July 21, 2000

Thue-Morse sequence and p-adic topology for the free monoid.

Jean Berstel, Maxime Crochemore* and Jean-Eric Pin†

Jean.Berstel@univ-mlv.fr,
Maxime.Crochemore@univ-mlv.fr,
Jean-Eric.Pin@liafa.jussieu.fr

Abstract

Given two words u and v, the binomial coefficient $\binom{u}{v}$ is the number of ways v appears as a subword (or subsequence) of u. The Thue-Morse sequence is the infinite word $t = abbabaab \cdots$ obtained by iteration of the morphism $\tau(a) = ab$ and $\tau(b) = ba$. We show that, for every prime p, and every positive integer n, there exists an integer $m = f(p, n)$, such that, for every non-empty word v of length less than of equal to n, the binomial coefficient $\binom{t[m]}{v}$ is congruent to 0 modulo p. In fact $f(p, n) = 2^n p^{1+\lfloor log_p n \rfloor}$ for $p \neq 2$ and $f(2, n) = 2^k$ if $F_{k-1} \leq n < F_k$, where F_k denotes the k-th Fibonacci number. It follows that, for each prime number p, there exists a sequence of left factors of t of increasing length, the limit of which is the empty word in the p-adic topology of the free monoid.

1 Introduction

The aim of this paper is to prove a new combinatorial property of a famous infinite word, called after his discoverers the word of Thue-Morse. This word has a great number of nice combinatorial properties, most of which can be found for instance in [3, Chap. 2]. It plays a central role in the study of square-free and cube-free words, and is also one of the "historical" examples of an infinite word defined by iteration of a morphism. As such, a number of papers have been devoted to the study of its factors, but Ochsenschlager [4] was the first to consider the subwords (or subsequences) of its left factors. One of the more useful tools in the study of subwords is the binomial coefficient introduced by Eilenberg [1], that counts, roughly speaking, the number of ways a given word v appears as a subword of another word u. Our main result shows that, given a prime number p and a positive integer n, one can find a left factor u of the word of Thue-Morse, such that all the binomial coefficients associated with the non-empty words v of length less than or equal to n are simultaneously congruent to 0 modulo p. As an application, we show that for each prime number p, there

*Institut Gaspard Monge, Université de Marne-la-Vallée, 5, Boulevard Descartes, Champs-sur-Marne, F-77454 Marne-la-Vallée Cedex 2

†LIAFA, Université Paris VII and CNRS, Case 7014, 2 Place Jussieu, 75251 Paris Cedex 05, France

exists a subsequence of the sequence of the left factors of the word of Thue-Morse, the limit of which is the empty word in the p-adic topology of the free monoid.

2 Counting the subwords of the Thue-Morse sequence

Let A be a finite alphabet. We denote by A^* the free monoid over A and by 1 the empty word. Let u and v be two words of A. A word $u = a_1 a_2 \cdots a_k$ is said to be a *subword* of v if v can be factorized as $v = v_0 a_1 v_1 \cdots a_k v_k$ where $v_0, v_1, \ldots, v_k \in A^*$. Following Eilenberg [1], we set

$$\binom{v}{u} = \mathrm{Card}\{(v_0, v_1, \ldots, v_k) \in A^* \times A^* \times \cdots \times A^* \mid v_0 a_1 v_1 \cdots a_k v_k = v\}$$

Thus $\binom{v}{u}$ is the number of distinct ways to write u as a subword of v. For instance, $\binom{aabbaa}{aba} = 8$ and $\binom{a^n}{a^m} = \binom{n}{m}$. The basic properties of these *binomial coefficients* are summarized by the following formulae

(a) For every word $u \in A^*$, $\binom{u}{1} = 1$.

(b) For every non empty word $u \in A^*$, $\binom{1}{u} = 0$.

(c) If $w = uv$, then $\binom{w}{x} = \sum_{x_1 x_2 = x} \binom{u}{x_1}\binom{v}{x_2}$.

