A NOTE ON INITIAL DIGITS OF RECURRENCE SEQUENCES

Siniša Slijepčević
Department of Mathematics, Bijenicka 30, University of Zagreb, 10000 Zagreb, Croatia
e-mail: slijepce@cromath.math.hr
(Submitted October 1996-Final Revision February 1997)

1. INTRODUCTION

The problem set in [3] is: What is the probability that initial digits of n^{th} Lucas and Fibonacci numbers have the same parity? We answer the problem and demonstrate a simple technique that provides answers on similar questions regarding relative frequency ("probability") of initial digits in almost any linear recurrence sequence.

The probability that a random number from the sequence X_n belongs to the set A (which has a certain property) is defined as the value of the limit (if it exists):

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}1_A(X_i),$$

where 1_A denotes the characteristic function of the set A: $1_A(x)=1$ if $x\in A$, $1_A(x)=0$ if $x\notin A$.

The main tool in the proofs will be the well-known Weyl-Sierpinski equidistribution theorem [1] in its simplest form.

Theorem: Let q be an irrational number, $\widetilde{T}_n=p+nq$ be a sequence and $T_n=\{\widetilde{T}_n\}$ its fractional part. Then the probability that T_n is in the interval $[a,b)$, $0\le a<b\le 1$, is $b-a$. (The fractional part of irrational translation is uniformly distributed on $[0,1)$.)

2. CALCULATION OF PROBABILITIES

The following two lemmas prove that anything that is close enough to irrational translation is uniformly distributed on $[0,1)$. We will apply it to the logarithms of linear recursive sequences.

Lemma 1: Let $\widetilde{T}_n=p+nq$, q irrational, $T_n=\{\widetilde{T}_n\}$ its fractional part, and $\widetilde{X}_n$, $X_n=\{\widetilde{X}_n\}$ another sequence such that $\lim_{n\to\infty}|\widetilde{X}_n-\widetilde{T}_n|=0$. Then the probability that some X_n falls in the interval $A=[a,b)$, $0\le a<b\le 1$ is $b-a$.

Proof: Given $\varepsilon>0$, there exists n_1 such that, for each $m>n_1$, $|\widetilde{X}_m-\widetilde{T}_m|<\frac{\varepsilon}{4}$. If

$$A_\varepsilon=\left[a+\frac{\varepsilon}{4},\,b-\frac{\varepsilon}{4}\right),$$

this means that, for each $m\ge n_1$, $T_m\in A_\varepsilon$ implies $X_m\in A$. Equivalently, for each $m\ge n_1$, $1_A(X_m)\ge 1_{A_\varepsilon}(T_m)$.

There exist $n_0\ge n_1$ such that, for each $n>n_0$, $\frac{1}{n}\sum_{m=0}^{n_1-1}1_{A_\varepsilon}(T_m)\le\frac{\varepsilon}{2}$ (the sum is constant, so we choose n_0 large enough).

For each $n > n_0$, we calculate

$$\begin{aligned}\frac{1}{n}\sum_{m=0}^{n-1} 1_A(X_m) &\ge \frac{1}{n}\sum_{m=n_1}^{n-1} 1_A(X_m) \ge \frac{1}{n}\sum_{m=n_1}^{n-1} 1_{A_\varepsilon}(T_m) \\ &\ge \frac{1}{n}\sum_{m=n_1}^{n-1} 1_{A_\varepsilon}(T_m) + \frac{1}{n}\sum_{m=0}^{n_1-1} 1_{A_\varepsilon}(T_m) - \frac{\varepsilon}{2} \\ &= \frac{1}{n}\sum_{m=0}^{n-1} 1_{A_\varepsilon}(T_m) - \frac{\varepsilon}{2}.\end{aligned} \tag{1}$$

Applying the equidistribution theorem, we get

$$\liminf_{n\to\infty} \frac{1}{n}\sum_{m=0}^{n-1} 1_A(X_m) \ge \lim_{n\to\infty} \frac{1}{n}\sum_{m=0}^{n-1} 1_{A_\varepsilon}(T_m) - \frac{\varepsilon}{2} = b - a - \varepsilon.$$

