SELF-REFERENCING SEQUENCE

SARAH QUINN, ANDREW L. TRATTNER, AND YUWEI ZHANG

Abstract. This paper explores features of the BMS sequence. We prove the uniqueness of the sequence, consider possible substitution rules to move between the BMS sequence and the block size sequence, look at which patterns can/cannot occur, and explore the density of the 1's and 2's respectively in the sequence.

1. Introduction

Sequences are enumerated, ordered strings of terms. Each term can occur multiple times in the sequence; in this paper we consider an infinite sequence of digits. Each numbered location of a sequence is called an index. For instance, the sequence "459" has a 4 at the first index, a 5 at the second index, and a 9 at the third index. We also group terms of the sequence in blocks. Blocks are subsequences of adjacent equal terms and are numbered with block indices. Consider "5668885." In this sequence, there are four blocks; the 5 at the first index is in the first block, the two 6's at the second and third indices comprise the second block, the three 8's at the fourth, fifth, and sixth indices comprise the third block, and the last 5 is in the fourth and final block. We include another depiction of indices and blocks below:

$$\begin{aligned} index &: 123456789\ldots \\ sequence &: 543662111\ldots \\ block &: 123445666\ldots \end{aligned}$$

Each block also has a length, which is defined as the number of terms in a block.

The focus of this paper is the "Buckle My Shoe," or, BMS sequence, which is an infinite sequence of 1's and 2's. The BMS sequence is self-referencing, meaning that the value of subsequent terms is determined by the value of previous terms of the sequence. The defining characteristic of the BMS sequence is that its sequence of block lengths is the same as its primary sequence. Thus, the value of the term at the nth index indicates the size of the nth block. Since the sequence is comprised of only 1's and 2's, if the current block is continuing, the digit at the previous index is repeated at the next. If the current block is ending, the digit at the next index is different from the previous one. The first nine terms of the sequence are shown below:

Date: December 12, 2016.

Sarah worked on the abstract, introduction, and uniqueness. Andy worked on limiting density and growth rate. Yuwei worked on substitution rules, restrictions on allowed patterns, the sequence of $\{1, 3\}$, and conclusion.

$$\begin{aligned} index &: 123456789\dots \\ sequence &: 122112122\dots \\ block &: 122334566\dots \end{aligned}$$

We explore the uniqueness of the BMS sequence in section 2, substitution rules in section 3, density of specific strings in section 4, growth rate in section 5, restrictions on allowed strings in section 6, and variations in section 7.

2. Uniqueness

A sequence is uniquely defined when there is only one possible iteration of that sequence.

Theorem 2.1. *The BMS sequence is* ***uniquely defined*** *when the first bit is 1.*

Proof. For the BMS sequence, the first bit determines the length of the first block, the second bit the length of second box, and so on. Hence, boxes, given by the groups of consecutive equal bits, can either be length 1 or 2 since those are the only digits in the sequence.

Set the first bit equal to 1. This implies that the first block is also size 1. Since the first bit is one and a member of a size one block, the second bit must be 2.

$$place : 1 \to 12 \tag{2.1}$$
$$sequence : 1 \to 12\dots \tag{2.2}$$
$$block : 1 \to 12 \tag{2.3}$$

By similar logic, we can say that the second block must be size 2 since the second bit is a 2. Therefore, the third bit must also be 2. Since that third bit ends the size 2 second block, we know the fourth bit must be a 1.

$$place : 12 \to 1234 \tag{2.4}$$
$$sequence : 12 \to 1221\dots \tag{2.5}$$
$$block : 12 \to 1223 \tag{2.6}$$

Likewise, we find that the fifth bit is 1, the sixth bit is 2, the seventh bit is 1, and so on. We can prove by strong induction that there is only one possible digit for each place in the sequence.

Consider the base case:

$b_1 = 1$, where b_1 is the variable representing the first bit. Since we set this bit equal to 1, it cannot be any other digit.

Now let's proceed to the inductive step:

Consider b_{n+1}, the $(n+1)$th bit of the sequence. Assume that for all $i < n+1$, the ith bit is uniquely determined (it can only be 1 or it can only be 2). We know that the $(n+1)$th bit is determined by the value of the nth bit and the value of the bit with a place value corresponding to the $(n+1)$th bit's block value. Since blocks have either size 1 or size 2 and there is at least one block of size 2, we know that the block of the $(n+1)$th bit must be less than the place of the $(n+1)$th bit. Therefore, we know by the induction hypothesis that since both the nth bit and the size of the $(n+1)$th bit's block are uniquely determined, the $(n+1)$th bit must be uniquely determined. $\square$

Theorem 2.2. *The BMS sequence is uniquely defined when the first bit 2.*

Proof. We can prove this by strong induction as we did above.

