The Ubiquitous Thue-Morse Sequence

Jeffrey Shallit

School of Computer Science, University of Waterloo
Waterloo, ON N2L 3G1 Canada
shallit@uwaterloo.ca
https://cs.uwaterloo.ca/~shallit/

Thue's Amazing Sequence

In 1912, the Norwegian mathematician Axel Thue discovered the following infinite binary sequence:

$$\mathbf{t} = t(0)t(1)t(2)\cdots = 0\,1\,1\,0\,1\,0\,0\,1\,1\,0\,0\,1\,\cdots$$

which he used to solve a problem in pattern avoidance.

Axel Thue (1863–1922)

- Now known as the *Thue-Morse sequence*.
- Rediscovered many times!
- Appears in many different areas: mathematics, computer science, fair division of assets, music, physics, . . .
- Lots of amazing properties

The Thue-Morse Sequence Defined

The Thue-Morse sequence

$$\mathbf{t} = t(0)t(1)t(2)\cdots = 0\,1\,1\,0\,1\,0\,0\,1\,1\,0\,0\,1\cdots$$

has many equivalent definitions.

The simplest is probably the following:

$t(n) =$ the sum of the bits, taken modulo 2, of n when written in base 2.

For example: $13 = 8 + 4 + 1$, so 13 in base 2 is 1101. The sum of the bits is 3, so $t(13) = 1$.

n	0	1	2	3	4	5	6	7	8	9	10	11	12	13
$t(n)$	0	1	1	0	1	0	0	1	1	0	0	1	0	1

Another Definition of The Thue-Morse Sequence

Here's another way to define the Thue-Morse sequence:

It is the unique infinite binary sequence that starts with 0 and is the fixed point of the map sending 0 to 01 and 1 to 10.

For example:

$$\overbrace{0\,1}^{0}\ \overbrace{1\,0}^{1}\ \overbrace{1\,0}^{1}\ \overbrace{0\,1}^{0}\ \overbrace{1\,0}^{1}\ \overbrace{0\,1}^{0}\ \overbrace{0\,1}^{0}\ \overbrace{1\,0}^{1}\ \cdots$$

Outline of the Talk

- The problem of the incompetent duelers
- The multigrades problem
- Pattern avoidance
- Frameless coloring of the plane
- Thue-Morse and music
- Fragility of the Thue-Morse sequence
- Generalization of the Thue-Morse sequence
- The Woods-Robbins infinite product

Alice and Bob Have a Duel

Consider two incompetent duellists: Alice and Bob.

Alice and Bob Have a Duel

Alice and Bob take turns shooting at each other until one is hit. The successful shooter is the winner.

But both of them are incompetent: their probability p of a successful shot is fairly small.

If Alice and Bob alternate shots, like $ABABAB\cdots$, then Alice has a significant advantage.

For example, if $p = \frac{1}{2}$, then Alice succeeds on the first shot with probability $\frac{1}{2}$.

In order that Alice have a second shot, she must miss on the first shot, and Bob must miss on his shot (which occurs with probability $\frac{1}{2}$), so the probability that Alice succeeds on her second shot is $\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = (\frac{1}{2})^3 = \frac{1}{8}$.

Continuing in this fashion, Alice's probability of winning is $\frac{1}{2} + (\frac{1}{2})^3 + (\frac{1}{2})^5 + \cdots = \frac{2}{3}$. This is not fair to Bob!

Alice and Bob Have a Duel

In fact, if $p = 1/2$, then the only fair strategy is for Alice to take a shot and then have Bob take infinitely many shots!

(Fair means both participants have probability $1/2$ of winning.)

More generally, if they alternate shots, Alice's probability of winning the duel is

$$p + q^2 p + q^4 p + \cdots = \frac{p}{1 - q^2}.$$

where $q = 1 - p$.

This equals $\frac{1}{2-p}$ and is *always* greater than $\frac{1}{2}$.

So alternating shots can never be fair.

Alice and Bob Have a Duel

So how can we make the duel more fair? Obviously it depends on the probability p of a successful shot.

Cooper and Dutle suggested constructing a schedule for the order of shots, in advance, using the greedy algorithm.

Alice will shoot first. Calculate the probability of having won after n shots for each combatant, and then assign the $(n+1)$st shot to the person whose probability is smaller.

