Saint Mary's College of California

Department of Mathematics

Senior Thesis

Pseudorandom Number Generation

Author:

Matthew Dami

Committee:

Dr. Andrew Conner

Dr. Hans de Moor

May 16, 2016

1 Introduction

Random number generation is the process of creating long uniform sequences of numbers where it is impossible to predict the next number in the sequence. There are two methods of generating random number sequences. The first method is True Random Number Generation which measures events in nature then a computer translates those events into a random sequence of numbers. A common way to do this is by measuring the radioactive decay from elements. The second method is Pseudorandom Number Generation which uses an algorithms to produce sequences of numbers that possess random-like qualities.

Pseudorandom number sequences are most commonly used in Monte Carlo simulations and other more general numerical simulations. These sequences are also heavily used in areas such as gambling and computer games.

In this paper we will concern ourselves with one common method used to generate such sequences while using the following definition for random number sequences. A **sequence of numbers can be considered random** if it meets both of the following criteria: the values are uniformly distributed over a defined interval or set, and it is impossible to predict future values based on past or present ones. While we don't have a strict mathematical definition for randomness we can test both of these criteria mathematically.

2 Linear Congruential Method

The linear congruential method, and its variations, are the most commonly used pseudorandom number generators. The linear congruential method can be easily implemented using a computer and can quickly generate sequences.

The linear congruential generator produces a sequence, $\{X_n\}$, defined by the following equation where a, c, m, X_0 are all real numbers and chosen in advance

$$X_{n+1} \equiv (aX_n + c) \text{ mod m}, \quad n \geq 0$$

We choose the four integers; m, a, c, and X_0 such that the following conditions are met:

Table 1: Choosing the constants

m	the modulus	$0 < m$
X_0	the starting value	$0 \leq X_0 < m$
a	the multiplier	$0 < a < m$
c	the increment	$0 \leq c < m$

The sequence created by the linear congruential generator is called a **linear congruential sequence**. While any combination of integers will produce a sequence, that sequence may not be pseudorandom. In this paper we will discuss the primary method of picking integers that will produce a pseudorandom sequence.

For an example let us consider the sequence when $X_0 = a = c = 7$, $m = 10$.

$$X_1 = a * X_0 + c \text{ mod } m = 7 * 7 + 7 \text{ mod } 10 = 6$$
$$X_2 = a * X_1 + c \text{ mod } m = 7 * 6 + 7 \text{ mod } 10 = 9$$

We continue this process to generate the sequence listed below:

$$7, 6, 9, 0, 7, 6, 9, 0, ...$$

These constants generate a sequence that illustrates two important issues with the Linear Congruential Method. First, we are not guaranteed to get a pseudorandom sequence for every choice of a, c, and m. Second, we will always have a repeating cycle of numbers. The **period** is the length of the shortest cycle of repeating numbers in a congruential sequence. So we would say that the sequence when $X_0 = a = c = 7$ and $m = 10$ has period of four. When we have a short period it becomes easy to predict what values come next.

Since we use modular arithmetic, (mod m) to generate a linear congruential sequence, $\{X_n\}$ we can have at most m different terms since each $X_n \in \mathbb{Z}_m$ where $\mathbb{Z}_m = \{0, 1, 2, ..., m-1\}$. In the event that $\{X_n\}$ has period m, we say that the sequence has a *maximum period.* A primary issue we will cover is how to choose our parameters to guarantee that our sequence will have maximum period.

We can immediately dismiss a few impractical values of a. If $a = 1$, then the resulting sequence would be generated by the equation $X_{n+1} \equiv (X_n + c)$ mod m, which would produce a sequence where each value increases by c making it very easy to predict what values will come next. The worst case situation would be $a = 0$, which would always produce a sequence with period two. Thus, we will assume that $a \geq 2$.

The letters a, c, m, and X_0 are used throughout this paper in the same manner as stated in Table 1. It is also useful to define $b = a - 1$ and to let $\mathbb{Z}^+$ denote all positive integers greater than 0, as they will be used in many equations and theorems. Following the logic from above we can say for practical reasons $b \geq 1$.

We will now prove an alternative non-recursive formulation of the linear congruential sequence which provides a tool for finding the n^{th} term of a sequence based on only a, c, m, and X_0.

