Average State Complexity of Operations on Unary Automata

Cyril Nicaud

LIAFA, CNRS and Université Paris VII
2 Place Jussieu 75251 Paris Cedex 05, FRANCE
e-mail: Cyril.Nicaud@liafa.jussieu.fr

Abstract. Define the *complexity* of a regular language as the number of states of its minimal automaton. Let $\mathcal{A}$ (respectively $\mathcal{A}'$) be an n-state (resp. n'-state) deterministic and connected unary automaton. Our main results can be summarized as follows:

1. The probability that $\mathcal{A}$ is minimal tends toward 1/2 when n tends toward infinity,
2. The average complexity of $L(\mathcal{A})$ is equivalent to n,
3. The average complexity of $L(\mathcal{A}) \cap L(\mathcal{A}')$ is equivalent to $\frac{3\zeta(3)}{2\pi^2}nn'$, where ζ is the Riemann "zeta"-function.
4. The average complexity of $L(\mathcal{A})^*$ is bounded by a constant,
5. If $n \leq n' \leq P(n)$, for some polynomial P, the average complexity of $L(\mathcal{A})L(\mathcal{A}')$ is bounded by a constant (depending on P).

Remark that results 3, 4 and 5 differ perceptibly from the corresponding worst case complexities, which are nn' for intersection, $(n-1)^2+1$ for star and nn' for concatenation product.

1 Introduction

This paper addresses a rather natural problem: find the average state complexity of the basic operations on automata. It is certainly an important question for both theorical and pratical reasons. It is a part of the subject founded by Knuth in the sixties [Knu68,Knu69,Knu73], the analysis of algorithms. A general presentation and a complete introduction of this kind to problems can be found in [SF96].

However, surprisingly, almost no result is available in the literature. The worst case complexity of most operations is known [YZS94,Yu97], but the average case analysis seems to be an extremely difficult problem. The main reason is that the number of non-isomorphic deterministic and connected automata with n states and say, two letters, is not even known!

This is why we restrict ourselves to the case of one-letter automata. But, even in this case, non-trivial arguments of number theory are required to analyze elementary looking operations.

Define the *complexity* of a regular language as the number of states of its minimal automaton. Let $\mathcal{A}$ (respectively $\mathcal{A}'$) be an n-state (resp. n'-state) deterministic and connected unary automaton. Our main results can be summarized as follows:

M. Kutyłowski, L. Pacholski, T. Wierzbicki (Eds.): MFCS'99, LNCS 1672, pp. 231–240, 1999.

1. The probability that $\mathcal{A}$ is minimal tends toward $1/2$ when n tends toward infinity,
2. The average complexity of $L(\mathcal{A})$ is equivalent to n,
3. The average complexity of $L(\mathcal{A}) \cap L(\mathcal{A}')$ is equivalent to $\frac{3\zeta(3)}{2\pi^2}nn'$, where ζ is the Riemann "zeta"-function.
4. The average complexity of $L(\mathcal{A})^*$ is bounded by a constant,
5. If $n \leq n' \leq P(n)$, for some polynomial P, the average complexity of $L(\mathcal{A})L(\mathcal{A}')$ is bounded by a constant (depending on P).

Remark that results 3, 4 and 5 differ perceptibly from the corresponding worst case complexities, which are nn' for intersection, $(n-1)^2+1$ for star and nn' for concatenation product.

The proofs are too long to be described in this paper. However, in Section 4, we present a sketch of one proof to illustrate the kind of technics used here.

2 Notations

If f, g are two functions from $\mathbb{N} \times \mathbb{N}$ into $\mathbb{R}$, we say that f is *equivalent* to g (denoted by $f \asymp g$) if there exists a function ε from $\mathbb{N} \times \mathbb{N}$ into $\mathbb{R}$ such that the two following statements hold:

- for all n, n' in $\mathbb{N}^2$, $f(n,n') = \big(1+\varepsilon(n,n')\big)g(n,n')$
- $\varepsilon(n,n') \to 0$ when $\min\{n,n'\} \to \infty$

If f is a function from $\mathbb{N} \times \mathbb{N}$ into $\mathbb{R}^+$, we say that f is *polynomially bounded* by a non negative real constant C (denoted by $f \preccurlyeq_P C$) if, for every polynomial $P \in \mathbb{N}[X]$, there exists an integer $N_P \in \mathbb{N}$ such that, for each n, $n' \in \mathbb{N}$ with $n \geq N_P$ and $n \leq n' \leq P(n)$, $f(n,n') \leq C$. Of course C depends on the choice of P.

