Probab. Theory Relat. Fields 100, 191–203 (1994)

Probability Theory and Related Fields

Random walks supported on random points of $\mathbf{Z}/n\mathbf{Z}$

Martin Hildebrand*

Institute for Mathematics and its Applications, University of Minnesota, Minneapolis, MN 55455-0436, USA

Received: 13 November 1993/In revised form: 21 March 1994

Summary. This paper considers random walks on the integers mod n supported on k points and asks how long does it take for these walks to get close to uniformly distributed. If k is a constant, Greenhalgh showed that at least some constant times $n^{2/(k-1)}$ steps are necessary to make the distance of the random walk from the uniform distribution small; here we show that if n is prime, some constant times $n^{2/(k-1)}$ steps suffice to make this distance small for almost all choices of k points. The proof uses the Upper Bound Lemma of Diaconis and Shahshahani and some averaging techniques. This paper also explores some cases where k varies with n. In particular, if $k = \lfloor (\log n)^a \rfloor$, we find different kinds of results for different values of a, and these results disprove a conjecture of Aldous and Diaconis.

Mathematics Subject Classification (1991): 60B15, 60J15

Introduction

Consider a random walk on the integers mod n as follows. Pick certain points on the integers mod n; the random walk will be "supported" on these points. Let the walk start at 0. Pick one of the points which supports the walk at random (according to a specified probability distribution) and add it to the current position of the walk. (The first time you do this, you will be adding to 0.) Repeat, picking the points independently of other picks but with the same probability distribution. How long does it take for this walk to get close to uniformly distributed on the integers mod n? If the certain points are -1 (i.e. $n-1$), 0, and 1, then it takes some constant times n^2 steps to get close to

* Research Supported in Part by a Rackham Faculty Fellowship at the University of Michigan

uniform. (See [CDG].) Here we explore this question for most choices of k points to support the random walk where k is a fixed constant or is a power of $\log n$. Note that our choices of the k points may vary with n even when k itself does not. Related questions have been explored by Greenhalgh in [Gr] and [Gr2] and by Dou in [Do], and the question is posed by Diaconis in [Di].

We define the distance of a probability distribution P on a finite group G from the uniform distribution U to be

$$\| P-U \| := \frac{1}{2} \sum_{s \in G} \left| P(s) - \frac{1}{|G|} \right| = \max_{A \subseteq G} |P(A)-U(A)| .$$

This distance is the same as in [Di].

Let P^{*m} be the probability distribution of the sum of m i.i.d. random variables distributed as P. In other words, if P gives the probability distribution of each step of the random walk, P^{*m} gives the probability distribution of the position of the random walk after m steps.

In this paper we show

Theorem 1 *Suppose k is a fixed positive integer which is at least 2. Let p_i, $i=1, \ldots, k$ be such that $p_i>0$ and $\sum_{i=1}^{k} p_i=1$. Given $\varepsilon>0$,*

$$E[\| P_{\tilde{a}}^{*m} - U \|] < \varepsilon$$

for $m=\lfloor \gamma n^{2/(k-1)} \rfloor$ for some constant $\gamma>0$ (which may depend on k and the values for p_i but not on n) and for sufficiently large primes n. We define $\tilde{a} := (a_1, a_2, \ldots, a_k)$, and we let

$$P_{\tilde{a}}(a) = \begin{cases} p_i & \text{if } a=a_i \text{ for some } i \\ 0 & \text{otherwise.} \end{cases}$$

The expectation is taken over a uniform choice of all possible $\tilde{a}$ such that $a_1, \ldots, a_k \in \mathbf{Z}/n\mathbf{Z}$ and such that all values of $a_1, \ldots, a_k$ are distinct.

Note that if n is not prime, there is a non-zero probability that a random walk supported on k points will stay on a proper subgroup of $\mathbf{Z}/n\mathbf{Z}$. For instance, if $k=3$ and $n=3p$ where p is a large integer, then with probability approximately $1/27$, the values a_1, a_2, and a_3 will all be multiples of 3.

Note that for $k=3$, Theorem 1 says that after γn steps for some constant γ, most walks supported on 3 points will be close to uniformly distributed. Since Greenhalgh's lower bound (to be described in a later section) is the same up to a constant multiple, most walks will take this number of steps (up to a constant multiple) to get close to uniformly distributed. There are, however, certain choices of 3 points which lead to a slow random walk (for example, -1, 0, and 1). Theorem 2, however, provides an upper bound for how long it takes any random walk on $\mathbf{Z}/n\mathbf{Z}$ supported on a constant k points to get close to uniformly distributed in the case where n is prime.

Theorem 2 *Let $\varepsilon>0$ be given. Suppose p_i is as in Theorem 1. If n is prime and $a_1, \ldots, a_k$ are distinct elements of $\mathbf{Z}/n\mathbf{Z}$, then for some constant $\gamma>0$,*

$$\| P_{\tilde{a}}^{*m} - U \| < \varepsilon$$

regardless of the choice of $a_1, \ldots, a_k$ if $m=\gamma n^2$.