Lemma 2.1. *Let $\{X_n\}$ be a linear congruential sequence. Assuming that $a \geq 2$ then*

$$X_{n+k} = (a^k X_n + (a^k - 1)c/b) \bmod m, \; k \geq 0, \; n \geq 0.$$

Proof. The proof is by induction on k. For the base case of $k = 0$ we can clearly see that the lemma is true since $X_{n+0} = a^0 X_n + (a^0 - 1)c/(a-1) \bmod m$. Assume the lemma is true for the $(n+k)^{th}$ term of $\{X_n\}$, and now consider the $(n+k+1)^{st}$ term.

$$\begin{aligned}
X_{n+(k+1)} &= aX_{n+k} + c \text{ (mod) } m, && \text{by the LCS} \\
&= a\left(a^k X_n + \frac{(a^k-1)c}{(a-1)}\right) + c \text{ (mod) } m, && \text{by induction assumption} \\
&= a^{k+1}X_n + \frac{a(a^k-1)c}{(a-1)} + c \text{ (mod) } m \\
&= a^{k+1}X_n + \left(\frac{a(a^k-1)}{(a-1)} + 1\right) c \text{ (mod) } m \\
&= a^{k+1}X_n + \frac{(a^{k+1}-1)c}{(a-1)} \text{ (mod) } m
\end{aligned}$$

We now have the required equation for the $(n+k+1)^{st}$ term of $\{X_n\}$. Therefore, by mathematical induction, the lemma is true for all $k \geq 0$. □

2.1 Choosing a Multiplier

When we choose a multiplier a, we need to keep in mind that we want to generate a sequence that has a maximum period. Note that having a sequence of maximum period doesn't mean that the sequence will be pseudorandom. For example we could have a sequence of maximum period but all the integers appear in linear order. Clearly, such a sequence would not be pseudorandom since it would be incredibly easy to predict. We will concern ourselves with the first part of this problem, obtaining a sequence with maximum period.

Theorem A. *The linear congruential sequence generated by (a, c, m, X_0) has a period of length m if and only if all the following hold:*
i) c is relatively prime to m;
ii) $b = a - 1$ is a multiple of p, for every prime p dividing m;
iii) b is a multiple of 4, if m is a multiple of 4.

Theorem A gives a very clear list of rules that we will follow in order choose the multiplier a. When we generate sequences that fulfill the conditions of Theorem A we will have a sequence that fulfills the first criteria of our definition of a random number sequence. Before we prove Theorem A we shall first go over several lemmas that are used in the final proof. The processes of proving this major result is modeled after the process used in The Art of Computer Programming.[1]

Lemma 2.2. *Let p be a prime number, and let $e \in \mathbb{Z}^+$, where $p^e > 2$. If*

$$X \equiv 1(\text{modulo } p^e) \text{ and } X \not\equiv 1(\text{modulo } p^{e+1}) \qquad \text{(A)}$$

then

$$X^p \equiv 1(\text{modulo } p^{e+1}) \text{ and } X^p \not\equiv 1(\text{modulo } p^{e+2}). \qquad \text{(B)}$$

Proof. Assume statement (A) is true, then $X = 1 + qp^e$ for some $q \in \mathbb{Z}$ and not a multiple of p. By the binomial theorem

$$\begin{aligned} X^p =&(1+qp^e)^p \\ =&1 + \binom{p}{1} qp^e + ... + \binom{p}{p-1} q^{p-1}p^{(p-1)e} + q^p p^{pe} \\ =&1 + qp^{e+1}\left(1 + \frac{1}{p}\binom{p}{2}qp^e + \frac{1}{p}\binom{p}{3}q^2p^{2e} ... + \frac{1}{p}\binom{p}{p}q^{p-1}p^{(p-1)e}\right) \end{aligned}$$

Since $e \geq 1$, the quantity contained in the parentheses (q') is an integer. If $1 < k < p$ then $\binom{p}{k}$ is divisible by p. Therefore $\frac{1}{p}\binom{p}{k}q^{k-1}p^{(k-1)e}$ is divisible by $p^{(k-1)e}$. The last term $\frac{1}{p}\binom{p}{p}q^{p-1}p^{(p-1)e} = q^{p-1}p^{(p-1)e-1}$ is divisible by p, because $(p-1)e > 1$ when $p^e > 2$. Thus every term, except for the first, is divisible by p. Therefore, $X^p = 1 + q'p^{e+1}$, where $q' \in \mathbb{Z}$ and not divisible by p, this implies that $X^p \equiv 1(\text{modulo } p^{e+1})$ and $X^p \not\equiv 1(\text{modulo } p^{e+2})$. □

Lemma 2.3. *Let $m = p_1^{e_1}p_2^{e_2}...p_h^{e_h}$ be the decomposition of a positive integer m into distinct prime factors. The length, α, of the period of the linear congruential sequence $\{X_f\}$ determined by (X_0, a, c, m) is the least common multiple of lengths α_g of the periods of the linear congruential sequences (X_0 mod $p_g^{e_g}$, a mod $p_g^{e_g}$, c mod $p_g^{e_g}$, $p_g^{e_g}$), $1 \leq g \leq h$.*

Proof. By mathematical induction on h, it is only necessary to prove that if i and u are relatively prime where GCD(i, u) = 1, then the period length, α, of the given sequence $\{X_f\}$ is the least common multiple of the lengths α_1, α_2 of the respective period lengths of the sequences respectively $\{Y_f\}$ and $\{Z_f\}$ given by (X_0 mod i, a mod i, c mod i, i) and (X_0 mod u, a mod u, c mod u, u).

