Mixing Points on a Circle

Dana Randall[1⋆] and Peter Winkler[2⋆⋆]

[1] College of Computing, Georgia Institute of Technology, Atlanta GA 30332-0280
[2] Dept. of Mathematics, Dartmouth College, Hanover NH 03755-3551

Abstract. We determine, up to a log factor, the mixing time of a Markov chain whose state space consists of the successive distances between n labeled "dots" on a circle, in which one dot is selected uniformly at random and moved to a uniformly random point between its two neighbors. The method involves novel use of auxiliary discrete Markov chains to keep track of a vector of quadratic parameters.

1 Introduction

Randomized approximation algorithms using Markov chains have proven to be powerful tools with an astounding variety of applications. Most notably, when the state space consists of configurations of a physical system, random sampling provides a method for estimating the thermodynamic properties of the model. Showing that these chains converge quickly to their stationary distributions is typically the vital step in establishing the efficiency of these algorithms.

Over the last 15 years there has been substantial progress in developing methods for bounding the mixing rates of finite Markov chains, including coupling, canonical paths, and various isoperimetric inequalities (see, e.g., [4], [6], and [11] for surveys). As a consequence, there are now provably efficient algorithms for many discrete sampling problems, such as matchings, independents sets and colorings. However, there has been a scarcity of results for problems defined on continuous spaces, mostly because methods used to analyze discrete chains seem to break down in the continuous setting. This is likely an artifact of the appearance of the "minimum stationary probability" parameter which often arises in the discrete cases, and not anything inherent in continuous chains, which often seem to be efficient. Thus, an obvious challenge to the mixing community is to develop new tools amenable to Markov chains on continuous spaces.

For example, the hard-core lattice gas model from statistical physics examines the possible configurations of n balls of radius r in a finite region R of $\mathbf{R}^n$, where balls are required to lie in non-overlapping positions. Although the centers of the balls can occupy any real point in $\mathbf{R}^n$, the model is typically discretized so that they lie on lattice points in $\Lambda_n \subseteq \mathbf{Z}^n$. In this setting, if r is chosen to be slightly more than $1/2$, then configurations correspond to independent sets on the lattice

⋆ Supported in part by NSF grant CCR-0105639.
⋆⋆ Completed during a term as Clay Senior Fellow at the Mathematical Sciences Research Institute.

graph. While much analysis has focused on rigorous means of efficiently sampling discrete models such as independent sets in Λ_n (e.g., [3], [8]) very little is known about sampling in the continuous setting.

One powerful tool for sampling points in a continuous state space is a walk used to estimate the volume of a convex body in $\mathbf{R}^n$ [2]. Notice that this chain provides a way to sample sets consisting of n real points on the unit interval because these configurations can be mapped to points in the unit simplex. However, the moves arising from walks in the simplex are less natural in the context of independent sets since they move all the points in the independent set in a single move.

Alternative chains for sampling lattice gas configurations were considered in [5]. The first variant is a single particle, global chain that connects all pairs of states that differ by the position of one point. It is shown that the chain is rapidly mixing in $\mathbf{R}^n$ if the density of points is sufficiently low. The second variant examined is a single particle, local chain where steps involve moving one point to a new position in its current neighborhood. Although it is believed that this chain mixes rapidly in many settings, it was only shown that a discretized version of the chain mixes rapidly in one dimension. In this chain, the unit interval is subdivided into m discrete points and at each step one of the n particles is chosen and moved to an adjacent point if that point is currently unoccupied. The mixing time for this chain was shown to be $O(n^3m^2)$.

In [12], the authors of this paper previously considered a natural chain whose stationary distribution is the set of configurations of n dots on the real interval $[0,1]$ and showed that the mixing time is $\Theta(n^3 \log n)$. The analysis was based on a "two phase coupling": the first phase brings the dots of one chain close to the corresponding dots of the other; the second aligns them exactly.

Here, we tackle another simple and natural chain whose stationary state is uniform on the unit n-simplex. A state of the chain consists of the successive distances (along the circle) between n sequentially-numbered points on a unit-circumference circle. Moves consist of moving a random such point to a spot on the circle chosen uniformly at random between its two neighbors. We formalize the moves in the next section. The mixing time τ of the chain is then shown to satisfy

$$cn^3 \leq \tau \leq c'n^3 \log n,$$

for constants c, c'.

