The Annals of Probability
1993, Vol. 21, No. 2, 710–720

RANDOM PROCESSES OF THE FORM $X_{n+1} = a_n X_n + b_n \pmod p$

BY MARTIN HILDEBRAND[1]

University of Michigan

This paper considers random processes of the form $X_{n+1} = a_n X_n + b_n \pmod p$, where $X_0 = 0$ and the sequences a_n and b_n are independent with a_n identically distributed for $n = 0, 1, 2, \ldots$ and b_n identically distributed for $n = 0, 1, 2, \ldots$. Chung, Diaconis and Graham studied such processes where $a_n = 2$ always; this paper considers more general distributions for a_n and b_n. The question is how long does it take these processes to get close to the uniform distribution? If a_n is a distribution on $\mathbf{Z}^+$ which does not vary with p and b_n is a distribution on $\mathbf{Z}$ which also does not vary with p, an upper bound on this time is $O((\log p)^2)$ with appropriate restrictions on p unless $a_n = 1$ always, $b_n = 0$ always or a_n and b_n can each take on only one value. This paper uses a recursive relation involving the discrete Fourier transform to find the bound. Under more restrictive conditions for a_n and b_n, this paper finds that a generalization of the technique of Chung, Diaconis and Graham shows that $O(\log p \log\log p)$ steps suffice.

1. Introduction. Computers often generate pseudorandom number sequences by using recurrences such as

$$X_{n+1} = aX_n + b \pmod p,$$

where p, a and b are integers. Although deterministic, the sequence $X_0 = 0, X_1, X_2, \ldots$ has some properties of a random number sequence. [8] contains a further discussion of these sequences. To increase "randomness," one may combine several generators. Thus we wish to investigate properties of the process

$$X_{n+1} = a_n X_n + b_n \pmod p,$$

where $X_0 = 0$ and a_n and b_n are independent random variables. Throughout this paper, we assume that a_n, $n = 0, 1, 2, \ldots$, are i.i.d. and that b_n, $n = 0, 1, 2, \ldots$, are i.i.d. Under many circumstances, elementary techniques, such as those found in [7], show that these processes converge to the uniform distribution. We ask for bounds, as a function of p, on the time it takes to get close to the uniform distribution.

An investigation of the properties of this process where $a_n = 2$ always appears in [2]. Such deterministic doubling makes the process converge to the uniform distribution much faster than where $a_n = 1$ always. Chapter 3 of [5] investigates cases where a_n always is a fixed integer $a > 1$; the techniques and

Received January 1991; revised November 1991.

[1]This paper is based on a portion of the author's Ph.D. dissertation at Harvard University.

AMS 1991 *subject classifications*. Primary 60B15; secondary 60J15.

Key words and phrases. Random processes, Fourier transform, convergence, uniform distribution, upper bound lemma, recursion.

results parallel those of [2]. That chapter also considers some cases where a varies with p. Chassaing [1] shows that a particular process, where $a_n = m$ and $b_n = 0, 1, \ldots, m-1$, uniformly is an optimal process on the integers mod p where there are m uniform choices at each step. Chassaing leaves open questions about processes with nonuniform choices as well as determining the actual rate of convergence for the random processes we are studying here.

One may wish to bound the rate of convergence of some processes where b_n has one value or a_n has different values such as 3 or 2. This paper presents a technique which enables one to get reasonable bounds. The key part of it is the use of a recursive formula for the discrete Fourier transform $\hat{P}_n(k)$ to bound $\max_{k \neq 0} |\hat{P}_{n+s}(k)|$ in terms of a constant times $\max_{k \neq 0} |\hat{P}_n(k)|$.

For each of the cases considered in this paper, a_n and b_n have distributions which do not depend on p. (Some cases where a_n or b_n have distributions which do depend on p are discussed in [5]. These processes sometimes have very different behavior.) a_n has finite support on $\mathbf{Z}^+$, and b_n has finite support on $\mathbf{Z}$. All constants in this paper may depend on the distributions for a_n and b_n. The result is the following theorem.

