PROPERTIES OF SEQUENCES

Simon Norton
Cambridge, England

Anil (WW 41 (2) p223) asks whether there is a theory which explains how the periodic behaviour of Fibonacci letter-sequences works. The answer is yes, there is indeed such a theory, and it has many other applications, both mathematical and practical. Though this theory has nothing to do with words, I hope that readers will be interested in its results. However, most of the proofs will be beyond the scope of this article.

We start by defining a **generalized Fibonacci number-sequence** as a doubly infinite sequence $A = (a_n)$, where n ranges over all integers (negative, zero and positive), such that for all n we have $a_{n+1} = a_n + a_{n-1}$. We regard two such sequences A and B as **equal** if they differ only in labelling, i.e. there is a k such that $a_n = b_{n+k}$ for all n, and we call A and B **similar** if there are q and k such that $a_n = q.b_{n+k}$ for all n.

If the sequence A consists of integers, then we call it **primitive** if these integers have no common factor greater than 1. It is in fact sufficient that any pair of consecutive terms, say a_0 and a_1, are integers with no common factor. A sequence of integers with a common factor is called **imprimitive,** and may be **reduced** to a similar primitive sequence by dividing every term by the highest common factor. If two primitive sequences are similar, then either they are equal or one is equal to the negation of the other.

Two sequences are of great importance: the **Fibonacci numbers** $F = f_n = \ldots, 2, -1, 1, 0, 1, 1, 2, \ldots$ and the **Lucas numbers** $L = l_n = \ldots, -4, 3, -1, 2, 1, 3, 4, \ldots$, which are named after the French mathematician Edouard Lucas. Both of these sequences have the property that if reversed only the signs of (some of) the terms are changed, and they and their negations are the only primitive sequences with this property. It is natural to label these sequences in such a way that the central elements (0 for F and 2 for L) have suffix zero, so that f_n and l_n are the terms n places after 0 and 2 (if n is positive) or $-n$ places before (if n is negative).

The Fibonacci and Lucas sequences have many interesting arithmetical properties; two that we shall use later are $F_n L_n = F_{2n}$ and $L_{4n} = L_{2n}^2 - 2$.

We say that a sequence has an **interval** of n if it has two equal terms n places apart. Then it can be seen that F has intervals of 1, 3 and twice any odd number, while L has intervals of any multiple of 4. For every n there is a unique primitive sequence (up to negation) with interval n; for example if $n = 5, 7, 9$ these sequences are $\ldots, 5, -2, 3, 1, 4, 5, \ldots$, $\ldots, 13, -7, 6, -1, 5, 4, 9, 13, \ldots$ and $\ldots, 17, -10, 7, -3, 4, 1, 5, 6, 11, 17, \ldots$ respectively. The repeated number is always a divisor of f_n.

If a is an integer and u is a positive integer, there are unique integers b and c such that $a = ub + c$ and c is between 0 and $u - 1$ inclusive. We call c the **remainder** obtained by dividing a by u.

If we start with a generalized Fibonacci number-sequence, replace each term by the remainder obtained by dividing it by a fixed number u, and then encode the possible remainders $0, 1, \ldots, u-1$ into an **alphabet** of u symbols, we get a **generalized Fibonacci letter-sequence.** We assume that for every u there is a **standard alphabet** of u symbols which we will use unless otherwise stated. If u is not more than 26, the symbols we will use are Z for 0 and the first $u - 1$ letters of the Roman alphabet for $1, \ldots, u - 1$.

The definition of equality carries over to letter-sequences, but the definitions of primitivity and similarity need to be changed: a letter-sequence is primitive unless there is an integer greater than 1 that divides not only two consecutive terms of any number-sequence A that gave rise to it, say a_0 and a_1, but also the alphabet-length u. The quotient of u by its highest common factor with a_0 and a_1 we call the **reduced alpabet-length.** For two letter-sequences to be similar, not only must it be possible to choose similar number-sequences that give rise to them, but the reduced alphabet-lengths must be the same.

