is a lamplighter process about which we know that at time k it is completely stationary except for that we know that one particular lamp is on. Then $||P(X_k \in \cdot) - \pi|| = \frac{1}{2}$.) If the state space, S, is large it is often hard to say very much about the total variation norm because of the enormous calculations such statements involve. In many situations, however, it is possible to let the size of the state space grow in a natural way, and in such cases one can sometimes make sharp statements of the asymptotic behavior of the total variation norm as $|S| \to \infty$, where $|S|$ is the cardinality of the state space. Such statements are often made in terms of *lower and upper bounds* and *thresholds* for the convergence rate.

DEFINITION. Let $\{S_N, N = 1, 2, \ldots\}$ be a sequence of state spaces and let $\{\{X_n^N\}_{n=0}^{\infty}, N = 1, 2, \ldots\}$ be a sequence of Markov chains on $S_1, S_2, \ldots$ respectively.

Further, let $\pi^1, \pi^2, \ldots$ denote the stationary distributions of the Markov chains respectively. If $\{k(N)\}$ is a sequence such that

$$\lim_{N\to\infty} ||P(X^N_{k(N)} \in \cdot) - \pi^N|| = 1$$

then we say that $k(N)$ is a *lower bound* for the convergence rate of the sequence of Markov chains. If, on the other hand,

$$\lim_{N\to\infty} ||P(X^N_{k(N)} \in \cdot) - \pi^N|| = 0$$

we say that $k(N)$ is an *upper bound.* If $\{k(N)\}$ is such that for any $\epsilon > 0$, $(1-\epsilon)k(N)$ is a lower bound and $(1+\epsilon)k(N)$ is an upper bound, then $k(N)$ is said to be a *threshold* for the convergence rate.

These definitions make sense due to the well known and easily proved fact that the total varaition norm is decreasing in n. When calculating lower bounds one method is to find a sequence $\{A^N\}$ of easily analyzed sets such that $P(X^N_{k(N)} \in A^N) \to 0$ whereas $\pi^N(A^N) \to 1$. Then $k(N)$ is a lower bound. For upper bounds, two useful tools are *coupling times* and *strong stationary times.*

DEFINITION. Let $\{X_n\}$ and $\{Y_n\}$ be two Markov chains on the same state space and with the same transition probabilities, and suppose that their joint behaviour is specified in such a way that for each i, $X_i = Y_i$ implies $X_{i+1} = Y_{i+1}$. Define the random variable

$$T = \inf\{n : X_n = Y_n\}.$$

We call T the *coupling time* for the two Markov chains.

DEFINITION. Let $\{X_n\}$ be a Markov chain on the state space S, with stationary distribution π. Assume that T is a stopping time such that $P(X_n = x|T = k) = \pi(x)$ for every n and k such that $k \leq n$ and every $x \in S$. Then T is said to be a *strong stationary time* for $\{X_n\}$.

The reason for the usefulness of these tools is the well known *coupling inequality* (see e.g. Lindvall [13]). We state it below and, for later purposes, we supply the proof.

LEMMA 2.1 *Let $\{X_n\}$ be a Markov chain with state space S and let $\{Y_n\}$ be another Markov chain with the same state space and the same transition probabilities starting in the stationary distribution, π. If T is a coupling time for the two chains, then*

$$||P(X_n \in \cdot) - \pi|| \leq P(T > n). \qquad (1)$$

If T is instead a strong stationary time for $\{X_n\}$ then (1) still holds.

Proof. Suppose that T is a coupling time. Then, for any subset, A, of S

$$P(X_n \in A) - \pi(A) = P(X_n \in A) - P(Y_n \in A)$$

$$= P(X_n \in A, T \leq n) - P(Y_n \in A, T \leq n)$$

$$+P(X_n \in A, T > n) - P(Y_n \in A, T > n)$$

$$\leq P(T > n)$$

since the events $\{X_n \in A, T \leq n\}$ and $\{Y_n \in A, T \leq n\}$ are the same. If T is instead a strong stationary time for $\{X_n\}$, the formalism is exactly the same, but the philosophy is a bit different since these events are no longer the same. They have, however, the same probability. □

For further discussion see e.g. [13], [3] or [6].

3 The Complete Graph

In this section we consider a (p_m, p_c, p_r)-lamplighter process on $G = \mathcal{K}_N$, the complete graph with N vertices. The probabilities p_m and p_c are assumed to be strictly positive. For convenience, we will use the convention that the embedded random walk is allowed to jump to the vertex it is presently at. Readers who dislike having loops in graphs may think of this as moving a probability $1/N$ from p_m to p_r. This makes no difference for our asymptotical considerations. The advantage of this convention is that the position of the particle is uniform immediately after its first move.

Now, to be formal, let $(\Omega, \mathcal{F}, P)$ be a probability space and let $\{X_n\}$ be a (p_m, p_c, p_r)-lamplighter process defined on Ω, starting in the state $X_0 = (g, 0, 0, \ldots, 0)$ for some fixed vertex g, i.e. with the particle in vertex g and with all vertices having mode 0. The stationary distribution of this process is the uniform distribution on $G \times \{0,1\}^G$. This means that the modes of the vertices are 0 or 1 with probability $\frac{1}{2}$ independently of each other and of the position of the particle which is uniformly distributed on G. We are going to show that a threshold for the convergence rate to stationarity is given by

$$\frac{2p_c + p_m}{4p_c p_m} N \log N. \tag{2}$$

It turns out that this is much simpler to show in the case when $p_c \leq p_r$, so we will do this case first and only then move on to the general case.

3.1 The case $p_c \leq p_r$

The main tool for determining the convergence rate in the case $p_c \leq p_r$ is the use of an auxiliary process $\{X_n^*\}$ with state space $G \times \{0, \sigma\}$, where σ is short for "stationary" (the reason for this notation will be evident from Lemma 3.1 below). We will run $\{X_n^*\}$ and $\{X_n\}$ simultaneously on the same probability space. Just as the original process, $\{X_n^*\}$ is a Markov chain which should be thought of as a particle moving in the graph G and switching the modes of the vertices it visits. The precise behaviour of $\{X_n^*\}$ is as follows. At each time n, the particle chooses either

- to move to a uniformly chosen vertex, with probability p_m, or
- to turn the state of the vertex it stands at into state σ (irrespective of its previous state), with probability $2p_c$, or
- to rest, with probability $p_r - p_c$.

Note that a vertex in state σ remains in this state forever.

We now construct a Markovian coupling of $\{X_n\}$ and $\{X_n^*\}$. Start $\{X_n^*\}$ at time 0 with the same particle position as for $\{X_n\}$, and with all vertices in state 0. The joint behaviour is then specified through the following conditional behaviour of $\{X_n\}$ given $\{X_n^*\}$.

- Whenever the $\{X_n^*\}$ particle moves, the $\{X_n\}$ particle moves to the same vertex.
- When the $\{X_n^*\}$ particle chooses to turn its vertex into state σ, then the $\{X_n\}$ particle chooses either to switch its vertex or to rest, each with conditional probability $\frac{1}{2}$.
- When the $\{X_n^*\}$ particle chooses to rest, then its $\{X_n\}$ colleague does the same.

It is easy to check that $\{X_n\}$ defined in this way has the correct marginal behaviour, and also that the following result holds.