THEOREM 1. *With certain restrictions (described below) on p, $O((\log p)^2)$ steps suffice to make $\|P_n - U\| \to 0$ unless $a_n = 1$ always, $b_n = 0$ always, or both a_n and b_n can take on only one value.*

The restrictions on p in Theorem 1 depend on the following conditions:

1. $(a, p) = 1$ if $P(a_n = a) > 0$.
2. $(b, p) = 1$ for some b such that $P(b_n = b) > 0$.
3. $(\tilde{a}_1 - \tilde{a}_2, p) = 1$ for some $\tilde{a}_1$ and $\tilde{a}_2$ such that $P(a_n = \tilde{a}_1), P(a_n = \tilde{a}_2) > 0$.
4. Let $f(x) = \sum_{b \in \mathbf{Z}} P(b_n = b) e^{2\pi i b x}$. If b_n has more than one possible value, $|f(x)| = 1$ at a finite number of points e_i/c with $0 < e_i < c$ for all e_i and c is a positive integer chosen to be as small as possible so that the e_i's are integers. (There may be no such e_i for some distributions for b_n. In this case let $c = 1$.) The condition on p is that $(p, c) = 1$. (If b_n has only one possible value, view this condition as false for all p.)

The restrictions on p in Theorem 1 are that p must satisfy the first three conditions if b_n can take on only one value, and otherwise p must either satisfy conditions 1, 2 and 3 or satisfy conditions 1 and 4.

The restrictions on p above enable the theorem to be proved. Furthermore, due to problems such as parity, some restrictions are necessary for the process to converge to the uniform distribution. For example, if p is even, $b_n = 1$ always and $a_n = 2$ or 1 with probability $1/2$ each, then odd numbers appear in the stationary distribution with probability $2/3$.

Furthermore, we prove the following theorem.

THEOREM 2. *If b_n has more than one possible value and it is not the case that $a_n = 1$ always, then, if c defined in condition 4 is 1 and p is restricted to satisfy condition 1, then $O(\log p \log\log p)$ steps suffice to make $\|P_n - U\| \to 0$.*

2. Background. This section gives some background found in [2], [3] and [5].

Let $P_n(j) := \text{Prob}\{X_n = j\}$, $0 \le j \le p-1$. Define the variation distance between P_n and U by

$$\|P_n - U\| := \frac{1}{2}\sum_j |P_n(j) - 1/p|.$$

It is easy to show $\|P_n - U\| = \max_{A \subseteq \mathbf{Z}/p\mathbf{Z}} |P_n(A) - U(A)|$.

Let $q = q(p) = e^{2\pi i/p}$. Define the Fourier transform $\hat{f}: \mathbf{Z}/p\mathbf{Z} \to \mathbf{C}$ by

$$\hat{f}(k) := \sum_{j \in \mathbf{Z}/p\mathbf{Z}} q^{jk} f(j).$$

The following lemma will be the starting point for finding our bounds.

UPPER BOUND LEMMA.

$$\begin{aligned}\|P - U\|^2 &= \frac{1}{4}\left(\sum_j |P(j) - U(j)|\right)^2 \\ &\le \frac{1}{4} p \sum_j |P(j) - U(j)|^2 \\ &= \frac{1}{4}\sum_{k \ne 0} |\hat{P}(k)|^2.\end{aligned}$$

To prove this, we have used the Cauchy–Schwarz inequality, the Plancherel theorem, $\hat{U}(k) = 0$ if $k \ne 0$, and $\hat{P}(0) = \hat{U}(0) = 1$.

A more general upper bound lemma is described and used in [3]–[5].

To use the upper bound lemma here, we need to bound $|\hat{P}_n(k)|$ for $k \ne 0$; we do so by using the recursive formula for $\hat{P}_n(k)$.

3. Recursive formula for the Fourier transform. Let $P_n(j) = P(X_n = j)$. Let $q := q(p) := e^{2\pi i/p}$. The following lemma relates $\hat{P}_{n+1}$ to $\hat{P}_n$.