For each n, $n' \in \mathbb{N}$, we denote respectively by $n \vee n'$ and $n \wedge n'$ the lcm and the gcd of n and n'. We denote by $d|n$ the fact that the integer d divides the integer n.

Given a deterministic automaton $\mathcal{A}$, $|\mathcal{A}|$ denotes the number of its states and $||\mathcal{A}||$ the number of states of its minimal automaton. By extension, if L is a regular language, we denote by $||L||$ the number of states of its minimal automaton, that is, its complexity. Note that if $\mathcal{A}(L)$ is any automaton recognizing L, then $||\mathcal{A}(L)|| = ||L||$.

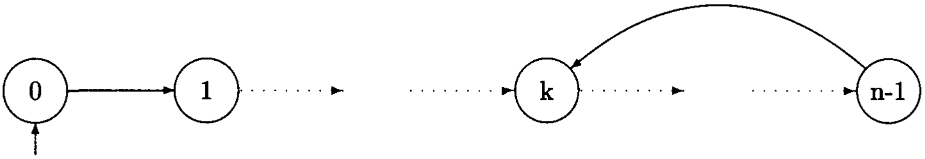
Let S be a finite subset of a set T. If f is a function from T into $\mathbb{R}$, we denote by $\langle f, S\rangle$ the sum $\sum_{s\in S} f(s)$.

3 The number of minimal automata

In this section we enumerate the minimal unary automata (see [Eil74] [HU79]) with n states. For this purpose, we establish and use the *characterization lemma* which is very useful for a combinatorial analysis of minimal unary automata.

To avoid any isomorphism problems, we fix a rule for the labels of the states. For every deterministic and connected automaton with n states, with initial state q_0, we label each state q with the smallest integer i such that $q_0.a^i = q$. This condition prevents two distinct automata from being isomorphic.

A deterministic complete and connected unary automaton is always of the following form, for some $k \in \{0, \dots, n-1\}$ (the final states are omitted):



Therefore, such an automaton is totally determined by the integer k and its set of final states. More precisely, it is equal to one of the automata $\mathbf{A}(n,k,F)$ defined as follows : given two integers k and n such that $0 \leq k \leq n-1$ and a subset F of $\{0, \dots, n-1\}$, $\mathbf{A}(n,k,F)$ is the unary automaton whose set of states is $Q = \{0, \dots, n-1\}$ and transition function is given by $q.a = q+1$ for $0 \leq q \leq n-2$ and $(n-1).a = k$. The initial state of this automaton is 0 and its set of final states is F.

The *loop* of $\mathcal{A} = \mathbf{A}(n,k,F)$, denoted by *loop*$(\mathcal{A})$, is the automaton $\mathbf{A}(n-k, 0, F')$ where $F' = \{i \in [\![0, n-k-1]\!] \mid i+k \in F\}$. The automaton $\mathcal{A}$ is simply called a *loop* if it is equal to its loop, that is, if and only if $k = 0$.

Loops play an important role in the next lemma, which characterizes minimal unary automata. Two states of an automaton are said to have *the same finality* if they are either both final or both non-final.

Lemma 1. (Characterization Lemma) *An automaton $A(n,k,F)$ is minimal if and only if the two following conditions hold:*

1. *its loop is minimal*
2. *if $k \neq 0$, the states $k-1$ and $n-1$ do not have the same finality.*

We are now ready to evaluate the average number of minimal automata. First, denoting by $\mathcal{U}_n$ the set of complete, deterministic and connected unary automata with n states (with the proper labels on their states), it is easy to see that $|\mathcal{U}_n| = n2^n$.

Next we enumerate the minimal n-loops (loops with n states). Fix an integer n. For every n-loop $\mathcal{L} = \mathbf{A}(n, 0, F)$ define

$$k_{min}(\mathcal{A}) = \min\{k \in [\![1, n]\!] \mid F.a^k = F\}$$

Note that $k_{min}(\mathcal{L})$ exists since $F.a^n = F$. A n-loop $\mathcal{L}$ is said to be *primitive* if $k_{min}(\mathcal{L}) = n$.

