A CRITERION FOR NON-AUTOMATICITY OF SEQUENCES

Jan-Christoph Schlage-Puchta

keywords: Automatic sequence, Finite automaton.

MSC 11B85

A finite automaton consists of a finite set $\mathcal{S}$ of states with a specified starting state s_0, an input alphabet $\mathcal{A}$, an output alphabet $\mathcal{B}$, and two functions $f : \mathcal{A} \times \mathcal{S} \to \mathcal{S}$, $g : \mathcal{S} \to \mathcal{B}$. Given a word w over $\mathcal{A}$, the output of the automaton is determined as follows: At first, the automaton is in s_0. Then the first letter a of w is read, and the new state of the automaton is changed to $s_1 = f(a, s_0)$. Then the next letter b of w is read, and the state of the automaton is changed to $s_2 = f(b, s_1)$. This is repeated untill all letters of w are read, and the procedure terminates. If the automaton ends in the state s, it returns the value $g(s)$.

Fix some integer $q \geq 2$. In our context, the alphabet $\mathcal{A}$ consists of the integers $0, 1, \dots, q-1$, and $\mathcal{B}$ consists of integers or elements in some fixed finite fields. Every integer $n \geq 1$ can be written in the form $n = \sum e_i(n)q^i$ with $e_i(n) \in \{0, 1, \dots, q-1\}$, hence n can be viewed as word over $\mathcal{A}$, and the automaton can be applied to this word. More precisely, write $n = \sum_{i=0}^k e_i q^i$ with $e_i \in \{0, 1, \dots, q-1\}$ and $e_k \neq 0$, and identify the integer n with the string $e_k e_{k-1} \dots e_1 e_0$. In this way, every automaton defines a sequence $(a_n)_{n\geq 0}$. An automaton with $\mathcal{A} = \{0, 1, \dots, q-1\}$ is called a q-automaton, and an arbitrary sequence is called q-automatic, if there exists a q-automaton which generates this sequence. More generally, a sequence is called automatic, if it is k-automatic for some integer $k \geq 2$.

Apart from intrinsic interest, the question whether a given sequence is automatic is of interest because of its number theoretical consequences. In fact, automaticity and algebraicity are linked via the following result of G. Christol, T. Kamae, M. Mendès France and G. Rauzy [2].

Theorem 1. *Let p be a prime number, $(a_n)_{n\geq 1}$ be a sequence of elements in $\mathbb{F}_p$. Then the series $\sum_{n=1}^\infty a_n x^n$ is algebraic over $\mathbb{F}_p(x)$ if and only if the sequence (a_n) is p-automatic.*

Hence, to prove the transcendence of a power series, we need only to show that a certain sequence is not automatic. This can for example be accomplished by the following theorem of A. Cobham [3].

Theorem 2. *Let $(a_n)_{n\geq 1}$ be an automatic seqence over an alphabet $\mathcal{B}$. Assume that for some $a \in \mathcal{B}$ the limit $\delta_a = \lim_{x\to\infty} \frac{1}{x}|\{n \leq x : a_n = a\}|$ exists. Then δ_a is rational.*

In [1], J.-P. Allouche used this to prove the following result.

Corollary 1. *The power series* $f(x) = \sum_{n\geq 1}(\mu(n) \bmod p)x^n$ *is transcendental over* $\mathbb{F}_p(x)$ *for all primes* p.

Here, $\mu(n)$ denotes the Möbius-function, i.e., the multiplicative function satisfying $\mu(p) = -1, \mu(p^k) = 0$ for all primes p and integers $k \geq 2$.

Proof. Since $\mu(n) = 0$ if and only if n is divisible by some square $a^2, a \geq 2$, we see that in the notation of Theorem 2,

$$\delta_0 = \lim_{x\to\infty} \frac{1}{x}|\{n \leq x : \exists a \geq 2, a^2|n\} = 1 - \prod_p \left(1 - \frac{1}{p^2}\right) = 1 - \frac{\pi^2}{6},$$

hence, the limit exists and is irrational. So by Theorem 2, the sequence $(\mu(n) \pmod p)_{n\geq 1}$ is not automatic. □

Albeit short and ingenious, the proof has the disadvantage that it is difficult to apply to other situations for two reasons. First, it requires to evaluate $\prod_p \left(1 - \frac{1}{p^2}\right)$ and prove that the result is irrational. This is equivalent to Euler's evaluation of $\zeta(2)$ and the fact that π^2 is irrational. In our case, these are well known, yet non-trivial facts. However, in other cases there might be no known formula for δ_a. The second, more fundamental problem is that in many cases $\delta_a = \frac{1}{|\mathcal{B}|}$ for all a, so Theorem 2 cannot be applied.