Joshua Cooper
American mathematician

Aaron Dutle
American mathematician

Alice and Bob Have a Duel

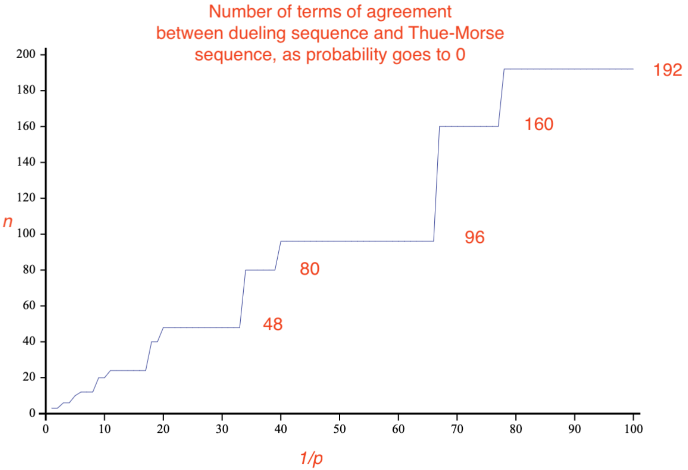
For example, if $p = 1/3$ we get

Turn	A's probability	B's probability	shooter
0	—	—	A
1	0.333	0	B
2	0.333	0.222	B
3	0.333	0.370	A
4	0.432	0.370	B
5	0.432	0.436	A
6	0.476	0.436	B
7	0.476	0.465	B
⋮	⋮	⋮	⋮

Cooper and Dutle proved that as $p \to 0$, the greedy sequence $ABBABAAB \cdots$ tends to the Thue-Morse sequence!

An Open Problem

How *quickly* does the greedy sequence converge to the Thue-Morse sequence? Some empirical data:



An Open Problem

As p goes to 0, it appears that the greedy sequence agrees with the Thue-Morse sequence on exactly either the first $3 \cdot 2^n$ or the first $5 \cdot 2^n$ terms.

And if $p = 2^{-n}$, it seems that the greedy sequence agrees with $\mathbf{t}$ on exactly the first $3 \cdot 2^{n-1}$ terms.

Can you prove that?

I offer US $25 to the first person to prove or disprove it.

(Warning: if you're doing computational experiments with floating-point arithmetic, round-off errors will quickly give you incorrect results. You need exact rational arithmetic to get the right answers.)

The Multigrades Problem

Consider a system of n equations:

$$\begin{aligned} x_1 + x_2 + \cdots + x_n &= y_1 + y_2 + \cdots + y_n \\ x_1^2 + x_2^2 + \cdots + x_n^2 &= y_1^2 + y_2^2 + \cdots + y_n^2 \\ x_1^3 + x_2^3 + \cdots + x_n^3 &= y_1^3 + y_2^3 + \cdots + y_n^3 \\ &\vdots \\ x_1^r + x_2^r + \cdots + x_n^r &= y_1^r + y_2^r + \cdots + y_n^r \end{aligned}$$

which we abbreviate by

$$(x_1, x_2, \ldots, x_n) \stackrel{r}{=} (y_1, y_2, \ldots, y_n).$$

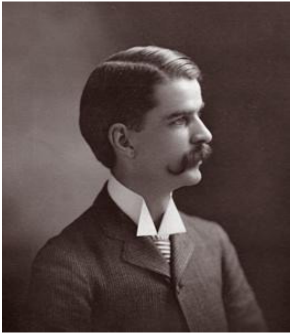
Such a system is called a *multigrade*.

We would like to find solutions in *natural numbers* (non-negative integers) where all the x_i, y_i are *distinct*. And it would be nice if n was as small as possible for a given r.

The Multigrades Problem

The multigrades problem was introduced by Tarry and Escott in 1910.

Gaston Tarry
(1843–1913)
French amateur mathematician



Edward Brind Escott, Jr.
(1868–1946)
American mathematician

The Multigrades Problem

A multigrade is called *ideal* if $n = r + 1$. Here are some examples of ideal solutions:

$$(1, 6, 8) \stackrel{2}{=} (2, 4, 9)$$
$$(1, 5, 8, 12) \stackrel{3}{=} (2, 3, 10, 11)$$
$$(2, 3, 11, 15, 19) \stackrel{4}{=} (1, 5, 9, 17, 18)$$
$$(2, 3, 13, 15, 25, 26) \stackrel{5}{=} (1, 5, 10, 18, 23, 27)$$

However, no ideal solutions are known for $r > 11$.

