Exploring the Thue-Morse sequence

Adam Hutchings, Jake Roggenbuck

July 22, 2022

1 Introduction

The Thue-Morse sequence can be defined in many ways, but it is the sequence whose nth term $(T(n))$ depends on the number of ones in the binary representation of n : 1 if odd, 0 if even. The sequence T_3, which is defined as the sequence whose nth term is $T(3n)$, has some interesting properties that we will investigate – among the most interesting of which are the persistent imbalance of zeros and ones and the lengths of runs in the sequence.

2 No-2 Theorem

We define a $motif$ in T_3 to be a sequence of the same digit not contained inside any larger motif. The authors observed that for the first 10^{11} terms of T_3, there are no motifs of length 2, which can be proven formally.

No-2 Theorem There are no motifs of length 2 in T_3.

Proof We begin by observing that a motif of zeros of length 2 is the sequence 00 surrounded by 1s, and a motif of ones of length 2 is the opposite. These translate to 1001 and 0110 in T_3, which must exist if the No-2 Theorem is false. Because we obtain the elements of T_3 from choosing every third element of the Thue-Morse sequence, the existence of 1001 or 0110 in T_3 implies the existence of $1XX0XX0XX1$ or $0XX1XX1XX0$ in the Thue-Morse sequence where X is either digit.

Lemma 2.1 The sequences 000 and 111 do not exist in the Thue-Morse sequence.

Proof The non-existence of repetitions of three in the sequence is a well-known fact, but may be established as follows: jumping from $2n$ to $2n + 1$ changes a 0 to a 1 and nothing else, which means that T_{2n} and T_{2n+1} are opposite. Therefore, the entire Thue-Morse sequence consists of repetitions of 01 and 10, which allows for no 000 or 111. $\square$

Lemma 2.1 allowed for us to create a computer program to comb through all ten-digit sequences of zeros and ones, with the constraints that they follow the pattern $0XX1XX1XX0$ and do not contain 000 or 111. The only ten-digit sequences that satisfy these constraints are 0011001010, 0011001100, 0011011010,0101001010, 0101001100, 0101011010, 0101101010, and 0101101100.

We do not check for any occurrences of 1001 in T_3, (which would follow the pattern $1XX0XX0XX1$ in T) but this is not necessary, as shown by:

Lemma 2.2 A sequence X exists in the Thue-Morse sequence *iff* the sequence $\overline{X}$ also exists, where $\overline{X}$ is X with 0 and 1 interchanged.

Proof Such a sequence X will occur entirely inside the first n digits of the sequence, where n is an arbitrarily large power of 2. The *next* n digits are the same as the first n with 0 and 1 interchanged, so the existence of X in the first block is equivalent to the existence of $\overline{X}$ in the second. $\square$

(This means that we do not need to check for any occurrences in 1001 because these exist *iff* occurrences of 0110 occur.)

Lemma 2.3 If n is some power of 2 and the sequence X is not more than n digits long, then if X does not occur in the first $8n$ digits of the Thue-Morse sequence, it never occurs.

Proof Call the first n digits of the Thue-Morse sequence A, and $\overline{A}$ (as defined above) B. Then by the same logic as the proof of Lemma 2.2, the first $8n$ digits are $ABBABAAB$. As the sequence n is no more than n digits long, any occurrence of it is either entirely within one n-long block or straddling two. Therefore, the only possible "environments" in which the pattern may occur are A, B, AA, AB, BA, or BB. All of these environments occur in $ABBABAAB$, so if the pattern never exists there, it never exists anywhere. $\square$

Now, Lemma 2.3 means we only need to check the first 128 digits (as the 10-digit sequences are less than 16 long, so checking 128 is sufficient.) Checking using a simple computer program rules these out. $\square$

3 The Imbalance Of Zeros And Ones

The T_3 sequence begins 00000001000..., and the imbalance between zeros and ones continues on. We denote the ratio of zeros to ones in every third digit of a stretch of the Thue-Morse sequence in the first n entries by $r(n)$, and note the following values calculated by a computer:

Ratio for the first 2^{10} digits: 2.8
Ratio for the first 2^{11} digits: 2.1045454545454545
Ratio for the first 2^{12} digits: 2.1045454545454545
Ratio for the first 2^{13} digits: 1.7282717282717284
Ratio for the first 2^{14} digits: 1.7282717282717284
[Twenty entries omitted]
Ratio for the first 2^{35} digits: 1.0228080103552828

The ratio does indeed appear to approach 1, but only very slowly – even at over 10^{10} entries of T_3 computed, there are still 2% more zeros than ones. Therefore, we conjecture the following:

Ratio Conjecture $r(n)$ is never less than 1.

