Observations on Every Third Digit of the Thue-Morse Sequence

Adam Hutchings   Henry Hutchings   Jake Roggenbuck

August 6, 2022

Abstract

We investigate some interesting properties of the sequence made up of every third term of the Thue-Morse sequence. Moreover, we create proofs for said properties.

Contents

1 Introduction

The Thue-Morse sequence, or T, can be defined in many ways, and we will use the following definition for the nth digit of the sequence, denoted by $T(n)$:

Definition 1.1 (Thue-Morse Sequence) *The Thue-Morse sequence is the sequence whose nth term $T(n)$ depends on the number of ones in the binary representation of n. If this number is odd, $T(n)$ is 1; otherwise, it is 0.*

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 lengths of runs in the sequence and the persistent imbalance of zeros and ones. We will prove two main results:

Theorem 1.1 (Motif Theorem) *Motifs in T_3 are only of lengths 1, 3, 6, 7, or 8.*

Theorem 1.2 (Ratio Theorem) *The number of zeros in the first n digits of T_3 always exceeds the number of ones for any n.*

We also develop two main techniques to prove these theorems and others – the segment-matching method and the alpha-beta-gamma method – which may be extended in the future to investigate T_n, which is defined so that $T_n(x) = T(nx)$.

2 Run-Length Of T_3

2.1 Motifs and the Λ Function

We define *segments*, *runs*, and *motifs* in T_3 as follows:

Definition 2.1 *A* segment *of length n is n consecutive terms in a sequence.*

Definition 2.2 *A* run *of length n is a segment of length n, all of whose terms are identical.*

Definition 2.3 *A* motif *of length n is a run of length n which is not contained inside a run of length $m > n$.*

Similar to the computer science notion of *run-length encoding*, we define a function Λ which takes as input a sequence of integers x and outputs another sequence of integers $\Lambda(x)$ whose nth term is denoted $\Lambda(x)(n)$. It is defined as follows:

- $\Lambda(x)(0)$ is the length of the motif in x starting at $x(0)$.
- $\Lambda(x)(n)$ is the length of the motif starting at the term in x indexed by $\sum_{k=0}^{n-1} \Lambda(x)(k)$.

In other words, Λ takes as input a sequence of numbers and transforms it into a sequence denoting the lengths of motifs in the input.

2.2 $\Lambda(T_3)$

The authors created a computer program to investigate the sequence $\Lambda(T_3)$ and noted that it consists, for all values checked, of only 1, 3, 6, 7, and 8. In other words, length of motifs in T_3 consist of only these values. The rest of Section 2 will be devoted to proving this.

2

2.3 No-2 Theorem

Theorem 2.1 (No-2 Theorem) *There are no motifs of length 2 in T_3. Equivalently, $\Lambda(T_3)$ contains no 2s.*

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, except at the beginning of the sequence. However, the sequence T_3 begins with a motif of length 7, so this caveat is not necessary. 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 runs of length 3 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 concatenations of 01 and 10, which allow 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.

3

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, we can conclude the proof of Theorem 2.1. 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 has ruled these out. $\square$

2.4   The Segment-Matching Method

Proving the No-2 Theorem left us with some useful tools to tackle future proofs, but the time taken to solve such a proof by brute force will grow exponentially with the length of the segment to check for. In particular, our naive approach to checking for the existence of a segment in T_3 has been to generate the equivalent segment in T that would have created it (such as translating 0110 in T_3 into $0XX1XX1XX0$ in T.) The number of possible segments to check in T grows roughly exponentially with the length of the original segment, each one to check being on the order of 3 times the size of the original one. The segment-matching method will reduce checking for a segment of length x in T_3 to checking for at most four segments of length at most x in T.