We can characterize minimal loops:

Lemma 2. *For each n-loop $\mathcal{L}$, the minimal automaton of $\mathcal{L}$ has $k_{min}(\mathcal{L})$ states and $k_{min}(\mathcal{L})$ divides n. In particular a loop $\mathcal{L}$ is primitive if and only if it is minimal.*

Denoting by μ the Möbius function, we have the following result:

Theorem 1. *There are exactly $\sum_{d|n} \mu(n/d)2^d$ minimal n-loops. This number is equivalent to 2^n. Furthermore, there are no more than $2^{(n/2)+1}$ non-minimal loops with n states.*

The proof of this theorem is very classical. Using Lemma 2, we can reduce the problem to the well-known problem of counting the number of primitive circular words on a two-letter alphabet, which justifies the definition of a primitive loop. This number is also n times the number of irreductible polynomials of degree n over $\mathbb{F}_2$, the field with two elements, and a natural bijection has recently been found, using Galois theory arguments [Del99]. For a survey of contexts where the same kind of numbers appear, see [All99].

This result is very important as it says that very few loops are not minimal. Thus, as a first approximation, we can consider that each loop is minimal. Indeed, for all the average analysis of this paper, unary automata behave as if their loops were minimal. Using the characterization lemma, we can give an equivalent to the number of minimal unary automata.

Theorem 2. *The number of minimal automata with n states is equivalent to $n2^{n-1}$.*

We define the average number of states of the minimal automaton of an n-state automaton as $\frac{1}{|\mathcal{U}_n|}\sum_{\mathcal{A}\in\mathcal{U}_n} \|\mathcal{A}\|$. The following theorem shows that the number of states of the minimal automaton of a deterministic connected automaton is very close to the number of states of this automaton:

Theorem 3. *The average number of states of the minimal automata of an n-state automaton is equivalent to n.*

The proof is not difficult, and the result claims that it is not often useful to spend time minimizing a unary automaton.

4 Intersection

In this section, we give the average and worst case complexity of the intersection on unary automata. Remark that the union has exactly the same behavior as the intersection since the minimal automaton of a regular language L has the same number of states as the minimal automaton of its complement.

Fix two integers n and n' greater than 2. For every $(\mathcal{A}, \mathcal{A}') \in \mathcal{U}_n \times \mathcal{U}_{n'}$, of respective initial states q_0 and q_0', define the product automaton $\mathcal{A} \times \mathcal{A}'$, as the automaton whose initial state is (q_0, q_0'), whose set of states is the set of reachable pairs from the initial states $(q_0.a^i, q_0'.a^i)$, $i \in \mathbb{N}$. The transition function of this automaton is defined by $(q, q').a = (q.a, q'.a)$ and a reachable pair (q, q') is final if and only if q (respectively q') is a final state of $\mathcal{A}$ (resp. $\mathcal{A}'$).

It is well-known that this automaton recognizes the intersection of $L(\mathcal{A})$ and $L(\mathcal{A}')$.

Our first result concerns the worst case complexity. It slighlty improves a result of [YZS94] (they only consider the case when n and n' are prime together), and use the fact that for n large enough, there always exist a prime number between $n - n^{\alpha}$ and n, for some real number $\alpha \in]0,1[$ [BH96,Dav74,Hux72]):

Proposition 1. *In the worst case, the complexity of the intersection is equivalent to nn'.*

Denote by $\mathcal{U}$ the set of all complete, deterministic and connected unary automata. The average complexity of the intersection is exactly

$$\frac{1}{|\mathcal{U}_n|\,|\mathcal{U}_{n'}|}\langle \| \times \|, \mathcal{U}_n \times \mathcal{U}_{n'}\rangle$$

where $\| \times \|$ is the function from $\mathcal{U} \times \mathcal{U}$ into $\mathbb{N}$ which maps $(\mathcal{A}, \mathcal{A}')$ onto $\|\mathcal{A} \times \mathcal{A}'\|$.

Our main result is a precise evaluation of the average complexity of the intersection:

Theorem 4. *The average complexity of the intersection of a n-state automaton and a n'-state automaton is equivalent to $\frac{3\zeta(3)}{2\pi^2} nn'$*

The proof requires a result from analytic number theory established by G. Tenenbaum [Ten97] along classical techniques (see, e.g., [Ten96]). The result is interesting on its own account and we now state it formally.

