Random Walks on Discrete and Continuous Circles

by

Jeffrey S. Rosenthal

School of Mathematics, University of Minnesota, Minneapolis, MN, U.S.A. 55455

(Appeared in *Journal of Applied Probability* **30** (1993), 780–789.)

Summary. We consider a large class of random walks on the discrete circle $\mathbf{Z}/(n)$, defined in terms of a piecewise Lipschitz function, and motivated by the "generation gap" process of Diaconis. For such walks, we show that the time until convergence to stationarity is bounded independently of n. Our techniques involve Fourier analysis and a comparison of the random walks on $\mathbf{Z}/(n)$ with a random walk on the continuous circle S^1.

Keywords. Random walk, Fourier analysis, Time to stationarity, Generation gap.

1. Introduction.

This paper considers certain random walks on the group $\mathbf{Z}/(n)$ of integers mod n. Our random walks will converge in total variation distance to the uniform distribution U_n on $\mathbf{Z}/(n)$. We shall be concerned with the *rate* of this convergence, as a function of n, the size of the group.

It is known that for *simple* random walk on $\mathbf{Z}/(n)$, where we move left or right one space, or remain don't move, each with (say) probability $\frac{1}{3}$, it takes $O(n^2)$ steps to approach the uniform distribution in total variation distance (see Diaconis [1], Chapter 3C, Theorem 2). Indeed, it is easily verified that for any random walk on $\mathbf{Z}/(n)$ in which the size of a single step is *bounded* independently of n, at least $O(n^2)$ steps are required to approach uniformity. Various faster convergence results have been obtained when the step distribution itself is chosen randomly; see Hildebrand [4] and Dou [3].

In this paper, we shall consider random walks whose step distribution grows linearly with n. This study was motivated by the following "generation gap" algorithm on $\mathbf{Z}/(n)$ suggested by Diaconis [2]. Consider the random walk on $\mathbf{Z}/(n)$ which begins at the identity, and at each step moves to another point with probability proportional to the *distance* (on $\mathbf{Z}/(n)$) between the two points. In other words, the further around the circle a point is, the more likely the process is to jump there on its next turn. If we continue to jump around on $\mathbf{Z}/(n)$ in this manner, how long (as a function of n) will it take until our distribution is roughly uniform?

The generation gap algorithm is a special case of the following set-up. We let f be a non-negative, real-valued Lipschitz function on the unit circle S^1, and we embed $\mathbf{Z}/(n)$ in S^1 in the obvious way. Then, for each n, we consider the probability distribution P_n on $\mathbf{Z}/(n)$ induced by using the restriction of f to $\mathbf{Z}/(n)$ to define the step distribution. We shall allow n to get large, but keep the function f fixed. We show that under such conditions, the total variation distance of the random walk on $\mathbf{Z}/(n)$ after m steps to the uniform distribution U_n can be bounded by a quantity which is independent of n, and which goes to zero exponentially quickly as a function of m. In this sense, we shall say that such random walks converge to uniform on $\mathbf{Z}/(n)$ in a *constant* number of steps.

This paper is organized as follows. Section 2 gives a precise statement of our main result. Section 3 gives two examples of the use of this result. Finally, Section 4 proves the result, using Fourier Analysis on abelian groups. The key idea is to relate the random walks (and their Fourier coefficients) on $\mathbf{Z}/(n)$ to the corresponding ones on the continuous circle S^1. The result will then follow from standard results about Fourier Analysis.

2. Definitions and Main Result.

We begin with some standard definitions. Given two probability distributions P and Q on $\mathbf{Z}/(n)$, we define their *variation distance* by

$$\|P - Q\| = \frac{1}{2} \sum_{j \in \mathbf{Z}/(n)} |P(j) - Q(j)| \ .$$

We define their *convolution* $P * Q$ by

$$P * Q(s) = \sum_{j \in \mathbf{Z}/(n)} P(j) Q(s - j) \ ;$$

$P * Q$ is thus a new probability distribution on $\mathbf{Z}/(n)$, which represents the distribution after starting at the identity, and taking one step according to P, then a second step according to Q.