LEMMA 3.1 *For each n, the conditional distribution of X_n given X_n^* is as follows. The positions of the particles are identical for the two processes, all vertices which are in mode 0 for X_n^* are in mode 0 for X_n as well, whereas the vertices which are in mode σ for X_n^* are for X_n independently in mode 0 or 1 with probability $\frac{1}{2}$ each.*

Now, for $a > 0$, define two stopping times for the $\{X_n^*\}$-process:

$$T_{-a} = \inf\{n : X_n^*(v) = \sigma \text{ for all but at most } N - N^{1/2+a} \text{ vertices } v\}$$

$$T_{+a} = \inf\{n : X_n^*(v) = \sigma \text{ for all but at most } N - N^{1/2-a} \text{ vertices } v\}.$$

Our candidates for lower and upper bounds for the $\{X_n^*\}$ process are

$$k_{-a}(N) = (1/2 - 2a)\frac{2p_c + p_m}{2p_c p_m} N \log N$$

and

$$k_{+a}(N) = (1/2 + 2a)\frac{2p_c + p_m}{2p_c p_m} N \log N$$

respectively. The following lemma relates these to the stopping times T_{-a} and T_{+a}.

LEMMA 3.2 *With the definitions above,*

$$P(T_{-a} > k_{-a}(N)) \to 1$$

and

$$P(T_{+a} > k_{+a}(N)) \to 0$$

as $N \to \infty$ for any $a > 0$.

Proof. In order to "stationarize" a vertex, the $\{X_n^*\}$ particle has to first arrive at a vertex and then turn its state into σ. The time taken between two such stationarizations is the sum of two independent geometric random variables, the first of which has mean $\frac{1}{p_m}$ and the second of which has mean $\frac{1}{2p_c}$. Hence this time has mean $\frac{2p_c+p_m}{2p_cp_m}$ and variance $\frac{1-p_m}{p_m^2} + \frac{1-2p_c}{4p_c^2}$. Let the random variable Z_- denote the number of such stationarizations needed to reach the stopping time T_{-a}. Since the vertices visited are iid uniform, we are in the well known coupon collector's context, and we get $\mathbf{E}[Z_-] \sim (1/2-a)N\log N$ and $\mathbf{Var}(Z_-) = O(N^2)$. (These calculations are elementary and have been made many times before, see e.g. [2] on the "top to random"-shuffle.) Using Chebyshev's inequality twice (first for Z_- and then for the time taken to do Z_- stationarizations), the first half of the lemma follows. The second half is completely analoguous. □

Next, we write X_n as $(\alpha_n, \beta_n, \gamma_n)$ where α_n is the mode of vertex g (recall that g is the vertex where the particle starts), β_n is the configuration of modes on $G\setminus\{g\}$ and γ_n is the position of the particle. The stationary distribution π can be decomposed as $\pi = \pi_\alpha \times \pi_\beta \times \pi_\gamma$ where π_α, π_β and π_γ are the uniform distributions on $\{0,1\}$, $\{0,1\}^{G\setminus\{g\}}$ and G, respectively. It is clear that if we know that there is a certain number of 1-vertices in $G\setminus\{g\}$, then the distribution of these is uniform on the family of subsets of $G\setminus\{g\}$ of the size in question. Stated formally we have:

LEMMA 3.3 *Let L_n denote the number of 1-vertices in $G\setminus\{g\}$ at time n. Then*

$$||P(\beta_n \in \cdot) - \pi_\beta|| = ||P(L_n \in \cdot) - \mathcal{B}(N-1, 1/2)||$$

where $\mathcal{B}(m,p)$ denotes the law of the binomial distribution with parameters m and p.

Now let D_n be the event that the particle at time n is at a vertex which has already been made stationary, and let D'_n be the event that $X_n^*(g) = \sigma$, i.e. that the starting vertex has become stationary. Clearly, $P(D_n|T_{+a} \le n) \ge \frac{N-N^{1/2}}{N} \to 1$ as $N \to \infty$ so that $P(D_{k_{+a}(N)} \cap \{T_{+a} \le k_{+a}(N)\}) \to 1$ as $N \to \infty$. Similarly, $P(D'_{k_{+a}(N)}) \to 1$ as $N \to \infty$. Hence,

$$P(D_{k_{+a}(N)} \cap D'_{k_{+a}(N)} \cap \{T_{+a} \le k_{+a}(N)\}) \to 1$$

as $N \to \infty$. Since, for any event E,

$$||P(X_n \in \cdot) - \pi|| \le ||P(X_n \in \cdot|E) - \pi|| + ||P(X_n \in \cdot|E^c) - \pi||$$

it is enough in order to establish $k_{+a}(N)$ as an upper bound to prove that

$$||P(X_{k_{+a}(N)} \in \cdot|D_{k_{+a}(N)} \cap D'_{k_{+a}(N)} \cap \{T_{+a} \le k_{+a}(N)\}) - \pi|| \to 0.$$

Under this conditioning, $\alpha_{k_{+a}(N)}$, $\beta_{k_{+a}(N)}$ and $\gamma_{k_{+a}(N)}$ are independent, and furthermore $\alpha_{k_{+a}(N)}$ and $\gamma_{k_{+a}(N)}$ have the desired (i.e. uniform) distributions. Lemma 3.3 thus implies that

$$P(X_{k_{+a}(N)} \in \cdot|D_{k_{+a}(N)} \cap D'_{k_{+a}(N)} \cap \{T_{+a} \le k_{+a}(N)\}) - \pi||$$

$$= P(\beta_{k_{+a}(N)} \in \cdot|D_{k_{+a}(N)} \cap D'_{k_{+a}(N)} \cap \{T_{+a} \le k_{+a}(N)\}) - \pi_\alpha||$$

$$\le ||\mathcal{B}(N - N^{1/2-a} - 1, 1/2) - \mathcal{B}(N-1, 1/2)|| \to 0$$

(the last two steps follow by conditioning on the number of stationarized vertices and using the Local Central Limit Theorem, see e.g. [7]). This shows that $k_{+a}(N)$ is an upper bound for $\{X_n\}$.

Now, if we can establish $k_{-a}(N)$ as a lower bound we will have proved the desired threshold result. This, however, follows immediately from the fact that the probability of having at least $N/2 - N^{1/2+a/2}$ vertices in mode 1 tends to 1 for the stationary distribution whereas it tends to 0 for the distribution of $X_{k_{-a}(N)}$. This fact in turn follows immediately from the Central Limit Theorem keeping in mind that at times earlier than T_{-a} there are $N^{1/2+a}$ vertices which are à priori known to have mode 0, and using Lemma 3.2. We have thus proved the desired result for the case $p_c \leq p_r$.

PROPOSITION 3.4 *For the (p_m, p_c, p_r)-lamplighter process with $p_c \leq p_r$ on the complete graph with N vertices starting in a fixed vertex with all vertices in mode 0, a threshold for the convergence rate is given by*

$$\frac{2p_c + p_m}{4p_c p_m} N \log N.$$

3.2 The general case

The task in this subsection is to build on the approach of Section 3.1 in order to extend Proposition 3.4 in such a way that the condition $p_c \leq p_r$ can be dropped. The critical use of this condition is in the definition of the $\{X_n^*\}$ process, where a probability mass of size p_c is moved (compared to the $\{X_n\}$ process) from the resting probability to the switching probability. This breaks down for $p_c > p_r$ because the resting probability would drop below 0.

