Automatic sequences, logarithmic density, and fractals

Jason Bell

A sequence is k-*automatic* if its n'th term is generated by a finite state machine with n in base k as the input.

Examples of automatic sequences

The Thue-Morse sequence

$$0110100110010110100 10 \cdots$$

This sequence is 2-automatic.

e.g., n=13, then $n = 1101$ in base 2. Output $= 1$.

The Rudin-Shapiro sequence.

$$1, 1, 1, -1, 1, 1, -1, 1, 1, 1, 1, -1, -1, \ldots.$$

The n'th term of this sequence is given by -1 to the number of occurrences of "11" in the binary exmpansion of n.

This sequence is 2-automatic.

The *kernel* of a k-automatic sequence,

$$\{f(n) \mid n \geq 0\},$$

(or k-kernel) is the set of all subsequences of the form

$$\{f(k^a n + b) \mid n \geq 0\}$$

with $a \geq 0$ and $0 \leq b < k^a$.

Example: Take the Thue-Morse sequence

$$0110100110010110100 10\cdots.$$

If $TM(n)$ denotes the n'th term of the sequence, then

$$TM(2n) = TM(n) \qquad TM(2n{+}1) = 1{-}TM(n).$$

Thus either

$$TM(2^a n + b) = TM(n)$$

for all n, or

$$TM(2^a n + b) = 1 - TM(n)$$

for all n.

The 2-kernel of the Thue-Morse sequence consists of only two sequences; namely, the Thue-Morse sequence and its "opposite."

In general, we expect the k-kernel of a sequence to be infinite.

In the case of k-automatic sequences, however, the k-kernel is finite.

THEOREM: A sequence is k-automatic if and only if its k-kernel is finite.

Allouche and Shallit used the k-kernel characterization of automatic sequences to give a natural generalization: Regular sequences.

Notice that the collection of (real) sequences forms a $\mathbb{Z}$-module, which we call S.

$$(1, 3, \pi, 2, 0, \ldots) + (0, 1, 1, -1, 2, \ldots)$$

$$= (1, 4, \pi + 1, 1, 2, \ldots).$$

$$3 \cdot (1, 3, \pi, 2, 0, \ldots) = (3, 9, 3\pi, 6, 0, \ldots).$$

Given a real sequence $\{f(n) \mid n \geq 0\}$, we let $M(f, k)$ denote the $\mathbb{Z}$-submodule of S generated by all sequences in the k-kernel of $\{f(n)\}$.

Definition: We say that a sequence $\{f(n)\}$ is k-*regular* if the the module $M(f,k)$ is finitely generated.

Remark: A k-automatic sequence is k-regular.

After all, if the k-kernel is finite, then clearly the module $M(f,k)$ is finitely generated.

Some examples of regular sequences

Let $p(x)$ be a polynomial with real coefficients. Then the sequence $p(0), p(1), \ldots$ is k-regular for every k.

More generally, if

$$a_0 + a_1 x + a_2 x^2 + \cdots$$

is a rational power series with no poles inside the unit disc, then $a_0, a_1, \ldots$ is k-regular for every k.

An example from history?

- In the first century, Josephus along with 40 other rebels were hiding in a cave from the Romans during the Roman-Jewish war.

- Faced with certain death, the 41 men decided killing themselves was preferable to being killed by the Romans.

- Suicide was considered much worse than murder in Judaism (possibly).

- The men decided to form a circle and kill every other person in the circle till only one was left, the last person would then commit suicide.

Let $J(n)$ denote the last person to die in the Josephus circle of size n.

Then one sees that

$$J(2n) \;=\; 2J(n) - 1$$

$$J(2n+1) \;=\; 2J(n) + 1.$$

These relations show that the sequence $\{J(n)\}$ is 2-regular.

Incidentally, when $n = 41$,

$J(41) = 2J(20) + 1 = 2(2J(10) - 1) + 1 = 19.$

A different example.

Let $f(n)$ count the number of 1's in the binary expansion of n. Then $f(n)$ is 2-regular.