LEMMA 3.1. *If P satisfies condition* 1, *then*

$$\begin{aligned}\hat{P}_{n+1}(k) &= \sum_{a \ge 1}\sum_{b \in \mathbf{Z}} P(a_n = a)P(b_n = b)\hat{P}_n(ak)q^{bk} \\ &= \left(\sum_{a \ge 1} P(a_n = a)\hat{P}_n(ak)\right)\left(\sum_{b \in \mathbf{Z}} P(b_n = b)q^{bk}\right).\end{aligned}$$

PROOF. Suppose $X_{n+1} = j$, $a_n = a$ and $b_n = b$. Then $j = aX_n + b$ and $X_n = (1/a)(j - b)$. Note that $1/a$ exists in $\mathbf{Z}/p\mathbf{Z}$ since $(a, p) = 1$. Thus

$$P_{n+1}(j) = \sum_{a \ge 1}\sum_{b \in \mathbf{Z}} P(a_n = a)P(b_n = b)P_n((1/a)(j - b)).$$

Thus

$$\begin{aligned}
\hat{P}_{n+1}(k) &= \sum_{j \in \mathbf{Z}/p\mathbf{Z}} P_{n+1}(j) q^{jk} \\
&= \sum_{a \geq 1} \sum_{b \in \mathbf{Z}} \sum_{j \in \mathbf{Z}/p\mathbf{Z}} P(a_n = a) P(b_n = b) P_n((1/a)(j-b)) q^{jk} \\
&= \sum_{a \geq 1} \sum_{b \in \mathbf{Z}} \sum_{j \in \mathbf{Z}/p\mathbf{Z}} P(a_n = a) P(b_n = b) P_n((1/a) j) q^{(j+b)k} \\
&= \sum_{a \geq 1} \sum_{n \in \mathbf{Z}} \sum_{j \in \mathbf{Z}/p\mathbf{Z}} P(a_n = a) P(b_n = b) P_n(j) q^{(aj+b)k} \\
&= \sum_{a \geq 1} \sum_{b \in \mathbf{Z}} \sum_{j \in \mathbf{Z}/p\mathbf{Z}} P(a_n = a) P(b_n = b) P_n(j) q^{j(ak)} q^{bk} \\
&= \sum_{a \geq 1} \sum_{b \in \mathbf{Z}} P(a_n = a) P(b_n = b) q^{bk} \hat{P}_n(ak). \qquad \square
\end{aligned}$$

Lemma 3.1 also provides an expression relating $\hat{P}_{n+2}$ to $\hat{P}_n$:

COROLLARY.

$$\begin{aligned}
\hat{P}_{n+2}(k) = &\sum_{\tilde{a}_1 \geq 1} \sum_{\tilde{a}_2 \geq 1} \sum_{\tilde{b}_1 \in \mathbf{Z}} \sum_{\tilde{b}_2 \in \mathbf{Z}} P(a_n = \tilde{a}_1) P(b_n = \tilde{b}_1) q^{\tilde{b}_1 k} P(a_n = \tilde{a}_2) \\
&\times P(b_n = \tilde{b}_2) q^{\tilde{b}_2 \tilde{a}_1 k} \hat{P}_n(\tilde{a}_1 \tilde{a}_2 k).
\end{aligned}$$

4. The strategy. We shall relate the upper bound lemma to the use of the discrete Fourier transform. Our strategy will involve partitioning the integers mod p into sets and bounding the length of the discrete Fourier transform in each set. Once this becomes small enough, we use the upper bound lemma to bound the distance from uniform.

Partition $\mathbf{Z}/p\mathbf{Z}$ into $\{0\}$, S_{m_1}, $S_{m_1+1}, \ldots, S_{m_2-1}$, S_{m_2} and T. Let

$$S = \{0\} \cup \bigcup_{i=m_1}^{m_2} S_i.$$

In other words, let S be the complement of T in $\mathbf{Z}/p\mathbf{Z}$. Let d be a constant less than 1 and a be a positive integer.

Examine Table 1. If

$$\left|\hat{P}_{n+rm}(k)\right| \leq (c' + x(k,m)(1-c'))M_n,$$

where

$$M_n := \max_{k \neq 0} |\hat{P}_n(k)|,$$

$k \neq 0$, r is a constant integer and $x(k,m)$ is the entry in the column containing k at time $n + mr$, then we can conclude the following.