Given a distribution P_n on $\mathbf{Z}/(n)$, it induces a random walk on $\mathbf{Z}/(n)$ which starts at the identity and has step distribution given by P_n. Thus, its distribution after m steps is given by P_n^{*m}, the m-fold convolution product of P_n with itself. We let U_n be the uniform distribution on $\mathbf{Z}/(n)$, and consider the variation distance

$$\|P_n^{*m} - U_n\|$$

as a function of m and n. (This assumes we have been given a distribution P_n on $\mathbf{Z}/(n)$ for *each* n.) The usual question is, as a function of n, how large must m be to make the above variation distance small?

It is easily seen using Fourier analysis that if P_n has bounded support (i.e. P_n is non-zero only on a neighbourhood of $0 \in \mathbf{Z}/(n)$, and the size of this neighbourhood is bounded as a function of n), then m must be of size $O(n^2)$ (for large n) to make the variation distance small. In this paper, we consider families of distributions P_n defined differently, and show that m need only be of size $O(1)$ to make the variation distance small.

To define our random walk, let f be a non-negative real-valued function on the continuous circle S^1, which satisfies the following "piecewise Lipschitz" and positivity conditions:

(A1) The circle S^1 can be decomposed into J intervals $I_1, I_2, \ldots, I_J$ such that for some positive constants L and α, and for $1 \le j \le J$,

$$|f(x) - f(y)| \le L|x-y|^\alpha \qquad \text{for all} \qquad x, y \in \mathrm{Int}(I_j)\ .$$

(A2) $f(x) > 0$ for some $x \in \mathrm{Int}(I_j)$, for some j;

Identify S^1 with the interval $[0, 2\pi)$ in the obvious way. For each n, define a measure P_n on $\mathbf{Z}/(n)$ by

$$P_n(j) = \frac{f\left(\frac{2\pi j}{n}\right)}{\sum\limits_{s=0}^{n-1} f\left(\frac{2\pi s}{n}\right)}\ , \qquad j \in \mathbf{Z}/(n)\,.$$

(Note that the hypotheses on f imply that $\sum\limits_{s=0}^{n-1} f\left(\frac{2\pi s}{n}\right)$ is positive for all but finitely many n.) The measure P_n is thus obtained by regarding $\mathbf{Z}/(n)$ as sitting inside S^1 and using the values of f on $\mathbf{Z}/(n)$ as weights for P_n.

Let P_n^{*m} be the m-fold convolution product of P_n with itself, and let U_n be the uniform distribution on $\mathbf{Z}/(n)$. The main result of this paper is the following.

Theorem 1. *Under the above assumptions, there are positive constants A and B (depending on f but not on n or m) such that the random walk on $\mathbf{Z}/(n)$ satisfies*

$$\|P_n^{*m} - U_n\| \le A\, e^{-Bm}\ ,$$

for all m and for all but finitely many n.

We shall prove the above Theorem using Fourier Analysis and the Upper Bound Lemma of Diaconis and Shashahani (see Diaconis [1]). The proof is presented in Section 4. The key idea of the proof is to relate the random walks on the discrete circles $\mathbf{Z}/(n)$ to a single random walk on the continuous circle S^1, induced by the same function f. For large n, the random walk on $\mathbf{Z}/(n)$ will be "similar" to the random walk on S^1, and thus the rates of convergence will be related to the *single* rate of convergence for the random walk on S^1.

In Section 3 below, we consider two examples of the use of Theorem 1. We conclude this Section with a remark about the necessity of the restrictions on f.

Remark. The requirement that f satisfy a Lipschitz condition (except at a finite number of points), or some similar condition, is indeed necessary. It is not sufficient that f be merely, say, a bounded L^1 function on S^1. Indeed, let $p_1 < p_2 < p_3 < \ldots$ be an increasing sequence of prime numbers, and define the function f on S^1 by

$$f(x) = \begin{cases} 0 & \text{if} \quad x = \dfrac{2\pi j}{p_i} \quad \text{for some} \quad i \quad \text{and some integer } j \neq 0, \pm 1 \\ 1 & \text{otherwise.} \end{cases}$$

Then f is a bounded L^1 function which does not satisfy (A1) above. On the other hand, on $\mathbf{Z}/(p_i)$, the induced random walk is simple random walk (it moves distance one to the right or left, or does not move, each with probability 1/3). Thus, for $n = p_i$, $O(n^2)$ steps are required to approach uniform. Hence, the conclusion of Theorem 1 is not satisfied.