Thus, if $a_1, \ldots, a_k$ are in an "arc" about the origin (e.g. if $k=3$, this arc could be $-1, 0$, and 1), then this choice leads to the slowest (up to a constant multiple) walk possible for the choice of $a_1, \ldots, a_k$.

This paper also explores cases where the value k varies with n. In particular, we shall show the following 2 theorems.

Theorem 3 *Suppose $k=\lfloor (\log n)^a \rfloor$ for some constant $a>1$. Suppose $p_i=1/k$ for $i=1, \ldots, k$. Let $\varepsilon>0$ be given. If $m=\left\lfloor \frac{a}{a-1}\frac{\log n}{\log k}(1-\varepsilon) \right\rfloor$, then $\| P_{\tilde{a}}^{*m} - U \| \to 1$ as $n\to\infty$ for all choices of $\tilde{a}$ where $\tilde{a}$ and $P_{\tilde{a}}$ are as in Theorem 1. If $m \geqq \frac{a}{a-1}\frac{\log n}{\log k}(1+\varepsilon)$, then for integers n, $E[\| P_{\tilde{a}}^{*m} - U \|] \to 0$ as $n\to\infty$ where the expectation is taken over a uniform choice of $\tilde{a}$ (as in Theorem 1).*

Theorem 4 *If $k=\lfloor (\log n)^a \rfloor$ where a is a constant less than 1, then for any fixed positive value b, the distance $\| P_{\tilde{a}}^{*m} - U \| \to 1$ as $n\to\infty$ when $m=\lfloor (\log n)^b \rfloor$.*

Theorems 3 and 4 show that if $k=\lfloor (\log n)^a \rfloor$, the length of time it takes for random walks to typically get close to uniformly distributed will be substantially different depending on the value of a.

Theorems 3 and 4 disprove a conjecture of Aldous and Diaconis [AD]; their conjecture stated that for an arbitrary group G with n elements if

$$m=\left\lfloor \frac{\log n}{\log k}(1+\varepsilon) \right\rfloor,$$

then $E[\| P_{\tilde{a}}^{*m} - U \|] \to 0$ as $n\to\infty$ provided that $k\to\infty$ and $\log n/\log k \to \infty$. However, if $a>1$, Theorem 3 shows that the conjecture is off by a factor of $a/(a-1)$ while if $a<1$, Theorem 4 shows that the conjecture is more substantially incorrect. Dou [Do] has also observed that k must be "rather large" for the conjecture of Aldous and Diaconis to hold. He cited the fact that random walks on $(\mathbf{Z}/2\mathbf{Z})^d$ require at least d points to even cover the group. However, our theorems show that even when there are much more than enough points to cover the group, a typical random walk still may converge slower than suggested by the conjecture of Aldous and Diaconis.

Theorem 3 also shows that a "cutoff" phenomenon occurs for typical walks being considered. In a relatively short period of time, the distance of the typical walk from uniform goes from close to 1 to close to 0. Such phenomena are relatively common; see Diaconis [Di] for further examples where such phenomena appear.

Background and previous results

In the case where k is fixed, Greenhalgh [Gr] has shown the following lower bound.

Theorem 5 *Suppose* $p_i>0$, $i=1, \ldots, k$, $\sum_{i=1}^{k} p_i=1$. *Then there exists a value* $\beta(p_1, \ldots p_k)>0$ *and* $n_0(p_1, \ldots, p_k)$ *such that for all choices of* $\tilde{a}$, $m=\lfloor \beta(p_1, \ldots, p_k) n^{2/(k-1)} \rfloor$, *and* $n>n_0(p_1, \ldots, p_k)$,

$$\| P_{\tilde{a}}^{*m}-U \| \geqq \tfrac{1}{4} .$$

Theorem 5 nicely complements Theorem 1. Theorem 5 says that for all random walks supported on k points, at least some constant multiple of $n^{2/(k-1)}$ steps are necessary for the random walk to get close to uniform. On the other hand, Theorem 1 says that if n is prime, some constant multiple times $n^{2/(k-1)}$ steps suffice to get most random walks supported on k points close to uniformly distributed.

The case of Theorems 1 and 2 where $p_i=1/k$ was proved in [Hi], and the proofs here are minor modifications of the proofs in [Hi].