Let $A = \{a, b\}$, and let $\tau : A^* \to A^*$ be the monoid morphism defined by $\tau(a) = ab$ and $\tau(b) = ba$. For every $n \geq 0$, we set $u_n = \tau^n(a)$ and $v_n = \tau^n(b)$. Then u_0, $u_1, \ldots$ is a sequence of words of A^* such that each u_i is a proper left factor of u_{i+1}. Therefore this sequence defines an infinite word $t = abbabaabbaababba\cdots$ called the *infinite word of Thue-Morse*. The aim of this section is the study of the binomial coefficients of the form $\binom{t[m]}{v}$ where $t[m]$ denotes the left factor of length n of t. We first recall the result of Ochsenschlager.

Theorem 2.1 [4] *For every word x such that $0 \leq |x| \leq n$, $\binom{u_n}{x} = \binom{v_n}{x}$. Furthermore there exists a word x of length $n+1$ such that $\binom{u_n}{x} \neq \binom{v_n}{x}$.*

If we count modulo some prime number p, we have the following main result.

Theorem 2.2 *For every prime number p, and for every positive integer n, there exists an integer $m = f(p, n)$ such that, for every non-empty word v of length less than or equal to n, $\binom{t[m]}{v} \equiv 0 \mod p$.*

Theorem 2.2 is the consequence of two more precise results, corresponding to the cases $p = 2$ and $p \neq 2$, respectively. We first treat the case $p = 2$. Let $(F_n)_{n \geq 0}$ be the Fibonacci sequence defined by $F_0 = 0$, $F_1 = 1$ and $F_{n+2} = F_{n+1} + F_n$ for every $n \geq 0$. Then we have

Theorem 2.3 *For every $n \geq 0$, and for every word x such that $0 < |x| < F_n$, $\binom{u_n}{x} \equiv \binom{v_n}{x} \equiv 0 \mod 2$.*

Proof. We show by induction on n that

$$\binom{u_n}{x} \equiv 0 \mod 2 \text{ for } 0 < |x| < F_n, \textit{and}$$
$$\binom{u_n}{x} \equiv \binom{v_n}{x} \mod 2 \text{ for } 0 < |x| < F_{n+1}.$$

There relations clearly hold for $n = 0, 1$. Thus let $n > 1$ and let x be a word with $0 < |x| < F_{n+1}$. Then

$$\binom{u_{n+1}}{x} = \sum_{x_1 x_2 = x} \binom{u_n}{x_1}\binom{v_n}{x_2} \equiv \binom{u_n}{x} + \binom{v_n}{x} \mod 2$$

Indeed, if $x = x_1 x_2$ and $x_1, x_2 \neq 1$, then $|x_1| < F_n$ or $|x_2| < F_n$ and consequently, either $\binom{u_n}{x_1} \equiv 0 \mod 2$ or $\binom{v_n}{x_2} \equiv 0 \mod 2$.

Next, by the induction hypothesis, $\binom{u_n}{x} \equiv \binom{v_n}{x} \equiv 0 \mod 2$, which shows that

$$\binom{u_n}{x} + \binom{v_n}{x} \equiv 0 \mod 2$$

Consider now a word x with $0 < |x| < F_{n+2}$. Then again

$$\binom{u_{n+1}}{x} = \sum_{x_1 x_2 = x} \binom{u_n}{x_1}\binom{v_n}{x_2} = \binom{u_n}{x} + \binom{v_n}{x} + \sum \binom{u_n}{x_1}\binom{v_n}{x_2}$$

the sum on the right hand side being restricted to all pairs (x_1, x_2) with $x = x_1 x_2$ and $0 < |x_1|, |x_2| < F_{n+1}$. Indeed, if x_1 or x_2 has length greater that F_{n+1} then the other word has length less than F_n and the corresponding term vanishes.

Using the induction hypothesis, we get that

$$\sum_{x_1 x_2 = x} \binom{u_n}{x_1}\binom{v_n}{x_2} = \sum_{x_1 x_2 = x} \binom{v_n}{x_1} + \binom{u_n}{x_2}$$

Theorem 2.2 follows immediately from Theorem 2.3 when $p = 2$. It suffices to put $f(2, n) = 2^k$ if $F_{k-1} \leq n < F_k$. The first values of $f(2, n)$ are given in Table 1.

n	1	2	3	4	5	6	7	8	9	10	11	12	13
f(2,n)	4	8	16	16	32	32	32	64	64	64	64	64	128

Table 1: The first values of $f(2, n)$.