Before giving an example to explain this let us note that every generalized Fibonacci letter-sequence is periodic. We therefore denote it by its period, so, for example, the sequence $\ldots ZAAZAAZAA \ldots$, obtained from F by using the standard alphabet of length 2, is denoted by (ZAA). This is the same sequence as (AZA) or (AAZ).

If we take the number-sequences $2F = \ldots, 4, -2, 2, 0, 2, 2, 4, 6, \ldots$ and $3F = \ldots, 6, -3, 3, 0, 3, 3, 6, 9, \ldots$, and convert them to letter-sequences using the standard alphabets of length $u = 3$ or 9 respectively, we get $\ldots AABZBBAZ \ldots$ and $\ldots FFCZCCFZ \ldots$ respectively. In fact the sequences $(AABZBBAZ)$ and $(FFCZCCFZ)$ are periods. These sequences are similar because the highest common factors of the two (a_o, a_1, u)'s, $(0, 2, 3)$ and $(0, 3, 9)$ respectively, are 1 and 3, giving a reduced alphabet-length of 3 in each case.

This illustrates an important result: if two letter-sequences are similar then they can be obtained from each other by a substitution cipher, in this case taking (Z, A, B) to (Z, F, C). There is also a partial converse result: if two letter-sequences are related by a substitution cipher and have the same reduced alphabet-length, then either every letter that appears in the period does so at most once, or the two letter-sequences are similar.

If we choose a u, we define three period numbers which together make up its **period pattern:** The **basic** period is the number of places separating repeats of letters in the letter-sequence corresponding to the Fibonacci number-sequence. Equivalently, if we are using a standard alphabet, it is the number of places separating Z's (as every repeated letter is immediately preceded by a Z). The **full** and **short** periods are the largest and smallest possible periods for any primitive letter-sequence with alphabet-length u. (Note that the last two are the same when all periods have the same length, as is often the case.)

We now write out all primitive letter-sequences for u between 1 and 8 inclusive using the standard alphabets of the relevant lengths.

$u = 1:\ (Z)$
$u = 2:\ (ZAA)$
$u = 3:\ (ZAABZBBA)$
$u = 4:\ (ZAABCA)\ (ZCCBAC)$
$u = 5:\ (ZAABCZCCADZDDCBZBBDA)\ (ACDB)$
$u = 6:\ (ZAABCEBACDAEZEEDCADECBEA)$
$u = 7:\ (ZAABCEAFZFFEDBFA)\ (ZBBDFCBEZEECADEB)\ (ZDDAEFDCZCCFBACD)$
$u = 8:\ (ZAABCEZEEBGA)\ (ZGGFECZCCFAG)\ (FGEDAEFCADEA)\ (BACDGCBEGDCG)$

This gives us a lot of examples to play with. The period-patterns for $u = 1, \ldots, 8$ can be seen to be $(1, 1, 1)$, $(3, 3, 3)$, $(4, 8, 8)$, $(5, 20, 4)$, $(12, 24, 24)$, $(8, 16, 16)$ and $(6, 12, 12)$ respectively. If $u = 4$ or 7 then all the sequences are similar, but this is not the case for $u = 5$ (obviously, as the two sequences have different periods) or $u = 8$ (when the first two are similar, as are the last two, but these pairs are not similar to each other). The first sequence shown is in each case the Fibonacci sequence; the last is the Lucas sequence.

The second sequence for $u = 5$ is special in another way: let's call it of **geometric** type. This means that u divides $a_0.a_2 - a_1^2$. We use this name because the above condition is equivalent to one stating that any finite part of the letter-sequence can be obtained from a geometric series. In this case the series $(1, 3, 9, 27, \ldots)$ gives rise to $(ACDB)$ repeated indefinitely.

We are now in a position to state some results about the period-pattern corresponding to any u.

1. If $u = 1$, 2 or 5, then the period pattern is as given above.

2. If u is a prime whose last digit is 3 or 7, then the full period is $2(u + 1)$ divided by an odd number. If u is a prime whose last digit is 1 or 9, then its full period is a divisor of $u - 1$ (or equivalently and analogously with the previous case, $2(u - 1)$ divided by an even number).