An obvious way to try to handle this problem for the $p_c > p_r$ case is to give $\{X_n^*\}$ state space $G \times \{0, \sigma, 1\}^G$, and to let the particle switch its vertex to σ with probability $2\min(p_c, p_r)$. The arguments giving the upper bound in Section 3.1 would then go through essentially unchanged to yield, for any $a > 0$, an upper bound of

$$(1+a)\frac{2\min(p_c, p_r) + p_m}{4\min(p_c, p_r)p_m} N \log N \tag{3}$$

in the general case. However, the arguments for the lower bound do not go through, and in fact the upper bound (3) is not sharp in the $p_c > p_r$ case. (This approach will be used in Section 4 for the lamplighter on $\mathbf{Z}$ mod N, where it, in contrast to on $\mathcal{K}_N$, gives the correct convergence rate.)

In order to find the correct convergence rate for the general case we shall take a different approach: namely to consider an embedded (time-transformed) process $\{Y_n\}$. This process has the same state space as $\{X_n\}$ and equals the $\{X_n\}$ process sampled at times at which the particle has just moved. More precisely, if k_n is the nth time that the particle decides to move in the $\{X_n\}$ process, then $Y_n = X_{k_n+1}$.

It is clear that $\{Y_n\}$ is a Markov chain, and in order to calculate its transition probabilities we shall calculate the probability that when the particle of the $\{X_n\}$ process leaves a vertex, the mode of that vertex is different from what it was when the particle first arrived. Denoting this probability d and conditioning on the first step after the arrival, we have

$$d = p_c(1-d) + p_r d$$

so that

$$d = \frac{p_c}{1 + p_c - p_r} = \frac{p_c}{2p_c + p_m}.$$

The transition probabilities for $\{Y_n\}$ are thus as follows. At each time point the particle chooses a new vertex uniformly at random, and independently of this choice the mode of the vertex it just left switches with probability d. The point of introducing the $\{Y_n\}$ process is that $d < \frac{1}{2}$ for all values of (p_m, p_c, p_r), so that the approach used in the $p_c \leq p_r$ case becomes applicable for $\{Y_n\}$. Indeed, we may define a Markov chain $\{Y_n^*\}$ which relates to $\{Y_n\}$ in the same way as $\{X_n^*\}$ relates to $\{X_n\}$. The state space of $\{Y_n^*\}$ is $G \times \{0, \sigma\}^G$, and its transition probabilities are such that the particle positions are iid uniform, and the vertex which the particle has just left turns into mode σ with probability $2d$, and keeps its value otherwise. We can couple $\{Y_n\}$ and $\{Y_n^*\}$ in a way analogous to the coupling of $\{X_n\}$ and $\{X_n^*\}$ in Section 3.1, and a $(\{Y_n\}, \{Y_n^*\})$ analogue of Lemma 3.1 then holds. Proceeding as in Section 3.1, we arrive at the following result (note that $\{Y_n\}$ obviously has the same unique stationary distribution π as $\{X_n\}$):

PROPOSITION 3.5 *Fix $a > 0$, and let $\hat{k}_{-a}(N) = (1/2 - 2a)\frac{2p_c + p_m}{2p_c} N \log N$ and $\hat{k}_{+a}(N) = (1/2 + 2a)\frac{2p_c + p_m}{2p_c} N \log N$. We then have*

$$||P(Y_{\hat{k}_{-a}(N)} \in \cdot) - \pi|| \to 1$$

and

$$||P(Y_{\hat{k}_{+a}(N)} \in \cdot) - \pi|| \to 0$$

as $N \to \infty$, so that $\frac{2p_c + p_m}{4p_c} N \log N$ is a threshold for $\{Y_n\}$. For $Y_{\hat{k}_{-a}(N)}$, we furthermore have that $P(Y_{k_{-a}(N)} \in E_a^N) \to 0$ and $\pi(E_a^N) \to 1$ as $N \to \infty$, where E_a^N is the event that at least $N/2 - N^{1/2 + a/2}$ vertices are in mode 1.

Next, write T_k for the random time point in the $\{X_n\}$ process corresponding to time k in the $\{Y_n\}$ process (i.e. T_k is the time immediately after the kth move of the $\{X_n\}$ particle). We have that T_k is the sum of k iid geometric random variables with mean $\frac{1}{p_m}$, whence $\mathbf{E}[T_k] = \frac{k}{p_m}$ and $\mathbf{Var}(T_k) = \frac{k(1-p_m)}{p_m^2}$. Using Chebyshev's inequality, we see that $T_{\hat{k}_{-a}(N)}$ and $T_{\hat{k}_{+a}(N)}$ are very well concentrated around $(1/2 - 2a)\frac{2p_c + p_m}{2p_c p_m} N \log N$ and $(1/2 + 2a)\frac{2p_c + p_m}{2p_c p_m} N \log N$, respectively. Naively, one might think that this together with Proposition 3.5 would immediately yield the desired threshold for $\{X_n\}$. This would indeed be the case if $T_{\hat{k}_{-a}(N)}$ and $T_{\hat{k}_{+a}(N)}$ were fixed times, or if we knew that $X_{T_{\hat{k}_{-a}(N)}}$ (resp. $X_{T_{\hat{k}_{+a}(N)}}$) was independent (or nearly independent) of $T_{\hat{k}_{-a}(N)}$ (resp. $T_{\hat{k}_{+a}(N)}$), because then we could treat $T_{\hat{k}_{-a}(N)}$ and $T_{\hat{k}_{+a}(N)}$ almost as strong uniform times. We do not know this, however, so some more work is needed. The following lemma supplies a lower bound rather painlessly.

LEMMA 3.6 *Fix $a > 0$, and let*

$$\hat{\hat{k}}_{-a}(N) = (1/2 - 2a)\frac{2p_c + p_m}{2p_c p_m} N \log N + \sqrt{N} \log N.$$

Then

$$||P(X_{\hat{\hat{k}}_{-a}(N)} \in \cdot) - \pi|| \to 1$$

as $N \to \infty$.

Proof. Fix $a > 0$, and define E_a^N as in Proposition 3.5. Note that since $\pi(E_{a/2}^N) \to 1$ as $N \to \infty$, it is sufficient to show that

$$P(X_{\hat{\hat{k}}_{-a}(N)} \in E_{a/2}^N) \to 0 \tag{4}$$

as $N \to \infty$.

Pick $\epsilon > 0$. By Chebyshev's inequality, we can find a constant K_1 such that the event

$$\begin{aligned} A_1 \quad = \quad & \Big\{(1/2 - 2a)\frac{2p_c + p_m}{2p_c p_m} N \log N - K_1\sqrt{N \log N} < T_{\hat{k}_{-a}(N)} \\ & < (1/2 - 2a)\frac{2p_c + p_m}{2p_c p_m} N \log N + K_1\sqrt{N \log N}\Big\} \end{aligned}$$

satisfies $P(A_1) > 1 - \epsilon$ for all sufficiently large N. We furthermore define the event

$$A_2 = \{X_{T_{\hat{k}_{-a}(N)}} \notin E_a^N\} = \{Y_{\hat{k}_{-a}(N)} \notin E_a^N\}.$$

By Proposition 3.5 we have that $P(A_2) \to 1$ as $N \to \infty$. Hence $P(A_1 \cap A_2) > 1 - 2\epsilon$ for large N. On the event A_1 we have for large N that $\hat{\hat{k}}_{-a}(N) > T_{\hat{k}_{-a}(N)}$ and that

$$\hat{\hat{k}}_{-a}(N) - T_{\hat{k}_{-a}(N)} \leq \sqrt{N} \log N + K_1\sqrt{N \log N},$$

so that the number of vertices that turn from state 0 to state 1 between time $T_{\hat{k}_{-a}(N)}$ and time $\hat{\hat{k}}_{-a}(N)$ is at most $\sqrt{N} \log N + K_1\sqrt{N \log N}$. Hence, for large N, the event $A_1 \cap A_2$ implies that at most

$$N/2 - N^{1/2+a/2} + \sqrt{N} \log N + K_1\sqrt{N \log N} \tag{5}$$

vertices are in state 1 at time $\hat{\hat{k}}_{-a}(N)$. The expression (5) is less than $N/2 - N^{1/2+a/4}$ for large N, whence

$$P(X_{\hat{\hat{k}}_{-a}(N)} \in E_{a/2}^N) \leq 1 - P(A_1 \cap A_2) < 2\epsilon$$

for large N. Since ϵ was arbitrary, (4) follows. □

The upper bound is given by the next lemma, whose proof requires a bit more machinery.