Greenhalgh [Gr 2] has shown a result which is related to Theorem 1. This result is

Theorem 6 *For* k *fixed,* $k>3$, p *prime, and* $m=\sigma(p)p^{2/(k-1)}$ *where* $\sigma(p)\to\infty$ *as* $p\to\infty$, *then* $Pr(\| P^{*m}-U \|\to 0)\to 1$ *as* $p\to\infty$ *where* $g_1, \ldots, g_k$ *are chosen i.i.d. uniformly from* G *and*

$$P(g)=\sum_{i: g_i=g} \frac{1}{k}.$$

Note that this theorem is essentially Theorem 1 in the case $p_i=1/k$ for $i=1, \ldots, k$; since few choices of $g_1, \ldots, g_k$ in Greenhalgh's theorem have $g_i=g_j$ for some $i\neq j$, the change in the choice of points to support the random walk does not change the conclusion. The proof in this paper does show some things the proof in [Gr2] does not; namely this paper does prove the results when $k=2$ and $k=3$ and when $p_i\neq 1/k$. Greenhalgh's proof does not readily adapt to such cases.

Dou [Do] has shown results for random walks supported on random points of an arbitrary set of finite groups provided that the size of the support grows sufficiently quickly. In particular, he has shown

Theorem 7 *Let* $\varepsilon>0$ *be given. Suppose that* G *is any finite group with* n *elements. Let* $a_1, \ldots, a_k$ *be chosen uniformly from* G *such that* $a_i\neq a_j$ *if* $i\neq j$, *and let*

$$P_{\tilde{a}}(g)=\begin{cases} 1/k & \text{if } g=a_i \text{ for some } i \\ 0 & \text{otherwise.} \end{cases}$$

Then if $k\geqq(\log n)^a$ *for some constant* $a>2$ *and*

$$m\geqq\frac{a}{a-2}\frac{\log n}{\log k}(1+\varepsilon),$$

then $E[\| P_{\tilde{a}}^{*m}-U \|]\to 0$ *as* $n\to\infty$.

Note that the upper bound in Theorem 3 is close to the result of Dou except that the $a-2$ term has been replaced by $a-1$.

In some of our proofs, we shall use the Upper Bound Lemma of Diaconis and Shahshahani (described in [Di]) applied to the integers mod n. Let P be a probability distribution on the integers mod n. Define the Fourier transform of P by

$$\hat{P}(j) := \sum_{k \in \mathbf{Z}/n\mathbf{Z}} P(k) q^{jk}$$

where $q := q(n) := e^{2\pi i/n}$.

The Upper Bound Lemma is

Lemma 1

$$\| P - U \|^2 \leqq \frac{1}{4} \sum_{j=1}^{n-1} |\hat{P}(j)|^2 .$$

Furthermore, it can be shown that $\hat{Q}(j) = (\hat{P}(j))^m$ if $Q = P^{*m}$. See [Di]. This property enables bounds on $\| P^{*m} - U \|$ to be determined from the Fourier transform of P.

Fourier transforms and the Upper Bound Lemma can be generalized to deal with any finite group. See [Di] for the details.

Results for fixed values of k

Throughout this section, k is assumed to be a constant integer greater than or equal to 2. Different values of k may change the value of "constants" introduced in this section. Furthermore, we assume that the probabilities $p_1, \ldots, p_k$ are positive, sum to 1, and are constant; changing the values of these probabilities also may change some the value of "constants."

In this section, we shall prove Theorems 1 and 2.

To prove Theorem 1, we shall prove a result on $E[\| P_{\tilde{a}}^{*m} - U \|^2]$. The following lemma relates this result to Theorem 1.

Lemma 2 *If given $\varepsilon' > 0$, there exists a value $\gamma > 0$ such that $E[\| P_{\tilde{a}}^{*m} - U \|^2] < \varepsilon'$ for sufficiently large primes n and $m = \lfloor \gamma n^{2/(k-1)} \rfloor$, then Theorem 1 holds.*

Proof. Let $\varepsilon > 0$ be given. Find values $a > 1$ and $\varepsilon' > 0$ such that $\sqrt{a\varepsilon'} + (1/a) < \varepsilon$. It is clear that such values exist. By the hypothesis of the lemma, there exists a value $\gamma > 0$ such that for sufficiently large primes n and $m = \lfloor \gamma n^{2/(k-1)} \rfloor$, $E[\| P_{\tilde{a}}^{*m} - U \|^2] < \varepsilon'$.

Since the distance is non-negative,

$$P[\| P_{\tilde{a}}^{*m} - U \|^2 > a\varepsilon'] < \frac{1}{a}$$

and

$$P[\| P_{\tilde{a}}^{*m} - U \| > \sqrt{a\varepsilon'}] < \frac{1}{a}.$$

Thus since $\| P_{\tilde{a}}^{*m} - U \| \leqq 1$, we conclude that

$$E[\| P_{\tilde{a}}^{*m} - U \|] < \sqrt{a\varepsilon'} + \frac{1}{a} < \varepsilon.$$

Hence Theorem 1 holds. ■

The following lemma provides a bound on the square of the distance for a given random walk on k points.