3.1 The Weak Ratio Theorem

Weak Ratio Theorem $r(2^k) > 1$ for all integers $k > 4$.

As a side note, the restriction $k > 4$ exists because the first one occurs at position 8 in $T3$, so $r(2^k)$ for $n < 5$ is undefined.

Proof To start, as in the proof of Lemma 2.3, we may consider that for some natural number k, the entire Thue-Morse sequence is composed of two blocks of length 2^k, which we may call A_k and B_k, such that B_k is A_k with every 0 and 1 swapped. Next, we will consider three pairs of functions: $\alpha_{0/1}$, $\beta_{0/1}$, and $\gamma_{0/1}$. $\alpha_0(k)$ is the number of zeros in the block A_k when we count every third element, *starting from element 1*, and $\alpha_1(k)$ is the number of ones when we count in the same manner. The β functions are the totals when taken starting on the second element of the block A_k, and the γ functions are counted starting on the third elements. Based on the definition of the T_3 sequence, we see that:

$$r(2^k) = \frac{\alpha_0(k)}{\alpha_1(k)}.$$

We also note that because $\alpha_0(k) > 0$ and $\alpha_1(k) > 0$ for $k > 4$, the Weak Ratio Theorem can be phrased in terms of a difference.

Alternative Weak Ratio Theorem Statement $\alpha_0(k) > \alpha_1(k)$ for all $k > 4$.

We may continue by deriving expressions for the values of each of these functions by stating them in terms of equivalent values for smaller blocks. Assuming k is even, 2^k is equal to 1 *mod* 3, and if k is odd then 2^k is equal to 2 *mod* 3. Now, considering $\alpha_0(k+1)$ for even k, we know that we count through the two blocks AB, and because block A has a remainder of 1 *mod* 3, we start counting zeros in block B at the third element, which is equivalent to counting ones starting at the third element in A (because B is equal to A with zeros and ones swapped.) Therefore, we see that

$$\alpha_0(k+1) = \alpha_0(k) + \gamma_1(k).$$

Using the same reasoning, for even k :

$$\alpha_0(k+1) = \alpha_0(k) + \gamma_1(k), \alpha_1(k+1) = \alpha_1(k) + \gamma_0(k),$$

$$\beta_0(k+1) = \beta_0(k) + \alpha_1(k), \beta_1(k+1) = \beta_1(k) + \alpha_0(k),$$

$$\gamma_0(k+1) = \gamma_0(k) + \beta_1(k), \gamma_1(k+1) = \gamma_1(k) + \beta_0(k).$$

And for odd k :

$$\alpha_0(k+1) = \alpha_0(k) + \beta_1(k), \alpha_1(k+1) = \alpha_1(k) + \beta_0(k),$$

$$\beta_0(k+1) = \beta_0(k) + \gamma_1(k), \beta_1(k+1) = \beta_1(k) + \gamma_0(k),$$

$$\gamma_0(k+1) = \gamma_0(k) + \alpha_1(k), \gamma_1(k+1) = \gamma_1(k) + \alpha_0(k).$$

Lemma 3.1.1 For even k, $\alpha_0(k-3) + \beta_1(k-3) > \alpha_1(k-3) + \beta_0(k+3)$ *iff* $r(2^k) > 1$.

Proof Noting that k is even, we see that

$$\alpha_0(k) = \alpha_0(k-1) + \beta_1(k-1) = \alpha_0(k-2) + \gamma_1(k-2) + \beta_1(k-2) + \alpha_0(k-2).$$

Expanding one more level, we get:

$\alpha_0(k-3)+\beta_1(k-3)+\gamma_1(k-3)+\alpha_0(k-3)+\alpha_0(k-3)+\beta_1(k-3)+\beta_1(k-3)+\gamma_0(k-3)$.