LEMMA 3.7 *For $a, K > 0$, define*

$$\hat{\hat{k}}_{+a,K}(N) = (1/2 + 2a)\frac{2p_c + p_m}{2p_c p_m} N \log N + KN.$$

For any $a, \epsilon > 0$, there exists a K such that

$$||P(X_{\hat{\hat{k}}_{+a,K}(N)} \in \cdot) - \pi|| < \epsilon$$

for all sufficiently large N.

Proof. Let us first consider yet another process $\{Z_n\}$ embedded in $\{X_n\}$. We give $\{Z_n\}$ the usual state space $G \times \{0,1\}^G$, and obtain it by sampling $\{X_n\}$ at times where the particle has just left a vertex leaving it in a different mode compared to its mode when the particle arrived (another way to say this is that $\{Z_n\}$ can be obtained by sampling $\{Y_n\}$ at times at which the number of 1-vertices has changed). The process $\{l(Z_n)\}$ is then identical to the Ehrenfest urn model, i.e. to the birth-and-death process on $\{0,1,\dots,N\}$ with transition probabilities

$$p_{xy} = \begin{cases} \frac{x}{N} & \text{if } y = x-1 \\ \frac{N-x}{N} & \text{if } y = x+1 \\ 0 & \text{otherwise} \end{cases}$$

(recall that the function $l : G \times \{0,1\}^G \to \{0,1,\dots,N\}$ simply counts the number of vertices in mode 1). Let $\{V^y(t)\}_{t\geq 0}$ denote the normalized Ornstein–Uhlenbeck process starting at y, i.e. the diffusion on $\mathbf{R}$ with $V^y(0) = y$, drift parameter $\mu(x) = -x$, and diffusion parameter $\sigma^2(x) = 1$. We shall exploit the well-known fact (see e.g. [10] or [12]) that the Ehrenfest urn model, suitably normalized, converges in distribution to the Ornstein–Uhlenbeck process as $N \to \infty$. Writing $\{l(Z_n^M)\}$ for the modification of $\{l(Z_n)\}$ obtained by starting with M 1-vertices (instead of 0), we have for any $y \in \mathbf{R}$ that

$$\left\{ \frac{l\left(Z_{[Nt/2]}^{N/2+y\sqrt{N/2}}\right) - \frac{N}{2}}{\sqrt{N/2}} \right\}_{t\geq 0} \xrightarrow{D} \{V^y(t)\}_{t\geq 0}\,.$$

The convergence is in the weak topology on the set of continuous functions on $\mathbf{R}_+$. Since $\{X_n\}$ is, basically, a slowdown of $\{Z_n\}$ by a factor $\frac{1}{p_m d} = \frac{2p_c+p_m}{p_c p_m}$, we get as a standard consequence that

$$\left\{ \frac{l\left(X_{[Nt/2p_m d]}^{N/2+y\sqrt{N/2}}\right) - \frac{N}{2}}{\sqrt{N/2}} \right\}_{t\geq 0} \xrightarrow{D} \{V^y(t)\}_{t\geq 0}\,. \tag{6}$$

The idea is now to use the recurrence of the Ornstein–Uhlenbeck process in order to find a coupling of $\{l(X_n)\}$ and a stationary version of the same process, with the property that the two trajectories meet before time $\hat{\hat{k}}_{+a,K}(N)$ with high probability. Pick $\epsilon > 0$, and set $\epsilon' = \frac{\epsilon}{10}$. Since

$$\mathbf{Var}(T_{\hat{k}_{+a}(N)}) = \frac{(1-p_m)k_{+a}(N)}{p_m^2} = (1/2+2a)\frac{2p_c+p_m}{2p_c p_m^3} N\log N,$$

we may, by Chebyshev's inequality, pick a constant K_2 such that

$$P(E'_N) > 1 - \epsilon' \tag{7}$$

for all N, where E'_N is the event that

$$(1/2+2a)\frac{2p_c+p_m}{2p_c p_m} N\log N - K_2\sqrt{N\log N} \;\leq\; T_{\hat{k}_{+a}(N)}$$

$$\leq (1/2+2a)\frac{2p_c+p_m}{2p_cp_m}N\log N + K_2\sqrt{N\log N}.$$

Similarly, we can pick K_3 such that for all N and a random variable ξ_N with law $\mathcal{B}(N,1/2)$ we have that

$$P\left(N/2 - K_3\sqrt{N/2} \leq \xi_N \leq N/2 + K_3\sqrt{N/2}\right) > 1-\epsilon' \tag{8}$$

By Proposition 3.5, we then have for all sufficiently large N that

$$P(E''_N) > 1-2\epsilon' \tag{9}$$

where E''_N is the event that

$$N/2 - K_3\sqrt{N/2} \ \leq \ l(X_{T_{\hat{k}_{+a}(N)}}) \ \leq \ N/2 + K_3\sqrt{N/2}.$$

By an application of (6), we have that

$$P\left(\sup_{n\leq 2K_2\sqrt{N\log N}}(|l(X_n^M) - N/2|) < 2K_3\sqrt{N/2}\right) > 1-\epsilon'$$

for large N and all $M \in [N/2 - K_3\sqrt{N/2}, N/2 + K_3\sqrt{N/2}]$. In conjunction with (7) and (9) with E'''_N being the event that

$$N/2 - 2K_3\sqrt{N/2} \leq l\left(X_{(1/2+2a)\frac{2p_c+p_m}{2p_cp_m}N\log N + K_2\sqrt{N\log N}}\right) \leq N/2 + 2K_3\sqrt{N/2}$$

this implies that for sufficiently large N,

$$P(E'''_N) > 1-4\epsilon'. \tag{10}$$

Now pick K_4 so large that if $\{V^y(t)\}$ and $\{\tilde{V}^{\tilde{y}}(t)\}$ are two independent Ornstein–Uhlenbeck processes we have

$$\inf_{y,\tilde{y}\in[-2K_3,2K_3]} P\left(V^y(t) = \tilde{V}^{\tilde{y}}(t) \text{ for some } t\in[0,K_4]\right) > 1-\epsilon'$$

(standard monotonicity arguments [13] imply that the infimum is attained for $y = -2K_3, \tilde{y} = 2K_3$, so such a K_4 can be found by recurrence and a.s. continuity of the Ornstein–Uhlenbeck process). For such a choice of K_4 we even have

$$\inf_{y,\tilde{y}\in[-2K_3,2K_3]} P\left(V^y(t) = \tilde{V}^{\tilde{y}}(t) \text{ for infinitely many } t\in[0,K_4]\right) > 1-\epsilon'. \tag{11}$$