Lemma 3

$$\| P_{\tilde{a}}^{*m} - U \|^2 \leqq \frac{1}{4} \sum_{j=1}^{n-1} \left| \left(\sum_{i=1}^{k} p_i^2 \right) + 2 \sum_{1 \leqq i_1 < i_2 \leqq k} p_{i_1} p_{i_2} \cos(2\pi(a_{i_1} - a_{i_2}) j/n) \right|^m.$$

This lemma can be readily proved from the Upper Bound Lemma, the equality $|\hat{P}(j)|^2 = \hat{P}(j)\overline{\hat{P}(j)}$, and the equality $\widehat{P^{*m}}(j) = (\hat{P}(j))^m$. ■

The expression inside the absolute value sign varies with the choice of k points. To explore the range of values of this expression, we wish to explore the range of the possible values of $\cos(2\pi(a_1 - a_i) j/n)$. The following definition will help us to do so. Let $g_n(x) = x_0$ where $x_0 \in (-n/2, n/2]$ so that $x \equiv x_0 \pmod{n}$. Note that $\cos(2\pi x/n) = \cos(2\pi g_n(x)/n)$.

The following lemma tells us about the probability that $g_n((a_1 - a_i) j)/n$ falls in certain ranges for $i = 2, 3, \ldots, k$ where a_1 and j take on any fixed value. We shall design the ranges so that the probability that

$$(g_n((a_1 - a_2) j)/n, \ldots, g_n((a_1 - a_k) j)/n)$$

falls in any one $k-1$-dimensional "cube" created from such ranges shall be small.

Lemma 4 *Given $m_2, m_3, \ldots, m_k$, and $\varepsilon > 0$,*

$$\begin{aligned} &P(m_i(\varepsilon/2)^{1/(k-1)} n^{(k-2)/(k-1)}/2n \\ &\quad \leqq g_n((a_1 - a_i) j)/n \\ &\quad \leqq (m_i + 1)(\varepsilon/2)^{1/(k-1)} n^{(k-2)/(k-1)}/2n,\ i = 2, \ldots, k) \\ &\quad \leqq \frac{1.1(\varepsilon/2)}{2^{k-1} n} \end{aligned}$$

for large enough prime numbers n.

Proof. The proof is fairly straightforward.

Note that if n is odd and prime, $g_n((a_1 - a_i) j)$ may take on any of the values $\frac{-n+1}{2}, \ldots, -2, -1, 1, 2, \ldots, \frac{n-1}{2}$ that have not already been taken

on by $g_n((a_1-a_{i'})j)$ for $i'<i$. Also note that different values of a_i correspond to different values of $g_n((a_1-a_i)j)$. Also note that the assumption that n is prime is crucial for this step; otherwise the lemma may not hold.

On the interval

$$\left[\frac{m_i(\varepsilon/2)^{1/(k-1)}n^{(k-2)/(k-1)}}{2},\frac{(m_i+1)(\varepsilon/2)^{1/(k-1)}n^{(k-2)/(k-1)}}{2}\right],$$

there are at most

$$1+\frac{(\varepsilon/2)^{1/(k-1)}n^{(k-2)/(k-1)}}{2}$$

values of $g_n((a_1-a_i)j)$.

Thus the probability in the lemma is less than or equal to

$$\frac{(1+((\varepsilon/2)^{1/(k-1)}n^{(k-2)/(k-1)}/2))^{k-1}}{(n-1)\dots(n-k+1)}$$

$$=\frac{((\varepsilon/2)^{1/(k-1)}n^{(k-2)/(k-1)}/2)^{k-1}}{n^{k-1}}\frac{(1+(2/(\varepsilon/2)^{1/(k-1)}n^{(k-2)/(k-1)}))^{k-1}}{(1-(1/n))(1-(2/n))\dots(1-((k-1)/n))}.$$

Note that since

$$\frac{(1+(2/(\varepsilon/2)^{1/(k-1)}n^{(k-2)/(k-1)}))^{k-1}}{(1-(1/n))(1-(2/n))\dots(1-((k-1)/n))}\to 1$$

as $n\to\infty$, the lemma follows by simple computation. ■

Next we shall look further at the values of the expression inside the absolute value in Lemma 3.

Lemma 5 *Let*

$$b_1:=\max\left(1-\left(\sum_{i=1}^{k}p_i^2\right),1-0.02\min_{i_1\neq i_2}p_{i_1}p_{i_2}\right).$$

Then

$$-b_1\leqq\left(\sum_{i=1}^{k}p_i^2\right)+2\sum_{1\leqq i_1<i_2\leqq k}p_{i_1}p_{i_2}\cos(2\pi(a_{i_1}-a_{i_2})j/n)\leqq 1$$

and if

$$\left(\sum_{i=1}^{k}p_i^2\right)+2\sum_{1\leqq i_1<i_2\leqq k}p_{i_1}p_{i_2}\cos(2\pi(a_{i_1}-a_{i_2})j/n)>b_1,$$

then $\cos(2\pi(a_1-a_i)j/n)>0.99$ *for* $i=2,3,\dots,k$.