3. Examples.

In this Section we present two simple examples of the use of Theorem 1.

Example 1. *The Generation Gap process.* Diaconis [2] has proposed the following process on $\mathbf{Z}/(n)$. At each step, move from a point x to a point y with probability proportional to the *distance* from x to y around the circle $\mathbf{Z}/(n)$. Intuitively, each step of this algorithm attempts to move as far as possible from the previous position (analogous to children attempting to be as different as possible from their parents). In terms of Theorem 1, we can formulate this process by defining a function f on $[0, 2\pi)$ by

$$f(x) = \min(x, 2\pi - x) \ .$$

The measures P_n induced by this function f are precisely those that generate the Generation Gap process. Thus, Theorem 1 shows that a constant number of steps suffices to approach the uniform distribution on $\mathbf{Z}/(n)$. In other words, the Generation Gap process mixes up very quickly.

Example 2. *Random walk with large step size.* Consider the random walk on $\mathbf{Z}/(n)$ which at each step moves to one of the dn nearest neighbours ($d \leq 1$) with equal probability. In the context of Theorem 1, this is the random walk on $\mathbf{Z}/(n)$ induced from the function

$$f(x) = \begin{cases} 1, & x \leq \pi d \qquad \text{or} \qquad x \geq 2\pi - \pi d \\ 0, & \pi d < x < 2\pi - \pi d \end{cases}$$

Thus, again only a constant number of steps are required to approach the uniform distribution on $\mathbf{Z}/(n)$. This is in contrast to the $O(n^2)$ steps that are required when the step size does not grow with n. Results of Dou [3] imply that for *most* choices of probability measures supported on dn points of $\mathbf{Z}/(n)$, a constant number of steps suffices to approach the uniform distribution.

There are obviously many other examples of uses of Theorem 1; new examples can be obtained simply by varying the function f.

4. Proof of Theorem 1.

In this Section we prove Theorem 1. We begin with a review of the relevant facts from Fourier analysis. Given a probability measure P_n on $\mathbf{Z}/(n)$, we define its *Fourier coefficients* by

$$a_{n,k} = \sum_{j=0}^{n-1} e^{2\pi ikj/n} P_n(j) \ , \qquad 0 \leq k \leq n-1 \ . \tag{1}$$

Similarly, given a probability measure P on S^1, we define its *Fourier coefficients* by

$$a_k = \int_{S^1} e^{ikx} P(dx) \ , \qquad k \in \mathbf{Z} \ . \tag{2}$$

We let U be the uniform distribution on S^1, and let U_n be the uniform distribution on $\mathbf{Z}/(n)$. The Upper Bound Lemma of Diaconis and Shashahani (see Diaconis [1], Chapter 3C) states that for the random walk on $\mathbf{Z}/(n)$ with step distribution given by P_n, its variation distance to uniform distribution U_n after m steps satisfies

$$\|P_n^{*m} - U_n\|_{\mathbf{Z}/(n)}^2 \leq \frac{1}{4} \sum_{k=1}^{n-1} |a_{n,k}|^{2m} \ . \tag{3}$$

Similarly, for the random walk on S^1 with step distribution given by P, its variation distance to the uniform distribution U on S^1 satisfies

$$\|P^{*m} - U\|_{S^1}^2 \le \frac{1}{4} \sum_{\substack{-\infty<k<\infty \\ k\neq 0}} |a_k|^{2m} \ . \tag{4}$$

The task at hand is to show that for the problem under consideration, the sum in (3) can be bounded by $A\, e^{-Bm}$ for some positive constants A and B independent of n.

To that end, we let f be a non-negative function on S^1 which satisfies (A1) and (A2) above. This implies that

$$M = \sup_{x,y\in S^1} |f(x) - f(y)|$$

is finite. Without loss of generality, we take $\alpha \le 1$ (which must be true if f is not piecewise constant). For convenience, we assume that $\int\limits_{S^1} f(x)dx = 1$; if not, we can divide f by its L^1 norm, and modify the constants L and M appropriately.