Let us now couple $\{X_n\}$ with an independent process $\{\tilde{X}_n\}$ with the same transition probabilities but started in stationarity. By (10) and (8) we have for large N that with probability at least $1-5\epsilon'$ both $X_{(1/2+2a)\frac{2p_c+p_m}{2p_cp_m}N\log N + K_2\sqrt{N\log N}}$ and $\tilde{X}_{(1/2+2a)\frac{2p_c+p_m}{2p_cp_m}N\log N + K_2\sqrt{N\log N}}$ are in the interval

$$[N/2 - 2K_3\sqrt{N/2},\ N/2 + 2K_3\sqrt{N/2}].$$

By invoking (6) and (11), we thus have for any K_5 and sufficiently large N that the trajectories of $\{l(X_n)\}$ and $\{l(\tilde{X}_n)\}$ cross each other at least K_5 times before time

$$\hat{\hat{k}}_{+a,K}(N) = (1/2+2a)\frac{2p_c+p_m}{2p_cp_m}N\log N + KN$$

with probability at least $1 - 6\epsilon'$; here we take $K = 4K_4 p_m d$. Each time the trajectories cross, their value has to differ by at most 1 at some time point, and on each such occasion the probability that the trajectories will coincide after a few (say 2) time units is bounded away from 0. The probability that this happens in such a way that the state of the lamp at which the particle stands is identical for the two processes, is also bounded away from 0. We may then modify the coupling of $\{X_n\}$ and $\{\tilde{X}_n\}$ so that $l(X_n) = l(\tilde{X}_n)$ from that time on. By picking K_5 large enough we can make this coalescence happen before time $\hat{\hat{k}}_{+a,K}(N)$ with probability at least $1 - 7\epsilon'$ for large N, so that by the coupling inequality we have

$$||P(l(X_{\hat{\hat{k}}_{+a,K}(N)}) \in \cdot) - \mathcal{B}(N, 1/2)|| > 1 - 7\epsilon'.$$

It is clear from the discussion in Section 3.1 that $\{X_n\}$ rapidly forgets the particle's starting position, so a straightforward analogue of Lemma 3.3 implies that

$$||P(X_{\hat{\hat{k}}_{+a,K}(N)} \in \cdot) - \pi|| > 1 - 8\epsilon' > 1 - \epsilon$$

for large N, and we are done. $\Box$

The desired theorem now follows immediately from Lemmas 3.6 and 3.7:

THEOREM 3.8 *For the (p_m, p_c, p_r)-lamplighter process on the complete graph with N vertices starting in a fixed vertex with all vertices in mode 0, a threshold for the convergence rate is given by*

$$\frac{2p_c + p_m}{4p_c p_m} N \log N.$$

4 Z mod N

We now turn our attention to the case $G = \mathbf{Z} \bmod N$, i.e. when the vertices are connected in a circle. Just like for the complete graph the state space of the process is $G \times \{0,1\}^G$ and it is assumed that the process starts in $(g, 0, 0, \ldots, 0)$ for some fixed g. Again it is clear that the stationary distribution, π, is the uniform distribution on $G \times \{0,1\}^G$. We let, as before, X_n denote the state of the process at time n. It will be shown that asymptotically as $N \to \infty$, the total variation norm $||P(X_n \in \cdot) - \pi||$ is completely determined by the distribution of the *cover time*, C^N, for the simple random walk of the particle, i.e. by the distribution of the first time the particle has visited all the vertices. This is stated in the following theorem. In particular, it means that a lower bound for the convergence rate must be of strictly smaller order than N^2 and that an upper bound must be of strictly larger order than N^2, whence there is no threshold.

THEOREM 4.1 *For the (p_m, p_c, p_r)-lamplighter process on* **Z** *mod N, with p_m, p_c and p_r all strictly positive, the following holds: For any $a > 0$,*

$$\lim_{N\to\infty} ||P(X_{aN^2} \in \cdot) - \pi|| = \lim_{N\to\infty} P(C^N > aN^2)$$
$$= Pr(M(ap_m) - m(ap_m) < 1)$$

and, consequently,

$$\lim_{N\to\infty} ||P(X_{k(N)} \in \cdot) - \pi|| = \begin{cases} 0 & \text{if } k(N) = \Omega(N^2) \\ 1 & \text{if } k(N) = o(N^2). \end{cases}$$

(It should be noted that the convergence to 1 for the total variation norm for the case $k(N) = o(N^2)$ can be obtained directly by a simple consideration of the particle position.)

Here $M(t) = \max_{0\le s\le t} B(s)$ and $m(t) = \min_{0\le s\le t} B(s)$, where $B(t)$ is standard Brownian motion. An explicit formula for $Pr(M(x) - m(x) > 1)$ is given below. It can be calculated by integrating the joint density M and m. A formula for this density can be found e.g. in Feller [8], page 342. We are grateful to Jean-Francois Le Gall for supplying these calculations, thereby saving us a great deal of effort.

$$Pr(M(x) - m(x) > 1) = \sum_{k=0}^{\infty} \left(\frac{8}{(2k+1)^2\pi^2} + 8x \right) e^{-\frac{1}{2}(2k+1)^2\pi^2 x}.$$

In Section 4.1 it is shown that the time, T^N, taken to visit all but $2\log_2 N$ vertices is asymptotically equal to C^N in distribution. Since the appearance of an unbroken row of $2\log_2 N$ 0-vertices is extremely unlikely for the stationary distribution, it will follow that $P(C^N > n)$ is an asymptotic lower bound for $||P(X_n \in \cdot) - \pi||$.

In Section 4.2 we introduce an auxiliary process $\{X_n^*\}$ similarly as in Section 3.1. We also argue that very soon after the time C^N (at an N^2-scale; it follows from well known random walk considerations that $\mathbf{E}[C^N] \sim \frac{1}{2p_m}N^2$, see e.g. [4, Chapter 6, page 8]), all vertices will have been visited at least $b \log N$ times with high probability, where b is abitrary but fixed. This will imply that the probability of having an unstationary vertex (i.e. a vertex g with $X_n^*(g) \neq \sigma$) soon after C^N is very low. This fact combined with a coupling argument on behalf of the position of the particle will give us $P(C^N > n)$ as an asymptotic upper bound for $||P(X_n \in \cdot) - \pi||$.

Theorem 4.1 will follow immediately from the lower bound in Section 4.1 and the upper bound in Section 4.2.

4.1 Lower Bound

The following simple lemma is a kind of reversed variant of the coupling inequality.

LEMMA 4.2 *Let T be a stopping time for a Markov chain, $\{Z_n\}$, with state space S and stationary distribution η. Assume that for a certain subset A of S it holds that*

$$P(Z_n \in A | T > n) = 1.$$

Assume further that for a given $\epsilon > 0$ we have that $\eta(A) \le \epsilon$. Then

$$||P(Z_n \in \cdot) - \eta|| \ge P(T > n) - \epsilon.$$

Proof. This is immediate since

$$||P(Z_n \in \cdot) - \eta|| = \sup_{B\subseteq S}(P(Z_n \in B) - \eta(B))$$

and

$$P(Z_n \in A) - \eta(A) \ge P(T > n) - \epsilon.$$

□

Apply the above lemma to $\{X_n\}$ with $T = T^N$, the first time all but $2\log_2 N$ vertices have been visited, and A being the set of states with at least one unbroken row of at least $2\log_2 N$ 0-vertices. We have that $P(X_n \in A|T > n) = 1$ because vertices not visited by the particle are in state 0. Since the stationary probability that an unbroken row of $2\log_2 N$ 0-vertices starts at a fixed position, i, is $2^{-2\log_2 N} = 1/N^2$, it follows that the expected number of such rows is $1/N$ whence $\pi^N(A) \leq 1/N$. We get that

$$||P(X_n^N \in \cdot) - \pi^N|| \geq P(T^N > n) - 1/N$$

so asymptotically, $P(T^N > n)$ is a lower bound for the total variation distance between the distribution of X_n and the stationary distribution.