TABLE 1

	Outside S	S_{m_2}	S_{m_2-1}	S_{m_2-2}	$\cdots$
n	1	1	1	1	$\cdots$
$n + 1r$	0	1	1	1	$\cdots$
$n + 2r$	0	d	1	1	$\cdots$
$n + 3r$	0	d^2	1	1	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$n + (a + 1)r$	0	d^a	1	1	$\cdots$
$n + (a + 2)r$	0	d^{a+1}	d	1	$\cdots$
$n + (a + 3)r$	0	d^{a+2}	d^2	1	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$n + (2a + 1)r$	0	d^{2a}	d^a	1	$\cdots$
$n + (2a + 2)r$	0	d^{2a+1}	d^{a+1}	d	$\cdots$

LEMMA 4.1. *If $m_2 - m_1$ is no larger than a constant times* $\log p$ *and n is larger than a constant times* $(\log p)^2$, *then* $\|P_n - U\| \to 0$ *as* $p \to \infty$.

PROOF. If s is larger than a large enough constant times $\log p$, then if $k \neq 0$,

$$\left|\hat{P}_{n+s}(k)\right| \leq (c' + (1 - c')d)M_n.$$

Let $c'' = (c' + (1 - c')d)$; c'' is a constant less than 1.

Let t be an integer such that $(c'')^t < 1/p$; t can be made smaller than an appropriate multiple of $\log p$. Then

$$|\hat{P}_{st}(k)| \leq (1/p)M_0 = 1/p,$$

and

$$\|P_{st} - U\|^2 \leq \tfrac{1}{4}p(1/p)^2 = 1/(4p) \to 0$$

as $p \to \infty$. □

5. Proof when p satisfies conditions 1, 2 and 3. Theorem 1 is true in cases where p satisfies the first three conditions. Let $\tilde{d}$ and $\tilde{e}$ be the values $\tilde{a}_1$ and $\tilde{a}_2$ in condition 2 with $\tilde{e} > \tilde{d}$. In the expansion for $\hat{P}_{n+2}(k)$ in the Corollary of Lemma 3.1, there are terms

$$P(a_n = \tilde{d})P(a_n = \tilde{e})(P(b_n = b))^2(q^{bk}q^{b\tilde{d}k} + q^{bk}q^{b\tilde{e}k})\hat{P}_n(\tilde{d}\tilde{e}k).$$

Choose a nonzero value of b such that $P(b_n = b) > 0$. Note that $(q^{bk}q^{b\tilde{d}k} + q^{bk}q^{b\tilde{e}k}) = q^{bk}q^{b\tilde{d}k}(1 + q^{b(\tilde{e}-\tilde{d})k})$. Let $h(k) := |1 + q^{b(\tilde{e}-\tilde{d})k}|$; $h(k)$ is close to 2 when k is close to $lp/(b(\tilde{e} - \tilde{d}))$ for some integer l. Since $(p, b(\tilde{e} - \tilde{d})) = 1$, $h(k)$ is not 2 unless l is a multiple of $b(\tilde{e} - \tilde{d})$. In that case, $k = 0 \pmod p$.

Write k as $lp/(b(\tilde{e}-\tilde{d})) + \delta$, where l is an integer chosen so that $|\delta|$ is as small as possible. Let S be the integers mod p such that

$$|\delta| \leq \frac{p}{2b(\tilde{e} - \tilde{d})a_{\max}^2},$$

with $a_{\max} := \max_{\{a:\ P(a_n = a) > 0\}} a$. There exists a constant $c' < 1$ such that if $|\hat{P}_{n+2}(k)| \geq c' M_n$, then k must be in S.

Partition $S - \{0\}$ into sets S_i as follows. Let

$$S_i = \left\{k: k \in \mathbf{Z}/p\mathbf{Z}, |\delta| \in \left[2^{i-1}, 2^i\right)\right\} \cap S.$$

Note that $|\delta| \geq 1/(b(\tilde{e} - \tilde{d}))$. Thus s, the number of nonempty sets, is no more than a constant multiple of $\log p$. Let m_1 be the smallest i such that S_i is nonempty and m_2 be the largest such i. If $k \in S_i$, note that $\tilde{a}_1 \tilde{a}_2 k \notin \{0\} \cup S_i \cup S_{i-1} \cup \cdots \cup S_{m_1}$ if $P(a_n = \tilde{a}_1), P(a_n = \tilde{a}_2) > 0$ and $1 \neq \tilde{a}_1 \tilde{a}_2$.

Let d_0 be the sum of the absolute values of the coefficients of the $\hat{P}_n(k)$ terms in the expression for $\hat{P}_{n+2}(k)$ in terms of $\hat{P}_n(\cdots)$. Since $P(a_n = 1) < 1$, then $d_0 < 1$. Let d be such that $d_0 < d < 1$. Let a be such that $d^a < d - d_0$. Note that a does not depend on p.