The Multigrades Problem: Applying the Thue-Morse Sequence

If we relax the “ideal” requirement, then we can use the Thue-Morse sequence to create solutions for arbitrarily large r, as discovered by Eugène Prouhet in 1851 (!). Namely, let

$$X = \{0 \leq i < 2^{r+1} \; : \; t(i) = 0\}$$
$$Y = \{0 \leq i < 2^{r+1} \; : \; t(i) = 1\}.$$

Then $X \stackrel{r}{=} Y$.

For example, for $r = 2$ we get

$$(0, 3, 5, 6) \stackrel{2}{=} (1, 2, 4, 7)$$

and for $r = 3$ we get

$$(0, 3, 5, 6, 9, 10, 12, 15) \stackrel{3}{=} (1, 2, 4, 7, 8, 11, 13, 14).$$

The Multigrades Problem: Going Further

Prouhet actually proved even more: the Thue-Morse solution can be generalized to arbitrary integer bases $k \geq 2$.

Define $s_k(n)$ to be the sum of the digits in the base-k representation of n.

Define $X_i = \{0 \leq i < k^{r+1} \, : \, s_k(n) \equiv i \pmod{k}\}$.

Then

$$X_0 \stackrel{r}{=} X_1 \stackrel{r}{=} \cdots \stackrel{r}{=} X_{k-1}.$$

For example, for $k = 3$ and $r = 2$ we have

$$(0,5,7,11,13,15,19,21,26) \stackrel{2}{=} (1,3,8,9,14,16,20,22,24) \stackrel{2}{=} (2,4,6,10,12,17,18,23,25).$$

Thue-Morse and Pattern Avoidance

Thue's 1912 paper on the Thue-Morse sequence was about pattern avoidance—more specifically, avoiding overlaps.

An *overlap* is a block of the form *axaxa*, where *a* is a single letter and *x* is a possibly empty block.

Examples in English of overlaps include `ululu` and `alfalfa`:

$$\overbrace{\texttt{u}}^{a}\overbrace{\texttt{l}}^{x}\overbrace{\texttt{u}}^{a}\overbrace{\texttt{l}}^{x}\overbrace{\texttt{u}}^{a} \quad \text{and} \quad \overbrace{\texttt{a}}^{a}\overbrace{\texttt{lf}}^{x}\overbrace{\texttt{a}}^{a}\overbrace{\texttt{lf}}^{x}\overbrace{\texttt{a}}^{a}.$$

Thue proved that $\mathbf{t} = 01101001\cdots$ contains no overlaps (is "overlap-free").

The proof is not very hard, but requires a bit of case analysis.

Thue-Morse and Pattern Avoidance

A *square* is two consecutive identical nonempty blocks, like the English word `murmur`.

Using the fact that the Thue-Morse sequence is overlap-free, we can (as Thue did) construct an infinite squarefree sequence, as follows:

Start with any binary overlap-free sequence, like Thue-Morse:

$$0\,1\,1\,0\,1\,0\,0\,1\,1\,0\,0\,1\,0\,1\,1\,0\,\cdots$$

Identify the positions of the 0's:

$$\color{red}{0}\,1\,1\,\color{red}{0}\,1\,\color{red}{0}\,\color{red}{0}\,1\,1\,\color{red}{0}\,\color{red}{0}\,1\,\color{red}{0}\,1\,1\,\color{red}{0}\,\cdots$$

Count the number of 1's between each occurrence of 0:

$$\color{red}{0}\overbrace{1\,1}^{2}\color{red}{0}\overbrace{1}^{1}\color{red}{0}\overbrace{\quad}^{0}\color{red}{0}\overbrace{1\,1}^{2}\color{red}{0}\overbrace{\quad}^{0}\color{red}{0}\overbrace{1}^{1}\color{red}{0}\overbrace{1\,1}^{2}\color{red}{0}\cdots$$

Thue-Morse and Pattern Avoidance

The resulting sequence

$$2\,1\,0\,2\,0\,1\,2\cdots$$

is squarefree.

To see this, suppose there were a square in it, say

$$\cdots a_1 a_2 \cdots a_i a_1 a_2 \cdots a_i \cdots .$$

Then this would correspond to the occurrence of

$$\cdots 0 \overbrace{1\cdots 1}^{a_1} 0 \cdots 0 \overbrace{1\cdots 1}^{a_i} 0 \overbrace{1\cdots 1}^{a_1} 0 \cdots 0 \overbrace{1\cdots 1}^{a_i} 0 \cdots .$$

in the binary sequence we started with.