To see that T^N and C^N are asymptotically equal in distribution we use the well known Brownian motion approximation of a simple random walk. By Donsker's Theorem (see e.g. Durrett [7])

$$\left\{\frac{1}{N}S([N^2t])\right\}_{t\geq 0} \xrightarrow{D} \{B(t)\}_{t\geq 0}$$

as $N \to \infty$, where $S(n)$ is a simple symmetric random walk starting at the origin and $B(t)$ is a standard Brownian motion. The mode of convergence is the same as in the proof of Lemma 3.7. Now, let $\tilde{M}(n) = \max_{0\leq r\leq n} S(r)$ and $\tilde{m}(n) = \min_{0\leq r\leq n} S(r)$ and recall that $M(t) = \max_{0\leq s\leq t} B(s)$ and $m(t) = \min_{0\leq s\leq t} B(s)$. Since the events $\{C^N > n\}$ and $\{T^N > n\}$ correspond to the events $\{\tilde{M}([np_m]) - \tilde{m}([np_m]) < N\}$ and $\{\tilde{M}([np_m]) - \tilde{m}([np_m]) < N - 2\log_2 N\}$ respectively (the motion of the particle is basically a slowdown of $\{S(r)\}$ by a factor $1/p_m$), we get for any positive number a that

$$P(C^N > aN^2) \to Pr(M(ap_m) - m(ap_m) < 1)$$

and

$$P(T^N > aN^2) \to Pr(M(ap_m) - m(ap_m) < 1).$$

Let us sum up the results of this subsection in a proposition.

PROPOSITION 4.3 *For any $a > 0$ we have*

$$\liminf_{N\to\infty} ||P(X_{aN^2} \in \cdot) - \pi|| \geq \lim_{N\to\infty} P(C^N > aN^2)$$

$$= Pr(M(ap_m) - m(ap_m) < 1)$$

where M and m are the maximum and minimum processes of a standard Brownian motion, respectively.

4.2 Upper bound

In order to derive an upper bound for the $\{X_n\}$ process, we will introduce an auxiliary process $\{X_n^*\}$ similarly as in Section 3.1. The state space of $\{X_n^*\}$ is $G \times \{0, \sigma, 1\}^G$, and the process is a Markov chain which as usual should be thought of as a particle walking on G and changing the modes of the vertices it visits. At each time n, the $\{X_n^*\}$ particle chooses either

- to move to a uniformly chosen vertex, with probability p_m, or

- to turn the state of the vertex it stands at into state σ (irrespective of its previous state), with probability $2\min(p_c, p_r)$, or
- if $p_c < p_r$: to rest, with probability $p_r - p_c$
- if $p_c > p_r$: to switch the the state of the vertex it stands at into state 1 (resp. 0, σ) if its previous state is 0 (resp. 1, σ), with probability $p_c - p_r$.

(Note that if $p_c \leq p_r$, this definition of $\{X_n^*\}$ is identical to that in Section 3.1.) We can now construct a coupling of $\{X_n\}$ and $\{X_n^*\}$ in exactly the same fashion as in Section 3.1, and the following analogue of Lemma 3.1 is then obvious:

LEMMA 4.4 *For each n, the conditional distribution of X_n given X_n^* is as follows. The positions of the particles are identical for the two processes, all vertices which are in mode 0 (resp. 1) for X_n^* are in mode 0 (resp. 1) for X_n as well, whereas the vertices which are in mode σ for X_n^* are for X_n independently in mode 0 or 1 with probability $\frac{1}{2}$ each.*

The main part of the game of deriving the upper bound for the $\{X_n\}$ process will be to show that with high probability, all vertices v satisfy $X_n^*(v) = \sigma$ shortly after cover time. To this end, we are now going to show that for any fixed positive b, all vertices have, with high probability, been visited at least $b\log N$ times shortly after cover time. Let for a fixed $a > 0$ the random variable U_a^b be the number of vertices which have not been visited at least $b\log N$ times at time $C^N + aN^2$.

LEMMA 4.5 $$\mathbf{E}[U_a^b] \leq \frac{b\log N}{a}.$$

Proof. Let, for all vertices, i, the time C_i be the first time that vertex i is visited and let A_i be the event that i is not visited at least $b\log N$ times before time $C^N + aN^2$. The Ergodic Theorem (see e.g. Durrett [7]) implies that the expected time between two visits at a vertex is N. Thus

$$P(A_i) \leq \frac{Nb\log N}{aN^2} = \frac{b\log N}{aN}$$

by Markov's inequality, using the fact that $C_i \leq C^N$. Therefore

$$\mathbf{E}[U_a^b] = \sum_{i=1}^{N} P(A_i) \leq \frac{b\log N}{a}$$

as desired. □

Now fix an $\epsilon \in (0, 1/2)$. Let us from now on call the vertices which have not been visited at least $b\log N$ times by time $C + aN^2$ "bad". Divide the bad vertices into two classes:

(1) the bad vertices at a distance at least ϵN from L, and

(2) the bad vertices closer than ϵN from L,

where L is the last vertex to be reached by the lamplighter, i.e. L is the position of the lamplighter at time C^N. Write

$$U_a^b = U_a^b(1) + U_a^b(2)$$

in obvious notation. We are going to prove that $P(U_a^b(k) > 0) \to 0$ as $N \to \infty$ for $k = 1, 2$. The key to our proof is the following result, which is of some independent interest and which also seems related to the study of cutpoints for simple random walk (see [15]).

THEOREM 4.6 *Consider a simple symmetric random walk, $\{S(n)\}_{n=0}^{\infty}$ on $\mathbf{Z}$ starting at the origin and with a reflecting boundary at N, where $N > 0$. Let T be the first time that $S_n = N$ and let $\tilde{U}_a^{b,\epsilon}$ be the number of vertices in $\{0, 1, \ldots, [(1-\epsilon)N]\}$ which have not been visited at least $b \log N$ times by the time $T + aN^2$. Then*

$$P(\tilde{U}_a^{b,\epsilon} > 0) \leq O(\frac{1}{\log N}).$$

The strong Markov property implies that as far as the points to the right of the origin are concerned we can regard the origin as a reflecting boundary so that by the same arguments as those used to prove Lemma 4.5 we have that

$$E[\tilde{U}_a^{3b,\epsilon/2}] \leq \frac{3b \log N}{a} = O(\log N).$$

Since it is true for any pair, (Z, W), of nonnegative random variables that

$$P(W > 0) \leq \frac{\mathbf{E}[Z]}{\mathbf{E}[Z|W > 0]},$$

we have that Theorem 4.6 will follow if it can be proved that

$$\mathbf{E}[\tilde{U}_a^{3b,\epsilon/2} | \tilde{U}_a^{b,\epsilon} > 0] \geq O((\log N)^2).$$

For this we need a couple of preliminary lemmas.