One can show the following.

LEMMA 5.1. *For $k \neq 0$,*

$$|\hat{P}_{n+2m}(k)| \leq (c' + x(k, m)(1 - c'))M_n,$$

where $x(k, m)$ is the entry in the column containing k at time $n + mr$ in Table 1.

PROOF. We proceed by verifying a column from top to bottom and considering the columns from left to right. Note that M_n is nonincreasing as n increases. Thus the lemma is true for k outside S, which corresponds to the first column. The lemma is true for any entry, including all in the first row, with value 1.

Let us verify the lemma for another entry. Note that each row is nondecreasing as you go from left to right. Suppose $k \in S_i$. Suppose we have verified the lemma for all entries in columns to the left and all entries in the column for S_i in rows above the row for time $n + (m + 1)r$. Suppose also that the entry just above the entry we are examining is d^{f-a} and the entry just to the upper left of the entry we are considering is d^f. Then

$$\begin{aligned}
|\hat{P}_{n+(m+1)r}(k)| &\leq d_0 |\hat{P}_{n+mr}(k)| \\
&\quad + (1 - d_0) \max_{k_0 \in S_{i+1} \cup \cdots \cup S_{m_2} \cup \{k:\ k \notin S\}} |\hat{P}_{n+mr}(k)| \\
&\leq c' M_n + (1 - c') M_n \left(d_0 d^{f-a} + (1 - d_0) d^f\right) \\
&\leq c' M_n + (1 - c') M_n d^{f-a+1}.
\end{aligned}$$

So we make this entry d^{f-a+1} (as in Table 1) and satisfy the lemma. □

Thus the theorem holds true in this case.

6. Proof when p satisfies conditions 1 and 4. This case is similar to the previous case. We develop sets S_i as before, and we use Table 1 except that we let $r = 1$ instead of letting $r = 2$.

Define $f(x)$ and c as in condition 4.

Let $S = \{k \in \mathbf{Z}/p\mathbf{Z}: (k/p) = (\tilde{c}/c) + \delta$ with $|\delta| \leq 1/(2a_{\max}c)\}$ where $\tilde{c}$ is an integer chosen so that $|\delta|$ is as small as possible. Note that $\delta \neq 0$ if $k \neq 0 \pmod p$ since $(c, p) = 1$. Divide S into sets $\{0\}$ and $S_i := \{k \in S: p|\delta| \in [2^{i-1}, 2^i)\} \cap S$. There are no more than a constant times $\log p$ nonempty sets. Let m_1 be the smallest i such that S_i is nonempty and m_2 be the largest such i. If $a > 1$, $P(a_n = a) > 0$ and $k \in S_i$, then $ak \notin \{0\} \cup S_{m_1} \cup \cdots \cup S_i$.

Note that by the continuity of $f(x)$ and by Lemma 3.1, if $k \notin S$, then $|\hat{P}_{n+1}(k)| \leq c'M_n$ for some constant $c' < 1$.

Let $d_0 = P(a_n = 1)$. Let d be such that $d_0 < d < 1$. Note that $P(a_n = 1) < 1$. Let a be such that $d^a < d - d_0$.

By using an argument similar to that of Lemma 5.1, we can show the following.

LEMMA 6.1. *For* $k \neq 0$,

$$|\hat{P}_{n+m}(k)| \leq (c' + x(k, m)(1 - c'))M_n,$$

where $x(k, m)$ *is the entry in the column containing* k *at time* $n + mr$ *in Table* 1.

This lemma completes the proof of Theorem 1. □

7. Proof of Theorem 2. Theorem 2 is a generalization of results found in [2] and in [5], Chapter 3. Those results deal with cases where a_n can take on only one possible value a. There we can use the "a-ary" expansion of a fraction to get our bound. In this example, we cannot use the a-ary expansion, but we still use many of the techniques of [2] and [5], Chapter 3.

Suppose b_n is such that c in condition 4 is 1 and that p is restricted by condition 1.