This chain seems superficially similar to the chain "n dots on an interval." However, the methods used there required a lattice structure for the state space and an eigenfunction for the chain, neither of which is available here. Instead we use an entirely different and apparently novel approach involving auxiliary discrete Markov chains, borrowing only the two phase coupling from [12].

2 n Dots on a Circle

To avoid confusion with other points on the circle, the n designated points defining the chain will be called *dots* and the chain itself will be called "n dots on a circle." The circle will be assumed to have circumference 1 and its points are permanently identified with points in the half-open interval $[0,1)$, modulo 1.

The *configuration* $\bar{x} = (x_1, \ldots, x_n)$ of chain $\mathbf{X}$ at any particular time is the vector of locations of dots 1 through n. The dots are assumed to be labeled clockwise around the circle.

A step of the chain is determined by a uniformly random integer k between 1 and n, and an independent real λ chosen from the Lebesgue distribution on $[0,1]$. If the previous configuration of the chain was $\bar{x}$, the new configuration will be

$$\bar{x}' = (x_1, \ldots, x_{k-1}, (1-\lambda)x_{k-1} + \lambda x_{k+1}, x_{k+1}, \ldots, x_n)$$

where the subscripts are taken modulo n.

For the *state* of $\mathbf{X}$ we take not $\bar{x}$, but the difference vector $\bar{u} = (u_1, \ldots, u_n)$ with $u_i := x_{i+1} - x_i \in [0,1)$. One reason for this reduction is that otherwise, the mixing time of the chain is dominated by the less interesting issue of the location of the dots' "center of gravity" along the circumference. It is not difficult to show that this point executes a driftless random walk around the cycle which takes time $\Theta(n^4)$ to mix.

However, it is convenient for us, at times, to consider the unreduced chain $\mathbf{X}^+$ whose state is the configuration $\bar{x}$.

Theorem 1. *The stationary state of the unreduced "n dots on a circle" chain $\mathbf{X}^+$ is uniform on $[0,1)^n$ subject to clockwise labeling modulo 1.*

Proof. This follows from detail balance (or simply noticing that the position of x_k in a uniform point is indeed uniformly random between x_{k-1} and x_{k+1}).

It is convenient here to define the *mixing time* of a Markov chain $\mathbf{X}$ to be the least number t of steps such that, beginning at any state, the total variation distance $\frac{1}{2}\|\mathbf{X}(t) - \sigma\|_1$ between the state of $\mathbf{X}$ at time t and the stationary distribution σ of $\mathbf{X}$ is less than $\frac{1}{4}$.

Theorem 2. *The mixing time of the reduced "n dots on an circle" chain $\mathbf{X}$ is $O(n^3 \log n)$.*

Proof. For the upper bound, it suffices to exhibit a coupling which will achieve coalescence of any chain $\mathbf{X}$ with a stationary chain $\mathbf{Y}$ in time $\mathrm{O}(n^3 \log n)$, with probability at least $\frac{3}{4}$. Our coupling proceeds in two phases, first bringing the dots of one chain close to the dots of the other, then aligning them exactly.

Our second ("exact alignment") phase is much like the corresponding phase in "n dots on a line." The first phase employs the familiar linear coupling, but our argument will rest on quite a different eigenfunction. Let $\bar{z} := \bar{x} - \bar{y}$, where $\bar{y}$ is the (current) configuration of the $\mathbf{Y}$ chain. Coupling will be achieved exactly when $\bar{z} = c\mathbf{1}$ for some $c \in [0,1)$, i.e. when $z_1 = z_2 = \cdots = z_n$.

The first-stage "linear" coupling of $\mathbf{X}$ and $\mathbf{Y}$ is achieved simply by using the same random k and λ at each step.

Let $s_0 := 3\sum_{i=1}^n z_i^2$, $s_1 = \sum_{i=1}^n z_i z_{i-1}$, $s_m := 2\sum_{i=1}^n z_i z_{i-m}$ for $m = 2, \ldots, q-1$ where $q = \lfloor n/2 \rfloor$, and $s_q := \beta \sum_{i=1}^n z_i z_{i-q}$, where $\beta = 2$ when n is odd but only 1 when n is even. Let us observe what becomes of $\bar{s} := (s_0, \dots, s_q)$, in expectation, as a step of the chain is taken.

