Cutoff for a Random Walk on the Integers mod n

Michael E. Bate and Stephen B. Connor

15th July 2014

Abstract

We analyse a random walk on the ring of integers mod n, which at each time point can make an additive 'step' or a multiplicative 'jump'. When the probability of making a jump tends to zero as an appropriate power of n we prove the existence of a total variation cutoff for this process, with cutoff time dependent on whether the step distribution has zero mean.

Keywords: random walk; mixing time; cutoff phenomenon; group representation theory; integers mod n; random number generation

2010 Mathematics Subject Classification:
Primary 60J10

1 Introduction

In this note we consider a random walk X on $\mathbb{Z}_n = \mathbb{Z}/n\mathbb{Z}$ (where n is *odd*) defined as follows:

$$X_t = \begin{cases} X_{t-1} + \xi'_t \mod n & \text{with probability } 1 - p_n \\ 2X_{t-1} \mod n & \text{with probability } p_n, \end{cases} \tag{1}$$

where $\{\xi'_t\}$ are a set of i.i.d. random variables with finite support $B \subset \mathbb{Z}$, whose distribution does not vary with n. We denote the mean and variance of ξ' by μ and σ^2 respectively. We will refer to an 'addition' move as a 'step', and to a 'multiplication' move as a 'jump'. To ensure that X is irreducible we assume that the group $\langle B_n, +\rangle$ is not a proper subgroup of $\mathbb{Z}_n$ for any odd n, where $B_n = \{z \mod n : z \in B\}$. Furthermore, since n is odd, multiplication by 2 is an invertible operation, and thus X is ergodic with uniform equilibrium distribution π_n on $\mathbb{Z}_n$.

Define the total variation distance from π_n of a probability distribution P on $\mathbb{Z}_n$ by

$$\|P - \pi_n\|_{\mathrm{TV}} = \max_{A \subset \mathbb{Z}_n} |P(A) - \pi_n(A)| = \frac{1}{2} \sum_{s \in \mathbb{Z}_n} |P(s) - 1/n| \; .$$

A number of authors have previously considered random processes of the form

$$X_t = a_t X_{t-1} + b_t \mod n;$$

these processes are similar to schemes used for random number generation, a link which has naturally motivated interest in bounding the time taken for the total variation distance from uniform to become suitably small (the so-called "mixing time", which is typically taken to be the first time at which the total variation distance drops below 1/4). A nice introduction to the area can be found in Terras (1999, Chapter 6). The earliest such work appears to be that of Chung, Diaconis, and Graham (1987), in which $a_t = a = 2$ and b_t is chosen uniformly from $\{-1, 0, 1\}$: they show that $O(\log n \log\log n)$ steps suffice for this walk to mix, and that $O(\log n \log\log n)$ steps are also necessary for n of the form $2^m - 1$; on the other hand, for almost all odd n, $1.02 \log_2 n$ steps suffice. This (deterministic) act of doubling each time causes the process to mix significantly faster than when $a_t = 1$ for all t where, if b_t is uniform on a finite set (and assuming that the resulting process is irreducible), the mixing time is of order n^2 (Diaconis, 1988; Saloff-Coste, 2004).

Rather more general results have been established in a series of works by Hildebrand. It is shown in his thesis (Hildebrand, 1992, Chapter 3) that if multiplication is deterministic ($a_t = a$ for all t) and for fairly general choices of b_t (which don't depend on n), $O(\log n \log\log n)$ steps suffice, and in fact for almost all n, $O(\log n)$ steps suffice; the method closely follows that of Chung et al. (1987). When a_t is allowed to vary with t, a general upper bound for the mixing time is proved in (Hildebrand, 1993): using a recursive relation involving discrete Fourier transforms (of which more below), he shows that (unless $a_t = 1$ always, $b_t = 0$ always, or a_t and b_t can each take on only one value) $O((\log n)^2)$ time steps are always sufficient. Other related results can be found in Hildebrand (1994a,b).

A particularly interesting feature of these processes is the quantitatively different behaviour that can be obtained by making small changes to the distribution of a_t and b_t. For example, Chung et al. (1987) remark upon the following curiosity to be found when $a_t = 2$ and b_t is supported on $\{-1, 0, 1\}$ with $\mathbb{P}(b_t = 1) = \mathbb{P}(b_t = -1) = q$. If $q = 1/4$ or $q = 1/2$ then $O(\log n)$ steps suffice to make the total variation distance small; however, if $q = 1/3$ then $O(\log n \log\log n)$ steps may be required. Similarly, Hildebrand (1992, Chapter 5) considers the situation where b_t is uniform on ± 1 and a_t is supported on $\{2, (n+1)/2\}$, with $\mathbb{P}(a_t = 2) = p \in (0, 1)$: the mixing time is shown to be at most $O((\log n)^m)$, where m is 2 if $p = 1/2$, and 1 otherwise. If the distribution of b_t is altered to uniform on $\{-1, 0, 1\}$ then $O((\log n \log\log n)^m)$ steps suffice.

The principal difference between these earlier works and the process defined in (1) is that we allow the probability of a 'jump', p_n, to depend on n. In particular, we are able to show that if p_n tends to zero as a power of n, then our process exhibits a total variation cutoff.