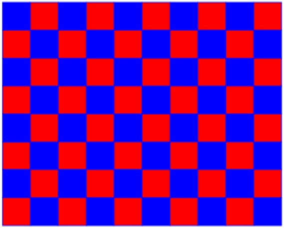
But this is an overlap *axaxa*! Take

- $a = 0$
- $x = \overbrace{1\cdots 1}^{a_1} 0 \cdots 0 \overbrace{1\cdots 1}^{a_1}$.

Frameless Colorings of the Plane

Let's consider the upper-right quarter plane and assign some colors to each lattice point (m, n).

For example, if we color a point red if $m + n$ is even and blue otherwise, we get an infinite checkerboard like

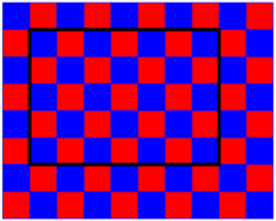


Frameless Colorings of the Plane

We want to assign the colors in such a way that we avoid some particular arrangements of colors.

We call a rectangular block a *picture frame* if it has at least two rows and two columns, and the top row equals the bottom row, and the left column equals the right column.

There are lots of picture frames in the checkerboard. Here is one:

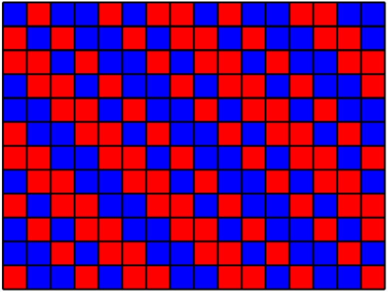


Frameless Colorings of the Plane

Question: can we color the lattice points of the upper right quarter plane so there are no picture frames at all?

We call such a coloring *frameless*.

Answer: yes, we can! In the Thue-Morse coloring, we color (m, n) with $f(m, n) = t(m + n)$. Here is an illustration, where we color a square red if $f(m, n) = 0$ and blue if $f(m, n) = 1$.

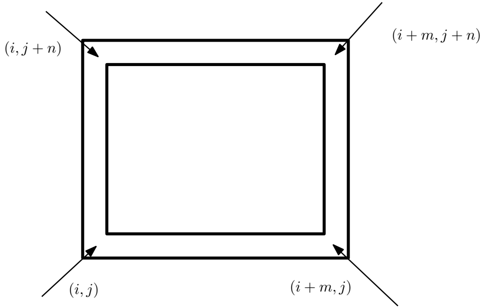


Frameless Colorings of the Plane

Theorem. (Wang 1965) The Thue-Morse coloring $f(m, n) = t(m + n)$ is frameless.

Proof. Suppose, contrary to what we want to prove, that the rectangular block whose corners are (i, j) and $(i + m, j + n)$ has a picture frame.

Then by comparing the left side to the right, we get $f(i, j + k) = f(i + m, j + k)$ for $0 \leq k \leq n$. By comparing the bottom to the top we get $f(i + l, j) = f(i + l, j + n)$ for $0 \leq l \leq m$.



Frameless Colorings of the Plane

We established that

$$f(i, j+k) = f(i+m, j+k) \quad \text{for } 0 \leq k \leq n \qquad (1)$$
$$f(i+l, j) = f(i+l, j+n) \quad \text{for } 0 \leq l \leq m.$$

Now by definition $f(w, z) = t(w+z)$.

WLOG $m \leq n$.

From (1) we get $t(x+k) = t(x+m+k)$ for $0 \leq k \leq n$ and $x = i+j$.

So the $2m+1$ symbols

$$t(x)t(x+1)\cdots t(x+2m)$$

form an overlap, a contradiction, since the Thue-Morse sequence has no overlaps.

Frameless Colorings of the Plane

We showed how to color the upper right quadrant of the plane to avoid frames.

How can we color the *entire* plane?

We leave this as a challenge (or see the paper of Wang cited at the end).

Thue-Morse and Music

Per Nørgård is the most famous living Danish composer. He independently discovered the Thue-Morse sequence and has used it, and variations of it, frequently in his music.

Because the Thue-Morse sequence avoids overlaps, it's a candidate for building self-similar but nonrepetitive music.