In case dot k is chosen, z_k will fall uniformly between z_{k-1} and z_{k+1}, hence

$$\mathrm{E}z_k^2 = \frac{1}{z_{k+1} - z_{k-1}} \int_{z_{k-1}}^{z_{k+1}} z^2 dz = \frac{1}{3}\left(z_{k-1}^2 + z_{k-1}z_{k+1} + z_{k+1}^2\right) .$$

It follows that over all choices of k (using primes to denote future values),

$$\mathrm{E}s_0' = \frac{n-1}{n}s_0 + 3\frac{1}{3n}\left(\frac{s_0}{3} + \frac{s_2}{2} + \frac{s_0}{3}\right) = \left(1 - \frac{1}{3n}\right) s_0 + \frac{1}{2n}s_2 .$$

Similarly,

$$\mathrm{E}s_1' = \left(1 - \frac{2}{n}\right) s_1 + \frac{1}{n}\frac{s_0}{3} + \frac{1}{n}\frac{s_2}{2} = \frac{1}{3n}s_0 + \left(1 - \frac{2}{n}\right) s_1 + \frac{1}{2n}s_2 ,$$

$$\mathrm{E}s_2' = \left(1 - \frac{2}{n}\right) s_2 + 2\frac{1}{n}s_1 + 2\frac{1}{n}\frac{s_3}{2} = \frac{2}{n}s_1 + \left(1 - \frac{2}{n}\right) s_2 + \frac{1}{n}s_3 ,$$

and for $2 < m < q-1$,

$$\mathrm{E}s_m' = \frac{1}{n}s_{m-1} + \left(1 - \frac{2}{n}\right) s_m + \frac{1}{n}s_{m+1} .$$

Finally,

$$\mathrm{E}s_{q-1}' = \frac{1}{n}s_{q-2} + \left(1 - \frac{2}{n}\right) s_{q-1} + \frac{2}{\beta n}s_q$$

and

$$\mathrm{E}s_q' = \frac{1}{n}s_{q-1} + \left(1 - \frac{2}{\beta n}\right) s_q .$$

Let us now define a completely separate, discrete Markov chain on the state space $\{0, 1, \dots, q\}$, which behaves according to the following transition matrix **D**:

$$\begin{pmatrix}
1 - \frac{1}{3n} & \frac{1}{3n} & 0 & 0 & 0 & \cdot & \cdot & \cdot & 0 \\
0 & 1 - \frac{2}{n} & \frac{2}{n} & 0 & 0 & \cdot & \cdot & \cdot & 0 \\
\frac{1}{2n} & \frac{1}{2n} & 1 - \frac{2}{n} & \frac{1}{n} & 0 & \cdot & \cdot & \cdot & 0 \\
0 & 0 & \frac{1}{n} & 1 - \frac{2}{n} & \frac{1}{n} & 0 & \cdot & \cdot & 0 \\
0 & 0 & 0 & \frac{1}{n} & 1 - \frac{2}{n} & \frac{1}{n} & 0 & \cdot & 0 \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{n} & 1 - \frac{2}{n} & \frac{1}{n} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{2}{\beta n} & 1 - \frac{2}{\beta n}
\end{pmatrix}$$

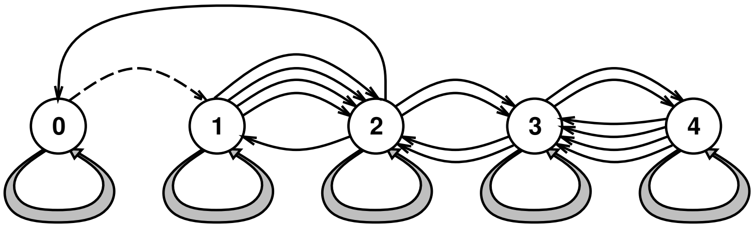

Fig. 1. The chain **D** for $n = 8$. The dotted arrow represents a transition of probability $1/3n$, other thin arrows $1/2n$.

See Fig. 1 for a schematic of the state space of **D**.