Proof. The proof is a straightforward exercise. ■

Note that there is a constant $c_1\in(0,1]$ such that if $\cos(2\pi k/n)>0.99$, then

$$\cos(2\pi k/n)\leqq 1-\frac{(c_1/2)(g_n(k))^2}{n^2}.$$

(There is nothing particularly special about the value 0.99; similar constants can be found if 0.99 is replaced by other positive values less than 1.) Also note that if $\cos(2\pi k/n)>0.99$ and $|m_i|(\varepsilon/2)^{1/(k-1)}n^{(k-2)/(k-1)}/2n \leqq |g_n((a_1-a_i)j)/n|$, then $\cos(2\pi(a_1-a_i)j/n) \leqq \cos(2\pi m_i(\varepsilon/2)^{1/(k-1)}n^{(k-2)/(k-1)}/2n)$. We shall use this bound in proving

Lemma 6 *There exists a value n_0 such that if n is a prime number greater than n_0 then for all a_1 and j,*

$$E\left[\left|\left(\sum_{i=1}^{k} p_i^2\right)+2\sum_{1\leqq i_1<i_2\leqq k} p_{i_1}p_{i_2}\cos(2\pi(a_{i_1}-a_{i_2})j/n)\right|^m\right]\leqq\frac{\varepsilon}{n}$$

where $m=\lfloor \gamma n^{2/(k-1)} \rfloor$ where γ is a positive constant not depending on a_1 and j.

Proof. First observe from Lemmas 4 and 5 that

$$\begin{aligned}
&E\left[\left|\left(\sum_{i=1}^{k} p_i^2\right)+2\sum_{1\leqq i_1<i_2\leqq k} p_{i_1}p_{i_2}\cos(2\pi(a_{i_1}-a_{i_2})j/n)\right|^m\right]\\
&\quad\leqq b_1^m+\sum_{\substack{m_i\in[-\ell_n-1,\ell_n]\\ i=2,\ldots,k}}\frac{1.1(\varepsilon/2)}{2^{k-1}n}\left(1-c_1\left(\min_{i_1\neq i_2}p_{i_1}p_{i_2}\right)\right.\\
&\qquad\left.\times\sum_{i=2}^{k}\frac{(\min(|m_i|,|m_i+1|))^2(\varepsilon/2)^{2/(k-1)}}{4n^{2/(k-1)}}\right)^m\\
&\quad=b_1^m+2^{k-1}\sum_{\substack{m_i\in[0,\ell_n]\\ i=2,\ldots,k}}\frac{1.1(\varepsilon/2)}{2^{k-1}n}\\
&\qquad\times\left(1-c_1\left(\min_{i_1\neq i_2}p_{i_1}p_{i_2}\right)\sum_{i=2}^{k}\frac{m_i^2(\varepsilon/2)^{2/(k-1)}}{4n^{2/(k-1)}}\right)^m
\end{aligned}$$

where ℓ_n is the largest positive integer ℓ such that

$$\cos\left(2\pi\left(\frac{\ell(\varepsilon/2)^{1/(k-1)}n^{(k-2)/(k-1)}}{2n}\right)\right)>0.99$$

and the argument of the cosine function is in the first quadrant. This inequality results from splitting the case where $\cos(2\pi(a_1-a_i)j/n)\geqq 0.99$ for all $i=2,\ldots,k$ from the other cases. By Lemma 5, in the other cases the term whose expected value we are taking will be less than b_1^m. Where $\cos(2\pi(a_1-a_i)j/n)\geqq 0.99$ for all $i=2,\ldots,k$, we may assume that the term inside the absolute value sign is positive; otherwise we know its absolute value is less than b_1. The remaining terms in the inequality come from the probability in Lemma 4 and the bound on $\cos(2\pi(a_1-a_i)j/n)$.

Since $c_1\leqq 1$, since $p_i\leqq 1/k$ for some i, and since as $n\to\infty$, $\max_{m_i\in[0,\ell_n]}m_i^2(\varepsilon/2)^{2/(k-1)}/(4n^{2/(k-1)})\to(\arccos 0.99/2\pi)^2<1$, we have

$$c_1\left(\min_{i_1\neq i_2}p_{i_1}p_{i_2}\right)\sum_{i=2}^{k}\frac{m_i^2(\varepsilon/2)^{2/(k-1)}}{4n^{2/(k-1)}}<1$$

for sufficiently large n and for all m_i in the range of the sum. Thus

$$0<1-c_1\left(\min_{i_1\neq i_2} p_{i_1}p_{i_2}\right)\sum_{i=2}^{k}\frac{m_i^2(\varepsilon/2)^{2/(k-1)}}{4n^{2/(k-1)}}\leqq 1\,.$$

Hence

$$\left(1-c_1\left(\min_{i_1\neq i_2} p_{i_1}p_{i_2}\right)\sum_{i=2}^{k}\frac{m_i^2(\varepsilon/2)^{2/(k-1)}}{4n^{2/(k-1)}}\right)^m\leqq\exp\left(-mc_2\sum_{i=2}^{k}\frac{m_i^2(\varepsilon/2)^{2/(k-1)}}{4n^{2/(k-1)}}\right)$$

for some constant $c_2>0$.