Note that the values for $b_0, b_1, b_2, \ldots$ are independent of the values chosen for $a_0, a_1, a_2, \ldots$. So we can consider a specific choice of values for a_n, for example, $a_0 = 2$, $a_1 = 3$, $a_2 = 2$, $a_3 = 2$, $a_4 = 3, \ldots$. Given this sequence for a_n, let b_n, $n = 0, 1, 2, \ldots$, be independent random variables with the same distribution as before:

$$X_{n+1} = a_n a_{n-1} \cdots a_1 b_0 + a_n \cdots a_2 b_1 + \cdots + a_n b_{n-1} + b_n \pmod p.$$

Thus, given $k \in \mathbf{Z}/p\mathbf{Z}$,

$$(*) \qquad |\hat{R}_{n+1}(k)|^2 = g(a_n a_{n-1} \cdots a_1 k/p) g(a_n \cdots a_2 k/p) \times \cdots g(a_n k/p) g(k/p),$$

where R_n is the probability distribution of X_n given the specific values of a_i, $i = 0, \ldots, n-1$, $g(x) := |f(x)|^2$, and $f(x)$ is as defined in condition 4; $f(x)$ is just the discrete Fourier transform of b_n.

We shall show the following.

LEMMA 7.1. *If there are more than* $c_1 \log p \log\log p$ *values for* $a_i \neq 1$ *for* $i = 1, \ldots, n$, *where* c_1 *is a constant which depends only on the probability distributions for* a_n *and* b_n, *then*

$$\|R_n - U\| < \varepsilon_1(p),$$

where $\varepsilon_1(p) \to 0$ *as* $p \to \infty$.

Theorem 2 will follow from this lemma and the following propositions.

PROPOSITION 7.1. *There is a constant* c_2 *which depends only on* c_1 *and* $P(a_n = 1)$ *such that if* $n > c_2 \log p \log\log p$, *then the probability there are no more than* $c_1 \log p \log\log p$ *values of* $a_i \neq 1$ *is less than* $\varepsilon_2(p)$, *where* $\varepsilon_2(p) \to 0$ *as* $p \to \infty$.

PROPOSITION 7.2. *If* $n > c_2 \log p \log\log p$, *then*

$$\|P_n - U\| < \varepsilon_1(p) + \varepsilon_2(p).$$

PROOF OF LEMMA 7.1. Let $x \in [0, 1)$. Although we cannot use the a-ary expansion, we can make an expansion based on the specific values for $a_1, \ldots, a_n$:

$$A_1 A_2 \cdots A_n.$$

Start with x. Multiply it by a_n. Let A_1 be the integer part of the result. Continue on with the fractional part of the result. Multiply it by a_{n-1}. Let A_2 be the integer part of the result, and continue on with the fractional part. Continue through a_1.

Certain values A_j may have been produced by multiplication by 1. (In these cases A_j will always be 0.) Remove A_j from the expansion in this case.

Let $a_{\max} = \max_{\{a:\ P(a_n = a) > 0\}} a$. Consider $g(x)$ when x is in $[1/a_{\max}^2, 1 - (1/a_{\max}^2)]$. There $|g(x)| \leq b$, where b is a constant less than 1. Define

$$h(x) := \begin{cases} b, & \text{if } x \in \left[1/a_{\max}^2, 1 - \left(1/a_{\max}^2\right)\right], \\ 1, & \text{otherwise;} \end{cases}$$

$g(x) \leq h(\{x\})$, where $\{x\}$ denotes the fractional part of x.

Let $t = \lceil \log_2 p \rceil$. Choose $m = rt$ so that r is an integer depending on t. (We shall define r more explicitly later.)

In the expansion $A_1 A_2 \cdots A_n$ (with terms corresponding to $a_i = 1$ removed), there may be under m terms left. We shall define n and m so that the probability of picking such a_i is low. So let us examine the case where there are at least m terms left.

Are there conditions which assure us that $\{a_n \cdots a_j k/p\} \in [1/a_{\max}^2, 1 - (1/a_{\max}^2)]$? If $a_j = 1$, we shall not look for such conditions. Otherwise look at A_{n-j+1} (which was not removed). If A_{n-j+1} is neither 0 nor as large as possible, $\{a_n \cdots a_j k/p\} \in [1/a_{\max}^2, 1 - (1/a_{\max}^2)]$. Furthermore, unless A_{n-j+1} is the last term not removed, it is followed by another term A_l with

$l > n - j + 1$. Unless both A_{n-j+1} and A_l are both 0 or both as large as possible, then $\{a_n \cdots a_j k/p\} \in [1/a_{\max}^2, 1 - (1/a_{\max}^2)]$.