Definition 1. (Levin, Peres, and Wilmer, 2009) A sequence of Markov chains $\{X^{(n)}\}_{n\in\mathbb{N}}$ is said to exhibit a *total variation cutoff* at time τ_n with *window size* w_n if $w_n = o(\tau_n)$ and

$$\lim_{c\to\infty} \liminf_{n\to\infty} \|\mathbb{P}(X^{(n)}_{\tau_n - cw_n} \in (\cdot)) - \pi_n(\cdot)\|_{\mathrm{TV}} = 1$$
$$\lim_{c\to\infty} \limsup_{n\to\infty} \|\mathbb{P}(X^{(n)}_{\tau_n + cw_n} \in (\cdot)) - \pi_n(\cdot)\|_{\mathrm{TV}} = 0\,.$$

Intuitively this says that as n gets large the convergence to equilibrium, measured using total variation distance, happens in a negligible window of order w_n around the cutoff time τ_n. We remark that it is possible for the 'right' and 'left' window sizes in the above definition to be of different orders – see Connor (2010) for an example. There has been much interest in studying the mixing times of Markov chains and proving the existence of cutoff phenomena: see Levin et al. (2009) and Diaconis (2011) for recent introductions to the area, or Saloff-Coste (2004) for a more analytical overview.

Our main result is the following.

Theorem 2. *Suppose that $p_n = 1/(2n^\alpha)$ for some $\alpha > 0$ such that $n/\sigma_{S'} \to \infty$, where*

$$\sigma^2_{S'} := \frac{(1-p_n)(\mu^2 + p_n\sigma^2)}{p_n^2}.$$

Then X exhibits a total variation cutoff at time $T_n^X = 2n^\alpha \log_2(n/\sigma_{S'})$, with window size $n^{\alpha/2}\sqrt{T_n^X}$.

This paper's contribution contrasts with the existing results mentioned above for processes of the type $X_t = a_t X_{t-1} + b_t$ (where the distribution of a_t is independent of n), for which (to the best of our knowledge) no cutoff results have been established. In the present setting the mixing time of our process X is also relatively insensitive to the distribution of the step lengths ξ'_t. Theorem 2 shows that the mixing time of X essentially depends on ξ' only through its mean, μ: in the case of zero drift the mixing time is $2(1-\alpha/2)n^\alpha \log_2 n$ (for $0 < \alpha < 2$), while if $\mu \neq 0$ the chain mixes slightly faster, with cutoff at $2(1-\alpha)n^\alpha \log_2 n$ (for $0 < \alpha < 1$).

2 Working with a subsampled chain

The main obstruction to analysing our process X using standard techniques for random walks on groups is that the distribution of X_k is not given by convolution of k independent increment distributions. This problem can be overcome by (initially) restricting attention to the process Y which is produced by subsampling X at jump times. Denote the jump times of X by $\tau_1, \tau_2, \ldots$, and let $\tau_0 = 0$; then $Y_k := X_{\tau_k}$. This process clearly satisfies $Y_k = Y'_k \mod n$, where

$$Y'_k = 2^k Y_0 + \sum_{i=1}^{k} 2^{k+1-i} S'_i, \tag{2}$$

and where

$$S'_i = \sum_{t=\tau_{i-1}+1}^{\tau_i - 1} \xi'_t. \tag{3}$$

Here (and throughout) we use the convention that random variables with a prime take values in $\mathbb{Z}$, while those without take values in $\mathbb{Z}_n$. Thus $S_i = S'_i \mod n$ is the change in X due to steps taken between jump times τ_{i-1} and τ_i. Like X, Y is ergodic with uniform equilibrium distribution. From (2) it is clear that the mixing time of Y is independent of its starting state, and so for ease of exposition we shall set $X_0 = Y_0 = 0$. It is also clear from (2) that the distribution of Y_k is given by convolution of the distributions corresponding to the independent increments $\{2^{k+1-i}S_i\}$, and this will prove essential to our method for establishing an upper bound on the mixing time in Section 3.

The length of time between jumps of X clearly has a Geometric(p_n) distribution:

$$\mathbb{P}(\tau_1 = j) = p_n(1-p_n)^{j-1}, \quad j = 1, 2, \ldots,$$

and a straightforward application of the conditional variance formula shows that

$$\sigma^2_{S'} := \mathrm{Var}\left[S'_i\right] = \frac{(1-p_n)(\mu^2 + p_n\sigma^2)}{p_n^2}. \tag{4}$$

Note in particular that when $p_n \to 0$,

$$\sigma_{S'}^2 \sim \begin{cases} \sigma^2/p_n & \text{if } \mu = 0 \\ \mu^2/p_n^2 & \text{otherwise.} \end{cases} \tag{5}$$

Theorem 3. *Suppose that $p_n = 1/(2n^\alpha)$ for some $\alpha > 0$ such that $n/\sigma_{S'} \to \infty$ as $n \to \infty$. Then Y exhibits a cutoff (in total variation distance) at time $T_n = \log_2(n/\sigma_{S'})$, with cutoff window of size $O(1)$. Indeed,*

$$\| \mathbb{P}\left(Y_{T_n+c} \in \cdot\right) - \pi_n(\cdot)\|_{\mathrm{TV}} \begin{cases} \geq 1 - 4^{1+c/3} & c < 0 \\ \leq O(4^{-c}) & \text{as } c \to \infty. \end{cases}$$

The fact that Y exhibits such a tight cutoff makes it simple to demonstrate a cutoff for X, as claimed in Theorem 2.