Thus we may conclude

$$\begin{aligned}
&E\left[\left|\left(\sum_{i=1}^{k}p_i^2\right)+2\sum_{1\leqq i_1<i_2\leqq k}p_{i_1}p_{i_2}\cos(2\pi(a_{i_1}-a_{i_2})j/n)\right|^m\right]\\
&\leqq b_1^m+\sum_{\substack{m_i\in[0,l_n]\\ i=2,\ldots,k}}1.1\frac{\varepsilon}{2n}\exp\left(-mc_2\sum_{i=2}^{k}m_i^2\frac{(\varepsilon/2)^{2/(k-1)}}{4n^{2/(k-1)}}\right)\\
&\leqq b_1^m+\sum_{\substack{m_i\in[0,\infty)\\ i=2,\ldots,k}}1.1\frac{\varepsilon}{2n}\exp\left(-mc_2\sum_{i=2}^{k}m_i\frac{(\varepsilon/2)^{2/(k-1)}}{4n^{2/(k-1)}}\right)\\
&=b_1^m+\left(\frac{1.1\varepsilon}{2n}\right)\Big/(1-\exp(-mc_2(\varepsilon/2)^{2/(k-1)}/4n^{2/(k-1)}))^{k-1}\,.
\end{aligned}$$

For $m=\lfloor \gamma n^{2/(k-1)}\rfloor$ for some constant $\gamma>0$, the denominator is greater than 0.7. Then

$$\begin{aligned}
E\left[\left|\left(\sum_{i=1}^{k}p_i^2\right)+2\sum_{1\leqq i_1<i_2\leqq k}p_{i_1}p_{i_2}\cos(2\pi(a_{i_1}-a_{i_2})j/n)\right|^m\right]&\leqq b_1^m+\frac{0.9\varepsilon}{n}\\
&<\frac{\varepsilon}{n}
\end{aligned}$$

for sufficiently large n since b_1 is a constant less than 1. ■

By using the Upper Bound Lemma, we may conclude that $E[\|P_{\tilde{a}}^{*m}-U\|^2]<(\varepsilon/4)<\varepsilon$ if m is as in Lemma 6. Thus the hypothesis of Lemma 2 is satisfied, and we have proved Theorem 1. ■

Proof of Theorem 2

This proof is quite straightforward.

Observe that $(a_1-a_2, n)=1$. Also observe that

$$\begin{aligned}
-1+p_1^2&\leqq\left(\sum_{i=1}^{k}p_i^2\right)+2\sum_{1\leqq i_1<i_2\leqq k}p_{i_1}p_{i_2}\cos(2\pi(a_{i_1}-a_{i_2})j/n)\\
&\leqq(1-2p_1p_2)+2p_1p_2\cos(2\pi(a_1-a_2)j/n)\,.
\end{aligned}$$

Note that $(a_1-a_2)j$ runs through $1, 2, \ldots, n-1 \pmod{n}$ for $j=1, 2, \ldots, n-1$. Thus, by arguments identical (up to a constant) to arguments in the proof of the right side of (9) in [CDG], Theorem 2 holds. (The proof in [CDG] simply uses the Upper Bound Lemma to bound the simple random walk on $\mathbf{Z}/n\mathbf{Z}$.) ■

Proof of Theorem 3

First off, we shall prove the portion where

$$m=\left\lfloor \frac{a}{a-1}\frac{\log n}{\log k}(1-\varepsilon)\right\rfloor.$$

We shall show that after this many steps, the random walk will, with probability approaching 1, be on a set which has probability approaching 0 on the uniform distribution on $\mathbf{Z}/n\mathbf{Z}$. Thus the distance from uniform will approach 1. We shall use

Lemma 7 *There is a function $f(n)$ with $f(n)\to 0$ as $n\to\infty$ such that with probability approaching 1 the proportion of the points picked more than once in the m trials is less than $f(n)$.*

Proof. The probability that the point chosen on the i-th trial is picked on one or more of the remaining $m-1$ trials is less than m/k. Hence the expected number of trials which are duplicated elsewhere is less than $(m^2/k)<(\log n)^{(2-a)}$ for sufficiently large n. By Markov's inequality, the probability that the number of trials which are duplicated elsewhere is greater than $(\log n)^{((2-a)+1)/2}$ will approach 0 as $n\to\infty$. (Recall that $a>1$.) So let

$$f(n)=\frac{(\log n)^{((2-a)+1)/2}}{m}.$$

Note that $f(n)\to 0$ as $n\to\infty$ and satisfies the condition of the lemma. ■

After m steps of the random walk, there are

$$(\lfloor (\log n)^a \rfloor)^{\lfloor (a/(a-1))(\log n/\log k)(1-\varepsilon)\rfloor}$$

choices to make. Since $\mathbf{Z}/n\mathbf{Z}$ is abelian, rearranging the trials in the random walk will not change the final position of the walk. Thus each walk (except for those walks where the proportion of trials being duplicated is over $f(n)$) has at least $(\lfloor \lfloor (a/(a-1))(\log n/\log k)(1-\varepsilon)\rfloor(1-f(n))\rfloor)!$ other walks with the same sum.