Define "blocks" B_{ki} each of length t on $A_1 A_2 \cdots A_n$ (with the terms corresponding to $a_i = 1$ removed). If A_i is as large as possible (i.e., $A_i = a_{n-i+1} - 1$), replace it with $a_{\max} - 1$:

$$(*) \le \prod_{i=1}^{r} b^{A(B_{ki})},$$

where $A(B)$ is the number of "generalized alternations" in the block B. A generalized alternation is either an alternation (i.e., where $\tilde{A}_i \neq \tilde{A}_{i+1}$, where $\tilde{A}_i$ and $\tilde{A}_{i+1}$ are adjacent values in the block) or a case where all values in the block are identical and not 0 or $a_{\max} - 1$. Thus

$$\|R_n - U\|^2 \le \frac{1}{4} \sum_{k \neq 0} \prod_{i=1}^{r} b^{A(B_{ki})}.$$

Note that, for any value of k/p, a block of this length must have at least one generalized alternation no matter what values we pick for a_i unless $k = 0 \pmod p$. Recall condition 1. Note also that all the blocks B_{ki} are distinct for a given i. The set of blocks $\{B_{ki}\}$ may in fact be different from the set of blocks $\{B_{k1}\}$. This poses a difference from the arguments of [2] and [5], Chapter 3.

Define C_{li}, $l = 1, 2, \ldots, (a_{\max}^t - (p-1) - 2)$ as follows. Let C_{li} be a t-tuple with entries in $\{0, 1, \ldots, a_{\max} - 1\}$ such that C_{li} is not B_{ki} for a nonzero $k \in \mathbf{Z}/p\mathbf{Z}$, C_{li} is not $C_{l'i}$ if $l' < l$, and C_{li} is not all 0's and is not all $(a_{\max} - 1)$'s. (Where there is an arbitrary choice for defining C_{li}, any choice suffices.) Thus, since $b \ge 0$,

$$\|R_n - U\|^2 \le \frac{1}{4} \sum_{k \neq 0} \sum_{i=1}^{r} b^{A(B_{ki})} + \frac{1}{4} \sum_{l=1}^{(a_{\max}^t - (p-1)-2)} \prod_{i=1}^{r} b^{A(C_{li})}.$$

Note that for a given i the set of all B_{ki} and C_{li} includes all t-tuples not identically equal to 0 or $a_{\max} - 1$.

To proceed, we use "interchanging," developed in [2].

If $a \le c$ and $b \le d$ are nonnegative real numbers, then $bc + ad \le cd + ab$. By applying this repeatedly, one can show the following.

LEMMA 7.2.

$$\sum_{j=1}^{s} \prod_{i=1}^{r} a_{\pi_i(j)} \le \sum_{j=1}^{s} \prod_{i=1}^{r} a_j = \sum_{j=1}^{s} a_j^r,$$

where π_i *is a permutation of* $1, \ldots, s$, *and* $a_j \ge 0$.

PROOF. To see this, consider the product whose first term is the smallest value. Now find the product whose second term is the smallest value. If the products are the same, perform no interchange. Otherwise, the two products look like

$$ace \quad \text{and} \quad bdf,$$

where $a \le b$ and $d \le c$.

If $e \le f$, then $ae \le bf$ and $ace + bdf \le ade + bcf$. Interchange just the terms d and c (here the second term) and increase the sum.

If $f < e$, then $df \le ce$ and $ace + bdf \le adf + bce$. Interchange everything after a and b (here only first term). After both kinds of interchange, the sum is larger and the smallest elements for the first two terms are in the same product.

Repeat the argument until the product has all terms which are the smallest. Note that the sum does not decrease at any point.