Consider the base case: $b_1 = 2$. Since we set this bit equal to 2, it cannot be equal to any other digit.

We follow with the same inductive step as in the previous proof. $\square$

We find that when the BMS sequence starts with a 1 it looks like:

$$1221121221221121122\ldots$$

And when the BMS sequence starts with a 2 it looks like:

$$221121221221121122\ldots$$

Although these sequences start with a different first bit, they appear very similar.

Corollary 2.2.1. If a 1 is appended to the beginning of the BMS sequence where $b_1 = 2$, it is equal to the BMS sequence where $b_1 = 1$. The BMS sequence beginning with 2 is a subsequence of the BMS sequence beginning with 1.

b_1 encodes both its value and its block size since it is in the first place and consequently the first block. When $b_1 = 1$, the first block only has size 1 so b_2 also encodes both its value and its block size since it is in the second place and in the second block. This essentially makes the first bit when $b_1 = 1$ "removable." When we remove $b_1 = 1$ and reset the place values, we get the BMS sequence for $b_1 = 2$.

3. Substitution Rules

Given a sequence of block lengths of 1 or 2, we can construct a sequence of 1's and 2's having those given block lengths. Suppose we would like to construct a sequence of block lengths $22112\ldots$, starting the sequence with a 1, we would get

$$\begin{array}{cccccc} 2 & 2 & 1 & 1 & 2 & \ldots \\ | & | & | & | & | & \\ 11 & 22 & 1 & 2 & 11 & \ldots \end{array}$$

Our BMS sequence can be described by substitution rules from each block length to a sequence of 1's or 2's with the given block length. Formally, we define a set of substitution morphisms for the sequence.

Definition 3.1. Let Σ and Δ be alphabets. A **morphism** is a map $\varphi : \Sigma^* \to \Delta^*$ that obeys the identity $\varphi(xy) = \varphi(x)\varphi(y)$ for all words $x, y \in \Sigma^*$ [1, Chapter 1.4]. In our case, a substitution morphism $\varphi : \Sigma^* \to \Sigma^*$ replaces each letter by a word.

If we mark the block lengths and observe its mapping to the sequence,

$$\begin{array}{cccccccccc} 1 & 2 & 2 & 1 & 1 & 2 & 1 & 2 & 2 & \ldots \\ | & | & | & | & | & | & | & | & | & \\ 1 & 22 & 11 & 2 & 1 & 22 & 1 & 22 & 11 & \ldots \end{array}$$

we find the set of substitution morphisms $\varphi = \{\varphi_1, \varphi_2\} : \Sigma^* \to \Sigma^*$ over the alphabet $\Sigma = \{1, 2\}$ as follows:

$$\begin{aligned} \varphi_1 \colon\ & 1 \to 1 \\ & 2 \to 11, \\ \varphi_2 \colon\ & 1 \to 2 \\ & 2 \to 22, \end{aligned}$$

We alternate between the two substitution morphisms, using φ_1 for the odd position and φ_2 for even positions. In other words, let $w = a_1 a_2 \cdots a_n$, where $a_i \in \Sigma$. We apply the substitution rules as

$$\varphi(w) = \varphi_{i_1}(a_1)\varphi_{i_2}(a_2)\cdots\varphi_{i_n}(a_n), \quad \text{where } i_k \equiv k \pmod 2.$$

For example, if we start with the seed sequence $w = 12$, by applying the substitution rules, we obtain a growing sequence,

$$\begin{aligned} \varphi(w) &= \varphi_1(1)\varphi_2(2) \\ &= 1\ 22 \\ \varphi^2(w) &= \varphi(122) \\ &= \varphi_1(1)\varphi_2(2)\varphi_1(2) \\ &= 1\ 22\ 11 \\ \varphi^3(w) &= \varphi_1(1)\varphi_2(2)\varphi_1(2)\varphi_2(1)\varphi_1(1) \\ &= 1\ 22\ 11\ 2\ 1 \\ \varphi^4(w) &= \varphi_1(1)\varphi_2(2)\varphi_1(2)\varphi_2(1)\varphi_1(1)\varphi_2(2)\varphi_1(1) \\ &= 1\ 22\ 11\ 2\ 1\ 22\ 1 \\ &\cdots \end{aligned}$$

We extend our self-referencing sequence by repeatedly iterating the application of the substitution rules φ to the current sequence starting from the seed w. After k iterations, the first $|\varphi^k(w)|$ terms of the sequence stop changing.