The aim of this note is to give another proof of Corollary 1. In fact, we have the following more general result.

Theorem 3. *Let* $(a_n)_{n\geq 1}$ *be an automatic sequence. Assume that for some letter* a *and for every integer* k *there exists an integer* n *such that* $a_n = a_{n+1} = \cdots = a_{n+k} = a$. *Then there is a constant* $c > 0$ *such that for an infinite number of integers* x *we have* $a_n = a$ *for all* $n \in [x, (1+c)x]$.

Other criteria involving strings of repeated values can be found in [4]. Before proving our theorem, we first give some corollaries.

Corollary 2. *Let* $(q_i)_{i\geq 1}$ *be a sequence of positive integers such that* $\sum_i \frac{1}{q_i} < \infty$. *Assume that for all* $k \geq 1$, *there are indices* $i_1, \dots, i_k$ *such that* $(q_{i_l}, q_{i_m}) = 1$ *for* $1 \leq l < m \leq k$. *For all integers* $n \geq 1$, *set* $a_n = 0$, *if there exists some* i *such that* $q_i|n$, *and* $a_n = 1$ *otherwise. Then the sequence* $(a_n)_{n\geq 1}$ *is not automatic.*

Proof. Assume that the sequence $(a_n)_{n\geq 1}$ is automatic, and let k be a given integer. Choose indices $i_1, \dots, i_k$ as in the Corollary. By the Chinese remainder theorem, there is some integer n solving $n+l \equiv 0 \pmod{q_{i_l}}$, that is, $a_{n+l} = 0$ for $1 \leq l \leq k$. Hence, the assumptions of Theorem 3 are satisfied, and we deduce that there exist some $c > 0$ and arbitrarily large integers x such that $a_n = 0$ for all $n \in [x, (1+c)x]$. On the other hand, we can bound from below the number of integers $n \in [x, (1+c)x]$ such that $a_n = 0$ in the following way. Let $\epsilon > 0$ be given, and let K be some constant with $\sum_{i>K} \frac{1}{q_i} < \epsilon$. Let L be the least common multiple of $q_1, \dots, q_K$. Then the set of all integers n such that $q_i \not| n$ for all $i \le K$ is periodic with period L, and has density $d_K \ge \prod_{i\le K}\left(1-\frac{1}{q_i}\right)$, with equality if and only if the q_i are pairwise coprime. Note that

$$d_K > \prod_{i=1}^{\infty}\left(1-\frac{1}{q_i}\right) > 0,$$

that is, d_K can be bounded away from 0 independently of K. Now for $x \to \infty$, we have

$$\begin{aligned} |\{n \in [x,(1+c)x] : a_n = 1\}| &\ge |\{n \in [x,(1+c)x] : \forall i \le K : q_i \not| n\}| \\ &\quad - \sum_{i>K} |\{n \in [x,(1+c)x] : q_i|n\}| \\ &\ge d_K cx - L - \epsilon cx - |\{i : q_i \le (1+c)x\}| \\ &\ge (d_K-\epsilon)cx + o(x), \end{aligned} \tag{1}$$

since $|\{i : q_i \le (1+c)x\}| = o(x)$, for otherwise the series $\sum \frac{1}{q_i}$ would diverge. In fact, if

$$\limsup_{x\to\infty} \frac{|\{i : q_i < x\}|}{x} > 0,$$

there exists some constant $k > 0$ such that

$$\limsup_{n\to\infty} \frac{|\{i : 2^n \le q_i < 2^{n+1}\}|}{2^{n+1}} > k,$$

thus, $\sum_i \frac{1}{q_i} = \infty$. By Theorem 3 we would find arbitrarily large integers x such that the left-hand side of (1) is zero, thus we arrive at a contradiction. So the sequence $(a_n)_{n\ge1}$ is not automatic. $\square$

Choosing $q_i = p_i^2$, with p_i the i-th prime number, we find that the sequence $(\mu(n)^2 \pmod p)_{n\ge1}$ is not automatic, which is slightly stronger then Corollary 1.