Corollary 4. *In the setting of Theorem 3, X exhibits a cutoff at time $T_n^X := T_n/p_n$, with window size $\sqrt{T_n}/p_n$.*

Proof. Let $w_n = \sqrt{T_n}/p_n$, and for $c \in \mathbb{R}$ let $J_n(c)$ denote the number of *jumps* in X before time $T_n^X + cw_n$. Then $J_n(c) \sim \mathrm{Poisson}(m_n(c))$, where $m_n(c) = p_n(T_n^X + cw_n) = T_n + c\sqrt{T_n}$. The proof essentially now follows from the observation that $J_n(c)$ concentrates in an interval of order $\sqrt{T_n}$ around $m_n(c)$ as $n \to \infty$. Indeed,

$$\limsup_{n\to\infty} \mathbb{P}\left(|J_n(c) - m_n(c)| > |c|\sqrt{T_n}\right) \leq \limsup_{n\to\infty} \frac{\mathrm{Var}\left[J_n(c)\right]}{c^2 T_n} = \frac{1}{c^2}.$$

To show that X has not mixed before time T_n^X we simply note that, since Y exhibits a cutoff at T_n with window size $O(1)$, for $c \geq 0$

$$\begin{aligned} \liminf_{n\to\infty} \|\mathbb{P}(X_{T_n^X - cw_n} \in \cdot) - \pi_n(\cdot)\|_{\mathrm{TV}} &\geq 1 - \limsup_{n\to\infty} \mathbb{P}\left(J_n(-c) \geq T_n - O(1)\right) \\ &\geq 1 - \limsup_{n\to\infty} \mathbb{P}\left(|J_n(-c) - m_n(-c)| > |c|\sqrt{T_n}\right) \geq 1 - \frac{1}{c^2}. \end{aligned}$$

Similarly,

$$\begin{aligned} \limsup_{n\to\infty} \|\mathbb{P}(X_{T_n^X + cw_n} \in \cdot) - \pi_n(\cdot)\|_{\mathrm{TV}} &\leq \limsup_{n\to\infty} \|\mathbb{P}(Y_{J_n(c)} \in \cdot) - \pi_n(\cdot)\|_{\mathrm{TV}} \\ &\leq \limsup_{n\to\infty} \mathbb{P}\left(J_n(c) \leq T_n + O(1)\right) \leq \frac{1}{c^2}. \end{aligned}$$

□

It therefore remains to prove Theorem 3. The left hand cutoff window follows relatively simply from an application of Chebychev's inequality, as the next result shows. In Section 3 we show how to use group representation theory to provide a proof of the matching upper bound.

Lemma 5. *For $c \geq 3$,*

$$\| \mathbb{P}\left(Y_{T_n-c} \in \cdot\right) - \pi_n(\cdot)\|_{\mathrm{TV}} \geq 1 - 4^{1-c/3}.$$

Proof. In order to lower bound the total variation distance between Y and its equilibrium distribution at time $T_n - c$ we use the fact that the total variation distance is the maximal difference between the distribution of Y_{T_n-c} and the uniform measure π_n on all possible subsets of $\mathbb{Z}_n$. So consider the set

$$A_n(c,\beta) = \{z \in \mathbb{Z}_n \,:\, |z - \mathbb{E}\left[Y_{T_n-c}\right]| > (3/8-\beta)n\}\,,$$

for some $\beta \in (0, 3/8)$ which we shall choose later. Note that this set satisfies $\pi_n(A_n(c,\beta)) = 1/4+2\beta$, and that (subject to this condition) it has been chosen to be as far away as possible from $\mathbb{E}\left[Y_{T_n-c}\right]$ (measured using the usual distance between two numbers mod n).

Using (2) we calculate the variance of Y'_k to be

$$\operatorname{Var}\left[Y'_k\right] = \sum_{i=1}^{k} 4^{k+1-i}\sigma^2_{S'} = \frac{4}{3}(4^k-1)\sigma^2_{S'}\,, \tag{6}$$

and so

$$\begin{aligned}\mathbb{P}\left(Y_{T_n-c} \in A_n(c,\beta)\right) &\le \mathbb{P}\left(\left|Y'_{T_n-c} - \mathbb{E}\left[Y'_{T_n-c}\right]\right| > (3/8-\beta)n\right)\\ &\le \frac{4^{1-c}n^2\sigma^2_{S'}}{3\sigma^2_{S'}\left((3/8-\beta)n\right)^2} = \frac{4^{1-c}}{3(3/8-\beta)^2}\,.\end{aligned}$$