We see that **D** is designed so that $\bar{s} \cdot \mathbf{D} = \mathrm{E}\bar{s}'$. The stationary distribution of **D** is

$$\bar{s}(\infty) := \frac{1}{n+1}(3, 1, 2, 2, 2, \ldots, 2, 2, \beta) \ .$$

If at time 0 the chains **X** and **Y** yield a vector $\bar{s}(0)$ whose coordinates sum to ρ, then after t steps we have $\mathrm{E}s = \bar{s}(0) \cdot \mathbf{D}^t \to \frac{\rho}{n+1}(3, 1, 2, 2, 2, \ldots, 2, 2, \beta)$ and in particular $f(t) := s_0(t)/3 - s_1(t) \to 0$. But $f(t) = \frac{1}{2}\sum_{i=1}^{n}(z_{i+1}(t) - z_i(t))^2$ so the z_i's are becoming equal; thus, we see that the mixing time of **X** is intimately tied to the mixing time of **D**.

Lemma 1. *The discrete chain* **D** *mixes in time* $O(n^3)$.

Proof. We employ a simple coupling on the state space. The coupling is defined so that pairs of points $(i, j) \in \{0, ..., q\}^2$ are updated simultaneously, ensuring that the two points never move in opposite directions. We now define a distance function on pairs of points by letting $d(i, j) = |i - j|$ if $\min(i, j) \geq 2$ or if $\max(i, j) \leq 1$ and letting $d(i, j) = |i - j| - 1/2$ otherwise. An upper bound on this distance function is $B = q$. For this distance function and the above coupling, it is easy to see that for any $(i, j) \in \{0, ..., q\}^2$, the distance between i and j will be non-increasing in expectation under moves of the coupled chain. Moreover, $\mathrm{E}(\Delta d(i, j))^2) \geq V = 1/n$, whenever $i \neq j$. Applying the coupling theorem 4.1 from [7], we find that the coupling time $T \leq d(i_0, j_0)(2B - d(i_0, j_0))/V = O(n^3)$, and consequently the mixing time is $O(n^3)$.

Observing that $\rho < n^2$, we deduce from Lemma 1 that

$$\|\bar{s}(0) \cdot \mathbf{D}^t, \bar{s}(\infty)\|_{\mathrm{TV}} < an^3\left(1 - \frac{1}{bn^3}\right)^t$$

for appropriate constants a and b. Thus, after $10bn^3 \log n$ steps,

$$\mathrm{E}f(t) < \|\mathrm{E}\bar{s}(t), \bar{s}(\infty)\|_{\mathrm{TV}} < an^3\left(1 - \frac{1}{(bn^3)}\right)^{10bn^3 \log n} = \mathrm{O}(n^{-8}) \ .$$

Since $f(t)$ is non-negative, we deduce from Markov's inequality that for some constant c, there is a z so that with probability at least .9, $|z_i - z| < cn^{-4}$ for every i.

We may now assume that $\mathbf{Y}$ is shifted so that $z = 0$; the rest of the argument proceeds just as in the "n dots on a line" chain, using the second coupling to get $x_k = y_k$ with high probability when dot k is moved, and running $n(\log n)^2$ steps to ensure with high probability that every k has been called.

To lower bound the mixing time of $\mathbf{X}$, we will use a somewhat different auxiliary chain.

Theorem 3. *The mixing time of the reduced "n dots on an circle" chain $\mathbf{X}$ is $\Omega(n^3)$.*

Proof. We now have only one chain, $\mathbf{X}$, and are directly concerned with its state $\bar{u}$. Recall that $u_i := x_{i+1} - x_i$ is the distance between dots i and $i+1$.

Put $q = \lfloor n/2 \rfloor$ as before, again with $\beta = 2$ for n odd and $\beta = 1$ for n even. Define $r_0 := \sum_{i=1}^n u_i^2$, $r_q = \beta \sum_{i=1}^n u_i u_{i-q}$, and $r_m := 2\sum_{i=1}^n u_i u_{i-m}$ for $0 < m < q$. Note that $\sum_{m=0}^q r_m = (\sum u_i)^2 = 1$.