LEMMA 4.7 *For a simple symmetric random walk on the integers, the number of excursions from 0 required to reach $+M$ has a geometric distribution with mean $2M$.*

Proof. The strong Markov property implies that the paths of two excursions are iid. A standard martingale argument shows that an excursion has probability $1/2M$ of reaching $+M$. □

LEMMA 4.8 *Let $\xi_1, \xi_2, \ldots$ be iid $\{0, 1, 2\}$-valued random variables with distribution $\{p_0, p_1, p_2\}$ and let $Z = \inf\{i : \xi_i \in \{1, 2\}\}$. Then Z is independent of the event $\{\xi_Z = 2\}$ so that, in other words, the distribution of Z given $\xi_Z = 2$ has a geometric distribution with parameter $p_1 + p_2$.*

Proof.

$$P\{\xi_Z = 2, Z = k\} = P\{\xi_k = 2, Z = k\}$$

$$= p_2 p_0^{k-1} = \frac{p_2}{p_1 + p_2}(p_1 + p_2)p_0^{k-1} = P\{\xi_Z = 2\}P\{Z = k\}$$

□

Proof of Theorem 4.6. Note that since $\tilde{U}_a^{b,\epsilon} \geq \tilde{U}_{a'}^{b,\epsilon}$ for $a < a'$, it is no restriction to assume that a is small. Let a be small enough to ensure that the probability that the particle hits $[(1-\epsilon/2)N]$ in the time interval $(T, T+aN^2)$ is less than $1/2$ uniformly in N. Let B be the event that this does not happen and note that B is obviously positively correlated with the event $\{\tilde{U}_a^{b,\epsilon} > 0\}$ upon which we are going to condition. Now, for $i = 1, \ldots, N$, let V_i be the number of visits at i before time $T + aN^2$. Let X be the leftmost bad vertex, i.e. $X = \inf\{i : V_i \leq b \log N\}$ and note that this is well defined under the condition $\tilde{U}_a^{b,\epsilon} > 0$. If we can show that

$$\mathbf{E}[\tilde{U}_a^{3b,\epsilon/2} | \tilde{U}_a^{b,\epsilon} > 0, X = x] \geq O((\log N)^2)$$

for any $x < (1-\epsilon)N$, we will be done. By the above arguments, it is therefore enough to prove that

$$\mathbf{E}[\tilde{U}_a^{3b,\epsilon/2} | \tilde{U}_a^{b,\epsilon} > 0, X = x, B] \geq O((\log N)^2).$$

Under these conditions we may write V_i, for $i = x+1, x+2, \ldots, (1-\epsilon/2)N$ as

$$V_i = V_i' + V_i''$$

where V_i' is the number of visits at i before the last visit at x and V_i'' is the number of visits at i after the last visit at x and before T. Note that V_i' and V_i'' are conditionally independent.

Now, unconditionally, we have that the expected number of visits at i between two successive visits at x is 1. Given $V_x = v$, write

$$V_i' = \sum_{k=1}^{v-1} \xi_k$$

where ξ_k is the number of visits at i between visits k and $k+1$ at x. The conditions $\tilde{U}_a^{b,\epsilon} > 0$, $X = x$, B and $V_x = v$ means for the kth excursion from x that

(i) the excursion ends before it hits N, and

(ii) all states to the left of x are visited at least $b \log N$ times before the end of the $(v-1)$th excursion.

Condition (ii) makes excursions more likely to go to the left than to go to the right, and condition (i) also biases the random walk leftwards. Thus the conditions decrease the probability of hitting i at all. Also, given that i is hit, Lemma 4.8 implies that the expected number of times i is visited before the particle is back at x is not increased through this conditioning. We can conclude that

$$\mathbf{E}[V_i' | \tilde{U}_a^{b,\epsilon} > 0, X = x, B] \leq b \log N - 1 \leq b \log N.$$

Using Markov's inequality yields

$$P(V_i' \leq 2b \log N | \tilde{U}_a^{b,\epsilon} > 0, X = x, B) \geq \frac{1}{2}.$$

The next issue is to come up with a corresponding result for V_i''. Let us classify the excursions from i after the last visit at x into three goups:

(1) those who hit N,

(2) those who hit x,

(3) the others.

Since we are considering what happens after the last visit to x, we know that (1) will happen before (2). Calculating the unconditional probabilities of excursions of type (1), (2) and (3), and using Lemma 4.8, we obtain that V_i'' has a geometric distribution with expectation

$$\frac{2}{\frac{1}{i-x}+\frac{1}{N-i}},$$

which is less than or equal to $2(i-x)$. Thus

$$P(V_i'' \leq b\log N | \tilde{U}_a^{b,\epsilon} > 0, X = x, B) \geq 1 - (1 - \frac{1}{2(i-x)})^{b \log N}$$

which is at least $\frac{b\log N}{4(i-x)}$ for $i - x = \Omega(\log N)$ and N large enough. Combining this with what we know about V_i' we get

$$\begin{aligned}
&P(V_i \leq 3b\log N | \tilde{U}_a^{b,\epsilon} > 0, X = x, B) \\
&= P(V_i'' \leq b\log N | \tilde{U}_a^{b,\epsilon} > 0, X = x, B) P(V_i' \leq 2b\log N | \tilde{U}_a^{b,\epsilon} > 0, X = x, B) \\
&\geq \frac{b\log N}{8(i-x)}
\end{aligned}$$

eventually for $i - x = \Omega(\log N)$. In terms of the expectation of $\tilde{U}_a^{3b,\epsilon/2}$ this becomes

$$\begin{aligned}
&\mathbf{E}[\tilde{U}_a^{3b,\epsilon/2} | \tilde{U}_a^{b,\epsilon} > 0, X = x, B] = \sum_{i=x+1}^{(1-\epsilon/2)N} P(V_i \leq 3b\log N | \tilde{U}_a^{b,\epsilon} > 0, X = x, B) \\
&\geq \sum_{i=x+(\log N)^2}^{x+\epsilon N/2} \frac{b\log N}{8(i-x)} \geq \frac{b\log N}{8} O(\log N) \\
&= O((\log N)^2)
\end{aligned}$$

as desired. □

Now let us move on to the random walk on $\mathbf{Z}$ mod N. Conditioning on the position of L, say $L = l$, we know that the particle will visit one of the neighbours of l and then turn and go all the way around to reach l from the other direction. Associate the vertex at distance $[\epsilon N]$ from l in the direction of that neighbour with the origin in a simple random walk. From the time this vertex is hit until cover time the walk behaves like a simple symmetric random walk under the condition that $[(1-\epsilon)N]$ is hit before $-[\epsilon N]$. Let us call this event E and note that $P(E) = \epsilon$. The occurrence of the event $\{U_a^b(2) = 0\}$ can now be associated with the occurrence of the event $\{\tilde{U}_a^{b,\epsilon} = 0\}$ given E. Since this event has an unconditional probability which goes to 1, we can for any $\delta > 0$ choose N large enough to have

$$P(U_a^b(2) = 0) \geq \frac{\epsilon - \delta}{\epsilon}$$

which proves that

$$P(U_a^b(2) > 0) \to 0$$

as $N \to \infty$. To see that $P(U_a^b(1) > 0) \to 0$ choose, for a fixed but arbitrary $c > 0$, ϵ so small that the expected cover time for $\mathbf{Z}$ mod $c\epsilon N$ is less than δN^2. By considering a random walk on $\mathbf{Z}$ mod $c\epsilon N$ starting at the position of the original particle at time $C^N + aN^2$ we get that

$$P(U_{3a}^b(1) > 0) \leq \frac{2}{c} + \frac{\delta}{a} + p$$

where p is the probability that $U_a^b(2) > 0$ for the random walk on $\mathbf{Z}$ mod $c\epsilon N$ given that its cover time is less than aN^2, a probabilty which we know goes to 0. The first term in the above inequality comes from the well known fact that the position of the vertex last reached by simple random walk on $\mathbf{Z}$ mod $c\epsilon N$ is uniform. The second term follows from Markov's inequality. Since δ and c were arbitrary, it follows that

$$P(U_{3a}^b > 0) \to 0.$$

We have proved the following theorem.