Rearranging, and remembering that $\gamma = \gamma_0 + \gamma_1$:

$$\alpha_0(k) = 3\alpha_0(k-3) + 3\beta_1(k-3) + \gamma(k-3).$$

The expression for $\alpha_1(k)$ is similar, but all subscript zeros and ones are flipped:

$$\alpha_1(k) = 3\alpha_1(k-3) + 3\beta_0(k-3) + \gamma(k-3).$$

Therefore, we have:

$$\alpha_0(k) - \alpha_1(k) = 3\alpha_0(k-3) - 3\alpha_1(k-3) + 3\beta_1(k-3) - 3\beta_0(k-3).$$

Dividing this expression by 3 will not change its sign, so if

$$\alpha_0(k-3) - \alpha_1(k-3) + \beta_1(k-3) - \beta_0(k-3) > 0,$$

Then $r(2^k) > 1$. Rearranging terms in the expansion gives us Lemma 3.1.1.

Lemma 3.1.2 For odd k, $\alpha_0(k-3) + \gamma_1(k-3) > \alpha_1(k-3) + \gamma_0(k-3)$ *iff* $r(2^k) > 1$.

Proof We proceed in the same way as last time, except switching the use of the odd and even expansions.

We can now restate Lemmas 3.1.1 and 3.1.2 slightly, remembering the Alternative Weak Ratio Theorem Statement and replacing k with $k+3$ and $k-3$ with k :

Lemma 3.1.1 Restated For odd k, $\alpha_0(k) + \beta_1(k) > \alpha_1(k) + \beta_0(k)$ *iff* $\alpha_0(k+3) > \alpha_1(k+3)$.

Lemma 3.1.2 Restated For even k, $\alpha_0(k) + \gamma_1(k) > \alpha_1(k) + \gamma_0(k)$ *iff* $\alpha_0(k+3) > \alpha_1(k+3)$.

We now remember that for odd k, $\alpha_0(k+1) = \alpha_0(k)+\beta_1(k)$, and $\alpha_0(k+1) = \alpha_0(k) + \gamma_1(k)$ for even k. Doing the analogous expansions for $\alpha_1(k)$, we can restate the two lemmas yet again.

Lemma 3.1.1 Restated For odd k, $\alpha_0(k+1) > \alpha_1(k+1)$ *iff* $\alpha_0(k+3) > \alpha_1(k+3)$.

Lemma 3.1.2 Restated For even k, $\alpha_0(k+1) > \alpha_1(k+1)$ *iff* $\alpha_0(k+3) > \alpha_1(k+3)$.

We can combine these two by forgetting the irrelevant odd/even distinction and replacing $k+1$ with k and $k+3$ with $k+2$:

Lemma 3.1.3 For all k, $\alpha_0(k) > \alpha_1(k)$ *iff* $\alpha_0(k+2) > \alpha_1(k+2)$.

This means that we can show $\alpha_0(k) > \alpha_1(k)$ for $k = 5, 6$ and be done. In fact, we can check by hand:

$$\alpha_0(5) = 10, \alpha_1(5) = 1, \alpha_0(6) = 20, \alpha_1(6) = 2.$$

This completes the proof.

4 No-9+ Conjecture

The authors observed that for the first 10^{10} terms of T3, there are not motifs of lengths 9 or greater, which can be proven formally.

No-9+ Theorem There are no motifs of length 9 or greater in T_3.

5 6,8 Synchronization Conjecture

The authors observed that the runlengths of 1s of T_3 or $\Lambda_1(T_3)$ have a pattern in the order of 6s and 8s. It appears that 6 always follow 8 and vice versa however, there is a caveat. In $\Lambda_1(T_3)$, 6 does in fact follow 8 and vice versa until the location 87384 (**T OR T3?**) where there is an anomaly of a 6 following another 6. We have found by observation that frequency of anomalies stays between 0.019% and 0.028% from 10^7 to 10^9 in the digits of $\Lambda_1(T_3)$. The frequency of anomaly increases from the location 87384 forever.

6 Uniform 3s Conjecture

The frquency of values in $\Lambda_1(T_3)$ is uniform.