We then have by definition of the three sequences have $Y_f \equiv X_f \pmod{i}$ and $Z_f \equiv X_f \pmod{u}$ for all $f \geq 0$. Because $(i, u) = 1$, then

$$X_f = X_g \text{ if and only if } Y_f = Y_g \text{ and } Z_f = Z_g. \qquad \text{(C)}$$

Let α' be the least common multiple of α_1 and α_2. Since the sequence $\{Y_f\}$ has period α_1, we can say $Y_f = Y_0$ if and only if f is is a multiple of α_1. By the same logic the sequence $\{Z_f\}$ has period α_2, we can say $Z_f = Z_0$ if and only if f is is a multiple of α_2. Notice that f must be a multiple of both α_1 and α_2. We will prove this with the argument below.

By definition, $\{Z_f\}$ has period α so $X_\alpha \equiv X_0$ mod m and $X_n \not\equiv X_0$ for all $0 < n < \alpha$. Since $X_\alpha \equiv X_0$ mod m then m divides $X_\alpha - x_0^x$. Because i divides m, then i also divides $X_\alpha - X_0$. Therefore $X_\alpha \equiv X_0$ mod i and by extension $Y_\alpha \equiv Y_0$ mod i. Let α_1 be the period of $\{Y_f\}$; and the argument above, $\alpha_1 \leq \alpha$.

By the Division Algorithm $\alpha = q\alpha_1 + r$ for $q \in \mathbb{Z}$, $0 \leq r < \alpha_1$. Also by lemma 2.1;

$$Y_0 \equiv Y_\alpha = Y_{q\alpha_1 + r} = a^{q\alpha_1} Y_r + \frac{(a^{q\alpha_1} - 1)c}{b} \text{ mod } i$$

but $a^{\alpha_1} \equiv 1$ mod i. To prove this, we want to show that $Y_{q'\alpha_1} \equiv Y_0$ mod i for any $q' \geq 1$. We know that $Y_{\alpha_1} \equiv Y_0$ is equivalent to $a^{\alpha_1} \equiv 1$ mod i by lemma 2.1. We can prove this by using induction on lemma 2.1:

$$
\begin{aligned}
Y_{q'\alpha_1} &= a^{\alpha_1} Y_{(q'-1)\alpha_1} + \frac{(a^{q\alpha_1}-1)c}{b} \bmod i \\
&= 1 Y_{(q'-1)\alpha_1} + 0 \bmod i \\
&= Y_0 \bmod i \qquad \text{by induction hypothesis}
\end{aligned}
$$

Thus $Y_0 \equiv Y_r \bmod i$ and recall with the Division Algorithm we stated that $0 \leq r < \alpha_1$. By definition of period, we can't have $0 < r < \alpha$, so $r = 0$. Therefore, $\alpha = q\alpha_1$.

We now wish to prove that $\alpha' = \alpha$. Since $X_f = X_{f+\alpha}$ for large values of f, we have $Y_f = Y_{f+\alpha}$ (therefore α is a multiple of α_1) and $Z_f = Z_{f+\alpha}$ (therefore α is a multiple of α_2), so α must be $\geq \alpha'$. Moreover, we also know that $Y_f = Y_{f+\alpha'}$ and $Z_f = Z_{f+\alpha'}$ for all large values of f; by equation (C) above, $X_f = X_{f+\alpha'}$ by a similar argument as above we find $\alpha | \alpha'$. Thus $\alpha = \alpha'$. □

Lemma 2.4. *If $a \equiv 3$ (mod 4), then $(a^{2^{e-1}} - 1)/(a-1) \equiv 0$ (mod 2^e), while $e > 1$.*

Proof. Suppose that $a \equiv 3$ (mod 4), then we can say that $a = 3 + 4l$ when $l \in \mathbb{Z}$. This is equivalent to $a = 1 + (2 + 4l)$ and also $a = 1 + 2(1 + 2l)$. This then implies that $a \equiv 1$ (mod 2).

We now have $a^2 = 9 + 24l + 16l^2 = 1 + 8(1 + 3l + 2l^2)$. Thus $a^2 \equiv 1$ (mod 8). By continued application we find that $a^4 \equiv 1$ (mod 16), $a^8 \equiv 1$ (mod 32), and so on. Let us show this by use of induction:

$$\begin{aligned} a^{2^n} &\equiv 1 \bmod 2^{n+2} && \text{Induction hypothesis} \\ a^{2^n} &= 1 + 2^{n+2}k && \text{Change form of equation} \\ a^{2^{n+1}} &= (1 + 2^{n+2}k)^2 && \text{Begin induction} \\ &= 1 + 2^{n+3}k + 2^{n+4}k^2 \\ &= 1 + 2^{n+3}(k + 2^{n+1}k^2) \\ a^{2^{n+1}} &= 1 \bmod 2^{n+3} && \text{Desired out come} \end{aligned}$$