To begin, we see that the sequence T_3 is made up of blocks of 4, for every n which is a multiple of 4, of $T_3(n), T_3(n+1), T_3(n+2), T_3(n+3)$, which is equal to:

$$T(3n), T(3n+3), T(3n+6), T(3n+9).$$

Remembering that n is a multiple of 4, we will denote $3n/4$ by m, and denote the binary expansion of any number x by $[x]$. In that case, we have the following binary expansions of the four numbers above:

$$[3n] = [m]00,$$
$$[3n+3] = [m]11,$$
$$[3n+6] = [m+1]10,$$
$$[3n+9] = [m+2]01.$$

Therefore, we know their values in the Thue-Morse sequence by our original definition of it.

$$T(3n) = T(m),$$
$$T(3n+3) = T(m),$$
$$T(3n+6) \neq T(m+1),$$
$$T(3n+9) \neq T(m+2).$$

This sequence of four values is equivalent to $T_3(n), T_3(n+1), T_3(n+2), T_3(n+3)$ by the definition of the T_3 sequence. Knowing that n is a multiple of 4, then,

4

and therefore that m is a multiple of 3, we see that the T_3 sequence is made up of blocks of length 4 for every multiple of 3 m. Denoting the binary complement of x by $\overline{x}$, these blocks go:

$$T(m), T(m), \overline{T(m+1)}, \overline{T(m+2)}.$$

Because this segment in T_3 continues every four entries with a new m, for any segment X we wish to check, we need only identify X with a tuple of consecutive elements of the sequence

$$T(m), T(m), \overline{T(m+1)}, \overline{T(m+2)}, T(m+3), T(m+3), \overline{T(m+4)}, \overline{T(m+5)}, \cdots .$$

For some m which is a multiple of 3. We only need to check four possible tuples – beginning the tuple at the first entry equal to $T(m+3)$ is the same as checking starting at $T(m)$ for a different value of m, so we only need to match up X with a tuple starting at each of the first four entries. In fact, we may need to check fewer than four, because some may be ruled out if two consecutive values in the segment are different but each match up to the same value in the infinite sequence. Removing the overlines and duplicate entries then gives us four sequences in T to check, each of which are at most as long as X.

For example, we can now prove the No-2 Theorem without resorting to brute force. Remembering that the No-2 Theorem implies the existence of $0, 1, 1, 0$ or $1, 0, 0, 1$ in T_3, we can reason about where these sequences must appear. Because $T_3(4n) = T_3(4n+1)$ for all natural numbers n, we can eliminate any possibility where the first two digits **01**10 **10**01 or the last two digits 01**10** 10**01** are at positions $4n, 4n+1$. Therefore, any 0110 or 1001 sequence would have to begin at position $4n+1$ or $4n+3$. We can make a table for each of the two possibilities, where the first two rows are the direct values in 0110 and 1001 and the last two rows contain the sequence in T we wish to look for. In fact, due to Lemma 2.2, we need only look for 0110.

Case 1 ($4n+1$):

$T(m)$	$\overline{T(m+1)}$	$\overline{T(m+2)}$	$T(m+3)$
0	1	1	0
$T(m)$	$T(m+1)$	$T(m+2)$	$T(m+3)$
0	0	0	0

Case 2 ($4n+3$):

$\overline{T(m+2)}$	$T(m+3)$	$T(m+3)$	$\overline{T(m+4)}$
0	1	1	0
$T(m+2)$	$T(m+3)$		$T(m+4)$
1	1		1

In either case, we have three consecutive entries of the same value in the Thue-Morse sequence, which by Lemma 2.1 never exist. Therefore, No-2 is proven again. □

2.5 The Motif Theorem

To prove Theorem 1.1, we need to prove that no other kinds of motifs exist. No-2 has been proven, so we will break this into three parts: proving that no motifs of length 4 exist, proving that none of length 5 exist, and proving that none of length 9 or greater exist.