Here the first inequality follows from Y being equal to Y' mod n, and the second from Chebychev's inequality and the definition of T_n. Thus $A_n(c,\beta)$ satisfies

$$\left|\mathbb{P}\left(Y_{T_n-c} \in A_n(c,\beta)\right) - \pi_n(A_n(c,\beta))\right| \ge \frac{1}{4} + 2\beta - \frac{4^{1-c}}{3(3/8-\beta)^2}\,. \tag{7}$$

This lower bound is maximised over values $\beta = \beta(c) \in (0, 3/8)$ when

$$\beta(c) = \frac{3}{8} - \left(\frac{4^{1-c}}{3}\right)^{1/3} \qquad (c \ge 3)$$

and using this value of $\beta(c)$ yields the claimed left hand window of the cutoff:

$$\|\,\mathbb{P}\left(Y_{T_n-c} \in \cdot\right) - \pi_n(\cdot)\|_{\mathrm{TV}} \ge 1 - \left(\frac{9}{4^{c-1}}\right)^{1/3} \ge 1 - 4^{1-c/3}\,, \quad c \ge 3.$$

$\square$

3 Upper bound

3.1 Upper bounds and representation theory

Our basic method for obtaining upper bounds on the mixing times of our processes is to employ the techniques developed by Diaconis and Shahshahani (1981) for analysing random walks on groups. We briefly recall the main details. Given a finite group G, a (complex) representation ρ of G is a group homomorphism $\rho : G \to \mathrm{GL}_n(\mathbb{C})$, where $\mathrm{GL}_n(\mathbb{C})$ denotes the group of $n \times n$ invertible complex matrices. We call the number n the degree of ρ, denoted $\deg(\rho)$, and we call the representation irreducible (or simple) if it cannot be decomposed into a direct sum of two representations of smaller degree. Up to isomorphism, there are only finitely many such irreducible representations, and these include the trivial representation of degree 1 which sends every element of G to the complex number 1.

Given a probability P on G and a representation ρ, we can form the Fourier transform $\hat{P}(\rho)$ of P at ρ by setting

$$\hat{P}(\rho) := \sum_{g \in G} P(g)\rho(g),$$

so $\hat{P}(\rho)$ is an $n \times n$ matrix, where $n = \deg(\rho)$. One of the most attractive features of this Fourier transform is that it is well-behaved with respect to convolution, in that $\widehat{(P * Q)}(\rho) = \hat{P}(\rho)\hat{Q}(\rho)$ for any probabilities P and Q on G and any representation ρ. The following Upper Bound Lemma (Diaconis, 1988) allows one to compute an explicit upper bound for the total variation distance between a probability Q on G and the uniform distribution π on G. Since the Fourier transform behaves well with respect to convolution, this lemma provides a practical tool for bounding the mixing time of a random walk on a group.

Lemma 6. *Given a probability Q on a finite group G, we have*

$$\|Q - \pi\|_{\mathrm{TV}}^2 \leq \frac{1}{4} \sum \deg(\rho) \mathrm{tr}(\hat{Q}(\rho)\hat{Q}(\rho)^*),$$

where $A^* = (\overline{a_{ji}})$ denotes the complex conjugate transpose of the matrix $A = (a_{ij})$, tr *denotes the trace function on square matrices, and the sum is taken over all non-trivial irreducible representations ρ of G.*

3.2 Application to our walks

Our initial walk X involves both the additive and multiplicative structure of the ring $\mathbb{Z}_n$, and the measure giving the distribution of X_k cannot conveniently be expressed as the convolution of measures. This is the main reason we introduce the subsampled walk Y; although Y is not strictly a random walk on the additive group $(\mathbb{Z}_n, +)$, the measure giving the distribution of Y_k *can* be expressed as the convolution of measures, and the techniques described in the previous section apply. Here the representation theory is particularly straightforward: there are precisely n irreducible representations $\rho_0, \rho_1, \ldots, \rho_{n-1}$, they all have degree 1, and they are completely determined by the following equations

$$\rho_s(1) := e^{i\frac{2\pi}{n}s} \text{ for } 0 \leq s \leq n-1$$

(note that ρ_0 is the trivial representation). Therefore, for any probability Q on G and for any $0 \leq s \leq n-1$, $\hat{Q}(\rho_s)$ is just a complex number, $\hat{Q}(\rho_s)^*$ is just the complex conjugate of $\hat{Q}(\rho_s)$, and hence $\mathrm{tr}(\hat{Q}(\rho_s)\hat{Q}(\rho_s)^*) = |\hat{Q}(\rho_s)|^2$. The Upper Bound Lemma becomes

$$\|Q - \pi\|_{\mathrm{TV}}^2 \leq \frac{1}{4} \sum_{s=1}^{n-1} |\hat{Q}(\rho_s)|^2. \tag{8}$$