3. If u is a prime other than 2 or 5, then from the full period P (which is always even) the period-pattern can be calculated as follows. If P is divisible by 2 but not 4, then the period-pattern is $(P, P, P/2)$; if P is divisible by 4 but not 8, the period pattern is $(P/4, P, P)$; and if P is divisible by 8, the period pattern is $(P/2, P, P)$. According to which of these holds we say that u belongs to the first, second and third cases respectively.

4. If u is a prime power, say $u = p^m$, then the period pattern is usually obtained from that for p by multiplying each period by p^{m-1}. There is an exception when $p = 2$ and m is at least 3, in that the basic period is 3.2^{m-2} rather than 3.2^{m-1}. If there are any other exceptions, they must involve primes greater than 100,000; for any such primes p all the periods for p^m are obtained from those for p by multiplying by a power of p less than p^{m-1}.

5. If u is divisible by more than one prime, then we split it into prime power factors and take the period-pattern for each of them. Then each period for u is the least common multiple of the corresponding periods for the factors of u. For example when $u = 15$ we consider the period-patterns for 3 and 5, namely $(4, 8, 8)$ and $(5, 20, 4)$, and take the least common multiplies to get $(20, 40, 8)$ as the period-pattern for 15. To take another example, the period-pattern for $u = 13$ is $(7, 28, 28)$ (an example of the second case), so taking the least common multiple with the period-pattern for $u = 2$, i.e. $(3, 3, 3)$, we get $(21, 84, 84)$. This, then, explains the occurrence of 84 as the period for every primitive letter-sequence using the Roman alphabet.

Anil also made some comments on the frequencies with which various letters occurred in periods. Before we proceed to this, let us make some remarks on the utility of the above theory.

Suppose u is a Mersenne prime, i.e. a prime of form $2^n - 1$. Suppose furthermore that the last digit of u is 3 or 7. Then its full period is 2^{n+1} divided by an odd number, which means that it must be 2^{n+1} itself. Its basic period is therefore 2^n. That means that f_{2^n} is divisible by u but $f_{2^{n-1}}$ is not, so that $l_{2^{n-1}} = f_{2^n}/f - 2^{n-1}$ is divisible by u. Lucas proved the converse of this, i.e. that if u ends in a 3 or 7 and divides $l_{2^{n-1}}$ then u is prime. This provides the basis for a highly efficient test for the primality of Mersenne numbers: start with $3 = l_2$, apply the operation $z \mapsto z^2 - 2$, which doubles the suffix of l_{2m}, $n-2$ times, and $2^n - 1$ is prime if and only if it divides the resulting number (assuming always that $2^n - 1$ ends in a 3 or 7).

If alternate numbers of the Lucas sequence are taken we get the sequence $\ldots, 47, 18, 7, 3, 2, 3, 7, 18, 47, \ldots$ in which each term when multiplied by 3 is the sum of the two terms on either side. A similar sequence can be constructed by replacing 3 by any other integer greater than 2; for example if we choose 4 we get the sequence $\ldots, 194, 52, 14, 4, 2, 4, 14, 52, 194, \ldots$ – which also obeys the property that its $2n$th term can be obtained from its nth by squaring and subtracting 2. And here the condition that the Mersenne number $2^n - 1$ is prime if and only if it divides the 2^{n-2}th term holds for all n except 2.

When one reads of the discovery of a new very large prime, it is almost always a Mersenne number, and its primality has been checked by using this test or a variation thereof.

Finally, primality testing is basic to Internet cryptography, so that the theory behind the periods of Fibonacci sequence has ramifications for our everyday lives.

Now we return to the subject of the composition of period-sequences. The theory gets very complicated when u is non-prime, so we assume that u is prime and therefore rename it p. Then the following facts are always true:

1. All periods for primitive letter-sequences have the same length except for at most one similarity class. Any period of this class, if it exists, is always of geometric type.
2. Periods of geometric type exist exactly when $p = 5$ or p ends in a 1 or 9. The number of similarity classes of such periods is 1 when $p = 5$, otherwise 2. So, if p is not 5, there is one period of geometric type of full length. (More generally, a period for any number u can have geometric type exactly when every prime factor of u is 5 or a number ending in 1 or 9; furthermore the factor 5 cannot occur more than once. And if the number of different prime factors other than 5 is m, then the number of similarity classes of geometric type is 2^{m-1}.)
3. Leaving aside any periods of geometric type, the number of similarity classes is $p - 1$ divided by the short period if p ends in a 1 or 9, and $p + 1$ divided by the short period if p ends in a 3 or 7.
4. The Fibonacci and Lucas sequences for p are similar when p is of the first or third type, but not when p is of the second type.
5. In any period for any prime p, every letter occurs at most four times.
6. In any period of geometric type, every letter occurs at most once.
7. In any period that is neither of Fibonacci nor Lucas type, every letter occurs at most twice.
8. In any period that is of Fibonacci but not Lucas type, every letter occurs an even number of times. (Such a period exists exactly when p belongs to the second case.)
9. In any period of Lucas type, every letter occurs at most three times if p belongs to the first case; otherwise exactly one of the two possible odd occurrence-frequencies (1 or 3) happens.

As we said earlier, it is usually hard to devise any general theories for what happens when u is not prime; however for the full Roman alphabet for which $u = 26$ the theory is much easier because the full periods for its prime factors, 3 for the prime 2 and 28 for the prime 13, have no common factor. What this means is that the occurrence-frequencies can be obtained by multiplying those for $u = 2$ (1 or 2, since the period is (ZAA) by those for $u = 13$. If $u = 13$, we either have the Fibonacci type (for which letters can occur 0, 2 or 4 times), or the Lucas type (for which letters can occur 0, 2 and 3 times). So, for $u = 28$, the occurrence-frequencies can be 0, 2, 4 or 8 for the Fibonacci sequence and 0, 2, 3, 4 or 6 for the Lucas sequence (and any sequence is similar to one of these two).

We now turn our attention to a completely different type of sequence, namely that described on page 170 of the same issue. It can be described as follows.

We start with a finite sequence of positive integers, say $(1,1,2,2,1)$. We split it into blocks of equal numbers, in this case $((1,1),(2,2),(1))$. We then replace a block of a b's by a, b, thus in this case getting $(2,1,2,2,1,1)$. (In other words, we read our original sequence as "two ones, two twos, one one" and take the integers in the above phrase.) Let us call the result the **derived sequence** of the one we started with.

We may then start with a random sequence and repeatedly replace it by its derived sequence. Let us call the sequence obtained by doing this n times the nth derived sequence of the original. On page 170 we were told that it was not Conway who discovered this idea. However he definitely did analyse what happens. Roughly speaking, the situation is as follows. For reasons that will become apparent, he called the theory "chemistry".

1. If A and B are sequences, we say that the sequence AB (i.e. A followed by B) **splits** into A and B if the nth derived sequence of AB can be obtained by concatenating that of A with that of B. This happens whenever the last term of A (which is the same as the last term of any of its derived sequences) never appears as a first term in the nth derived sequence of B.

2. If we take any sequence and split it as far as we can, we call the components **elements.**

3. Most elements can only occur in the first few derivations, whatever sequence we start with. Those that can continue to occur indefinitely we call **stable.**

4. If the sequence A ends in $(1,n)$, where n is an integer greater than 1, then the behaviour of A under derivation is not changed if one replaces n by any other integer greater than 1. In particular, if A is an element so is any of the replacement sequences. However, it is not true that if A is stable then the replacement sequences are necessarily stable.

5. To be precise, for any n greater than 3, there are exactly two stable elements ending in $(1,n)$, and these remain stable whatever n we use (including 2 and 3). All other stable elements, and there are 92 of them, only involve the integers 1, 2 and 3.

6. If we call two elements **isotopes** if they differ only in the last term as described above, then the 92 above elements include 20 pairs of isotopes, plus 52 other elements, making a total of 72 elements up to isotopy. (The numbers 72, and 92 above, compare with the 81 stable elements in "real" chemistry.)

7. There is a fixed real number h such that if we start with any sequence other than $(2,2)$ (which is its own derivation), and divide the length of its nth derivation by h^n, then the result tends to a limit, depending on the initial sequence, as n tends to infinity. In other words, all sequences other than $(2,2)$ grow exponentially at approximately the same rate.