Remark 3.2. If w is self-referencing, then $\varphi(w)$ contains w as its prefix, i.e.

$$\varphi(w) = wx.$$

If $\varphi^j(x) \neq \epsilon\ \forall j$, the sequence of words, w, $\varphi(w)$, $\varphi^2(w)$, $\ldots$ converges, in the limit, to the infinite word

$$\varphi^\infty(w) = wx\varphi(x)\varphi^2(x)\cdots,$$

which is the unique fixed point of φ which starts with w, that is, $\varphi(\varphi^\infty(w)) = \varphi^\infty(w)$.

4. Restrictions on Allowed Patterns

As our sequence is self-referencing, we would imagine that not all subsequences are allowed. Indeed, there are certain subsequences that never occur in our self-referencing sequence.

Since our sequence is over the alphabet $\{1, 2\}$, we do not have 3's in the sequence, which means we have no runs of 3 of the same digits in a row. So 111 and 222 are not allowed. By the same reasoning, if we had 112211, then we would have 222 elsewhere in the sequence by its self-referencing property, so 112211 is not possible either. We can thus deduce the groups of disallowed sequences in the next level from the current level:

$$\begin{aligned} \text{No } 3 &\Rightarrow \text{No } 111,\ 222 \\ \text{No } 111 &\Rightarrow \text{No } 2\ 121\ 2,\ 1\ 212\ 1 \\ \text{No } 222 &\Rightarrow \text{No } 112211,\ 221122 \\ &\dots \end{aligned}$$

Given our substitutions rules, we have

$$w \text{ is disallowed } \Rightarrow \varphi'(w) \text{ is disallowed,}$$

where φ' denotes the same substitution rule as φ with an extra step at the end - for a disallowed sequence w that starts or ends with 1, we pad the implied disallowed sequence with a differing number before or after it respectively. For example, if we have 111, we add 2 to both ends of the implied sequence 121, so that $\varphi'(111) = 21212$ is not allowed, where as 11211, 11212, and 21211 are allowed.

We list the first few disallowed sequences as generated by our computer program:

(1) (111, 222);
(2) (12121, 21212), (112211, 221122);
(3) (121121121, 212212212), (11211211, 22122122), (1212211212, 2121122121), (1122121122, 2211212211);
(4) (12112122121121, 21221211212212),
(112112212211211, 221221121122122),
(121221211212, 212112122121),
(11221221121122, 22112112212211),
(1211211221211211, 2122122112122122),
(1121121221121121, 2212212112212212),
(1212211211212211, 2121122122121122),
(1122121121122121, 2211212212211212);
...

Each tuple above is a pair of sequences generated from the same seed sequence in the preceding level, with opposite numbers at every position - 1 in one and 2 in the other. Any sequence that contains any of the above sequences as a subsequence is not allowed.

Figure 1 shows the number of disallowed sequences of length n with no disallowed subsequences for n up to 500. We see that though the plot looks a bit irregular, it exhibits an approximate repeating pattern of peaks with increasing magnitude and period. This is due to the fact that we generate the next set of disallowed sequences from the seeding disallowed sequences in the previous iteration.