Recall from (2) that (with $Y_0 = 0$), $Y_k = \sum_{j=1}^{k} 2^j S'_{k+1-j} \mod n$. The measure P_k giving the distribution of Y_k is the convolution of the measures λ_j given by $\lambda_j(2^j a \mod n) = \mathbb{P}(S_1 = a)$ for every j, a, so we begin by calculating the Fourier transforms of the λ_j. To ease notation, for each $1 \leq j \leq k$ and $0 \leq s \leq n-1$, set

$$\omega_{sj} = \rho_s(2^j) = e^{i\frac{2\pi}{n}2^j s}$$

and note that for any j, s we have $\omega_{sj}^n = 1$. Then for each $0 \le s \le n-1$,

$$\begin{aligned}\hat{\lambda}_j(\rho_s) &= \sum_{a=0}^{n-1} \omega_{sj}^a \, \mathbb{P}\left(S_1 = a\right) = \sum_{a=0}^{n-1} \omega_{sj}^a \sum_{d \in \mathbb{Z}} \mathbb{P}\left(S_1' = a + dn\right) \\ &= \sum_{d \in \mathbb{Z}} \sum_{a=0}^{n-1} \omega_{sj}^{a+dn} \, \mathbb{P}\left(S_1' = a + dn\right) = \sum_{a \in \mathbb{Z}} \omega_{sj}^a \, \mathbb{P}\left(S_1' = a\right) = G_{S'}(\omega_{sj}) \,,\end{aligned}$$

where $G_{S'}$ is the probability generating function (PGF) of S'. It follows from its definition in (3) as a random sum of random step lengths that this satisfies

$$G_{S'}(\omega_{sj}) = \frac{p_n}{1 - (1 - p_n) G_\xi(\omega_{sj})} \,, \tag{9}$$

where G_ξ is the PGF of ξ.

When we substitute into the Upper Bound Lemma 6, we are interested in the modulus squared of such expressions, by Equation (8). The modulus of the top line squared is p_n^2, and the modulus of the bottom line squared is

$$\begin{aligned}&(1 - (1 - p_n) G_\xi(\omega_{sj}))(1 - (1 - p_n)\overline{G_\xi(\omega_{sj})}) \\ &\qquad = 1 - (1 - p_n)\left(G_\xi(\omega_{sj}) + \overline{G_\xi(\omega_{sj})}\right) + (1 - p_n)^2 G_\xi(\omega_{sj}) \overline{G_\xi(\omega_{sj})} \\ &\qquad = 1 - 2(1 - p_n) \mathrm{Re}\left(G_\xi(\omega_{sj})\right) + (1 - p_n)^2 \left|G_\xi(\omega_{sj})\right|^2 \,.\end{aligned}$$

Combining all of the above leads to the following upper bound for the total variation distance at time k:

$$\| \, \mathbb{P}\left(Y_k \in \cdot\right) - \pi(\cdot) \|_{\mathrm{TV}}^2 \le \frac{1}{4} \sum_{s=1}^{n-1} \prod_{j=1}^{k} \frac{p_n^2}{1 - 2(1 - p_n) \mathrm{Re}\left(G_\xi(\omega_{sj})\right) + (1 - p_n)^2 \left|G_\xi(\omega_{sj})\right|^2} \,. \tag{10}$$

3.3 Strategy for analysing the upper bound

In order to establish a cutoff for Y, we need to control the right hand side of (10) around time $T_n = \log_2(n/\sigma_{S'})$. To that end, we define for $c \in \mathbb{N}$ a function $U_n(c)$ by

$$U_n(c) = \sum_{s=1}^{n-1} \prod_{j=1}^{T_n + c} \phi_n(s, j) \tag{11}$$

where

$$\phi_n(s, j) := \frac{p_n^2}{1 - 2(1 - p_n) \mathrm{Re}\left(G_\xi(\omega_{sj})\right) + (1 - p_n)^2 \left|G_\xi(\omega_{sj})\right|^2} \in (0, 1] \,, \tag{12}$$

and note that our cutoff will be proved if we can show that (for odd n) $\limsup_{n \to \infty} U_n(c) \le U(c)$ for some function U satisfying $U(c) \to 0$ as $c \to \infty$.

Our strategy for bounding $U_n(c)$ involves identifying for each $1 \le s \le n-1$ enough values j for which $\phi_n(s, j)$ is sufficiently small to provide a useful upper bound. In order to do this, it is convenient to first reparametrise, so we let Z_n be a random variable uniformly distributed on the set $\{s/n : s = 1, \ldots, n-1\} \subset [0, 1]$. Then we may write

$$U_n(c) = \mathbb{E}\left[f_n(Z_n, T_n + c)\right] , \quad \text{where } f_n(x, t) := (n - 1) \prod_{j=1}^{t} \phi_n(nx, j) \,. \tag{13}$$

The second step is to split the analysis of the function f_n into two stages by splitting the range of x into two pieces. In order to do this, let L be an integer satisfying $2\alpha L > 1$, and let b be an integer satisfying $B \subseteq [-2^b, 2^b]$, where (recall that) B is the support of ξ. We define a finite lattice $\mathcal{L}$ of points in $[0,1]$ by

$$\mathcal{L} = \left\{ \frac{k}{2^{L+b}} \;:\; k = 0, \ldots, 2^{L+b} \right\} .$$

Now choose some $\varepsilon \in (0, 1/(2^{L+b}))$, and define the set $\mathcal{L}_\varepsilon$ to be the intersection of $[0,1]$ with

$$\bigcup_{x \in \mathcal{L}} \left[x - \frac{\varepsilon}{2}, x + \frac{\varepsilon}{2} \right] .$$

Importantly, $\mathcal{L}_\varepsilon$ depends only on α, B and ε, but not on n. We now proceed to bound $f_n(x, T_n + c)$ by considering in turn the cases where x does and does not belong to the set $\mathcal{L}_\varepsilon$.