We let P be the probability measure on S^1 defined by

$$dP = f(x)dx \ ,$$

and (for sufficiently large n) let P_n be the probability measure on $\mathbf{Z}/(n)$ defined by

$$P_n(j) = \frac{f\left(\frac{2\pi j}{n}\right)}{\sum\limits_{s=0}^{n-1} f\left(\frac{2\pi s}{n}\right)} \ .$$

We define the Fourier coefficients $a_{n,k}$ and a_k by (1) and (2) above. The plan will be to show that for large n, $a_{n,k}$ will be close to a_k. This will allow us to bound the sum in (3) independently of n.

Remark. The remainder of this section essentially amounts to obtaining various bounds on $a_{n,k}$ and a_k and on the relationship between them. There is of course a long history of bounds on Fourier coefficients, and we do not claim any great novelty in our methods or results. In particular, the bounds of Lemma 2 (a) and Propostion 4 (b) follow easily from standard techniques such as those in [5] (see Theorem 2.5.1 therein). However, for the discrete Fourier coefficients $a_{n,k}$ and the connection between $a_{n,k}$ and a_k, our bounds such as Lemma 2 (b) and Proposition 4 (c) do not appear to follow immediately from standard results.

We proceed as follows. For each $n > 0$, we define the operator T_n on the set of functions on S^1 by the equation

$$(T_n g)(x) = g\left(\frac{[xn/2\pi]}{n/2\pi}\right) ,$$

where $[y]$ denotes the greatest integer not exceeding y. Thus, $(T_n g)$ is a slight modification of the function g, which is constant on intervals of the form $\left[\frac{2\pi j}{n}, \frac{2\pi(j+1)}{n}\right]$. The benefit of the operator T_n comes from noting that

$$\int_{S^1} e^{ikx}(T_{|k|}f)(x)dx = 0 .$$

Furthermore, $(T_n g)$ provides a link between the function g on S^1, and the restriction of the function g to $\mathbf{Z}/(n)$, as the following Lemma shows.

Lemma 2. *For any function g on S^1 with $|g(x)| \leq 1$ and $|g(x) - g(y)| \leq |x - y|$ for all x and y, if $n \geq 3$,*

(a)

$$\int_{S^1} |(T_n(gf))(x) \; - \; g(x)f(x)|\, dx \; \leq \; \frac{4MJ\pi}{n} + 2\pi(L+M)\left(\frac{\pi}{n}\right)^{\alpha} .$$

(b)

$$\left|\int_{S^1} (T_n(gf))(x)dx \; - \; \sum_{j=0}^{n-1} g\left(\frac{2\pi j}{n}\right) P_n(j)\right| \; \leq \; \frac{4MJ\pi}{n} + 2\pi(L+M)\left(\frac{\pi}{n}\right)^{\alpha} .$$

Proof. For (a), we break up S^1 into n intervals, each of length $\frac{2\pi}{n}$, with midpoints at $\frac{2\pi j}{n}$. On J of these intervals, f will have a discontinuity, and we can only bound $|(T_n(gf))(x) - g(x)f(x)|$ by $2M$. On the other pieces, the function gf is easily seen to satisfy a Lipschitz condition with L replaced by $L+M$, so $|(T_n(gf))(x) - g(x)f(x)| \leq (L+M)\left(\frac{\pi}{n}\right)^{\alpha}$. We conclude that

$$\int_{S^1} |(T_n(gf))(x) \; - \; g(x)f(x)|\, dx$$

$$\leq (2M)(J)\left(\frac{\pi}{n}\right) + \int_{S^1} (L+M)\left(\frac{\pi}{n}\right)^{\alpha} dx$$

$$= \frac{4MJ\pi}{n} + 2\pi(L+M)\left(\frac{2\pi}{n}\right)^{\alpha} .$$

For (b), we first note that

$$\sum_{j=0}^{n-1} g(\frac{2\pi j}{n}) P_n(j) = \frac{\sum\limits_{j=0}^{n-1} g(\frac{2\pi j}{n}) f(\frac{2\pi j}{n})}{\sum\limits_{j=0}^{n-1} f(\frac{2\pi j}{n})}$$

$$= \frac{\frac{n}{2\pi}\left(\int\limits_{S^1} (T_n(gf))(x)dx\right)}{\frac{n}{2\pi}\int\limits_{S^1} (T_n f)(x)dx}$$

$$= \frac{\left(\int\limits_{S^1} (T_n(gf))(x)dx\right)}{\int\limits_{S^1} (T_n f)(x)dx} .$$