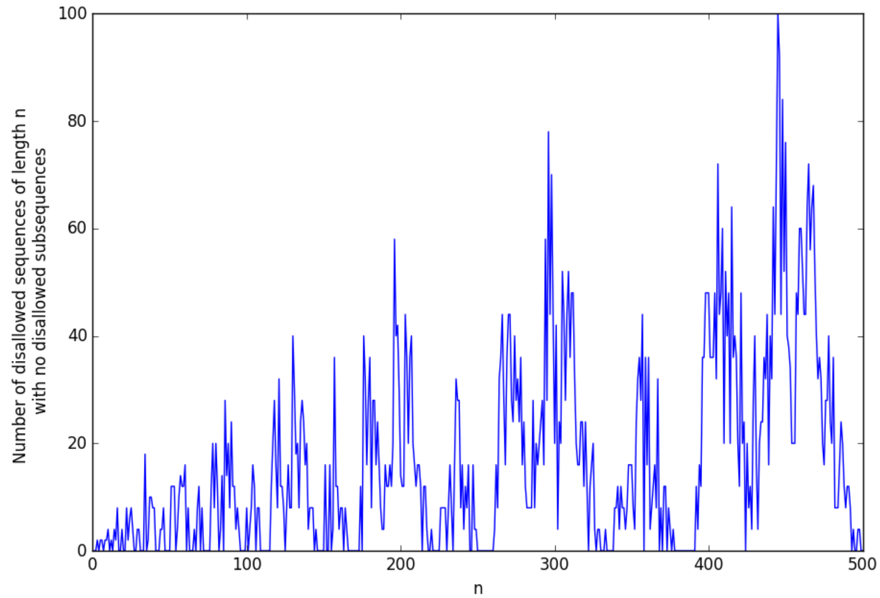

Figure 1. Plot of the number of disallowed sequences of length n with no disallowed subsequences for $n \in [0, 500]$

Given a sequence w, we can check if it contains disallowed subsequences using the following procedure:

Algorithm 1 TestIfAllowed(w)

Input: A sequence $w = a_1 a_2 \cdots a_n$.
Output: `TRUE/FALSE`: whether sequence w is allowed.

procedure TestIfAllowed(w)
  while len(w)> 2 **do**
    $w \leftarrow$ MarkBlockLengths(w)
  if $a_i \notin \{1, 2\}$ for some $1 \leq i \leq n$ **then**
    return `FALSE` ▷ If we end up with a digit ≥ 3, then it is disallowed!
  else
    return `TRUE`

The procedure MarkBlockLengths(w) works by marking each block of consecutive terms with its length. Note that it ignores a beginning or ending block of length 1, since that might have been part of a longer block in the full sequence. We first show an example below:

$$\begin{array}{ccccccccccccc}
\dots & 11 & 22 & 1 & 22 & 11 & 2 & 1 & 22 & 1 & 22 & 11 & \dots \\
 & | & | & | & | & | & | & | & | & | & | & | & \\
\dots & 2 & 2 & 1 & 2 & 2 & 1 & 1 & 2 & 1 & 2 & 2 & \dots \\
 & | & & | & | & & | & & | & | & | & & \\
\dots & 2 & & 1 & 2 & & 2 & & 1 & 1 & 2 & \dots & \\
 & & & | & | & & & & | & & & & \\
 & & \dots & 1 & 2 & & & & 2 & \dots & & & \\
 & & & & | & & & & & & & & \\
 & & & \dots & 2 & & & & & \dots & & &
\end{array}$$

The sequence given in this example does not contain disallowed subsequences and the procedure returns `TRUE`.

On the other hand, consider the example below:

$$\begin{array}{ccccccccccccc}
\dots & 2 & 1 & 22 & 1 & 22 & 11 & 2 & 1 & 22 & 1 & 22 & \dots \\
 & & | & | & | & | & | & | & | & | & | & | & \\
 & \dots & 1 & 2 & 1 & 2 & 2 & 1 & 1 & 2 & 1 & 2 & \dots \\
 & & & | & | & | & & | & & | & | & & \\
 & & \dots & 1 & 1 & 2 & & 2 & & 1 & 1 & \dots & \\
 & & & | & & | & & & & | & & & \\
 & & \dots & 2 & & 2 & & & & 2 & \dots & & \\
 & & & | & & & & & & & & & \\
 & & \dots & 3 & & & & & & & \dots & &
\end{array}$$

The sequence given in this second example boils down to 222 which in turn gives rise to the number 3. This means that the original sequence contains a disallowed subsequence. The procedure thus returns `FALSE`.

All sequences that are not disallowed can actually occur. In the next section, we will see some examples that show this, including all 6 of the sequences of length 3 and all 10 of the sequences of length 4 that are not disallowed.

5. Limiting density

For this section, we use a computer simulation to generate the BMS sequence in the obvious way to find interesting empirical evidence for the density of various words.

Definition 5.1. The **density** of a length n substring word $w = s_i s_{i+1} \dots s_{i+n-1}$ of some length N string $S = s_1 s_2 s_3 \dots s_N$ is the number of distinct appearances of w in S. The **limiting density** of w is the density in the limit as $N \to \infty$.

Conjecture 5.1.1. The limiting densities of words composed of one, two, three, or four symbols are as listed in the Guess column of Tables 1, 2, 3, and 4.

The following densities are obtained with $N = 10^8$.