Since it is valid for each ε, $\liminf_{n\to\infty} \frac{1}{n}\sum_{m=0}^{n-1} 1_A(X_m) \ge b-a$. We apply the same reasoning for intervals $[0, a)$ and $[b, 1)$. Since $1_{[0,a)}(x) + 1_{[a,b)}(x) + 1_{[b,1)}(x) = 1$, we get

$$\begin{aligned}\limsup_{n\to\infty} \frac{1}{n}\sum_{m=0}^{n-1} 1_{[a,b)}(X_m) &\le 1 + \limsup_{n\to\infty}\left(-\frac{1}{n}\sum_{m=0}^{n-1} 1_{[0,a)}(X_m)\right) + \limsup_{n\to\infty}\left(-\frac{1}{n}\sum_{m=0}^{n-1} 1_{[b,1)}(X_m)\right) \\ &= 1 - \liminf_{n\to\infty} \frac{1}{n}\sum_{m=0}^{n-1} 1_{[0,a)}(X_m) - \liminf_{n\to\infty} \frac{1}{n}\sum_{m=0}^{n-1} 1_{[b,1)}(X_m) \le b - a.\end{aligned} \tag{2}$$

Now we have $\liminf_{n\to\infty} \frac{1}{n}\sum_{m=0}^{n-1} 1_A(X_m) = \limsup_{n\to\infty} \frac{1}{n}\sum_{m=0}^{n-1} 1_A(X_m) = b - a$, so it converges and the lemma is proved.

The following lemma is a simple generalization that can be proved using the same technique (the proof is omitted).

Lemma 2: Let $\widetilde{T}_n = p + nq$, and let $\widetilde{X}_n^1, \widetilde{X}_n^2, \ldots, \widetilde{X}_n^k$ be k sequences such that, for each i, we have $\lim_{n\to\infty} |\widetilde{X}_n^i - \widetilde{T}_n| = 0$. Let q be irrational, and let $X_n^1, \ldots, X_n^k, T_n$ be the fractional parts of the sequences. Then the probability that, for random n, $X_n^1 \in [a_1, b_1), \ldots, X_n^k \in [a_k, b_k)$ is $b-a$, where

$$\bigcap_{i=1}^{k} [a_i, b_i) = [a, b).$$

Example 1: Probability that the first digit of F_n and that of L_n have the same parity is $\log_{10} \frac{648}{245}$.

Proof: Let $\widetilde{X}_n = \log_{10} F_n - \log_{10} p$, $\widetilde{Y}_n = \log_{10} L_n$, X_n, Y_n their fractional parts, $p = 1/\sqrt{5}$, and $q = (\sqrt{5}+1)/2$. As an example, we calculate the probability that, for given n, F_n begins with 1 and L_n begins with 3.

F_n begins with 1 if and only if, for some $k \in \mathcal{N}$, $F_n \in [10^k, 2\cdot 10^k)$, which is equivalent to

$$\begin{aligned} \log_{10} F_n &\in [k, \log_{10} 2 + k) \\ \Leftrightarrow \widetilde{X}_n = \log_{10} F_n - \log_{10} p &\in [k + \log_{10}\sqrt{5}, k + \log_{10} 2\sqrt{5}) \\ \Leftrightarrow X_n = \{\widetilde{X}_n\} &\in [\log_{10}\sqrt{5}, \log_{10} 2\sqrt{5}).\end{aligned} \tag{3}$$

L_n begins with 3 if and only if

$$Y_n = \{\log_{10} L_n\} \in [\log_{10} 3, \log_{10} 4). \tag{4}$$

Since $\widetilde{X}_n$ and $\widetilde{Y}_n$ asymptotically converge to $\widetilde{T}_n = n\log_{10} q$, $\log_{10} q$ irrational, we can apply Lemma 2. The probability is $\log_{10} 4/3$.

In the following table, we calculated all nonzero probabilities that, for random n, F_n begins with i and L_n begins with j (probability is $\log_{10} x$).