Per Nørgård (b. 1932)
Photo courtesy of Laivakoira2015
CC BY-SA 4.0,
`https://creativecommons.org/licenses/by-sa/4.0`
via Wikimedia Commons

Thue-Morse and Music

One of Nørgård's inventions is the so-called "infinity series" defined by

$$s(n) = \begin{cases} 0, & \text{if } n = 0; \\ s(\frac{n-1}{2}) + 1, & \text{if } n \text{ odd}; \\ -s(n/2), & \text{if } n > 0 \text{ is even.} \end{cases}$$

Here are the first few terms:

n	0	1	2	3	4	5	6	7	8	9
$s(n)$	0	1	-1	2	1	0	-2	3	-1	2

In Nørgård's piece *Voyage into the Golden Screen*, the first 512 notes of the infinity series appear, using the code $0 =$ C, $1 =$ C#, $2 =$ D, etc.

Notice that $s(n) \bmod 2 = t(n)$.

Thue-Morse and Music

Instead of playing Nørgård's piece, here is one inspired by the same ideas, where we use the major scale instead of the chromatic scale.

Play

Thue-Morse and Music

Yu Hin Au
Univ. of Saskatchewan

Christopher Drexler-Lemire

In a joint paper we proved

Theorem. (Au, D-L, & JOS) If the infinity series contains a block of the form xyx, with x non-empty, then y is at least twice as long as x. In particular, the infinity series does not contain two consecutive identical blocks.

Fragility of the Thue-Morse Sequence

The Thue-Morse sequence $\mathbf{t}$ has the following "fragility" property.

Theorem. (Rampersad) If we flip any finite number of bits of the Thue-Morse sequence, but at least one, then the resulting sequence has an overlap.

Narad Rampersad
Canadian computer scientist

For example, if we start with

$$\mathbf{t} = 0\,1\,1\,0\,\mathbf{1}\,\mathbf{0}\,0\,1\,1\,0\,0\,1\,0\,1\,1\,0\,\cdots,$$

and flip bits at index 4 and 5 (in red) it becomes

$$0\,1\,1\,0\,\mathbf{0}\,\mathbf{1}\,0\,1\,1\,0\,0\,1\,0\,1\,1\,0\,\cdots,$$

which produces an overlap as follows:

$$\overbrace{0}^{a}\;\overbrace{1\,1\,0\,0\,1}^{x}\;\overbrace{0}^{a}\;\overbrace{1\,1\,0\,0\,1}^{x}\;\overbrace{0}^{a}\;1\,1\,0\,\cdots.$$

A Generalization

The fact that **t** is overlap-free was generalized in a 2001 result of Anna Frid.

Recall that **t** is a fixed point of the map sending $0 \to 01$ and $1 \to 10$.

We can generalize this to a map that sends 0 to a block $\varphi(0)$ of m distinct numbers in $\{0, 1, \ldots, k-1\}$, and sends each number i to the block $(\varphi(0) + i) \bmod k$. Such a map is called *symmetric*.

Theorem. (Frid, 2001) If an infinite sequence **w** is the fixed point of a symmetric map, then **w** is overlap-free.

Anna Frid
Russian-French mathematician

The Woods-Robbins Infinite Product

E 2692. *Proposed by Donald R. Woods, Stanford University*

Show that the sequence of increasingly complex fractions

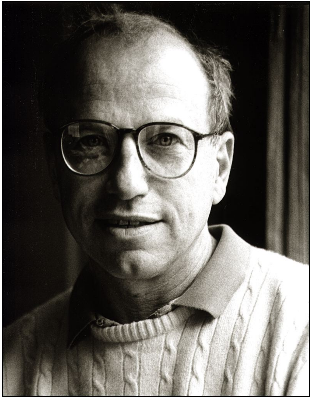
$$\frac{1}{2},\ \left(\frac{1}{2}\right)\Big/\left(\frac{3}{4}\right),\ \frac{\left(\frac{1}{2}\right)\Big/\left(\frac{3}{4}\right)}{\left(\frac{5}{6}\right)\Big/\left(\frac{7}{8}\right)},\quad \frac{\left(\frac{1}{2}\right)\Big/\left(\frac{3}{4}\right)}{\left(\frac{5}{6}\right)\Big/\left(\frac{7}{8}\right)}\Bigg/\frac{\left(\frac{9}{10}\right)\Big/\left(\frac{11}{12}\right)}{\left(\frac{13}{14}\right)\Big/\left(\frac{15}{16}\right)},\ \ldots,$$

approaches a limit, and find that limit.