To see this, notice that $f(2^a n+b)$ is just the number of 1's in the binary expansion of b added to the number of 1's in the binary expansion of n. Thus $f(2^a n+b) = f(n)+f(b)$.

It follows that $M(f,2)$ is generated by

$$(f(0), f(1), \ldots)$$

and the constant sequence

$$(1, 1, 1, \ldots).$$

As we have seen, a k-regular sequence may be unbounded, while a k-automatic sequence only takes on finitely many values.

THEOREM: A k-regular sequence is k-automatic if and only if it only takes on finitely many values.

A CHARACTERIZATION OF REGULAR SEQUENCES

We will give a ring-theoretic characterization of regular sequences. To do this, let us think of a natural number as being a word on the alphabet $\{1, 2, \ldots, k\}$, by using a modified base k-representation.

e.g., $k = 3$, $n = 13$. Then

$$20 = 1 \times 3^2 + 3 \times 3 + 2,$$

and so we associate the word 132 to $n = 13$.

We can therefore think of a sequence taking values in an abelian group A as being a map from $f : \{1, 2, \ldots, k\}^*$ into A.

We can characterize the f that are k-regular using two key properties: the iteration property and the shuffle property.

THE ITERATION PROPERTY

We say that f has the *iteration property* if for any m and any words $W_1, \ldots, W_m,$ there exist integer polynomials $\Phi_1, \ldots, \Phi_m$ with constant coefficient 1 such that for any words W, W' we have

$$\prod_{i=1}^{m} \Phi_i(t_i) \sum_{i_1=0}^{\infty} \cdots \sum_{i_m=0}^{\infty} f(W W_1^{i_1} \cdots W_m^{i_m} W') t_1^{i_1} \cdots t_m^{i_m},$$

is a polynomial in $A[t_1, \ldots, t_m] = A \otimes_{\mathbb{Z}} \mathbb{Z}[t_1, \ldots, t_m]$.

THE SHUFFLE PROPERTY

we say that f has the d-*shuffle property* if for any words $W, W_1, \ldots, W_d, W'$ we have

$$\sum_{\sigma \in S_d} \mathrm{sgn}(\sigma) f(W W_{\sigma(1)} W_{\sigma(2)} \cdots W_{\sigma(d)} W') = 0.$$

THEOREM: A sequence is k-regular if and only if it has the iteration and the shuffle property for some number d.

COROLLARY: A sequence is k-automatic if and only if it has the iteration and the shuffle property for some number d and takes on only finitely many values.

BACK TO AUTOMATIC SEQUENCES:

Using the ideas presented before we can give a new approach to looking at the *logarithmic density* with which a letter occurs in a k-automatic sequence.

First, what is the logarithmic density?

Given a subset S of the natural numbers, the ordinary density of S is just

$$\lim_{n\to\infty} \frac{\#\{s \in S \mid s \leq n\}}{n}$$

if the limit exists.

The problem is that the limit often doesn't exist.

The *logarithmic density* of S is just

$$\lim_{x\to\infty} \frac{1}{\log x} \sum_{s\in S, s\leq x} 1/s$$

if the limit exists.

The problem is that the limit often doesn't exist. But if the ordinary density exists, then the logarithmic density also exists.

SOME HISTORY:

Cobham (1972) proved that if $f(n)$ is a k-automatic sequence and a is a value assumed by f, then the logarithmic density of

$$S = \{n \mid f(n) = a\}$$

exists.

Later Allouche, Mendès France and Peyrière found a formula for the logarithmic density of S. This formula is hard to digest on a slide, so we'll look at some examples.

Let $f : \mathbb{N} \to \{0,1\}$ be the map which sends any string whose ternary expansion ends in $100\cdots01$ to 1 and all other strings to 0. Let S be the set of n whose ternary expansion ends in $100\cdots0$.

Then the logarithmic density of S is given by

$$\frac{C}{\log 3}\sum_{n\in S}\frac{1}{(3n+2)(3n+3)},$$

for some rational constant C.

Our approach is to work from the "left" rather than from the "right", using the shuffle and iteration properties.

Doing things this way we get a formula in terms of products rather than in terms of sums. There is no obvious way of (to me) of showing that the result obtained this way is the same as the result obtained by Allouche et al. by just doing some manipulations of series.