Therefore $a^{2^{e-1}} - 1 \equiv 0 \pmod{g2^{e+1}}$. So $a^{2^{e-1}} - 1 = 2^{e+1}$ for some $g \in \mathbb{Z}^+$. Therefore $(a^{2^{e-1}} - 1)/2 = g2^e$, thus $(a^{2^{e-1}} - 1)/2 = 0 \pmod{2^e}$ and this yields the desired result. □

We now have the tools required to begin proving Theorem A. Due to Lemma 2.3 it is sufficient to prove Theorem A if we can prove that m is a power of a prime number. If $\{X_n \bmod p_i^{e_i}\}$ has maximal period for all i then $\{X_n\}$ has maximal period. Let $p_1^{e_1}...p_h^{e_h} = lcm(\alpha_1, ..., \alpha_h) \leq \alpha_1...\alpha_h \leq p_1^{e_1}...p_h^{e_h}$ be true if and only if $\alpha_g = p_g^{e_g}$ for $1 \leq g \leq h$.

We can then assume that $m = p^e$ when p is prime and $e \in \mathbb{Z}^+$. Because of our restrictions upon a it suffices to look at $a > 1$. Since the sequence is periodic there is no loss of generality in claiming that $X_0 = 0$

By Lemma 2.1

$$X_f = \left(\tfrac{a^f - 1}{a - 1}\right) c \bmod m.$$

If c is not relatively prime to m, then X_f can never be equal to 1. So condition (1) from theorem A is required. The sequence has maximal period if and only if the smallest positive value of f for which $X_f = X_0 = 0$ is $n = m$. Using lemma 2.1 and condition (1), it is sufficient to prove the following lemma.

Lemma 2.5. *Assume that $1 < a < p^e$, where p is prime and $3 > 1$. If α is the smallest positive integer for which $(a^\alpha - 1)/(a-1) \equiv 0$ (modulo p^e), then $\alpha = p^e$ if and only if $a \equiv 1$ (mod p) when $p > 2$, and $a \equiv 1$ (mod 4) when $p = 2$.*

Proof. Assume that $\alpha = p^e$.

$$(a^f - 1)/(a-1) \equiv 0 \text{ mod } p^e$$

implies that

$$a^{p^e} \equiv 1 \text{ mod } p^e$$

hence

$$a^{p^e} \equiv 1 \text{ mod } p$$

If $\alpha \not\equiv 1$ (mod p), then $(a^f - 1)/(a-1) \equiv 0$ (mod p^e) if and only if $a^f - 1 \equiv 0$ (mod p^e). The condition $a^{p^e} - 1 \equiv 0$ (mod p^e) then implies that $a^{p^e} \equiv 1$ (mod p); but $a^{p^e} \equiv a$ (mod p); hence $a \not\equiv 1$ (mod p) leads to a contradiction.

If $p = 2$ and $a \equiv 3$ (mod 4), then $(a^{2^{e-1}} - 1)/(a-1) \equiv 0$ (mod 2^e) by Lemma 2.4. Thus $\alpha \equiv 1$ mod 4 when $p = 2$, completing the first part of the proof.

These arguments show that $a \equiv 1$ mod p, when $p > 2$ and additionally $a \equiv 1$ mod 4, when $p = 2$ is equivalent to the statement that $a = 1 + qp^w$ where $p^w > 2$ and q is not a multiple of p.

It still needs to be shown that this condition is sufficient to make $\alpha = p^e$. By Lemma 2.2, we see that for any positive integer c, $a^{p^c} \equiv 1$ (mod p^{w+c}), and $a^{p^c} \not\equiv 1 \pmod{p^{w+c+1}}$. Therefore, $(a^{p^c} - 1)/(a - 1) \equiv 0 \pmod{p^c}$ and $(a^{p^c} - 1)/(a - 1) \not\equiv 0 \pmod{p^{c+1}}$, respectively.

In particular, $(a^{p^e} - 1)/(a - 1) \equiv 0 \pmod{p^e}$. Now the congruential sequence $(0, a, 1, p^e)$ has $X_f = (a^f - 1)/(a - 1) \pmod{p^e}$. Therefore it has a period length of α and $X_n = 0$ if and only if n is a multiple of α. Then p^e is a multiple of α. This can only happen when $\alpha = p^c$ for some c. While $(a^{p^c} - 1)/(a - 1) \equiv 0 \pmod{p^c}$ and $(a^{p^c} - 1)/(a - 1) \not\equiv 0 \pmod{p^{c+1}}$ implies that $\alpha = p^e$. Thus completing the proof. □

This also serves as the proof for Theorem A. Now let's take a look at an example.