Numerically we find

$$\begin{aligned} 1/2 &= 0.500 \\ (1/2)/(3/4) &= 0.666 \\ \frac{(1/2)/(3/4)}{(5/6)/(7/8)} &= 0.700 \\ &\cdots \to 0.70710678\cdots. \end{aligned}$$

The Woods-Robbins Infinite Product



David Peter Robbins
(1942–2003)
American mathematician
Solver of the Woods problem

After simplification, it is not hard to see that the fraction $(2n+1)/(2n+2)$ appears in the numerator if $t(n) = 0$ and in the denominator if $t(n) = 1$.

Hence, without worrying too much about convergence, the limit in the Woods-Robbins problem can be written as

$$\prod_{n \geq 0} \left(\frac{2n+1}{2n+2} \right)^{(-1)^{t(n)}},$$

where $(t(n))$ is the Thue-Morse sequence.

The Woods-Robbins Infinite Product

Theorem. We have

$$\prod_{n\geq 0}\left(\frac{2n+1}{2n+2}\right)^{(-1)^{t(n)}} = \frac{\sqrt{2}}{2}.$$

Proof. (Allouche) Define

$$A = \prod_{n\geq 0}\left(\frac{2n+1}{2n+2}\right)^{(-1)^{t(n)}}, \quad B = \prod_{n\geq 1}\left(\frac{2n}{2n+1}\right)^{(-1)^{t(n)}}.$$

Then

$$\begin{aligned} AB &= \frac{1}{2}\prod_{n\geq 1}\left(\frac{n}{n+1}\right)^{(-1)^{t(n)}} \\ &= \frac{1}{2}\prod_{n\geq 0}\left(\frac{2n+1}{2n+2}\right)^{(-1)^{t(2n+1)}} \prod_{n\geq 1}\left(\frac{2n}{2n+1}\right)^{(-1)^{t(2n)}} = \frac{1}{2}A^{-1}B. \end{aligned}$$

Cancelling the B from both sides gives $A = \frac{1}{2}A^{-1}$, and so $A^2 = 1/2$.

The Greedy Algorithm and Woods-Robbins

Let's try to write $\sqrt{2}/2$ as a product of terms of the form $(2n+1)/(2n+2)$ or $(2n+2)/(2n+1)$ by choosing greedily, at each step, the next term to make the product so far most closely approximate $\sqrt{2}/2$:

Start with $1/2 = 0.500\cdots$. This is smaller than $\sqrt{2}/2 = 0.7071\cdots$, so multiply by $4/3$ to make the product larger.

Now we have $1/2 \cdot 4/3 = 0.666\cdots$. This is still smaller than $\sqrt{2}/2 = 0.7071\cdots$, so multiply by $6/5$ to make the product larger.

Now we have $1/2 \cdot 4/3 \cdot 6/5 = 0.800$. This is larger than $\sqrt{2}/2 = 0.7071\cdots$, so multiply by $7/8$ to make the product smaller.

Theorem. (Allouche & Cohen, 1985) If we keep following the greedy algorithm, the terms chosen are exactly those in the Woods-Robbins product!

The Greedy Algorithm and Woods-Robbins

Jean-Paul Allouche
b. 1953
French mathematician

Henri Cohen
b. 1947
French mathematician
Image courtesy of MFO

A connection with π

Since this is π day, we expect some connection between the Thue-Morse sequence and π.

Here are two formulas discovered by Jean-Paul Allouche:

$$\prod_{n\geq 0} \left(\frac{2n+1}{2n+2}\right)^{2t(n)} \left(\frac{2n+3}{2n+2}\right) = \frac{2\sqrt{2}}{\pi}$$

$$\prod_{n\geq 0} \left(\frac{2n+1}{2n+2}\right)^{2(1-t(n))} \left(\frac{2n+3}{2n+2}\right) = \frac{\sqrt{2}}{\pi}$$

For Further Reading

Allouche & Shallit, The ubiquitous Prouhet-Thue-Morse sequence, available at
https://cs.uwaterloo.ca/~shallit/Papers/ubiq15.pdf.

Au & Drexler-Lemire & Shallit, Notes and note pairs in Nørgård's infinity series, *Journal of Mathematics and Music* **11** (2017), 1–19.

Cooper & Dutle, Greedy Galois games, *Amer. Math. Monthly* **120** (2013), 441–451.

Propp, Avoiding chazakah with the Prouhet-Thue-Morse sequence, available at https://mathenchant.wordpress.com/2017/01/16/avoiding-chazakah-with-the-prouhet-thue-morse-sequence/.

Wang, Games, logic, and computers, *Scientific American* **213** (5) (November 1965), 98–107.