Asymptotic evaluation in a regular language of the number of words of given length with a fixed number of occurrences of a symbol

Alberto Bertoni (*), Christian Choffrut (+),
Massimiliano Goldwurm (*), Violetta Lonati (*)

(*) Università degli Studi di Milano, Dipartimento di Scienze dell'Informazione,
via Comelico 39, 20135 Milano, Italy

(+) Université Paris VII, L.I.A.F.A.,
2 Place Jussieu,75221 Paris, France

Abstract

Given a regular language $L \subseteq \{a,b\}^*$, let $N_L^n(k)$ be the cardinality of the set $\{w \in L \mid |w| = n, |w|_a = k\}$ and $d_L(n) = \max\{N_L^n(k) \mid k = 0, 1, \ldots, n\}$. Our aim is to give an asymptotic evaluation of $N_L^n(k)$ and $d_L(n)$. In this work we solve the problem for a subclass of regular languages which can be defined as follows.

Consider the minimum deterministic automaton recognizing L having say m states and let A and B be the $m \times m$ transition matrices associated with symbols a and b, respectively. For every integer $h > 0$ and every $1 \leq i \leq m$, consider the polynomial in the variable x

$$((Ax+B)^h)_{ii} = \eta_{k_1} x^{k_1} + \cdots + \eta_{k_\ell} x^{k_\ell}$$

where $k_1 < k_2 < \cdots < k_\ell$ and $0 < \eta_{k_j}$ for every $j = 1, 2, \ldots, \ell$. If $\ell > 1$ we define $M_L(h,i) = \mathrm{GCD}\{k_j - k_1 \mid j = 2, \ldots, \ell\}$.

We prove that, if $A + B$ is a primitive matrix and $M_L(h,i) = 1$ for some integers h and i, then there exist positive constants a, b, c, λ such that:

1. $N_L^n(k) \approx \frac{c\lambda^n}{\sqrt{2\pi a n}} \cdot e^{-\frac{(k-bn)^2}{2an}}$ for $n \to +\infty$,
2. $d_L(n) = \Theta\left(\frac{\lambda^n}{\sqrt{n}}\right)$ for $n \to +\infty$.

To obtain this result, we compute the Discrete Fourier Transform of the array $(N_L^n(0), N_L^n(1), \ldots, N_L^n(n))$ giving a suitable approximation based on Perron–Frobenius theory. This method seems to be applicable also to solve the problem for arbitrary regular languages.