We now consider the case $p \neq 2$. Put $f(p, n) = 2^n p^{1+\lfloor log_p n \rfloor}$. Then we can state

Theorem 2.4 *For every prime number $p \neq 2$, for every positive integers i and n, and for every word x such that $0 < |x| \leq n$, if $m = f(p, n)$, then $\binom{t[im]}{x} \equiv 0 \mod p$.*

We fix p and prove the theorem by induction on n. For $n = 1$, $f(p, 1) = 2p$, and we have for every $i > 0$,

$$\binom{t[2ip]}{a} = \binom{t[2ip]}{b} = ip \equiv 0 \mod p.$$

Assume that the theorem holds for some $n \geq 1$, and let x be a word such that $0 < |x| \leq n+1$. Put $f(p, n+1) = m$ and let i be a fixed positive integer. Then $t[im] = \tau(u)$ where $u = t[\frac{1}{2}im]$, and since $m/2$ is a multiple of $f(p, n)$, we have by induction, $\binom{u}{s} \equiv 0 \mod p$ for every word s such that $0 < |s| \leq n$.

At this point, we need some algebraic tools to conclude the proof. Let $k = \mathbb{Z}/p\mathbb{Z}$ and let $k\langle A^* \rangle$ be the algebra of polynomials in non-commutative variables over A with coefficients in k.

The monoid morphism $\tau : A^* \to A^*$ can be extended to an endomorphism $\tau : k\langle A^* \rangle \to k\langle A^* \rangle$. Another useful endomorphism is the Magnus transformation $\mu : k\langle A^* \rangle \to k\langle A^* \rangle$, defined by $\mu(a) = 1 + a$ and $\mu(b) = 1 + b$. As is well-known [3, p.123] , we have, for every word $s \in A^*$,

$$\mu(s) = \sum_{x \in A^*} \binom{s}{x} x$$

Let γ be the morphism defined by $\gamma(a) = a + b + ab$ and by $\gamma(b) = a + b + ba$. Then $\mu\tau = \gamma\mu$. Put, for every $s \in A^*$,

$$\gamma(s) = \sum_{x \in A^*} \langle \gamma(s), x \rangle x$$

Then we have

$$\mu\tau(u) = \gamma\mu(u) = \gamma\Big(\sum_{s \in A^*} \binom{u}{s} s \Big) = \sum_{s \in A^*} \binom{u}{s} s\gamma(s).$$

and hence

$$\mu\tau(u) = \sum_{x \in A^*} \sum_{s \in A^*} \binom{u}{s} \langle \gamma(s), x \rangle x$$

On the other hand,

$$\mu\tau(u) = \sum_{x \in A^*} \binom{\tau(u)}{x} x$$

Therefore

$$\binom{\tau(u)}{x} = \sum_{x \in A^*} \sum_{s \in A^*} \binom{u}{s} \langle \gamma(s), x \rangle.$$

Now, it follows immediately from the definition of γ that $\langle \gamma(s), x \rangle = 0$ if $|s| > |x|$ and $\langle \gamma(s), x \rangle = 1$ if $|s| = |x|$. Furthermore, by the induction hypothesis, $\binom{u}{s} \equiv 0$ mod p for every word s such that $0 < |s| < |x|$. Thus

$$\binom{\tau(u)}{x} \equiv \sum_{|s|=|x|} \binom{u}{s} \equiv \binom{|u|}{|x|} \equiv \binom{im}{|x|} \mod p.$$

Thus it remains to prove that $\binom{im}{|x|} \equiv 0 \mod p$. Given an integer i, denote by $\nu_p(i)$ the greatest integer such that $p^{N_p(i)}$ divides i. Observe that if $i > j$ and $\nu_p(i) > \nu_p(j)$, then $\nu_p(i-j) = \nu_p(j)$. Clearly, it suffices to show that

$$S = \Big(\sum_{0 \leq k < |x|} \nu_p(im - k)\Big) - \Big(\sum_{0 < k \leq |x|} \nu_p(k)\Big) > 0.$$

But since $m = 2^{n+1}p^{1+\lfloor \log_p(n+1)\rfloor}$, we have for $0 < k \leq |x| \leq (n+1)$, $\nu_p(im) \geq 1 + \lfloor \log_p(n+1)\rfloor > \nu_p(k)$, and hence $\nu_p(im - k) = \nu_p(k)$. Therefore $S = \nu_p(im) - \nu_p(|x|) > 0$ as required. □