Example: Let $m = 63$ and $X_0 = 0$. To meet the conditions of our theorem we first find the prime factors of m, which are 3^2 and 7.

Table 2: Candidates for the value of a

factors of b	value of b	value of a
$3 * 3 * 7$	63	64
$2 * 3 * 7$	42	43
$3 * 7$	21	22

We can clearly see that $a = 64$ won't work as it is greater than our value for m. This leaves us with the values 43 and 22. These are the only values of a that will produce a sequence with maximum period. It is important to notice that $b = 42$ contains a prime factor, 2, which is not a prime factor of 63. This is completely acceptable as long as 42 contains at least one of every prime factor of m. This satisfies condition (2) of our theorem. To meet condition (1) of the theorem we can simply set c equal to a prime number that is not a prime factor of m.

3 Testing a Sequence

A common way to test a given sequence for pseudorandomness is with a χ^2 test, which takes a sample sub-sequence and compares the frequency at which numbers appear with what would be expected in a true random sequence. This method, has a major drawback; it requires the sequence to already be generated and requires multiple tests of the sub-sequences. This method tends to be time consuming when we deal with sequences that have large periods. A better test to use on a Linear Congruential Sequence would to test the sequence based solely on the parameters a, m, and c.

We can begin by examining the *a priori* test presented in Theorem B and the corresponding proof. The idea of the theorem is that a truly random sequence will have $X_{n+1} < X_n$ about half the time.

Theorem B. *Let X_0, a, c, and m generate a linear congruential sequence with maximum period; let $b = a - 1$ and let d be the greatest common divisor (GCD) of m and b. The probability that $X_{n+1} < X_n$ is equal to $\frac{1}{2} + r$, where*

$$r = (2(c \ mod \ d) - d)/2m;$$

hence $|r| < d/2m$.

It is important to note that this theorem only applies when we choose X_0, a, c, and m such that the sequence generated has the maximum period length of m. Theorem A provides the tools necessary to choose our parameters so that we can apply Theorem B.

To prove Theorem B we require several techniques, some of which are interesting on their own. First we need to define a function $s : \mathbb{Z}_m \to \mathbb{Z}_m$ by

$$s(X) = (aX + c) \bmod m. \qquad (1)$$

Then we have, $X_{n+1} = s(X_n)$, and we can reduce the theorem to counting the number of $X \in \mathbb{Z}_m$ such that $s(X) < X$. We want to show that this number is

$$\tfrac{1}{2}(m + 2(c \bmod d) - d). \qquad (2)$$

When $a = 1$ we can see that $s(X) < X$ only when $X + c \geq m$, and that there are c such cases.

For a real number x; the floor operation, $\lfloor x \rfloor$, is equal to the largest integer less than or equal to x and the ceiling operation, $\lceil x \rceil$, is equal to the smallest integer greater than or equal to x.

Lemma 3.1. *Let α and β be real numbers and let n be an integer. Then*

$$\begin{aligned} \alpha < n < \beta \quad &\textit{if and only if} \quad \lfloor \alpha \rfloor < n \leq \lfloor \beta \rfloor; && (3) \\ \alpha < n < \beta \quad &\textit{if and only if} \quad \lceil \alpha \rceil \leq n < \lceil \beta \rceil. && (4) \end{aligned}$$

These formulas come directly from the definitions of the floor and ceiling operations.

For any X such that $0 \leq X < m$, let $k(X) = \lfloor (aX + c)/m \rfloor$. Then $s(X) = aX + c - k(X)m$ and $k(X) \leq (aX + c)/m$; therefore it follows that $s(X) < X$ is equivalent to

$$(k(X)m - c)/a \leq X < (k(X)m - c)/b. \qquad (5)$$

With Lemma 3.1 and equation (5) we have the following inequalities.

Lemma 3.2. *We find that $s(X) < X$ and $0 \leq X < m$ if and only if:*

i) $\lceil (km - c)/a \rceil \leq X < \lceil (km - c)/b \rceil$ *for some k,* $0 < k < a$*:*

or

ii) $\lceil m - c/a \rceil \leq X < m.$

Furthermore, the conditions i) and ii) are mutually exclusive.

Using these inequalities we are able to obtain the following the equation.