Hence,

$$\left|\sum_{j=0}^{n-1} g(\frac{2\pi j}{n}) P_n(j) - \int\limits_{S^1} (T_n(gf))(x)dx\right|$$

$$= \left|1 - \int\limits_{S^1} (T_n f)(x)dx\right| \left|\sum_{j=0}^{n-1} g(\frac{2\pi j}{n}) P_n(j)\right|$$

$$\leq \left|1 - \int\limits_{S^1} (T_n f)(x)dx\right| .$$

Now,

$$\left|1 - \int\limits_{S^1} (T_n f)(x)dx\right| = \left|\int\limits_{S^1} (f - (T_n f)(x))\, dx\right| \leq \int\limits_{S^1} |f - (T_n f)(x)|\, dx ,$$

so the result follows from part (a) by setting $g(x) = 1$. ∎

The Lemma and the triangle inequality immediately imply

Corollary 3.

$$\left|\int\limits_{S^1} g(x)f(x)dx - \sum_{j=0}^{n-1} g\left(\frac{2\pi j}{n}\right) P_n(j)\right| \leq \frac{8MJ\pi}{n} + 4\pi(L+M)\left(\frac{\pi}{n}\right)^{\alpha} .$$

Using the Lemma and the Corollary, we prove

Proposition 4. *The Fourier coefficients $a_{n,k}$ and a_k defined by equations (1) and (2) satisfy*

(a) *For all $k \neq 0$,* $\quad |a_k| < 1$.

(b) *For all $k \neq 0$,*

$$|a_k| \leq \frac{4MJ\pi}{|k|} + 2\pi(L+M)\left(\frac{\pi}{|k|}\right)^{\alpha} .$$

(c) *For any $n \geq 3$ and $0 \leq k \leq n-1$,*

$$|a_{n,k} - a_k| \leq \frac{8MJ\pi}{n} + 4\pi(L+M)\left(\frac{\pi}{n}\right)^{\alpha}$$

(d) *For any $n \geq 3$ and $0 < k \leq n-1$,*

$$|a_{n,k}| < \frac{12MJ\pi}{k} + 6\pi(L+M)\left(\frac{\pi}{k}\right)^{\alpha} .$$

(e) *For each $k > 0$, there is a number b_k, $0 < b_k < 1$, such that $|a_{n,k}| < b_k$ for all sufficiently large n.*

Proof. For (a), we recall that

$$a_k = \int_{S^1} e^{ikx} f(x) dx \ ,$$

and note that by the assumptions on f, it is positive on some open interval, on which e^{ikx} does not have constant argument. Thus, the inequality in the statement

$$|a_k| < \int_{S^1} |e^{ikx} f(x)| dx = \int_{S^1} f(x) dx = 1$$

is strict.

For (b), we recall that

$$\int_{S^1} e^{ikx} (T_{|k|} f)(x) dx = 0 \ ,$$

so that

$$\begin{aligned} |a_k| \ &= \left| \int_{S^1} e^{ikx} \left(f(x) - (T_{|k|} f)(x) \right) dx \right| \\ &\leq \int_{S^1} \left| f(x) - (T_{|k|} f)(x) \right| dx \ , \end{aligned}$$

and the bound now follows from Lemma 2 (a), with $g(x) = 1$ for all x.

For (c), recall that

$$|a_{n,k} - a_k| \;=\; \left| \sum_{j=0}^{n-1} e^{2\pi ikj/n} P_n(j) \;-\; \int_{S^1} e^{ikx} P(dx) \right| \;,$$

and use Corollary 3 with $g(x) = e^{ikx}$.