Word	Density	Guess
1	0.50000676	1/2
2	0.49999324	1/2

TABLE 1. Densities of Length 1 Words

Word	Density	Guess
11	0.16667268	1/6
22	0.16665918	1/6
12	0.33333408	1/3
21	0.33333408	1/3

TABLE 2. Densities of Length 2 Words

Word	Density	Guess
112	0.16667268	1/6
221	0.16665918	1/6
121	0.16667489	1/6
212	0.16666138	1/6
122	0.16665918	1/6
211	0.16667268	1/6

TABLE 3. Densities of Length 3 Words

Word	Density	Guess
1122	0.05555351	1/18
2211	0.05555519	1/18
1212	0.0555574	1/18
2121	0.05555572	1/18
1211	0.11111749	1/9
2122	0.11110566	1/9
1121	0.11111917	1/9
2212	0.11110399	1/9
1221	0.16665918	1/6
2112	0.16667268	1/6

TABLE 4. Densities of Length 4 Words

We also find empirical evidence for the deviation, or error, from the guesses for various N. As seen in Figure 2, the density of ones (aka length 1 words) oscillates about $1/2$ and converges as $N \to \infty$.

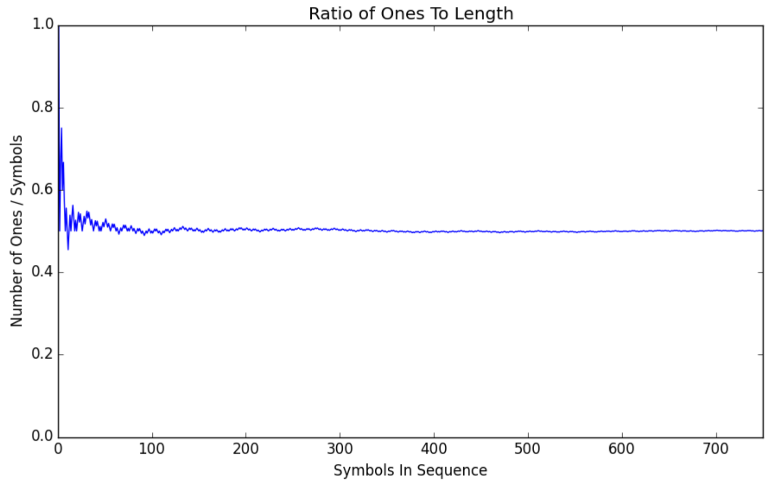

FIGURE 2. Density of Ones

The error from the conjectured density of ones in the BMS sequence is listed in Table 3.

N	Error	Error Magnitude
10^1	0.1	10^{-1}
10^2	0.01	10^{-2}
10^3	0.004	10^{-3}
10^4	0.0003	10^{-4}
10^5	0.00026	10^{-4}
10^6	1.3×10^{-5}	10^{-5}
10^7	4.8×10^{-6}	10^{-6}
10^8	6.76×10^{-6}	10^{-6}
10^9	1.225×10^{-6}	10^{-6}

FIGURE 3. Limiting Density of Ones

Conjecture 5.1.2. The difference, or error, between the density of ones in the BMS sequence up to index $N = 10^k$ and the limiting density of ones is between $10^{-k/2}$ and 10^{-k}.

We plot the errors and conjectured bounds in Figure 4.

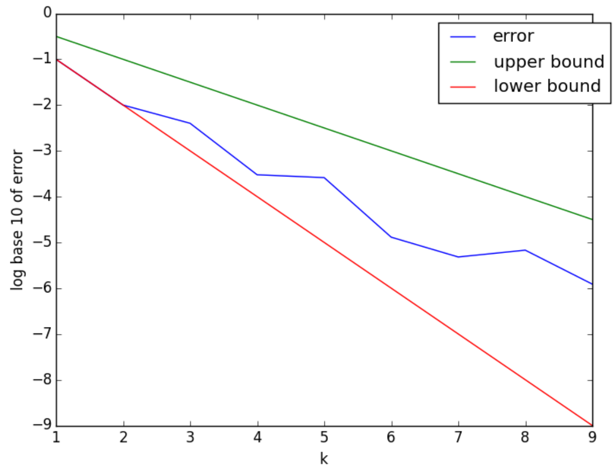

FIGURE 4. Error Of Limiting Density With $\log_{10}$ Scale

6. Growth Rate

If the limiting density of ones is indeed $1/2$, then we can say something about the growth rate of the BMS sequence as it grows infinitely long.