F_n	4	5	6	7	8	8	9	1	1	1	2	2	2	3	3	3	4	4
L_n	1	1	1	1	1	2	2	2	3	4	4	5	6	6	7	8	8	9
x	$\frac{\sqrt{5}}{2}$	$\frac{6}{5}$	$\frac{7}{4}$	$\frac{8}{7}$	$\frac{\sqrt{5}}{2}$	$\frac{9\sqrt{5}}{20}$	$\frac{10}{9}$	$\frac{3\sqrt{5}}{5}$	$\frac{4}{3}$	$\frac{\sqrt{5}}{2}$	$\frac{\sqrt{5}}{2}$	$\frac{6}{5}$	$\frac{\sqrt{5}}{2}$	$\frac{7\sqrt{5}}{15}$	$\frac{8}{7}$	$\frac{\sqrt{5}}{2}$	$\frac{9\sqrt{5}}{20}$	$\frac{10}{9}$

Summing the probabilities from the appropriate columns, we prove the formula. This probability (approximately 0.42241) is in accordance with the numerical test from [3]—4232 out of 10000.

In this example, we can avoid using Lemma 2, noting the fact that the initial digits of F_n and L_n are the same as the initial digits of $p \cdot q^n, q^n$. However, using the described technique, we can answer the same question about, e.g., 5th leftmost digits of F_n and L_n.

It can easily be proved (checking that $[(1-\sqrt{5})/2]^n$ is small enough for large n) that the entries in the table are the only possible ones (and not only with positive probability) [2].

Example 2: We will call a linear recurrence sequence Y_n *random enough* if the root q_1 of the characteristic polynomial that has the largest absolute value is real, positive, not a rational power of 10, unique and has multiplicity 1, and P_1 in equation (5) is positive.

The probability that a random enough recursive sequence begins with the digits 1997 is $\log_{10}(1+\frac{1}{1997})$.

Proof: We can then write the sequence in explicit form [4]:

$$Y_n = P_1 q_1^n + P_2(n) q_2^n + \cdots + P_k(n) q_k^n, \tag{5}$$

where P_1 is a real number and $P_2, \ldots, P_k$ are polynomials. Y_n begins with 1997 if and only if, for some $k \in \mathcal{N}$,

$$Y_n \in [1997 \cdot 10^k, 1998 \cdot 10^k) \Leftrightarrow \tag{6}$$

$$\Leftrightarrow \{\log_{10} Y_n\} \in [\log_{10} 1.997, \log_{10} 1.998). \tag{7}$$

Since $\lim_{n\to\infty} |\log_{10} Y_n - (\log_{10} P_1 + n \cdot \log_{10} q_1)| = 0$, we can apply Lemma 1. The probability is the length of the interval in (7).

We can prove the following formula in the same way.

Example 3: The probability that the i^{th} leftmost digit of a random enough recursive sequence is j obeys the generalized Benford's law (see [3] and [5]):

$$P = \log_{10} \prod_{k=10^{i-2}}^{10^{i-1}-1} \left(1 + \frac{1}{10k+j}\right)$$

for $i \geq 2$, and $P = \log_{10}(1+\frac{1}{j})$ for $i = 1$.

Lemma 1 implies as well that the fractional part of the logarithm of the random enough recursive sequence is uniformly distributed on $[0, 1)$.

REFERENCES

1. R. Durett. *Probability: Theory and Examples.* Belmont, CA: Wadsworth-Brooks, 1991.
2. P. Filipponi. "F_n and L_n Cannot Have the Same Initial Digit." *Pi Mu Epsilon Journal* **10.1** (1994):5-6.
3. P. Filipponi & R. Menicocci. "Some Probabilistic Aspects of the Terminal Digits of Fibonacci Numbers." *The Fibonacci Quarterly* **33.4** (1995):325-31.
4. T. N. Shorey & R. Tudeman. *Exponential Diophantine Equations.* Cambridge: Cambridge University Press, 1986.
5. L. C. Washington. "Benford's Law for Fibonacci and Lucas Numbers." *The Fibonacci Quarterly* **19.2** (1981):175-77.

AMS Classification Numbers: 11B39, 11N37

❖❖❖