For example, if $f : \mathbb{N} \longrightarrow \{0, 1\}$ is the map which sends any string ending in $100 \cdots 01$ to 1 and all other strings to 0 and S is the set where f is equal to 1. Then we get that the logaritmic density of S is

$$\frac{1}{6 \log 3}\Big(\log \prod_{n=1}^{2} (1 + 1/n) \Big) \; = \; 1/6.$$

In general, how do we get the logaritmic density from the finite state machine?

For each state s in the automaton, associate a subset $S(s)$ of the natural numbers.

Then there exist rational number C_s for each state in the automaton such that the logarithmic density is given by

$$\sum_s \frac{C_s}{\log\, k}\Big(\log \prod_{n \in S(s)} (1+1/n)\Big).$$

Let $f(n)$ be k-automatic. Allouche and Shallit asked whether the log density of the set S where $f(n) = a$ has logarithmic density $\log\, a/\log\, b$.

I don't know if it is true. We do get this result for simple automata in which the sets $S(s)$ are finite. (This seems to be the case for most *naturally occurring* automatic sequences.)

It is doubtful (to me), that the following example will simplify into a ratio of logarithms of rational numbers. Of course, it could be true for some other reason.

CHALLENGE PROBLEM: Let $f(n)$ be the 3-automatic sequence which is 1 if the base 3 expansion of n begins $10\cdots01$ and is zero otherwise.

Then the logarthimic density of the set S where f is 1 is given by

$$\frac{1}{\log 3}\Big(\prod_{i=1}^{\infty}(1+1/(3^i+1))\Big).$$

Show that

$$\Big(\prod_{i=1}^{\infty}(1+1/(3^i+1))\Big)$$

is not the d'th root of some rational number.

(Harder) Show that the above expression cannot be written as a ratio of logs of rational numbers.

Incidentally, if we use the formula of Allouche et al. we get the following expression for the logarithmic density:

$$\frac{C}{\log 3}\Big(\sum_{k=0}^{\infty} 1/(3^{k+1}+2)(3^{k+1}+3)\Big)$$

$$+\frac{C}{\log 3}\Big(\sum_{n=0}^{\infty} f(n)/(3n+3)(3n+2)\Big)$$

$$+\frac{C}{\log 3}\Big(\sum_{n=0}^{\infty} 2f(n)/(3n+1)(3n+3)\Big)$$

Here C is a rational constant.

FRACTALS

Allouche and Shallit observed that the logarithmic density of sets arising from k-automatic sequences are often of the form $C \log a / \log k$. Many fractals coming from looking at base k expansions have a *fractal dimension* of this form.

e.g. The Cantor set. If we magnify the Cantor set by a factor of 3, we obtain two copies of the original Cantor set. What can we say the *dimension* of the Cantor set should be?

Let d denote its dimension. Then $3^d = 2$ and so $d = \log 2 / \log 3$.

This is a good heuristic, but we need something more rigorous. Fortunately, there are many good ways of defining the dimension.

Let's look at the Cantor set again. Notice that if we regard the Cantor set as a map $f : [0,1] \to \{0,1\}$. Then

$$f(.0x_1x_2\cdots) = f(.x_1\cdots).$$

$$f(.2x_1x_2\cdots) = f(.x_1\cdots).$$

$$f(.1x_1\cdots) = 0.$$

This reminds us of the automatic property.

Given a subset S of the natural numbers and a positive integer k, we can associate a subset $X(S) \subseteq [0,1]$ by taking all numbers whose base k expansion $.a_1a_2\cdots$ has $a_i = 0$ if $i \notin S$.

THEOREM: If the ordinary density of S exists, then the fractal dimension of $X(S)$ exists; moreover the two quantities are equal.

For most automatic sequences the sets S we obtain do not have ordinary density. Nevertheless, we can "glue" fractals of the form $X(S)$ together using our automaton to obtain a fractal whose fractal dimension is the same as the logarithmic density of S.

Moreover, this fractal will be like the Cantor set in that it will have the automatic-like property.

I know I'm being vague, but it's complicated ....