So except for events with probability approaching 0, there are no more than

$$\frac{(\lfloor (\log n)^a \rfloor)^{\lfloor (a/(a-1))(\log n/\log k)(1-\varepsilon)\rfloor}}{(\lfloor \lfloor (a/(a-1))(\log n/\log k)(1-\varepsilon)\rfloor(1-f(n))\rfloor)!}$$

possible values. By Stirling's formula, this value can be shown to be $n^{1-\varepsilon+o(1)}$, and hence this portion of Theorem 3 is true. ■

Next we shall show that if n is prime then as $n\to\infty$, $E[\| P_{\tilde{a}}^{*m}-U\|]\to 0$ if $a>1$ and

$$m\geqq\frac{a}{a-1}\frac{\log n}{\log k}(1+\varepsilon).$$

First, we shall establish relations between certain expected values.

Lemma 8 *If* $E[\| P_{\tilde{a}}^{*m}-U\|^2]\to 0$ *as* $n\to\infty$, *then* $E[\| P_{\tilde{a}}^{*m}-U\|]\to 0$ *as* $n\to\infty$.

Proof. The proof is similar to the proof of Lemma 2. ■

It is slightly easier to deal with the case where each of the k values is picked i.i.d. from $\mathbf{Z}/n\mathbf{Z}$. In this case, we shall let $P_{\tilde{a}}(\ell)=|\{i: a_i=\ell\}|/k$.

Lemma 9 *If the expected value of* $\| P_{\tilde{a}}^{*m}-U\|^2\to 0$ *as* $n\to\infty$ *where the expected value is over all choices of* $\tilde{a}=(a_1,\ldots,a_k)$ *where* a_i *is i.i.d. uniform on* $\mathbf{Z}/n\mathbf{Z}$, *then the expected value of* $\| P_{\tilde{a}}^{*m}-U\|^2\to 0$ *as* $n\to\infty$ *where the expected value is over all choices of* $\tilde{a}=(a_1,\ldots,a_k)$ *where* $\tilde{a}$ *is uniform over all elements of* $(\mathbf{Z}/n\mathbf{Z})^k$ *with* $a_i\neq a_j$ *when* $i\neq j$.

Proof. The probability $(a_1,\ldots,a_k)$ will have a duplication when a_i is i.i.d. uniform on $\mathbf{Z}/n\mathbf{Z}$ is less than $(k^2/n)\to 0$ as $n\to\infty$. Since the distance is always between 0 and 1, the result on expected values follows. ■

With each of the k values picked i.i.d. from $\mathbf{Z}/n\mathbf{Z}$, note that $k\hat{P}_{\tilde{a}}(j)$ is the sum of k mutually independent random variables with expected value 0 and length 1; the expected value results from the fact that $\sum_{s\in G}\rho(s)=0$ when ρ is a non-trivial irreducible representation of degree 1. Thus, by Theorem A.16 of [AS] (and symmetry considerations), we have

$$\begin{aligned}
&P(|k\hat{P}_{\tilde{a}}(j)|>2c(\log n)^{(a+1)/2})\\
&\quad\leqq P(Re(k\hat{P}_{\tilde{a}}(j))>c(\log n)^{(a+1)/2})+P(Re(k\hat{P}_{\tilde{a}}(j))<-c(\log n)^{(a+1)/2})\\
&\qquad+P(Im(k\hat{P}_{\tilde{a}}(j))>c(\log n)^{(a+1)/2})+P(Im(k\hat{P}_{\tilde{a}}(j))<-c(\log n)^{(a+1)/2})\\
&\quad\leqq 4\exp(-c^2(\log n)^{a+1}/2\lfloor(\log n)^a\rfloor)\\
&\quad<1/n^2
\end{aligned}$$

by appropriate choice of the constant c. Thus

$$P(|\hat{P}_{\tilde{a}}(j)|>2c/(\log n)^{(a-1)/2})<1/n^2$$

and

$$\begin{aligned}
E[|\hat{P}(j)|^{2((1+\varepsilon)/(a-1))(\log n/\log\log n)}]&\leqq\frac{(2c)^{((1+\varepsilon)/(a-1))(\log n/\log\log n)}}{(\log n)^{(1+\varepsilon)(\log n/\log\log n)}}+\frac{1}{n^2}\\
&\leqq\frac{(2c)^{((1+\varepsilon)/(a-1))(\log n/\log\log n)}}{n^{1+\varepsilon}}+\frac{1}{n^2}\\
&\leqq\frac{c_1}{n^{1+\varepsilon'}}
\end{aligned}$$

for sufficiently large n where c_1 and ε' are positive constants.

