PROCEEDINGS OF THE
ROMAN NUMBER THEORY ASSOCIATION
Volume 1, *Number* 1, *March* 2016, *pages* 23–27

Christian Mauduit

Automata and Number Theory

written by Valerio Dose

Many natural questions in number theory arise from the study of the multiplicative representations of integers, and they are often at the origin of many important open problems in Mathematics and Computer Science. Among these questions, a simpler family consists of those which can be formulated by means of functions defined using an algorithm which is "simple" enough, in a way that will be made clear below.

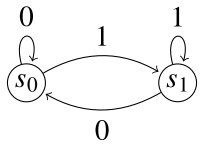
The study of a finite number of subsets of the natural numbers $\mathbb{N}$, can be related to the study of sequences of symbols in a finite set. For example, we can associate to even and odd numbers the set of symbols $\{0, 1\}$ and the infinite sequence $010101\ldots$. Also, we can associate to any subset $E \subseteq \mathbb{N}$ and to any integer $q \geq 2$, the language

$$L_q(E) = \{\text{rep}_q(n), n \in E\}$$

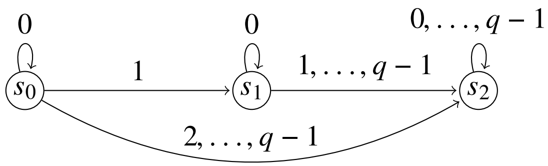
where $\text{rep}_q(n)$ is the representation of n in base q, which makes $L_q(E)$ a set of words on the alphabet $\{0, 1, \ldots, q-1\}$. This relation allows to express many questions about arithmetic sequences in the framework of the theory of formal languages, thus establishing a link between number theory, language theory and combinatorics on words.

In virtue of this connection, it is interesting to consider automatic sequences of integers, which are those recognizable by finite automata (an introduction to finite automata can be found in the book of Allouche and Shallit [1]).

For example, the graph associated to the finite 2-automaton that recognizes even numbers represented in base 2 and read from the left to the right can be represented by the diagram:



where s_0 is the initial and unique final state. A more elaborated example if given by the graph associated to the finite q-automaton that recognizes numbers in the sequence $\{q^n, n \in \mathbb{N}\}$, written in base q:



where s_0 is the initial state and s_1 is the unique final state.

A fundamental result that relates Number Theory to Finite Automata was proven if 1980:

Theorem 1 (Christol, Kamae, Mendès France and Rauzy, [2]) *Let $E \subseteq \mathbb{N}$ and $\mathbb{F}_q$ a finite field. The formal power series*

$$\sum_{n \in E} X^{-n} \in \mathbb{F}_q \left[\left[X^{-1}\right]\right]$$

is algebraic over $\mathbb{F}_q(X)$ if and only if E is recognizable by a finite q-automaton.

To understand how simple automatic sequences are, we define, for any sequence $w = \{w_n\}_{n\in\mathbb{N}}$ on a finite alphabet A, the function p_w : $\mathbb{N} \to \mathbb{N}$ by

$$p_w(n) = \#\{(b_1, \dots, b_n), \exists k \text{ s.t. } w_k = b_1, \dots, w_{k+n+1} = b_n\}$$

(i.e. $p_w(n)$ is the number of distinct factors of lenght n in the sequence w).

The connection between automaton and the function p_w was established in 1972:

Theorem 2 (Cobham [3]) *If w is recognizable by a finite q-automaton, then $p_w(n) = O(n)$.*

From a number theoretic point of view, it is a natural question to ask whether is it possible to recognize prime numbers with a finite automaton. The answer to this question is negative, as shown by Minsky and Papert (1966), in a result then generalized by Hartmanis, Shank and Schützenberger (1968), by Mauduit (1992) and by Cassaigne and Le Gonidec (2006).

Other natural questions regard the existence of prime numbers in a given automatic sequence. For examples the sequences $\{2^n + 1\}_{n\in\mathbb{N}}$ and $\{2^n - 1\}_{n\in\mathbb{N}}$ are both recognizable by a finite 2-automaton, and the problems associated correspond respectively to the search of Fermat and Mersenne primes.

In the case when E is a set recognized by a finite automaton whose associated graph is strongly connected, it follows from a remark by Fouvry and Mauduit (1996) that E contains infinitely many almost primes (see [4]). On the other hand it is still an open problem to find an asymptotic estimate for the number of primes less than a certain bound x in the set E.

For general automatic sets the situation is more complicated. One of the first problems to be considered in this direction concerns the search of primes with missing digits. Though some results on integers with missing digits were obtained by Erdős, Mauduit and Sárközy (1998), the problem of finding an asymptotic estimate of the quantity

$$\#\{p \leq x, p \text{ prime}, \mathrm{rep}_q(p) \in D^*\}$$

where D^* is the set of words on any given subset $D \subset \{0, \ldots, q-1\}$, is still open.

Other famous automatic sequences are the Thue-Morse sequences and the Rudin-Shapiro sequences. For these two examples it can be proved the following results appeared respectively in 2010 and 2015 (see also [5]):

Theorem 3 (Mauduit and Rivat [6]) *Let*

$$(t_n)_{n\in\mathbb{N}} = 01101001100101101001011001101001\ldots$$

be the Thue-Morse sequence and let $\mathbb{P}$ *be the set of prime numbers. The frequences of 0 and 1 in the sequence* $(t_p)_{p\in\mathbb{P}}$ *is* $\frac{1}{2}$.

Theorem 4 (Mauduit and Rivat [7]) *Let*

$$(r_n)_{n\in\mathbb{N}} = 0001001000011101000100 10\ldots$$

be the Rudin-Shapiro sequence and let $\mathbb{P}$ *be the set of prime numbers. The frequences of 0 and 1 in the sequence* $(r_p)_{p\in\mathbb{P}}$ *is* $\frac{1}{2}$.

References

[1] J.-P. Allouche and J. Shallit, *Automatic sequences*, Cambridge University Press, Cambridge, 2003. Theory, applications, generalizations.

[2] G. Christol, T. Kamae, M. Mendès France, and G. Rauzy, *Suites algébriques, automates et substitutions*, Bull. Soc. Math. France, **108** n. 4, 401–419 (1980).

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

[4] E. Fouvry and C. Mauduit. *Sommes des chiffres et nombres presque premiers*, Math. Ann. **305** n. 3, 571–599 (1996).

[5] B. Martin, C. Mauduit, and J. Rivat, *Théorème des nombres premiers pour les fonctions digitales*, Acta Arith. **165** n. 1, 11–45 (2014).

[6] C. Mauduit and J. Rivat, *Sur un problème de Gelfond: la somme des chiffres des nombres premiers*, Ann. of Math. (2), **171** n. 3, 1591–1646 (2010).

[7] C. Mauduit and J. Rivat, *Prime numbers along Rudin-Shapiro sequences*, J. Eur. Math. Soc., **17** 2595–2642 (2015).

Valerio Dose
Dipartimento di Matematica
Università di Roma "Tor Vergata"
Via della Ricerca Scientifica 1
00133 Roma, Italy
e-mail address: dose@mat.uniroma2.it