2.5.1 No-4

Theorem 2.2 *There are no motifs of length 4.*

Proof A counterexample, a motif of length 4, will follow the pattern 011110 or 100001. As we established in the No-2 Theorem proof above, we need only check the first example: 011110. For Case 1 (starting at 0 *mod* 4):

$T(m)$	$T(m)$	$\overline{T(m+1)}$	$\overline{T(m+2)}$	$T(m+3)$	$T(m+3)$
0	1	1	1	1	0
$T(m)$	$T(m)$	$T(m+1)$	$T(m+2)$	$T(m+3)$	$T(m+3)$
0	1	0	0	1	0

A contradiction is immediately apparent: $T(m)$ may not equal both 0 and 1, so Case 1 is ruled out.

- Case 2 has the sequence starting at 1 *mod* 4:

$T(m)$	$\overline{T(m+1)}$	$\overline{T(m+2)}$	$T(m+3)$	$T(m+3)$	$\overline{T(m+4)}$
0	1	1	1	1	0
$T(m)$	$T(m+1)$	$T(m+2)$	$T(m+3)$		T(m+4)
0	0	0	1		1

A sequence of 000 (which starts at $T(m)$) does not exist in Thue-Morse, so Case 2 is also ruled out.

Case 3:

$\overline{T(m+1)}$	$\overline{T(m+2)}$	$T(m+3)$	$T(m+3)$	$\overline{T(m+4)}$	$\overline{T(m+5)}$
0	1	1	1	1	0
$T(m+1)$	$T(m+2)$	$T(m+3)$		T(m+4)	T(m+5)
1	0	1		0	1

Such a subsequence is also impossible. Recall from Lemma 2.2 that the Thue-Morse sequence is composed of two concatenated blocks of length 2^k for any given k, which are opposites of each other. For $k = 1$, these two sequences are 01 and 10. So there are two cases.

- Sub-case 1 The arbitrary m we chose is even. Therefore, the sequence starting at position m of 10101 may be bracketed into blocks of 01 and 10 as [10][10][1. This third block must be finished by a 0, so we see that we will have a sequence of [10][10][10].

- Sub-case 2 If m is odd, the sequence is instead 1][01][01]. This is completed to be [01][01][01].

Now we can introduce one lemma to let us make short work of this case:

Lemma 2.4 $T(n) = T(2n)$.

Proof The binary representation of $2n$ is the binary representation of n with a 0 appended to the end. As per our definition of the values of the Thue-Morse sequence, the values $T(n)$ and $T(2n)$ are therefore equal. $\square$

Remark 2.1 *By the previous lemma, $T_{2n} = T_n$.*

In sub-case 1, therefore, we have $T(m) = T(m + 2) = T(m + 4) = 1$, and because m is even we may divide by 2 to get $T(m/2) + T(m/2 + 1) = T(m/2 + 2) = 1$. We know that no sequences of 000 or 111 may exist in T, so sub-case 1 does not exist. Using similar logic for sub-case 2, we see that it is also impossible. Therefore, Case 3 as a whole is impossible.

- Case 4:

$\overline{T(m+2)}$	$T(m+3)$	$T(m+3)$	$\overline{T(m+4)}$	$\overline{T(m+5)}$	T(m+6)
0	1	1	1	1	0
$T(m+2)$	$T(m+3)$		T(m+4)	T(m+5)	T(m+6)
1	1		0	0	0

This contains a subsequence of 000, which is not allowed. Therefore, Case 4 is impossible.

As none of the cases 1-4 are possible, we see that no sequence of 011110 may exist. Therefore, no subsequence of 100001 exists either, so No-4 is proven. $\square$

2.5.2   No-5

Theorem 2.3 *There are no motifs of length 5.*

Similarly to last time, we need only disprove the existence of a sequence 0111110 in T_3. We need only check two cases this time: knowing that $T_3(4m) = T_3(4m + 1)$, we cannot begin on a number which is 0 mod 4, because the initial 01 would fall across $4m$ and $4m + 1$ for some m. We also cannot begin on a number which is 3 mod 4, because the final 10 would fall across $4m$ and $4m + 1$.

Therefore, we only have two cases to check. In the checking of each one, we first display the sequence 0111110 matched up to the $T(m),\ldots$ sequence, and then below we remove overlines and duplicate values to arrive at an actual (hypothetical) sequence in T.