Definition 6.1. The **growth rate** of the first B block words of the BMS sequence is the number of symbols contained in the blocks divided by B.

We can conceptualize the growth rate as the number of symbols each block contributes, on average. So the growth rate converges to $3/2$ as $B \to \infty$.

This makes sense, since the first and second cases of φ_1 and φ_2 appear with about the same frequency for large B.

Another way to see this, and perhaps a useful technique to help prove the conjecture for limiting densities, is to observe the length five substrings generated by words of length three and count the occurrences of the words in the substrings. For example, 11211 contains $112, 121, 211$.

$$
\begin{aligned}
112 &\longrightarrow 21211, 12122\\
122 &\longrightarrow 11212, 22121\\
121 &\longrightarrow 21221, 21121\\
211 &\longrightarrow 11221, 22112\\
221 &\longrightarrow 11211, 22122\\
212 &\longrightarrow 12211, 21122
\end{aligned}
$$

It is clear that each length three word appears precisely six times in the substrings to the right of the arrows. This suggests that the BMS sequence obtains the stated growth rate.

7. Sequence of $\{1,3\}$

What if we allow the alphabet $\{1,3\}$ for the self-referencing sequence instead? We would then get the sequence

$$1333111333131333\ldots$$

or

$$333111333131333\ldots$$

What can we say about the limiting density for the sequence of $\{1,3\}$? We note that unlike the sequence of $\{1,2\}$, the numbers 1 and 3 are of the same parity. In this case, we are in fact able to derive the limiting density of 3's and 1's in the sequence.

First, if we distinguish the odd and even digits by marking the even digits with a bar, the sequence becomes

$$3\bar{3}3\bar{1}1\bar{1}3\bar{3}3\bar{1}3\bar{1}\ldots$$

We observe that a pattern arises - we define the morphism for this sequence $\varphi : \Sigma^* \to \Sigma^*$ over the alphabet $\Sigma = \{3, \bar{3}, 1, \bar{1}\}$ as:

$$\begin{aligned} \varphi: \quad & 3 \to 3\bar{3}3 \\ & \bar{3} \to \bar{1}1\bar{1} \\ & 1 \to 3 \\ & \bar{1} \to \bar{1}. \end{aligned}$$

Let the alphabet be in the order $a_1 = 3$, $a_2 = \bar{3}$, $a_3 = 1$, $a_4 = \bar{1}$. We then have the following incidence matrix [1, Chapter 8.2] associated with the morphism φ,

$$M(\varphi) = \begin{array}{c c} & \begin{array}{cccc} 3 & \bar{3} & 1 & \bar{1} \end{array} \\ \begin{array}{c} 3 \\ \bar{3} \\ 1 \\ \bar{1} \end{array} & \begin{bmatrix} 2 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 2 & 0 & 1 \end{bmatrix} \end{array},$$

where each entry M_{ij} represents the resulting number of occurrences of the letter a_i from applying the morphism φ to the letter a_j.

Now, we let $c_n(a)$ denote the count of the number of occurrences of the letter a in our current sequence of length n. We then have x_n, the vector of counts for each letter in the alphabet,

$$x_n = \begin{bmatrix} c_n(3) \\ c_n(\bar{3}) \\ c_n(1) \\ c_n(\bar{1}) \end{bmatrix}, \quad \text{and } x_0 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

as we start the sequence with 3.

The matrix $M(\varphi)$ is useful, because if we apply matrix M to the vector x_n,

$$x_{n+1} = Mx_n$$

represents the consequent vector of the number of occurrences for each letter after applying the morphism to the current sequence.

As M is a non-negative square matrix, we have the Perron-Frobenius eigenvalue $r \geq 0$ of M such that $|\lambda_j| \leq r$ for all λ_j of M [1, Theorem 8.3.7, Theorem 8.3.11].

We compute the characteristic polynomial of the matrix M:

$$\begin{aligned}
P(\lambda) &= \det(\lambda I - M) \\
&= \begin{vmatrix} \lambda-2 & 0 & -1 & 0 \\ -1 & \lambda & 0 & 0 \\ 0 & -1 & \lambda & 0 \\ 0 & -2 & 0 & \lambda-1 \end{vmatrix} \\
&= \lambda^4 - 3\lambda^3 + 2\lambda^2 - \lambda + 1.
\end{aligned}$$

Solving the equation

$$P(\lambda) = \lambda^4 - 3\lambda^3 + 2\lambda^2 - \lambda + 1 = (\lambda-1)(\lambda^3 - 2\lambda^2 - 1) = 0,$$

we get the eigenvalues

$$\lambda_1 \approx 2.2055694304, \lambda_2 = 1, \lambda_3, \lambda_4 \approx -0.1027847152 \pm 0.6654569115i.$$

We thus have $r = \lambda_1 = 2.2055694304$.