When dot k is moved, u_{k-1} and u_k are both affected. Each lands uniformly between 0 and $u_{k-1} + u_k$, thus $\mathrm{E}(u'_{k-1})^2 = \mathrm{E}(u'_k)^2 = \frac{1}{3}(u_{k-1} + u_k)^2$ and consequently, over all choices of k,

$$\mathrm{E}r'_0 = (1 - \frac{2}{n})r_0 + 2\frac{1}{3n}(r_0 + 2(r_1/2) + r_0) = (1 - \frac{2}{3n})r_0 + \frac{2}{3n}r_1 \ .$$

Similarly,

$$\mathrm{E}r'_1 = \frac{2}{3n}r_0 + \left(1 - \frac{5}{3n}\right) r_1 + \frac{1}{n}r_2 \ ,$$

for $2 < m < q-1$

$$\mathrm{E}r'_m = \frac{1}{n}r_{m-1} + \left(1 - \frac{2}{n}\right) r_m + \frac{1}{n}r_{m+1} \ .$$

for $1 < m < q-1$,

$$\mathrm{E}r'_{q-1} = \frac{1}{n}r_{q-2} + \left(1 - \frac{2}{n}\right) r_{q-1} + \frac{2}{\beta n}r_q \ ,$$

and

$$\mathrm{E}r'_q = \frac{1}{n}r_{q-1} + \left(1 - \frac{2}{\beta n}\right) r_q \ .$$

We now define another discrete, auxiliary chain on $\{0, 1, \ldots, q\}$ by the following transition matrix $\widehat{\mathbf{D}}$:

$$\begin{pmatrix}
1-\frac{2}{3n} & \frac{2}{3n} & 0 & 0 & 0 & \cdot & \cdot & \cdot & 0 \\
\frac{2}{3n} & 1-\frac{5}{3n} & \frac{1}{n} & 0 & 0 & \cdot & \cdot & \cdot & 0 \\
0 & \frac{1}{n} & 1-\frac{2}{n} & \frac{1}{n} & 0 & \cdot & \cdot & \cdot & 0 \\
0 & 0 & \frac{1}{n} & 1-\frac{2}{n} & \frac{1}{n} & 0 & \cdot & \cdot & 0 \\
0 & 0 & 0 & \frac{1}{n} & 1-\frac{2}{n} & \frac{1}{n} & 0 & \cdot & 0 \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{n} & 1-\frac{2}{n} & \frac{1}{n} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{2}{\beta n} & 1-\frac{2}{\beta n}
\end{pmatrix}$$

See Fig. 2 below, for a schematic of the state space of $\widehat{\mathbf{D}}$.

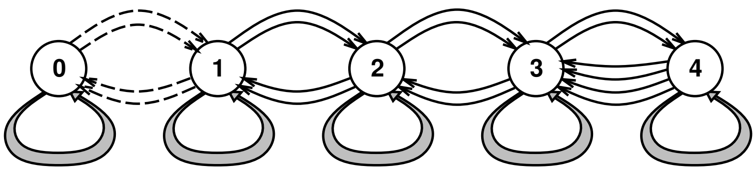

Fig. 2. The chain $\widehat{\mathbf{D}}$ for $n = 8$. Each dotted arrow represents a transition of probability $1/3n$, other thin arrows $1/2n$.

As before, $\widehat{\mathbf{D}}$ has been designed so that $\widehat{\mathbf{D}} \cdot \bar{r} = \mathrm{E}\bar{r}'$. Note that $\widehat{\mathbf{D}}$ is reversible and very close to being a simple random walk on a path. Its stationary distribution is

$$\bar{r}(\infty) := \frac{1}{n+1}(2, 2, 2, 2, 2, \ldots, 2, 2, \beta) \ .$$

and, as in the case of $\mathbf{D}$:

Lemma 2. *The discrete chain* $\widehat{\mathbf{D}}$ *mixes in time* $\Theta(n^3)$.

Proof. For the upper bound, we follow the approach in Lemma 1 and define the same coupling on the state space. This time we use the standard distance function $d(i, j) = |i - j|$ and again find that $E(\Delta d(i, j)) \geq 0$ and $E(\Delta d(i, j))^2 \geq 1/n$. This shows that the mixing time is $O(n^3)$.