Case 1:

$T(m)$	$\overline{T(m+1)}$	$\overline{T(m+2)}$	$T(m+3)$	$T(m+3)$	$\overline{T(m+4)}$	$\overline{T(m+5)}$
0	1	1	1	1	1	0
$T(m)$	$T(m+1)$	$T(m+2)$	$T(m+3)$		$T(m+4)$	$T(m+5)$
0	0	0	1		0	1

This implies a sequence of 000 in T, which is not allowed.

Case 2:

$\overline{T(m+1)}$	$\overline{T(m+2)}$	$T(m+3)$	$T(m+3)$	$\overline{T(m+4)}$	$\overline{T(m+5)}$	$T(m+6)$
0	1	1	1	1	1	0
$T(m+1)$	$T(m+2)$	$T(m+3)$		$T(m+4)$	$T(m+5)$	$T(m+6)$
1	0	1		0	0	0

This again results in a 000 sequence in T, which is not allowed.

Therefore neither case is allowed, so No-5 is proven. $\square$

2.5.3   No-9+

Theorem 2.4 *There are no motifs of length 9 or greater in T_3.*

Proving this requires us only to check for a *run* of length 9 or greater. That is, we need to check for 000000000 or 111111111 (and by Lemma 2.2, only 000000000.) We use the segment-matching method to check all 4 possible cases.

Case 1:

$T(m)$	$T(m)$	$\overline{T(m+1)}$	$\overline{T(m+2)}$	$T(m+3)$	$T(m+3)$	$\overline{T(m+4)}$	$\overline{T(m+5)}$	$T(m+6)$
0	0	0	0	0	0	0	0	0
$T(m)$		$T(m+1)$	$T(m+2)$	$T(m+3)$		$T(m+4)$	$T(m+5)$	$T(m+6)$
0		1	1	0		1	1	0

As in the proof of Lemma 2.1, we know that the T sequence is composed of 01 and 10, which rules out the sequence 0110110 at any point. This is because the breakdown into such blocks is either [01][10][11][0 or 0][11][01][10], both of which contain blocks of 11.

Case 2:

8

$T(m)$ 0	$\overline{T(m+1)}$ 0	$\overline{T(m+2)}$ 0	$T(m+3)$ 0	$T(m+3)$ 0	$\overline{T(m+4)}$ 0	$\overline{T(m+5)}$ 0	$T(m+6)$ 0	$T(m+6)$ 0
$T(m)$ 0	$T(m+1)$ 1	$T(m+2)$ 1	$T(m+3)$ 0		$T(m+4)$ 1	$T(m+5)$ 1	$T(m+6)$ 0	

This results in the same sequence 0110110 as before, so we see that it is also not possible.

Case 3:

$\overline{T(m+1)}$ 0	$\overline{T(m+2)}$ 0	$T(m+3)$ 0	$T(m+3)$ 0	$\overline{T(m+4)}$ 0	$\overline{T(m+5)}$ 0	$T(m+6)$ 0	$T(m+6)$ 0	$\overline{T(m+7)}$ 0
$T(m+1)$ 1	$T(m+2)$ 1	$T(m+3)$ 0		$T(m+4)$ 1	$T(m+5)$ 1	$T(m+6)$ 0		$T(m+7)$ 1

The sequence 1101101 is also disallowed for the same reason – both possible decompositions into blocks of length 2, which are 1][10][11][01] and [11][01][10][1, contain a block of 11.

Case 4:

$\overline{T(m+2)}$ 0	$T(m+3)$ 0	$T(m+3)$ 0	$\overline{T(m+4)}$ 0	$\overline{T(m+5)}$ 0	$T(m+6)$ 0	$T(m+6)$ 0	$\overline{T(m+7)}$ 0	$\overline{T(m+8)}$ 0
$T(m+2)$ 1	$T(m+3)$ 0		$T(m+4)$ 1	$T(m+5)$ 1	$T(m+6)$ 0		$T(m+7)$ 1	$T(m+8)$ 1

This gives the sequence 1011011, which can only be broken into [10][11][01][1 and 1][01][10][11], both of which have a 11.