Recursively repeat the procedure by considering only the portion of the sum which does not include the product of the smallest value multiplied by itself r times. If there is only one term left, it is the product of the largest term multiplied by itself r times. □

We can use Lemma 7.2 to conclude that

$$\|R_n - U\|^2 \le \frac{1}{4} \sum_{\substack{\text{length } B = t \\ A(B) > 0}} b^{rA(B)}.$$

Let $M(j)$ denote the number of blocks B of length t with $A(B) = j$. Then

$$\begin{aligned} M(j) &\le a_{\max} \binom{t-1}{j} (a_{\max} - 1)^j + \delta_{j1}(a_{\max} - 2) \\ &\le 2a_{\max} \binom{t-1}{j} (a_{\max} - 1)^j \\ &\le 2a_{\max} \binom{t}{j} (a_{\max} - 1)^j, \end{aligned}$$

where δ_{ij} is the Kronecker delta function.

Thus

$$\begin{aligned} \|R_n - U\|^2 &\le \frac{1}{4} \sum_{j=1}^{t} M(j) b^{rj} \\ &\le \frac{2a_{\max}}{4} \sum_{j=1}^{t} \binom{t}{j} (b^r (a_{\max} - 1))^j \\ &= \frac{a_{\max}}{2} \left((1 + ((a_{\max} - 1) b^r))^t - 1 \right) \\ &\le \frac{a_{\max}}{2} (\exp(t(a_{\max} - 1) b^r) - 1). \end{aligned}$$

For

$$r \ge \frac{\log(t(a_{\max} - 1))}{-\log b} + d,$$

$\|R_n - U\|^2 \le (a_{\max}/2)(\exp(b^d) - 1)$. As $d \to \infty$, the right-hand side goes to 0. Thus Lemma 7.1 is proved, and Theorem 2 follows from Lemma 7.1 and Propositions 7.1 and 7.2. □

8. Questions for further study. This work still leaves open a number of questions. What is the "correct" answer? Are there examples where $O((\log p)^2)$ steps are necessary? Could there be different orders for the correct answers for different p? For example, [2] shows that $O(\log p)$ steps suffice for almost all p but that $O(\log p \log\log p)$ steps are necessary for certain p if $a_n = 2$ and $b_n = 1, 0, -1$ uniformly. Computer results suggest that a similar thing happens for $a_n = 2, 1$ uniformly and $b_n = 1$. Recent work [6] expanding upon the arguments in this paper shows such results.

Some random processes on finite groups exhibit a cutoff phenomenon, where the distance from uniform is close to 1 before a transition point and close to 0 afterwards. Formally, if n is the size of the group, then if, for all $\varepsilon > 0$, there is a function $f(n)$ such that

$$\| P_{[(1-\varepsilon)f(n)]} - U \| \to 1,$$

$$\| P_{[(1+\varepsilon)f(n)]} - U \| \to 0,$$

then the process has a cutoff phenomenon. For examples of this phenomenon, see [3] and [5]. It is still open whether there is a cutoff phenomenon for the processes studied in this paper and, if so, which processes exhibit it.

Acknowledgments. The author would like to thank Persi Diaconis for suggesting a problem which led to this paper as well as other suggestions, and John Imbrie for his encouragement and suggestions. The author would also like to thank the referees for a number of suggestions on style and for a reference.

REFERENCES

[1] CHASSAING, P. (1989). An optimal random number generator on $\mathbf{Z}_p$. *Statist. Probab. Lett.* **7** 307–309.

[2] CHUNG, F., DIACONIS, P. and GRAHAM, R. L. (1987). A random walk problem arising in random number generation. *Ann. Probab.* **15** 1148–1165.

[3] DIACONIS, P. (1988). *Group Representations in Probability and Statistics*. IMS, Hayward, Calif.

[4] DIACONIS, P. and SHAHSHAHANI, M. (1981). Generating a random permutation with random transpositions. *Z. Wahrsch. Verw. Gebiete* **57** 159–179.

[5] HILDEBRAND, M. (1990). Rates of convergence of some random processes on finite groups. Ph.D. dissertation, Dept. Mathematics, Harvard Univ.

[6] HILDEBRAND, M. (1992). Random processes of the form $X_{n+1} = a_n X_n + b_n \pmod p$ where b_n takes on a single value. Unpublished manuscript.

[7] KEMENY, J. G. and SNELL, J. L. (1976). *Finite Markov Chains*. Springer, New York.

[8] KNUTH, D. (1981). *The Art of Computer Programming* **2**, 2nd ed. Addison-Wesley, Reading, MA.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF MICHIGAN
ANN ARBOR, MICHIGAN 48109-1003