Out next result deals with the automaticity of multiplicative functions.

Corollary 3. *Let $f : \mathbb{N} \to \mathbb{Z}$ be a multiplicative function. Let $q \ge 2$ be an integer. Assume that the following conditions hold.*

(1) *There exist infinitely many primes p such that there exists some $h_p \ge 1$ with $q|f(p^{h_p})$.*
(2) *If b_n denotes the n-th integer with $f(b_n) \not\equiv 0 \pmod q$, we have $\frac{b_{n+1}}{b_n} \to 1$.*

Then the sequence $(f(n) \pmod q)_{n\ge1}$ is not automatic.

Proof. As in the proof of Corollary 2, the first condition implies that for every k there exist some n with $f(n) \equiv f(n+1) \equiv \cdots \equiv f(n+k) \equiv 0 \pmod q$, while the second condition means that for every $c > 0$ there are only finitely many integers x such that $f(n) \equiv 0 \pmod q$ holds for all $n \in [x,(1+c)x]$. Together with Theorem 3, we obtain the desired conclusion. $\square$

This result generalizes a theorem of S. Yazdani [5, Theorem 2]. In fact, the conditions on f are relaxed in two aspects: First, the integers h_p are allowed to depend on p. More important, the lower bound for the density of the set of integers n satisfying $q \nmid f(n)$ in the second condition of Corollary 3 is smaller then in [5, Theorem 2]. The latter theorem requires $q \nmid f(p)$ for all primes in a residue class, which by the prime number theorem for arithmetic progressions implies condition (2) of Corollary 3.

Now we return to the proof of Theorem 3.

Proof of Theorem 3. Assume that $(a_n)_{n\geq 1}$ is a sequence satisfying the conditions of Theorem 3, and it is generated by a q-automaton. For every integer l, let n be an integer such that $a_{n+i} = a$ for all $i \leq q^l$. We may assume that n is divisible by q^l. Indeed, by hypothesis, for all integers $l \geq 1$ there exists an integer m such that $a_{m+i} = a$ for $0 \leq i < 2q^l$. Let n_l be the least integer such that $n \geq m$ and $q^l|n$. Then $n < m + q^l$, and therefore $a_{n+i} = a$ for $0 \leq i < q^l$. Let s be the state of the automaton reached when reading all digits of n except the last l digits. Then the definition of s implies that all states accessible from s within precisely l steps return a. To every state s define a set $\mathcal{N}_s \subseteq \mathbb{N}$ such that $l \in \mathcal{N}_s$ if and only if all states accessible from s within precisely l steps return a. Our argument above shows that for every $l \in \mathbb{N}$ there is some state s accessible from the starting state such that $l \in \mathcal{N}_s$. Hence, since there are only finitely many states, there exists some s_0 such that s_0 is accessible from the starting state, and there are infinitely many l such that all states accessible from s_0 in precisely l steps return a. Let d be some integer such that when reading d, the automaton stops in the state s_0. Then we claim that for all $l \in \mathcal{N}_{s_0}$ and all $n \in [dq^l, dq^l + q^l - 1]$, we have $a_n = a$. In fact, after reading d, the automaton is in state s_0, then, after reading l arbitrary digits, it is in some state returning a. Hence, our theorem follows with $c = (1 - q^{-1})d^{-1}$. $\square$

References

[1] J.-P. Allouche, Note on the transcendence of a generating function, in *New trends in probability and statistics, Vol. 4* (Palanga, 1996), VSP, Utrecht, 1997, 461–465.

[2] G. Christol, T. Kamae, M. Mendès France and G. Rauzy, Suites algébriques, automates et substitutions, *Bull. Soc. Math. France* **108** (1980), 401–419.

[3] A. Cobham, Uniform tag sequences, *Math. Systems Theory* **6** (1972), 164–192.

[4] J.-Y. Yao, Critères de non-automaticité et leurs applications, *Acta Arith.* **80** (1997), 237–248.

[5] S. Yazdani, Multiplicative functions and k-automatic sequences, *J. Théor. Nombres Bordeaux* **13** (2001), 651–658.

Jan-Christoph Schlage-Puchta
Mathematisches Institut
Eckerstr. 1
79111 Freiburg
Germany
jcp@arcade.mathematik.uni-freiburg.de