3 The p-adic topology of the free monoid

Recall that a p-group is a finite group of order p^k for some $k > 0$. One can show that two distinct words u and v of A^* can always be separated by a p-group in the following sense : there exists a p-group G and a monoid morphism $\varphi : A^* \to G$ such that $\varphi(u) \neq \varphi(v)$. Set, for every $u, v \in A^*$

$$r(u,v) = \min\{\mathrm{Card}(G) \mid G \text{ is a } p\text{-group that separates } u \text{ and } v\}$$

and

$$d(u,v) = e^{-r(u,v)}$$

with the usual conventions $\min \emptyset = \infty$ and $e^{-\infty} = 0$. Then d is a metric (in fact an ultrametric) which defines a topology on A, called the *p-adic topology* of the free monoid. This topology is the analoguous for the free monoid of the topology of the free group introduced by M. Hall [2]. It is the coarsest topology such that every monoid morphism from A into a discrete p-group is continuous. A^*, equipped with this topology, is a topological monoid. The interested reader is referred to [2,5,6] for a more detailed study of this topology. An example of converging sequence is given by the following proposition.

Proposition 3.1 *For every word $w \in A^*$, $\lim_{n\to\infty} w^{p^n} = 1$*

Proof. By the definition of the topology, it suffices to show that if $\varphi : A^* \to G$ is a monoid morphism onto a discrete p-group G, then $\lim_{n\to\infty} \varphi(g^{p^n}) = 1$. But if $\mathrm{Card}(G) = p^k$, then for $n \geq k$, $\varphi(g^{p^n}) = 1$ since the order of $\varphi(g)$ divides p^k. □

Since the multiplication is continuous, we also have, for every $x, y \in A^*$, $\lim_{n\to\infty} xg^{p^n}y = xy$, but it is less obvious to find an example of converging sequence that is not directly related to Proposition 3.1.

Theorem 3.2 *For every prime number p, there exists a strictly increasing sequence $m_1 < m_2 < \cdots$ such that $\lim_{n\to\infty} t[m_n] = 1$.*

Proof. Fix a prime number p, and set $m_n = f(p,n)$, where $f(p,n)$ is the function introduced after the proof of Theorem 2.3. Let $\sim_n$ be the congruence over A^* defined by

$u \sim_n v$ if and only if,

$$\text{for every word } x \text{ such that } |x| \leq n, \binom{u}{x} \equiv \binom{v}{x} \mod p.$$

By Theorem 2.2, $t[m_n] \sim_n 1$. Denote by $\varphi_n : A^* \to A^*/{\sim_n} = G_n$ the natural morphism. It is shown in [1] that G_n is a p-group and that for every monoid morphism from A^* into a p-group G, there exists a positive integer $k = k(\varphi)$ and a group morphism $\alpha_k : G_k \to G$ such that $\varphi = \alpha_k \varphi_k$. Now, if $n \geq k$, $t[m_n] \sim_k 1$ and hence $\varphi_k(t[m_n]) = 1$. It follows that $\varphi(t[m_n]) = 1$ for every $n \leq k$ and thus $\lim\limits_{n\to\infty}(t[m_n]) = 1$ in the discrete p-group G. Therefore $\lim\limits_{n\to\infty}(t[m_n]) = 1$.

Note that Theorem 3.2 cannot be deduced from Proposition 3.1 since the Thue-Morse doesn't contain any factor of the form u^3, where u is a non-empty word [3].

References

[1] S. Eilenberg, Automata, Languages and Machines, Vol B, Academic Press, New-York (1976).

[2] M. Hall Jr, A topology for free groups and related groups, *Ann. Math.* **52** (1950) 127-139.

[3] M. Lothaire, Combinatorics on Words, Encyclopedia of Mathematics 17, Addison Wesley, New- York (1983).

[4] P. Ochsenschlager, Binomialkoeffizenten und Shuffle-Zahlen, Technischer Bericht, Fachbereicht Informatik, T.H. Darmstadt (1981).

[5] J.E. Pin, Finite group topology and p-adic topology for free monoids, *12th ICALP, Lecture Notes in Computer Science* **194** (1985) 445-455.

[6] Ch. Reutenauer, Une topologie du monode libre, *Semigroup Forum* **18**, (1979), 33-49.