To establish the matching lower bound, we turn to the spectral gap of the chain. Lemma 3.1 in [10] states that if $\widehat{\mathbf{D}}$ and $\widetilde{\mathbf{D}}$ are Markov chains on the same state space with the same stationary distribution, and if there are constants c_1 and c_2 such that

$$c_1 \widehat{\mathbf{D}}(i,j) \le \widetilde{\mathbf{D}}(i,j) \le c_2 \widehat{\mathbf{D}}(i,j) \tag{1}$$

for all $i \neq j$, then

$$c_1 \operatorname{Gap}(\widehat{\mathbf{D}}) \le \operatorname{Gap}(\widetilde{\mathbf{D}}) \le c_2 \operatorname{Gap}(\widehat{\mathbf{D}}).$$

Let $\widetilde{D}$ be the standard symmetric random walk with transitions $\widetilde{D}(i, i \pm 1) = 1/n$, and self-loops elsewhere. This well-studied chain has spectral gap $\operatorname{Gap}(\widetilde{\mathbf{D}}) = O(\frac{1}{n^3})$. Comparing the two chains and noting that the condition (1) is satisfied with constants c_1 and c_2, we find that $\operatorname{Gap}(\widehat{\mathbf{D}}) = O(\frac{1}{n^3})$ as well. Using upper bounds on the spectral gap to lower bound the mixing rate (see, e.g., [13]), we can conclude that the chain mixes in time $\Omega(n^3)$.

Now we start the chain $\mathbf{X}$ with all dots coincident, aiming to show that after $t = cn^3$ steps (for some constant c) the total variation distance of its state distribution from the stationary distribution will still exceed $1/4$. To do this we exhibit a state-dependent event whose probability differs from a stationary chain's by more than $1/4$. This event will itself be based on a total variation distance, namely the value $d_t := \|\bar{r}(t), \bar{r}(\infty)\|_{\mathrm{TV}}$.

Let us consider the behavior of the auxiliary chain $\widehat{\mathbf{D}}$ while the chain $\mathbf{X}$ is running. When $\mathbf{X}$ has all dots coincident, the chain $\widehat{\mathbf{D}}$ starts at state $r_0 = (1, 0, 0, \ldots, 0)$ and has total variation distance from $\bar{r}(\infty)$ close to 1. We know from Lemma 2 that for some constant c, $\|\bar{r}(0)\widehat{\mathbf{D}}^{cn^3}, \bar{r}(\infty)\|_{\mathrm{TV}}$ still exceeds, say, 0.6.

Then

$$\mathrm{E} d_t = \mathrm{E}\|\bar{r}(t), \bar{r}(\infty)\|_{\mathrm{TV}} \ge \|\mathrm{E}\bar{r}(t), \bar{r}(\infty)\|_{\mathrm{TV}} = \|\bar{r}(0)\widehat{\mathbf{D}}^t, \bar{r}(\infty)\|_{\mathrm{TV}} > .6$$

for $t \le cn^3$. Since d_t can never exceed 1, it follows that $\Pr(d_t > .2) > .5$. It therefore suffices to show that for the stationary chain $\mathbf{Y}$, $\Pr(d_t > .2) < .25$.

In $\mathbf{Y}$, however, the distribution of $\bar{r} := \bar{r}(t)$ does not depend on t, and is instead determined by a uniformly random vector $\bar{u} = (u_1, \ldots, u_n)$ in the simplex

$$S := \{\bar{u} :\ \bar{u}_1, \ldots, u_n \ge 0,\ \sum_{i=1}^{n} u_i = 1\}\ .$$

We know that $\mathrm{E}\bar{r} = \bar{r}(\infty)$, so we need only that $\bar{r}$ becomes concentrated about its mean as $n \to \infty$. To establish this, we represent $\bar{u}$ by a sequence of independent exponentially distributed random variables $y_1, \ldots, y_n$ with $u_i = y_i/S$ where $S = \sum_{j=1}^{n} y_j$. We now consider the random variables $w_{i,m} := y_i y_{i-m}$, $0 \le m \le q = \lfloor n/2 \rfloor$.

For each fixed $m > 0$, we define a "dependence graph" G_m on $V = \{1, ..., n\}$ by $i \sim j$ iff $|i - j| = m$; since G_m is just a union of cycles, it has an equitable 3-colorable, i.e., there is a partition $V = A_m \cup B_m \cup C_m$ of the vertices into three sets of the same size (give or take one vertex) none of which contains an edge. But then no variable y_i appears in more than one term of the sum

$a_m = \sum_{i \in A_m} w_{i,m}$, so the terms in a_m are i.i.d.; thus from the Central Limit Theorem, we have that $a_m/\sqrt{n/3}$ is normally distributed in the limit. The same applies of course to b_m and c_m, and it follows that for n larger than some n_0, $a_m + b_m + c_m = \sum_{i=1}^{n} w_{i,m}$ lies within, say, $n^{2/3}$ of its mean (n) with probability greater than $1 - 1/10n$.