3.4 Controlling f_n for $x \notin \mathcal{L}_\varepsilon$

For $x \notin \mathcal{L}_\varepsilon$ we see that $2\pi 2^j ax \neq 0 \mod 2\pi$ for any $j = 1, 2, \ldots, L$ and $a \in B$. Thus $\cos(2\pi 2^j ax)$ is bounded away from 1 for all such x and j, and we can write

$$\mathrm{Re}\left(G_\xi(e^{i2\pi 2^j x}) \right) = \sum_{a=0}^{2^b} \mathbb{P}\left(|\xi| = a \right) \cos\left(2\pi 2^j ax \right) \leq 1 - \kappa(x) ,$$

for all $j = 1, \ldots, L$, where $\kappa(x)$ is strictly positive.

Substituting this into the expression for ϕ_n in (12), and lower-bounding the modulus squared of a complex number by the square of its real part, we obtain:

$$\begin{aligned}
\phi_n(nx, j) &\leq \frac{p_n^2}{1 - 2(1-p_n)\mathrm{Re}\left(G_\xi(e^{i2\pi 2^j x}) \right) + (1-p_n)^2 \left| G_\xi(e^{i2\pi 2^j x}) \right|^2} \\
&\leq \frac{p_n^2}{1 - 2(1-p_n)\mathrm{Re}\left(G_\xi(e^{i2\pi 2^j x}) \right) + (1-p_n)^2 \mathrm{Re}\left(G_\xi(e^{i2\pi 2^j x}) \right)^2} \\
&= \left(\frac{p_n}{1 - (1-p_n)\mathrm{Re}\left(G_\xi(e^{i2\pi 2^j x}) \right)} \right)^2 \\
&\leq \left(\frac{p_n}{1 - (1-p_n)(1-\kappa(x))} \right)^2 = O(p_n^2) .
\end{aligned}$$

Since $\phi_n(nx, j) \in (0, 1]$, it follows that for $x \notin \mathcal{L}_\varepsilon$,

$$f_n(x, T_n + c) = (n-1) \prod_{j=1}^{T_n + c} \phi_n(nx, j) \leq (n-1) \prod_{j=1}^{L} \phi_n(nx, j) \leq O(n^{1 - 2\alpha L}) .$$

Thanks to our choice of $L > 1/2\alpha$ we can now use Fatou's Lemma to deduce that

$$\limsup_{n \to \infty} \mathbb{E}\left[f_n(Z_n, T_n + c); Z_n \notin \mathcal{L}_\varepsilon \right] = 0 . \tag{14}$$

3.5 Controlling f_n for $x \in \mathcal{L}_\varepsilon$

It remains to deal with $\mathbb{E}\left[f_n(Z_n, T_n + c); Z_n \in \mathcal{L}_\varepsilon\right]$. We begin by writing (for any $t \in \mathbb{N}$)

$$\begin{aligned}\mathbb{E}\left[f_n(Z_n, t); Z_n \in \mathcal{L}_\varepsilon\right] = \frac{1}{n-1} \sum_{k=1}^{2^{L+b}-1} \sum_{r \geq 1} f_n\left(\tfrac{r}{n}, t\right) \mathbf{1}_{\left[\left|\frac{k}{2^{L+b}} - \frac{r}{n}\right| \leq \frac{\varepsilon}{2}\right]} \\ + \frac{1}{n-1} \sum_{r \geq 1} \left(f_n\left(\tfrac{r}{n}, t\right) + f_n\left(1 - \tfrac{r}{n}, t\right)\right) \mathbf{1}_{\left[\frac{r}{n} \leq \frac{\varepsilon}{2}\right]}, \qquad (15)\end{aligned}$$

where the last sum deals with the two end intervals in $\mathcal{L}_\varepsilon$. Now (15) can be bounded as follows:

$$\mathbb{E}\left[f_n(Z_n, t); Z_n \in \mathcal{L}_\varepsilon\right] \leq \frac{1}{n-1} \sum_{k=1}^{2^{L+b}-1} \sum_{r=-\infty}^{\infty} f_n\left(\frac{(n-1)\frac{k}{2^{L+b}} - r}{n}, t\right) + \frac{2}{n-1} \sum_{r=1}^{\infty} f_n\left(\tfrac{r}{n}, t\right), \qquad (16)$$

where we have used the symmetry of the functions f_n at either end of the interval $[0, 1]$ to rewrite the expression for the end intervals. Now replace t by $T_n + c$, and consider the function f_n in the double sum above:

$$\begin{aligned}f_n\left(\frac{(n-1)\frac{k}{2^{L+b}} - r}{n}, T_n + c\right) &= (n-1) \prod_{j=1}^{T_n+c} \phi_n\left((n-1)\tfrac{k}{2^{L+b}} - r, j\right) \\ &\leq (n-1)\phi_n\left((n-1)\tfrac{k}{2^{L+b}} - r, T_n + c\right). \qquad (17)\end{aligned}$$