THEOREM 4.9 *Let U_a^b be the numer of vertices of a simple symmetric random walk on $\mathbf{Z}$ mod N which have not been visited at least $b \log N$ times by time $C^N + aN^2$. Then*

$$P(U_a^b > 0) \to 0$$

for any positive numbers a and b.

We are now finally ready to go back to the lamplighter process $\{X_n\}$ and its auxiliary process $\{X_n^*\}$. Let S^N denote the the first time all the vertices have become stationary, i.e. have turned into state σ for the $\{X_n^*\}$ process. For any a, b and $\epsilon > 0$ choose N so large that $P(U_a^b) < \epsilon$ and note that the probability that a given vertex is unstationary after $b \log N$ visits is $q^{b \log N}$, where $q = \frac{p_m}{2\min(p_c, p_r) + p_m} < 1$. Thus the probability of having any unstationary vertex at time $C^N + aN^2$ is at most $\epsilon + Nq^{b \log N}$. Since b and ϵ are arbitrary we get

$$P(S^N - C^N > aN^2) \to 0$$

as $N \to \infty$ for any $a > 0$. Since

$$P(\frac{C^N}{N^2} > a - \delta) - P(\frac{C^N}{N^2} > a)$$

$$\to Pr(M((a-\delta)p_m) - m((a-\delta)p_m) < 1 \leq M(ap_m) - m(ap_m))$$

where as before M and m are the maximum and minimum processes of a standard Brownian motion, we can make this difference arbitrarily small by choosing N large and δ small. Combining these two facts yields

$$P(S^N > aN^2) - P(C^N > aN^2) \to 0$$

as $N \to \infty$.

We are, however, not finished yet. By Lemma 4.4, the time S^N is obviously a strong stationary time for the part of the process concerning the modes of the vertices, but it is not obviously so for the whole process including the the position of the particle. Write X_n as $X_n = (\alpha_n, \beta_n)$ where α_n is the configuration of modes and β_n is the position of the particle. Let $X'_n = (\alpha'_n, \beta'_n)$ be a copy of X_n starting in stationarity. Allowing ourselves to make a slight lie, we claim that by letting the particle of X'_n move in the opposite direction of the original particle until they meet, β_n and β'_n will with certainty have coupled before covering. This is completely true only when N is even and the two particles happen to start an even number of steps apart. Otherwise the coupling can be made to yield coalesence of the particle positions with a probability which tends to 1 as $N \to \infty$ by modifying it e.g. in such a way that the two processes evolve completely independently at times at which the particles are at an odd distance less than $\frac{N}{10}$ apart. Define

$$\tilde{S}^N = \inf\{n : S^N \le n \text{ and the two particles have coalesced}\}.$$

With the above coupling we have that $P(\tilde{S}^N = S^N) \to 1$ as $N \to \infty$, so that given $\epsilon > 0$ we can pick N sufficiently large so that $P(\tilde{S}^N \ne S^N) < \epsilon$. By imitating the proof of the coupling inequality (Lemma 2.1) and using Lemma 4.4 we get, for such N,

$$\begin{aligned}
&\sum_{w,z} |P(\alpha_n = w, \beta_n = z) - P(\alpha'_n = w)P(\beta'_n = z)| \\
&\le \sum_{w,z} P(\tilde{S}^N > n) \left| P(\alpha_n = w, \beta_n = z | \tilde{S}^N > n) - P(\alpha'_n = w)P(\beta'_n = z | \tilde{S}^N > n) \right| + \\
&+ P(\tilde{S}^N \le n) \left| P(\alpha_n = w | \tilde{S}^N \le n) P(\beta_n = z | \tilde{S}^N \le n) - P(\alpha'_n = w)P(\beta'_n = z | \tilde{S}^N \le n) \right| \\
&\le 2P(\tilde{S}^N > n) \\
&\le 2(P(S^N > n) + \epsilon)
\end{aligned}$$

so that

$$||P(X_n \in \cdot) - P(X'_n \in \cdot)|| = ||P(X_n \in \cdot) - \pi|| \le P(S^N > n) + \epsilon.$$

Since ϵ was arbitrary, and by the Brownian motion considerations in Section 4.1, we have established the following result:

PROPOSITION 4.10 *For any $a > 0$ we have*

$$\begin{aligned}
&\limsup_{N\to\infty} ||P(X_{aN^2} \in \cdot) - \pi|| \le \lim_{N\to\infty} P(C^N > aN^2) \\
&= Pr(M(ap_m) - m(ap_m) < 1)
\end{aligned}$$

where M and m are the maximum and minimum processes of a standard Brownian motion, respectively.

Theorem 4.1 now follows without further ado from Propositions 4.3 and 4.10.

REMARK. If instead $p_r = 0$, Proposition 4.10 may fail due to reducibility of the Markov chain. For instance, consider the $(2/3, 1/3, 0)$-lamplighter. If N is even and the particle starts in a known position, the present position of the particle will immediately tell us whether there is an odd or even number of 0-vertices.

ACKNOWLEDGMENT. We thank the referee for valuable comments and corrections.

References

1. D. Aldous, An Introduction to Covering Problems for Random Walks on Graphs, *J. Theor. Probab.* **2** (1989), 87–89.
2. D. Aldous and P. Diaconis, Shuffling Cards and Stopping Times, *Amer. Math. Monthly* **93** (1986), 333–348.
3. D. Aldous and P. Diaconis, Strong Stopping Times and Finite Random Walks, *Adv. Appl. Math.* **8** (1987), 69-97.
4. D. Aldous and J. A. Fill, "Reversible Markov Chains and Random Walks on Graphs," Draft book, December 1996.
5. D. Bayer and P. Diaconis, Trailing the Dovetail Shuffle to its Lair, *Ann. Appl. Probab.* **2** (1992), 294–313.
6. P. Diaconis, "Group Representations in Probability and Statistics," Harvard University, 1988.
7. R. Durrett, "Probability: Theory and Examples," Wadsworth & Brooks/ Cole, Pacific Grove, California, 1991.
8. W. Feller, "An Introduction to Probability Theory and its Applications, Vol. II," 3rd edition, Wiley, New York, 1966.
9. A. S. Garjat, Random Walk of an Active Particle in a Non-Homogeneous Environment, *Markov Proc. Relat. Fields* **1** (1995), 91–112.
10. M. Kac, Brownian Motion and the Theory of Random Walk, *Amer. Math. Monthly* **54** (1947), 369–391.
11. V. Kaimanovich and A. Vershik, Random Walks on Discrete Groups: Boundary and Entropy, *Ann. Probab.* **11** (1983), 457–490.
12. S. Karlin and H. M. Taylor, "A Second Course in Stochastic Processes", Academic Press, New York, 1981.
13. T. Lindvall, "Lectures on the Coupling Method," Wiley, New York, 1992.
14. R. Lyons, R. Pemantle and Y. Peres, Random Walks on the Lamplighter Group, *Ann. Probab.* **24** (1996), 1993-2006.
15. Y. Peres, Points of Increase for Random Walks, *Israel J. Math.,* to appear.
16. J. S. Rosenthal, Convergence Rates for Markov Chains, *SIAM Review* **37** (1995), 387–405.