Lemma 3.3. *The number of integers X such that $s(X) < X$ is:*

$$|\{X | 0 \leq s(X) < X < m\}| = m + \sum_{0<k\leq b} \lceil (km - c)/b \rceil - \sum_{0<k\leq b} \lceil (km - c)/a \rceil. \quad (7)$$

Theorem B will be proved if we can show that the equations (7) and (2) are equal to each other. It will be helpful to introduce the following functions; the latter of which is the "sawtooth" function often used in the Fourier series. We will now define the δ and ζ functions as they will be used in the following proofs.

$$\delta(z) = \lfloor z \rfloor + 1 - \lceil z \rceil = \begin{cases} 1, & \text{if } z \in \mathbb{Z} \\ 0, & \text{if } z \notin \mathbb{Z} \end{cases} \qquad (8)$$

$$\zeta(z) = z - \lfloor z \rfloor - \frac{1}{2} + \frac{1}{2}\delta(z) = z - \lceil z \rceil + \frac{1}{2} - \frac{1}{2}\delta(z) \qquad (9)$$

The function $\zeta(z)$ has several useful properties:

$$\zeta(-z) = -\zeta(z) \text{ because } \lfloor -z \rfloor = -\lceil z \rceil; \quad (10)$$

$$\zeta(z+n)=\zeta(z),\ \text{if } n\in\mathbb{Z}; \quad (11)$$

The third property requires a short proof to show how we arrive at it as demonstrated in Lemma 3.4.

Lemma 3.4. *If $n \in \mathbb{Z}^+$ then;*

$$\zeta(z)+\zeta\left(z+\frac{1}{n}\right)+...+\zeta\left(z+\frac{n-1}{n}\right)=\zeta(nz) \quad (12)$$

Proof. Let $d = z - \lfloor z \rfloor$ be the decimal part of z. fix $n \in \mathbb{Z}$. There exists $k \in \mathbb{Z}$ such that

$$\frac{k}{n} \leq d < \frac{k+1}{n} \equiv 0 \leq nd-k < 1.$$

Then $\zeta(nz) = nd - k - \frac{1}{2} = nd - \frac{2k+1}{2}$. Also

$$\delta\left(z+\frac{i}{n}\right)=\begin{cases} d+\frac{i}{n}-\frac{1}{2}, \text{ for } 0\leq i\leq n-1-k \\ d+\frac{i}{n}-1-\frac{1}{2} \text{ for } n-k-1<i\leq n-1 \end{cases}$$

Therefore

$$\begin{aligned}\sum_{i=0}^{n-1}\zeta\left(z+\frac{i}{n}\right) &= nd+\sum_{i=0}^{n-1}\frac{i}{n}-k-\frac{n}{2} \\ &= nd+\frac{1}{n}*\frac{n(n-1)}{2}-\frac{2k+n}{2} \\ &= nd-\frac{2k+1}{2}\end{aligned}$$

□

Lemma 3.5. *By changing the notation of equation (7) we can get:*

$$\frac{1}{2}\left(m-1+\sum_{0<k\leq a}\delta\left(\frac{km-c}{a}\right)-\sum_{0<k\leq b}\delta\left(\frac{km-c}{b}\right)\right)$$
$$+\sum_{0<k\leq a}\zeta\left(\frac{km-c}{a}\right)-\sum_{0<k\leq b}\zeta\left(\frac{km-c}{b}\right). \qquad (13)$$

Proof. We first observe that: $\left\lceil \frac{km-c}{b} \right\rceil = \frac{km-c}{b} - \zeta\left(\frac{km-c}{b}\right) + \frac{1}{2} - \frac{1}{2}\delta\left(\frac{km-c}{b}\right)$ and we find that the equation for $\left\lceil \frac{km-c}{a} \right\rceil$ is similar. We can apply these observations to equation (7) to transform it into:

$$m + \sum_{0<k\leq b} \frac{km-c}{b} - \sum_{0<k\leq b} \zeta\left(\frac{km-c}{b}\right) + \frac{b}{2} - \frac{1}{2}\sum_{0<k\leq b} \delta\left(\frac{km-c}{b}\right)$$

$$- \sum_{0<k\leq a} \frac{km-c}{a} - \sum_{0<k\leq a} \zeta\left(\frac{km-c}{a}\right) - \frac{a}{2} + \frac{1}{2}\sum_{0<k\leq a} \delta\left(\frac{km-c}{a}\right)$$

$$= m + \frac{m}{b} * \frac{b(b+1)}{2} - c + \frac{b}{2} - \frac{m}{a} * \frac{a(a+1)}{2} + c - \frac{b}{2}$$

$$- \sum_{0<k\leq b} \zeta\left(\frac{km-c}{b}\right) - \frac{1}{2}\sum_{0<k\leq b} \delta\left(\frac{km-c}{b}\right) - \sum_{0<k\leq a} \zeta\left(\frac{km-c}{a}\right) + \frac{1}{2}\sum_{0<k\leq a} \delta\left(\frac{km-c}{a}\right)$$

For simplicity we will now omit the summations from the following simplifications, but they will be re-added in the final equation. Recall that $b = a - 1$.