Here we have bounded the product by a single term, once again making use of the fact that ϕ_n takes values in $(0, 1]$. Since $\phi_n(s, j)$ involves s and j only through the function $G_\xi(\omega_{sj})$, where $\omega_{sj} = \exp(2\pi i 2^j s/n)$, we have (for sufficiently large n) that the bound in (17) is a function of

$$\begin{aligned}\exp\left(2\pi i \frac{2^{T_n+c}}{n}\left((n-1)\frac{k}{2^{L+b}} - r\right)\right) &= \exp\left(2\pi i \frac{2^{T_n+c}}{n}(k 2^{-(L+b)} + r)\right) \\ &= \exp\left(\frac{2^{1+c}\pi i (k 2^{-(L+b)} + r)}{\sigma_{S'}}\right).\end{aligned}$$

The second equality simply uses the definition of T_n, while the first results from shifting the argument of the exponential function by $2\pi i k 2^{T_n+c-(L+b)}$. (For large enough n this is an integer multiple of $2\pi i$, thanks to the finiteness of L and b and the assumption that $T_n \to \infty$.)

Writing

$$\theta_{krc} = \frac{2^{1+c}\pi(k 2^{-(L+b)} + r)}{\sigma_{S'}},$$

we therefore need to upper bound the function

$$\phi_n\left((n-1)\tfrac{k}{2^{L+b}} - r, T_n + c\right) = \frac{p_n^2}{1 - 2(1-p_n)\mathrm{Re}\left(G_\xi(e^{i\theta_{krc}})\right) + (1-p_n)^2 \left|G_\xi(e^{i\theta_{krc}})\right|^2}.$$

Now note that

$$G_\xi\left(e^{i\theta_{krc}}\right) = \sum_{a \in B} \mathbb{P}\left(\xi = a\right) e^{ia\theta_{krc}}, \qquad \text{and thus} \qquad \mathrm{Re}\left(G_\xi\left(e^{i\theta_{krc}}\right)\right) = \mathbb{E}\left[\cos\left(\xi\theta_{krc}\right)\right].$$

Similarly,

$$\left|G_\xi(e^{i\theta_{krc}})\right|^2 = \mathbb{E}\left[\cos\left(\xi\theta_{krc}\right)\right]^2 + \mathbb{E}\left[\sin\left(\xi\theta_{krc}\right)\right]^2 .$$

Since $p_n \to 0$ as $n \to \infty$, we see from (5) that $\sigma_{S'} \to \infty$ and thus $\theta_{krc} \to 0$. Using the Taylor expansions of cosine and sine the above can be approximated by

$$\begin{aligned}
\mathbb{E}\left[\cos\left(\xi\theta_{krc}\right)\right] &= 1 - \frac{(\mu^2+\sigma^2)\theta_{krc}^2}{2} + O(\theta_{krc}^4)\,;\\
\mathbb{E}\left[\sin\left(\xi\theta_{krc}\right)\right] &= \mu\theta_{krc} + O(\theta_{krc}^3)\,.
\end{aligned}$$

Neglecting terms of $O(\theta_{krc}^3)$ we arrive at

$$\begin{aligned}
\phi_n\left((n-1)\tfrac{k}{2^{L+b}} - r, T_n + c\right) &\sim \frac{p_n^2}{1-(1-p_n)\left[2-(\mu^2+\sigma^2)\theta_{krc}^2\right] + (1-p_n)^2\left[1-(\mu^2+\sigma^2)\theta_{krc}^2 + \mu^2\theta_{krc}^2\right]}\\
&= \frac{p_n^2}{p_n^2 + (1-p_n)(\mu^2+\sigma^2 p_n)\theta_{krc}^2}\\
&= \frac{1}{1+\sigma_{S'}^2\theta_{krc}^2} = \frac{1}{1+4^{1+c}\pi^2(k2^{-(L+b)}+r)^2}\,.
\end{aligned}$$

We now combine this bound with that in (17) and insert into (16) (using an identical argument for the second sum there):

$$\mathbb{E}\left[f_n(Z_n,t); Z_n \in \mathcal{L}_\varepsilon\right] \le \sum_{k=1}^{2^{L+b}-1}\sum_{r=-\infty}^{\infty}\frac{1}{1+4^{1+c}\pi^2(k2^{-(L+b)}+r)^2} + 2\sum_{r=1}^{\infty}\frac{1}{1+4^{1+c}\pi^2 r^2}\,.$$

Summing over r this bound becomes

$$\mathbb{E}\left[f_n(Z_n,T_n+c); Z_n \in \mathcal{L}_\varepsilon\right] \le \sum_{k=1}^{2^{L+b}-1}\frac{2^{-(1+c)}\sinh(2^{-c})}{\cosh(2^{-c})-\cos(2^{1-(L+b)}k\pi)} + \Big(2^{-(1+c)}\coth(2^{-(1+c)})-1\Big)\,.$$