We are now ready to define the limiting frequencies of the letters: let $f(a)$ denote the limiting frequency of the letter a in the sequence, our vector of frequencies is

$$v = \begin{bmatrix} f(3) \\ f(\bar{3}) \\ f(1) \\ f(\bar{1}) \end{bmatrix}.$$

We show that the limiting frequencies $f(a)$ are defined [1]. Our matrix M has 4 distinct eigenvalues as we found out above, and is thus diagonalizable. With $r = \lambda_1$, $|\lambda_1| > |\lambda_j|$ for $j > 1$, we write

$$\begin{aligned}
x_0 &= c_1 v_1 + c_2 v_2 + c_3 v_3 + c_4 v_4, \\
x_n &= M^n x_0 \\
&= S \Lambda^n S^{-1} x_0 \\
&= c_1 \lambda_1^n v_1 + c_2 \lambda_1^n v_2 + c_3 \lambda_3^n v_3 + c_4 \lambda_1^n v_4 \\
&= c_1 \lambda_1^n \left(v_1 + \frac{c_2}{c_1} \left(\frac{\lambda_2}{\lambda_1} \right)^n v_2 + \frac{c_3}{c_1} \left(\frac{\lambda_3}{\lambda_1} \right)^n v_3 + \frac{c_4}{c_1} \left(\frac{\lambda_4}{\lambda_1} \right)^n v_4 \right).
\end{aligned}$$

As $\left|\frac{\lambda_j}{\lambda_1}\right| < 1$ for $j > 1$, $\lim_{n\to\infty} x_n = c_1 \lambda_1^n v_1$. Therefore, x_n converges to a multiple of the dominant eigenvector v_1 with the ratio of convergence $\left|\frac{\lambda_2}{\lambda_1}\right|$, where λ_2 is the second largest eigenvalue. Starting with a non-negative vector x_0, the limiting frequency vector is then the normalized non-negative eigenvector of M associated with λ_1.

[1] Although our morphism is not primitive (as $\bar{1} \to \bar{1}$ only), the Perron-Frobenius eigenvalue r is a simple root of the characteristic polynomial of M; in other words, the number of occurrences of the dominating Jordan block in $J(\varphi)$ of M is 1. This alone is a necessary and sufficient condition for the existence of the frequencies of all letters in a pure morphic sequence like ours.

Thus, solving for v_1 (without normalizing), we have

$$\begin{aligned} Mv_1 &= \lambda v_1 \\ (\lambda_1 I - M)v_1 &= 0 \\ v &= \begin{bmatrix} \lambda_1^2 - \lambda_1 \\ \lambda_1 - 1 \\ \lambda_1^3 - 3\lambda_1^2 + 2\lambda_1 \\ 2 \end{bmatrix}. \end{aligned}$$

We hence have the formulas for the densities of 3 and 1 as follows:

$$\begin{aligned} \rho_3 &= \frac{f(3) + f(\bar{3})}{f(3) + f(\bar{3}) + f(1) + f(\bar{1})} \\ &= \frac{\lambda_1^2 - 1}{\lambda_1^3 - 2\lambda_1^2 + 2\lambda_1 + 1}, \\ \rho_1 &= \frac{f(1) + f(\bar{1})}{f(3) + f(\bar{3}) + f(1) + f(\bar{1})} \\ &= \frac{\lambda_1^3 - 3\lambda_1^2 + 2\lambda_1 + 2}{\lambda_1^3 - 2\lambda_1^2 + 2\lambda_1 + 1}. \end{aligned}$$

Substituting $r = \lambda_1 = 2.2055694304$ into the equations for ρ_3 and ρ_1, we obtain

$$\rho_3 \approx 0.6027847152,$$

$$\rho_1 \approx 0.3972152848.$$

For the density of 3 up to the nth digit in the sequence, the numerical values obtained from the formulas are confirmed by the results using direct counts of 3's in the sequence from our computer implementation, which we will present below.