For $m = 0$, the terms $w_{i,0} = y_i^2$ are already i.i.d. so again by CLT, there is some $n_1 > n_0$ such that for $n > n_1$, $\sum_{i=1}^{n} w_{i,0}$ lies within $n^{2/3}$ of its mean ($2n$) with probability greater than $1 - 1/10n$.

We have

$$\bar{r} = \Big(\sum_{i=1}^{n} w_{i,0} + 2\sum_{i=1}^{n} w_{i,1} + \cdots + 2\sum_{i=1}^{n} w_{i,q-1} + \beta \sum_{i=1}^{n} w_{i,q}\Big)/S^2$$

with $\mathrm{E}(S^2) = n(n+1)$ and $\sigma(S^2) = \sqrt{2n(n+1)(2n+3)}$. For $n > n_1$, the numerator lies within $n \times n^{2/3} = n^{5/3}$ of its mean with probability at least $9/10$, and by Chebyshev's inequality the denominator lies within distance $3n^{3/2}$ of its mean with probability at least $8/9$.

Putting this all together, we have that $\Pr\big(\bar{r}, \bar{r}(\infty)\|_{\mathrm{TV}} > .2\big) < .25$ with plenty to spare, and the proof is (finally) complete.

We remark that it is equally easy to use just $|r_0(t) - r_0(\infty)|$ instead of the total variation distance to prove the lower bound, using special properties of the Markov chain $\widehat{\mathbf{D}}$. The above method was chosen since it appears to be more generally applicable.

Acknowledgments This work began at the Banff International Research Station and has benefited from a conversation with Graham Brightwell and several with David Aldous.

References

1. P. Diaconis and M. Shashahani, Time to reach stationarity in the Bernoulli-Laplace diffusion model, *SIAM J. Math. Anal.* **81** #1 (1987), 208–218.
2. M. Dyer, A. Frieze, and R. Kannan, A random polynomial time algorithm for approximating the volume of convex sets, *J. for the Association for Computing Machinery* **38** (1991), 1–17.
3. M. Dyer and C. Greenhill, On Markov Chains for Independent Sets, *J. Algorithms,* **35** #1 (2000), 17–49.
4. R. Kannan, Markov chains and polynomial time algorithms. *Proc. 35th FOCS,* (1994), 656–671.
5. R. Kannan, M.W. Mahoney, and R. Montenegro, Rapid mixing of several Markov chains for a hard-core model, *Proc. 14th Annual ISAAC* (2003), 663–675.
6. L. Lovász and P. Winkler, Mixing times, *Microsurveys in Discrete Probability*, D. Aldous and J. Propp, eds., *DIMACS Series in Discrete Math. and Theoretical Computer Science,* (1998) **41**: 85-134.
7. M. Luby, D. Randall and A. Sinclair, Markov chain algorithms for planar lattice structures, *SIAM Journal on Computing.* **31** (2001), 167–192.

8. M. Luby and E. Vigoda, Fast convergence of the Glauber dynamics for sampling independent sets, *Random Structures and Algorithms*, **15** #3-4 (1999), 229–241.
9. T.-Y. Lee and H.-T. Yau, Logarithmic Sobolev inequalities for some models of random walks, *Ann. Prob.* **26** #4 (1998), 1855–1873.
10. N. Madras and D. Randall, Markov chain decomposition for convergence rate analysis, *Ann. Appl. Prob* **12** (2002), 581–606.
11. D. Randall, Mixing, *Proc. 44th Annual FOCS* (2004), 4–15.
12. D. Randall and P. Winkler, Mixing points in an interval, *Proc. ANALCO 2005*, Vancouver BC.
13. A. Sinclair, *Algorithms for Random Generation & Counting: A Markov Chain Approach.* Birkhäuser, Boston, 1993.
14. D.B. Wilson, Mixing times of lozenge tiling and card shuffling Markov chains, *Annals of Appl. Probab.* **14** (2004), 274–325.
15. D.B. Wilson, Mixing time of the Rudvalis shuffle, *Elect. Comm. in Probab.* **8** (2003), 77–85.