Finally, taking the worst possible value $k=1$ for *every* term in the sum, we get

$$\begin{aligned}
\mathbb{E}\left[f_n(Z_n,T_n+c); Z_n \in \mathcal{L}_\varepsilon\right] &\le \frac{2^{L+b-c}\sinh(2^{-c})}{\cosh(2^{-c})-\cos(2^{1-(L+b)}\pi)} + \Big(2^{-(1+c)}\coth(2^{-(1+c)})-1\Big)\\
&\sim \frac{2^{L+b-2c}}{1-\cos(2^{1-(L+b)}\pi)} + \frac{4^{1-c}}{3} \quad \text{as } c\to\infty.
\end{aligned} \tag{18}$$

Combining (14) and (18) yields the required result

$$\limsup_{n\to\infty} U_n(c) = \limsup_{n\to\infty}\mathbb{E}\left[f_n(Z_n,T_n+c); Z_n\in\mathcal{L}_\varepsilon\right] \le O(4^{-c}) \quad \text{as } c\to\infty,$$

and this completes the proof of Theorem 3.

4 Concluding remarks

We have shown that the process X exhibits a cutoff phenomenon when the probability p_n of jumping takes the form $p_n = 1/(2n^\alpha)$, for a range of α which depends upon the mean of our step distribution ξ ($\alpha \in (0,2)$ when $\mu = 0$, and $\alpha \in (0,1)$ otherwise). We have not yet said anything about the mixing time when α takes values on the boundary of these intervals, however.

If $\alpha = 0$ then our argument for upper bounding the mixing time breaks down (since a sufficiently fine lattice $\mathcal{L}$ does not exist). In this situation Lemma 5 is still applicable, showing that Y has not mixed by time $\log_2 n$, and an upper bound of $O(\log_2 n \log_2 \log_2 n)$ can be obtained by employing the method of Chung et al. (1987). On the other hand, if α takes the value at the upper boundary of the relevant interval then $n/\sigma_{S'} = O(1)$, and thus T_n is asymptotically independent of n: in this case our upper bound analysis still holds, and we see that Y mixes in constant time.

It is of course possible to generalise the process considered in this paper in a number of ways. Changing the form of p_n to $1/(cn^\alpha)$ for some constant $c > 1$ has no significant effect on the mixing time of Y, and so X will exhibit a cutoff at time $cn^\alpha \log_2(n/\sigma_{S'})$. Similarly, changing the transitions of X so that jumps involve multiplying by some (fixed) $k \geq 2$ (and considering only those n for which the resulting process still has a uniform equilibrium distribution) presumably just has the effect of changing the cutoff time for Y to $\log_k(n/\sigma_{S'})$. More interesting would be an analysis of a process X for which the multiplication factor is not deterministic, and for which the resulting subsampled chain Y does not have a distribution given by simple convolution; for example where $X_t = a_t X_{t-1}$ with probability p_n, with a_t being uniformly chosen from $\{2, (n+1)/2\}$.

Acknowledgements

Some of the ideas in this work arose during investigations into random walks on $\mathbb{Z}_n$ by Sam Wright, who was supported by Nuffield Science Undergraduate Research Bursary URB/40605. The authors would also like to express their gratitude to John Payne, whose numerical calculations for a particular instance of our process provided some useful early insights into the behaviour of the upper bound in Section 3.

References

Chung, F., P. Diaconis, and R. Graham (1987). Random walks arising in random number generation. *The Annals of Probability 15*(3), 1148–1165.

Connor, S. B. (2010). Separation and coupling cutoffs for tuples of independent Markov processes. *Latin American Journal of Probability and Mathematical Statistics 7*, 65–77.

Diaconis, P. (1988). *Group representations in probability and statistics*, Volume 11 of *Lecture Notes - Monograph Series*. Institute of Mathematical Statistics.

Diaconis, P. (2011). The Mathematics of Mixing Things Up. *Journal of Statistical Physics 144*(3), 445–458.

Diaconis, P. and M. Shahshahani (1981). Generating a random permutation with random transpositions. *Z. Wahrsch. Verw. Gebiete 57*, 159–179.

Hildebrand, M. (1992). *Rates of convergence of some random processes on finite groups*. Ph. D. thesis, Dept. Mathematics, Harvard University.

Hildebrand, M. (1993). Random processes of the form $X_{n+1} = a_n X_n + b_n$ $(mod\ p)$. *Annals of Probability 21*(2), 710–720.

Hildebrand, M. (1994a). Random walks supported on random points of $\mathbb{Z}/n\mathbb{Z}$. *Probab. Theory Related Fields 100*, 191–203.

Hildebrand, M. (1994b). *Some random processes related to affine random walks*. IMA Preprint Series, 1210.

Levin, D. A., Y. Peres, and E. L. Wilmer (2009). *Markov chains and mixing times*. American Mathematical Soc.

Saloff-Coste, L. (2004). Random walks on finite groups. In *Probability on discrete structures*, Volume 110 of *Encyclopaedia Math. Sci.*, pp. 263–346. Springer.

Terras, A. (1999). *Fourier Analysis on Finite Groups and Applications*. London Mathematical Society Student Texts. Cambridge University Press.

Department of Mathematics
University of York
York, YO10 5DD
UK