None of the four cases are allowed, so No-9 is proven. $\square$

2.6   6,8 Synchronization Remark

We observe that the runlengths of 1s of T_3 or $\Lambda_1(T_3)$ have a pattern in the order of 6s and 8s. In the sequence $\Lambda_1(T_3)$, runlengths of 6 almost always follow runlengths of 8 and vice versa. In sequence, 6 does in fact follow 8 and vice versa until the location 87384 in T 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.

2.7 Uniform 3s Conjecture

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

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. We will evaluate the ratio function for all powers of 2, and then reverse the segment-matching method to prove the whole Ratio Theorem.

3.1 The Weak Ratio Theorem

Theorem 3.1 (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 T_3, 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.

Theorem 3.2 (Alternative Statement Of 3.1) $\alpha_0(k) > \alpha_1(k) \; \forall \; k > 4$.

10

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 *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 3.2 *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 and 3.2 slightly, remembering the Alternative Weak Ratio Theorem Statement and replacing k with $k+3$ and $k-3$ with k :

Lemma 3.3 (3.1 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 3.4 (3.2 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 3.5 (3.1 Restated) *For odd k, $\alpha_0(k+1) > \alpha_1(k+1)$ iff $\alpha_0(k+3) > \alpha_1(k+3)$.*

Lemma 3.6 (3.2 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 3.7 *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. $\square$

3.2 Reversing Segment-Matching

In Section 2, we developed the notion of the segment-matching method, which may be thought of more generally as a *transformation* from a sequence in T_3 to one or more possible sequences in T which would be its equivalent. However, we can also develop the opposite notion – a transformation from T to T_3. Of course this is already possible:

$$[0]11[0]10[0]11[0]01[0]\cdots \rightarrow 00000\cdots.$$

But we can invert the method to obtain a more useful transformation which uses *every* digit in T to create the corresponding sequence in T_3. We may then

apply this transformation to the *whole* of the Thue-Morse sequence to better understand the T_3 sequence itself.

We recall that the sequence

$$T_3(n), T_3(n+1), T_3(n+2), T_3(n+3), ...$$

for some n which is a multiple of 4, is equivalent to

$$T(m), T(m), \overline{T(m+1)}, \overline{T(m+2)}, ...$$

For $m = 3n/4$. Setting $n = m = 0$, we see that the T_3 sequence itself is equal to:

$$T(0), T(0), \overline{T(1)}, \overline{T(2)}, ...$$

Now suppose we want to calculate the ratio function $r(n)$ for some n which is not a power of 2. We will define three new pairs of functions $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$ which are similar to the plain $\alpha\beta\gamma$ defined above, but which calculate the corresponding values for the first n instead of the first 2^n. In other words,

$$\alpha_0(k) = \hat{\alpha}_0(2^k)$$

and so forth.

We now note that the sequence $T(0), T(0), ...$ above may be reordered as:

$$2 * (T(0) + T(3) + ...) + (\overline{T(1)} + \overline{T(4)} + ...) + (\overline{T(2)} + \overline{T(5)} + ...).$$

For some number n, the first $4n$ digits of the T_3 sequence are equal to:

$$2 * \sum_{k=0}^{n-1} T(3k) + \sum_{k=0}^{n-1} \overline{T(3k+1)} + \sum_{k=0}^{n-1} \overline{T(3k+2)}.$$

In the same manner, we can express the number of zeros in the first $4n$ digits of T_3 (or the first $12n$ digits of T) as:

$$\hat{\alpha}_0(12n) = 2 * \hat{\alpha}_0(3n) + \hat{\beta}_1(3n) + \hat{\gamma}_1(3n).$$