Thus

$$E[\| P_{\hat{a}}^{*m} - U \|^2] \to 0$$

as $n \to \infty$ if n is prime and

$$m \geqq (1+\varepsilon) \frac{a}{a-1} \frac{\log n}{\log k}.$$

By Lemmas 8 and 9, our proof of Theorem 3 is now complete. ■

Note that, since any irreducible representation on any element of an abelian group has length 1 and since there are n irreducible representations (all with degree 1) of an abelian group of order n, Theorem 3 applies to any abelian group of order n and not just $\mathbf{Z}/n\mathbf{Z}$.

Proof of Theorem 4

The proof of Theorem 4 is relatively short.

Let $k = \lfloor (\log n)^a \rfloor$ with $a < 1$. Since we are performing a random walk on an abelian group, the position of the random walk after a certain number of steps depends only on the number of times we pick each of the k points, which "support" the walk. A given point, after $(\log n)^b$ steps can be picked $0, 1, 2, \ldots, (\log n)^b$ times. So there are less than $(1+(\log n)^b)^k$ values at this stage of the random walk. Note that

$$\begin{aligned}(1+(\log n)^b)^k &= e^{(c+b \log\log n)k} \\ &\leqq \exp((c+b\log\log n)(\log n)^a) \\ &= n^{(c+b \log\log n)(\log n)^a/(\log n)}\end{aligned}$$

where $c \to 0$ as $n \to \infty$. Note that $(c+b\log\log n)(\log n)^a/(\log n) \to 0$ as $n \to \infty$ if $a < 1$. Thus after this many steps, the distance from uniform approaches 1. ■

Problems for further study

In the case where $k = \lfloor (\log n)^a \rfloor$, $a > 1$, most walks showed a sharp cutoff where the distance from uniform rapidly goes from near 1 to near 0. If $k = 2$, such a cutoff does not exist. (See [CDG] for an example; here the distance from uniform does not depend on the 2 points chosen if n is prime.) It would be interesting to see more general circumstances where typical random walks do or do not have this cutoff.

Another problem for further study is to get a clearer picture of the transition around $a = 1$ when $k = \lfloor (\log n)^a \rfloor$. What happens when $a = 1 + o(1)$?

Dou's results involving "random random walks" on arbitrary finite groups were described earlier; Theorem 3 improved Dou's results for abelian groups.

Are there some finite groups for which Dou's result is the best possible? If not, can we always improve the constant $a/(a-2)$ in Dou's result to $a/(a-1)$? Subsequent work has found answers to these 2 questions and will be described in [DH].

Acknowledgements. The author wishes to thank Persi Diaconis for suggesting the problem described in this paper. The author also wishes to thank Carl Dou for e-mail discussions which inspired the author to find Theorem 3. The author also wishes to thank David Wilson for pointing out the reference [AR]; this reference pointed out the theorem in [AS], which simplified an earlier version of the proof of Theorem 3. The author also wishes to thank the referee for some suggestions on style.

References

[AD] Aldous, D., Diaconis, P.: Shuffling Cards and Stopping Times. (Techn. Rep. No. 231) Department of Statistics, Stanford University 1985

[AR] Alon, N., Roichman, Y.: Random Cayley Graphs Expanders. Random Struct. Algorithms **5**, 271–284 (1994)

[AS] Alon, N., Spencer, J.: The Probabilistic Method. New York: Wiley 1992

[CDG] Chung, F., Diaconis, P., Graham, R.L.: A random walk problem arising in random number generation. Ann. Probab. **15**, 1148–1165 (1987)

[Di] Diaconis, P.: Group Representations in Probability and Statistics. Hayward, California: Institute of Mathematical Statistics 1988

[Do] Dou, C.: Studies of Random Walks on Groups and Random Graphs. Ph.D. thesis. Department of Mathematics, Massachusetts Institute of Technology 1992

[DH] Dou, C., Hildebrand, M.: Enumeration and Random Random Walks on Finite Groups. (in preparation)

[Gr] Greenhalgh, A.: Random walks on groups with subgroup invariance properties. Ph.D. thesis. Department of Mathematics, Stanford University 1989

[Gr2] Greenhalgh, A.: On a Model for Random Random-Walks on Finite Groups. (Preprint 1990)

[Hi] Hildebrand, M.: Rates of Convergence of Some Random Processes on Finite Groups. Ph.D. thesis. Department of Mathematics, Harvard University 1990