$$m + \frac{m(b+1)}{2} + \frac{b}{2} - \frac{m(a+1)}{2} - \frac{a}{2}$$

$$m + \frac{ma}{2} + \frac{a-1}{2} - \frac{m(a-1)}{2} - \frac{a}{2}$$

$$m - \frac{m}{2} - \frac{1}{2}$$

We will now re-add the summations:

$$\frac{m}{2} - \frac{1}{2} - \sum_{0<k\leq b} \zeta\left(\frac{km-c}{b}\right) - \frac{1}{2}\sum_{0<k\leq b} \delta\left(\frac{km-c}{b}\right)$$

$$- \sum_{0<k\leq a} \zeta\left(\frac{km-c}{a}\right) + \frac{1}{2}\sum_{0<k\leq a} \delta\left(\frac{km-c}{a}\right)$$

Which we find is equal to equation (13):

$$= \frac{1}{2}\left(m - 1 + \sum_{0<k\leq a} \delta\left(\frac{km-c}{a}\right) - \sum_{0<k\leq b} \delta\left(\frac{km-c}{b}\right)\right)$$

$$+\sum_{0<k\leq a}\zeta\left(\frac{km-c}{a}\right)-\sum_{0<k\leq b}\zeta\left(\frac{km-c}{b}\right)$$

Thus we have shown that by changing notation we can transform equation (7) into equation (13). □

While this equation may look more complicated than what we began with, all of the sums that occur can be evaluated easily. This is due to the fact that m is relatively prime to a and we know that $(km-c) \bmod a$ takes on each of the values 0, 1,..., $a-1$ in some order since $0<k\leq a$. Therefore:

$$\sum_{0<k\leq a}\delta\left(\frac{km-c}{a}\right)=1, \text{ and } \sum_{0<k\leq a}\zeta\left(\frac{km-c}{a}\right)=0.$$

Let's show that the second formula holds true.

$$\begin{aligned}\sum_{0<k\leq a}\zeta\left(\frac{km-c}{a}\right) &= \sum_{0<k\leq a}\zeta\left(\frac{(km-c) \bmod a}{a}\right) && \text{by equations (10) and (11)}\\ &= \sum_{0<i\leq a-1}\zeta\left(\frac{i}{a}\right) && \text{by comments preceding the formula}\\ &= \left(\sum_{i=0}^{a-1}\frac{i}{a}\right)-\frac{(a-1)}{2}\\ &= \frac{1}{a}*\frac{a(a-1)}{2}-\frac{(a-1)}{2}\\ &= 0\end{aligned}$$

Since m is not relatively prime to b but c is, then $(km-c)/b \notin \mathbb{Z}$ for any k. Therefore we can say that the second sum in equation (13) is equal to zero. Finally we can see that for any $0<k\leq b$

$$\begin{aligned}\zeta\left(\frac{km-c}{b}\right) &= \zeta\left(\frac{km-d\lfloor c/d\rfloor - c \bmod d}{b}\right)\\ &= \zeta\left(\frac{km/d-\lfloor c/d\rfloor}{b/d}\right) - \frac{c \bmod d}{b} + \frac{1}{2}\delta\left(\frac{km/d-\lfloor c/d\rfloor}{b/d}\right),\end{aligned}$$

since $0 \leq (c \bmod d)/b < 1(b/d)$. The sum

$$\sum_{0<k\leq b} \zeta\left(\frac{km-c}{b}\right)$$

therefore becomes $\frac{1}{2}d - c \bmod d$. This is because

$$\zeta\left(\frac{km-c}{b}\right) = \zeta\left(\frac{km/d-\lfloor c/d\rfloor}{b/d}\right) - \frac{c \bmod d}{b} + \frac{1}{2}\delta\left(\frac{km/d-\lfloor c/d\rfloor}{b/d}\right)$$

and since m/d is relatively prime to b/d. This establishes Theorem B.

Trying to implement Theorem B and generate a LCS by hand can be a time consuming process. It is more efficient to use a programming language to test our choices for a, c, m, and X_0. An example of a program in Python can be found in Appendix A. This code computes the r value and can also generate the sequence to test that the expected probability of $X_{n+1} < X_n$ is equal to the actual results. This also allows the user to make sure that the generated sequence has maximal period.

From Theorem B we can see that when we have a sequence of maximum period, nearly any choice of a and c will give a good probability that $X_{n+1} < X_n$, except when we have a large d value.

4 Potency

For a LCS to be pseudorandom the value of a must produce a large amount of mixing between random numbers. When we have several options for our value of a we can examine each one to determine their potency.

Potency is a measurement of the mixing of numbers in a linear congruential sequence. This measurement is the smallest positive integer s that satisfies the equation:

$$(a-1)^s = b^s = 0 \pmod{m}$$

We are guaranteed an integer s exists when the multiplier a meets all the conditions of Theorem A. This is due to the fact that b is divisible by every prime factor of m. Because potency is a measure of mixing, we want to pick an a that has the largest s. The higher the potency, the better the mixing. In general. we want a potency of at least four in order to obtain minimal acceptable levels of mixing.[1]

There is an important connection between potency and Theorem B. In Theorem B, We want the value of d to be as close to one as possible as it will produce an r value that is closer to zero because $|r| < d/2m$. We have high potency when d is small compared to m, but we know that when the integer s is large we also have high potency. So if two potential values for a have the same potency we can also test the d value to determine if one is better than the other.