Recall that

$$x_n = \sum_{j=1}^{4} c_j \lambda_j^n v_j.$$

The density of 3 up to the nth digit is thus

$$\rho_3(n) = \frac{\sum_{i=1}^{2} \sum_{j=1}^{4} c_j \lambda_j^n v_{ij}}{\sum_{i=1}^{4} \sum_{j=1}^{4} c_j \lambda_j^n v_{ij}},$$

where v_{ij} denote the ith entry in eigenvector v_j, and c_j's are obtained by solving the simultaneous equations in $x_0 = c_1v_1 + c_2v_2 + c_3v_3 + c_4v_4 = [1 \quad 0 \quad 0 \quad 0]^T$.

The error term is given by

$$\begin{aligned} \varepsilon(n) &= |\rho_3(n) - \rho_3| \\ &= \left| \frac{\sum_{i=1}^{4} \sum_{j=1}^{2} c_i \lambda_i^n v_{ij}}{\sum_{i=1}^{4} \sum_{j=1}^{4} c_i \lambda_i^n v_{ij}} - \frac{\lambda_1^2 - 1}{\lambda_1^3 - 2\lambda_1^2 + 2\lambda_1 + 1} \right|. \end{aligned}$$

We tabulate the results below:

n	$\rho_3(n)$	$\varepsilon(n)$
10^1	0.6	2.785×10^{-3}
10^2	0.61	7.215×10^{-3}
10^3	0.603	2.153×10^{-4}
10^4	0.6027	8.472×10^{-5}
10^5	0.60279	5.285×10^{-6}
10^6	0.602785	2.848×10^{-7}
10^7	0.6027848	8.480×10^{-8}
10^8	0.60278472	4.800×10^{-9}
10^9	0.602784716	8.000×10^{-10}

TABLE 5. Density of 3 in the first n digits of the sequence of $\{1, 3\}$

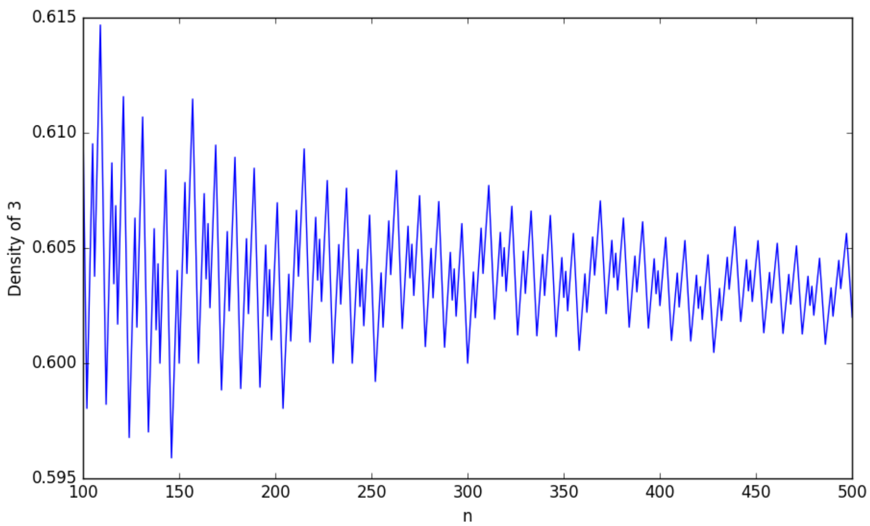

FIGURE 5. Plot of the density of 3 against n for the first $n \in [100, 500]$ digits of the sequence of $\{1, 3\}$

From the above figures, we see that the density of 3 in the first n digits of the sequence oscillates about the limiting density with decreasing amplitude and with a period of about 10. This oscillation is due to the fact that we have two complex eigenvalues for our matrix M. Overall, the density of 3 in the first n digits indeed gradually converges to the limiting density as we increase n.

8. Conclusion

We have seen that we are not able to obtain the limiting density for the sequence of $\{1, 2\}$, but are able to obtain it for the sequence of $\{1, 3\}$. Our conjecture is that this is because 1 and 3 are both odd, and the substitution rules preserve the parity, whereas 1 and 2 do not have the same parity. Future research could explore sequences of any two arbitrary numbers, sequences with more than two numbers, and general rules for determining the limiting densities of numbers in a sequence if they exist.

References

[1] Jean-Paul Allouche and Jeffrey Shallit. *Automatic sequences: theory, applications, generalizations*. Cambridge University Press, 2003.