Theorem 5. *[Tenenbaum] The following result holds:*

$$\sum_{\substack{1 \leq i \leq n \\ 1 \leq i' \leq n'}} i \vee i' = \frac{3\zeta(3)}{2\pi^2}(nn')^2\Big(1 + O\big(\frac{\log z}{z}\big)\Big)$$

with $z = \min\{n, n'\}$. Thus

$$\sum_{\substack{1 \leq i \leq n \\ 1 \leq i' \leq n'}} i \vee i' \asymp \frac{3\zeta(3)}{2\pi^2}(nn')^2$$

Sketch of the proof of Theorem 4: We exhibit an upper and a lower bound to the average of the intersection, both equivalent to $\frac{3\zeta(3)}{2\pi^2} nn'$. For the upper bound, we use the fact that if $\mathcal{A}$ is a n-state automaton and $\mathcal{A}'$ a n'-state automaton, $\|L(\mathcal{A}) \cap L(\mathcal{A}')\| \leq |\mathcal{A} \times \mathcal{A}'|$. We can compute exactly the number of states of $\mathcal{A} \times \mathcal{A}'$. Moreover the loop of $\mathcal{A} \times \mathcal{A}'$ contains $|loop(\mathcal{A})| \vee |loop(\mathcal{A}')|$ states and thus we can prove that

$$\sum_{\mathcal{A} \in \mathcal{U}_n} \sum_{\mathcal{A} \in \mathcal{U}_{n'}} |\mathcal{A} \times \mathcal{A}'| \asymp \sum_{\mathcal{A} \in \mathcal{U}_n} \sum_{\mathcal{A} \in \mathcal{U}_{n'}} |loop(\mathcal{A})| \vee |loop(\mathcal{A}')|$$

After some calculii, we conclude by Theorem 5 that the average of the intersection is bounded by a function equivalent to $\frac{3\zeta(3)}{2\pi^2} nn'$.

For the lower bound, we construct a set $G(l,l')$ of pairs $(\mathcal{L},\mathcal{L}')$ where $\mathcal{L}$ is an l-loop and $\mathcal{L}'$ is an l'-loop. This set is such that for every $(\mathcal{L},\mathcal{L}') \in G(l,l')$, $\mathcal{L} \times \mathcal{L}'$ is minimal. Hence as

$$\langle \| \times \|, \mathcal{U}_n \times \mathcal{U}_{n'} \rangle \geq \sum_{l=1}^{n} \sum_{l'=1}^{n'} \sum_{(\mathcal{L},\mathcal{L}') \in G(l,l')} \sum_{\substack{\mathcal{A} \in \mathcal{U}_n \\ loop(\mathcal{A})=\mathcal{L}}} \sum_{\substack{\mathcal{A}' \in \mathcal{U}_n \\ loop(\mathcal{A}')=\mathcal{L}'}} l \vee l'$$

and since for every l-loop $\mathcal{L}$ with $1 \leq l \leq n$, there are exactly 2^{n-l} n-state automata whose loop is $\mathcal{L}$,

$$\langle \| \times \|, \mathcal{U}_n \times \mathcal{U}_{n'} \rangle \geq 2^{n+n'} \sum_{l=1}^{n} \sum_{l'=1}^{n'} |G(l,l')| \, 2^{-l} 2^{-l'} (l \vee l')$$

Hence, to prove the theorem we have to construct a large enough set $G(n,n')$ so that

$$\sum_{l=1}^{n} \sum_{l'=1}^{n'} |G(l,l')| \, 2^{-l} 2^{-l'} (l \vee l') \asymp \sum_{l=1}^{n} \sum_{l'=1}^{n'} l \vee l'$$

and we conclude using Theorem 5.

To construct $G(l,l')$, we remove some subsets from $B(l,l')$, the set of all pairs of loops $(\mathcal{L},\mathcal{L}')$ such that $\mathcal{L}$ is a l-loop and $\mathcal{L}'$ is a l'-loop. We first remove all the pairs of loops $(\mathcal{L},\mathcal{L}')$ such that $L(\mathcal{L})$ or $L(\mathcal{L}')$ is either finite or cofinite. It is not difficult to see that there are no more than $2^l + 2^{l'}$ such pairs in $B(l,l')$. Define $H(l,l')$ the subset of $B(l,l')$ obtained after removing such pairs of loops.

Define the property $\mathcal{P}(l,l')$ that is true if and only if $l \wedge l' > \frac{\min\{l,l'\}}{5}$. For technical reasons we want that $G(l,l') = \emptyset$ if $\mathcal{P}(l,l')$ is satisfied. This is not restrictive since they are not a lot of (l,l') that satisfy $\mathcal{P}$.