Statement (d) is immediate from statements (b) and (c), the triangle inequality, and the observation that $k = |k| < n$.

For (e), we note that (a) and (b) imply that $|a_k| < 1$ and $a_k \to 0$. Thus, if we let $a_* = \max\{|a_k|,\, k > 0\}$, then $a_* < 1$. Hence, if we set

$$b_k = |a_k| + \frac{1 - a_*}{2} \;,$$

then $0 < b_k < 1$. Also, part (c) implies that $|a_{n,k}| < b_k$ provided n is chosen large enough that

$$\frac{8MJ\pi}{n} + 4\pi(L+M)\left(\frac{2\pi}{n}\right)^{\alpha} \;<\; \frac{1 - a_*}{2} \;.$$

■

Proposition 4 allows us to complete the proof of Theorem 1, but we first record a corollary about the random walk on S^1 itself.

Corollary 5. *The random walk on S^1 induced by the measure P converges to the uniform distribution U on S^1 exponentially quickly in total variation distance.*

Proof. From equation (4) above, the variation distance $\|P^{*m} - U\|^2_{S^1}$ is bounded by the sum

$$\frac{1}{4} \sum_{\substack{k \in \mathbf{Z} \\ k \neq 0}} |a_k|^{2m} \;;$$

it suffices to bound this sum by an expression of the form $C_1 e^{-C_2 m}$. From part (b) of Proposition 4, there is a constant C_3 such that $|a_k| < \frac{C_3}{|k|^\alpha}$. Then, writing

$$\sum_{\substack{k \in \mathbf{Z} \\ k \neq 0}} |a_k|^{2m} \;=\; \sum_{0 \neq |k| \leq (2C_3)^{1/\alpha}} |a_k|^{2m} \;+\; \sum_{|k| > (2C_3)^{1/\alpha}} |a_k|^{2m} \;,$$

we see that the second sum can be easily be bounded by an integral, and shown to be less than an exponentially decaying function of m. There are only a finite of terms in the first sum, so the first sum decays exponentially by part (a) of Proposition 4. ■

We now proceed to the proof of Theorem 1. From part (d) of Proposition 4, there is a constant C_4 such that $|a_{n,k}| < \frac{C_4}{k^\alpha}$ for $1 < k < n$. Then using equation (3) above, and using part (e) of Proposition 4, the variation distance $\|P_n^{*m} - U_n\|$ is bounded by

$$\begin{aligned}\frac{1}{4}\sum_{k=1}^{n-1}|a_{n,k}|^{2m} &\le \frac{1}{4}\sum_{k=1}^{n-1}\left(\min(\frac{C_4}{k^\alpha}, b_k)\right)^{2m} \\ &\le \frac{1}{4}\sum_{0<k\le(2C_4)^{1/\alpha}}(b_k)^{2m} \;+\; \frac{1}{4}\sum_{(2C_4)^{1/\alpha}<k<\infty}\left(\frac{C_4}{k^\alpha}\right)^{2m}\end{aligned}$$

where in this last expression we sum over all integers k, including those greater than n. This last expression is clearly independent of n. Furthermore, the expression can be bounded as in the proof of Corollary 5. Indeed, the first sum in the expression consists of a finite number of terms, and clearly decays exponentially with m. The second can easily be bounded by an integral and shown to also decay exponentially with m. Hence the sum decays exponentially quickly in m, uniformly in n (for n sufficiently large), proving Theorem 1.

Acknowledgements. I am very grateful to Persi Diaconis for suggesting the Generation Gap process, and for many helpful discussions. I thank the referee for useful comments.

REFERENCES

[1] P. Diaconis (1988), *Group Representations in Probability and Statistics,* IMS Lecture Series volume **11**, Institute of Mathematical Statistics, Hayward, California.

[2] P. Diaconis, personal communication.

[3] C.C. Dou (1991), *Ennumeration and Random Walk on Groups,* preprint.

[4] M.V. Hildebrand (1990), *Rates of Convergence of Some Random Processes on Finite Groups,* Ph.D. dissertation, Mathematics Department, Harvard University.

[5] T. Kawata (1972), *Fourier Analysis in Probability Theory.* Academic Press, New York.