For an example consider $m = 4862025$ which has the prime factorization $\{3^4 5^2 7^4\}$. The table below shows some of the possible values of b and their corresponding values for s. For simplicity we have chosen values of b that will also be equal to its corresponding d value.

Table 3: Five Potential values for b and the corresponding potency

factors of b	value of b	$b^s = 0 \pmod{m}$
$3*5*7$	105	$105^4 = 0 \pmod{4862025}$
$3*3*3*5*7$	945	$945^4 = 0 \pmod{4862025}$
$3*3*5*7*7$	2205	$2205^2 = 0 \pmod{4862025}$
$3*3*3*5*5*7$	4725	$4725^4 = 0 \pmod{4862025}$
$3*3*3*5*5*7*7$	231525	$231525^2 = 0 \pmod{4862025}$

We can clearly see that options three and five are poor choices as they have s values of 2. This leaves us with three options for our multiplier, but due to Theorem B we can narrow it down further. By Theorem B, option one should also produce an r value that is closer to .5 than the other options. We compute the r values of options one, two, and four to be certain that our intuition is correct. We will be setting the increment as $c = 11$.

When $b = 105$:

$$r = \frac{2(11 \bmod 105) - 105}{2*4862025} = -8.53553817*10^{-6}$$

When $b = 945$:

$$r = \frac{2(11 \bmod 945) - 945}{2*4862025} = -9.4919298*10^{-5}$$

When $b = 4725$:

$$r = \frac{2(11 \bmod 4725) - 4725}{2*4862025} = -4.83646217*10^{-4}$$

This shows that option one is the best choice for our multiplier for when $m = 4862025$. It satisfies the conditions listed out in theorem A, has a large value for s, and has the closest d value that is close to zero. All of these tools are needed in order to choose good parameters for generating a linear congruential sequence. Theorem A gives us a way to produce a sequence with maximal period so our sequences will be uniform, potency allows us to measure if our sequences will be predicable, and with Theorem B we can test our choices of a, c, and m to help us determine if our parameters were well chosen. While Theorem A and potency allow us to test a and m, only Theorem B gives us a method for determining if our choice for c is good.

References

[1] Knuth, D. E. (1998) The Art of Computer Programming Volume 2: Semi-numerical Algorithms. Boston: Addison-Wesley, 1998. Print.

[2] ”Linear Congruential Generator.” Wikipedia. Wikimedia Foundation, 13 Mar. 2016. Web. 20 Apr. 2016.

[3] Nguyen, Hubert. (2008) GPU Gems 3. Upper Saddle River, NJ: Addison-Wesley, 2008. GPU Gems. Web. 20 Apr. 2016.

[4] ”True Random Number Service.” RANDOM.ORG. N.p., n.d. Web. 20 Apr. 2016.

Appendix A

```
import math
import fractions
def main():
    ##Counts the amount of times the sequence increased.
    shiftup = 0
    ##Counts the amount of times the sequence decreases.
    shiftdown = 0
    ##The initial number in the sequence.
    ##Since the sequence will have a maximum period,
    ##where we begin our sequence is trivial.
    seed = int()
    ##The multiplier.
    mult = input(”What␣is␣the␣multiplier?␣”)
    mult = int(mult)
    #The increment.
    incre = input(”What␣is␣the␣increment?␣”)
    incre = int(incre)
    ##The modulus.
    mod = input(”What␣is␣the␣modulus?␣”)
    mod = int(mod)
    ##Initializes the list that will hold the sequence.
    ##Places the seed into the sequence.
    holdinglist = [seed]
    #Theoretical tests to determine if sequence will be psudorandom.
    #Setting peramaters.
    b = mult - 1
```

```
d = fractions.gcd(b, mod)
print ("The␣value␣of␣d␣is␣" , d)
r = (2*(incre % d) - d)/(2*mod)
##The while loop stops once a number has been repeated.
while ( holdinglist.count(seed) !=2) :
    xseed = seed
    seed = (((mult*seed)+incre)% mod)
    holdinglist.append(seed)
    #The if statement checks to see if the sequence
    ##has increased or decreased
    if xseed < seed:
        shiftup += 1
    else:
        shiftdown += 1
#Removes the repeated number.
k = holdinglist.pop()
## Zed is the length of the period.
zed = len(holdinglist)
print ("This␣sequence␣has␣a␣period␣of␣", zed)
##The probability that X_(n+1) < X_n
prob = shiftdown / zed
print (r)
r += 0.5
print ("The␣probability␣that␣X_(n+1)␣<␣X_n␣=␣", r)
##Prints the generated sequence
## Causes problems with longer sequences.
print(holdinglist)
```