Replacing $3n$ with n, we see:

$$\hat{\alpha}_0(4n) = 2 * \hat{\alpha}_0(n) + \hat{\beta}_1(n) + \hat{\gamma}_1(n)$$

for all n which is a multiple of 3. In fact, we can add one further restriction: if n is also even, then the total numbers of zeros in the first n digits of the Thue-Morse sequence is equal to the total number of ones, which are each equal to $n/2$. We can then restate the above expression as:

$$\hat{\alpha}_0(4n) = 2 * \hat{\alpha}_0(n) + \frac{n}{2} - \hat{\alpha}_1(n).$$

Correspondingly, for $\hat{\alpha}_1$ we have:

$$\hat{\alpha}_1(4n) = 2 * \hat{\alpha}_1(n) + \frac{n}{2} - \hat{\alpha}_0(n).$$

Subtracting:

Lemma 3.8 $\hat{\alpha_0}(4n) - \hat{\alpha_1}(4n) = 3 * \hat{\alpha_0}(n) - 3 * \hat{\alpha_1}(n)$ *for n divisible by 6.*

This may be viewed as a generalized form of Lemma 3.7.

Remark 3.1 *We can use this result to show roughly how fast the ratio function $r(n)$ decays. As the total size of $\alpha_0(k) + \alpha_1(k)$ is at most 1 away from $\frac{2^k}{3}$, we see that in the limit, $r(2^{k+2}) - 1 \simeq \frac{3}{4} * (r(2^k) - 1)$.*

3.3 Ratio Theorem

Lemma 3.9 *The difference $(\hat{\alpha}_0(x) - \hat{\alpha}_1(x)) - (\hat{\alpha}_0(x+n) - \hat{\alpha}_1(x+n))$ has absolute value at most $n/3 + 1$.*

Proof Because the $\hat{\alpha}(n)$ functions are defined as the number of zeros and ones in every third term of the first n terms of the Thue-Morse sequence, going from x to $x + n$ adds at most $n/3 + 1$ terms to consideration, which adds at most $n/3 + 1$ terms to either the zero or the one tally. $\square$

Lemma 3.10 *If the smaller value of $\hat{\alpha_0} - \hat{\alpha_1}$ is n for two numbers which are d apart, then the lowest value of $\hat{\alpha_0} - \hat{\alpha_1}$ possible between them is $n - d/6 - 1$.*

Proof The halfway point between the two numbers must be at most $d/6 + 1$ less than the lowest value, because it is $d/2$ between them. Any other number is closer to a value which is at least n, so it must be higher than this. $\square$

Now, consider two integers 6 apart x, y which are both multiples of 6, and the lower value of $\hat{\alpha}_0 - \hat{\alpha}_1$ evaluated at x and y is n. Then, the lowest possible value of $\hat{\alpha}_0 - \hat{\alpha}_1$ between $4x$ and $4y$ by Lemma 3.10 is equal to $3n - d/6 - 1$, which is equal to $3n - 5$. Now, if n is at least 3, then $3n - 5 > n$, so all values between (and including) $4x$ and $4y$ have $\hat{\alpha}_0 - \hat{\alpha_1}$ equal to at least n, so we can show that every number m between $16x$ and $16y$ has $\hat{\alpha}_0 - \hat{\alpha_1}$ at least n by choosing two adjacent multiples of 6 x' and y' which surround m and applying Lemma 3.10 once again. Therefore, we have that:

Lemma 3.11 *If x and y are both multiples of 6, and $\hat{\alpha}_0 - \hat{\alpha}_1$ evaluated at x and y is at least 3, then the value of $\hat{\alpha}_0 - \hat{\alpha}_1$ is at least 3 when evaluated at any m such that $4x \le m \le 4y$.*

By choosing some x and y such that $x = 4y$ and $\hat{\alpha}_0(n) - \hat{\alpha}_1(n) \ge 3$ for every n such that $x \le n \le y$, we can apply Lemma 3.11 repeatedly to show that there is an excess of zeros over ones for every $m > x$. We evaluated the difference of zeros and ones for $x = 6, y = 24$ and found an excess of at least three zeros as required above, so the Ratio Theorem is proven. $\square$