We first work in the case when $\mathcal{P}$ is not satisfied by l and l': we want to remove from $H(l,l')$ the pairs $(\mathcal{L},\mathcal{L}')$ such that $\mathcal{L} \times \mathcal{L}'$ is not minimal. Define $\mathcal{B} = \mathcal{L} \times \mathcal{L}'$. We distinguish two kinds of pairs, according to whether $l \wedge l'$ divides $\|\mathcal{B}\|$ or not.

- If $l \wedge l'$ divides $\|\mathcal{B}\|$: we exhibit a condition sufficient to ensure that a pair of loops is such that its product is not minimal. The following lemma characterizes non-minimal loops in the particular case $l \wedge l' = 1$:

  Lemma 3. *Let $l, l' \geq 1$ be two integers such that $l \wedge l' = 1$. If $\mathcal{L}$ is a l-loop and $\mathcal{L}'$ a l'-loop then $\mathcal{L} \times \mathcal{L}'$ is minimal if and only if both $\mathcal{L}$ and $\mathcal{L}'$ are minimal.*

  We want to use this lemma even if $l \wedge l' \neq 1$. We have to introduce some new notations. For every $i \in \{0, \cdots, d-1\}$, define the loop $\mathcal{L}_d^{(i)} = \mathbf{A}(l/d, 0, F^{(i)})$ where

  $$F^{(i)} = \{j \in \{0, \cdots, (l/d) - 1\} \mid dj + i \text{ is a final state of } \mathcal{L}\}$$

The construction of $\mathcal{L}_d^{(i)}$ is motivated by the following property, which holds for every d dividing l and l' and every $i \in \{0, \cdots, d-1\}$:

$$\mathcal{L}_d^{(i)} \times \mathcal{L}'^{(i)}_d = (\mathcal{L} \times \mathcal{L}')_d^{(i)}$$

Fix $d = l \wedge l'$. The integer d divides $||\mathcal{B}||$ by hypothesis. If $\mathcal{B}$ is not minimal then, by Lemma 2, $||\mathcal{B}||$ strictly divides $|\mathcal{B}|$. Hence for every $i \in \{0, \cdots, d-1\}$, $(\mathcal{L} \times \mathcal{L}')_d^{(i)}$ is not minimal. Thus $\mathcal{L}_d^{(i)} \times \mathcal{L}'^{(i)}_d$ is not minimal and as $l/d \wedge l'/d = 1$, $\mathcal{L}_d^{(i)}$ or $\mathcal{L}'^{(i)}_d$ is not minimal, by Lemma 3.
Therefore, if there exists $i \in \{0, \cdots, d-1\}$ such that both $\mathcal{L}_d^{(i)}$ and $\mathcal{L}'^{(i)}_d$ are minimal then $\mathcal{L} \times \mathcal{L}'$ is minimal. Using the fact that we are working on l, l' which does not satisfy $\mathcal{P}$, we can bound the number of pairs of loops such that $\mathcal{B}$ is not minimal and $l \wedge l'$ divides $||\mathcal{B}||$ by $2^{9l/10}2^{l'}$, for $l \leq l'$.
- If $d = l \wedge l'$ does not divides $||\mathcal{B}||$: the characterization of minimal products of loops is completely different in this case. We introduce an equivalence relation $\equiv$ on $\{0, \cdots, n-1\}$ defined by

$$i \equiv j \Leftrightarrow \text{ there exists } k,\ (i = j + k||\mathcal{B}||) \mod d$$

We first prove that every equivalence class contains the same number m of elements and that each class contains at least two elements. Moreover if $i \equiv j$ then $(\mathcal{L} \times \mathcal{L}')_d^{(i)}$ and $(\mathcal{L} \times \mathcal{L}')_d^{(j)}$ recognize the same language. Hence, since they have the same number of states, $(\mathcal{L} \times \mathcal{L}')_d^{(i)} = (\mathcal{L} \times \mathcal{L}')_d^{(j)}$. With this considerations we can prove that there are at most $2^{l/m}2^{l'/m} \leq 2^{l/2}2^{l'/2}$ pairs of loops such that their product is not minimal and such that $d = l \wedge l'$ does not divides $||\mathcal{A} \times \mathcal{A}'||$.

Hence we construct $G(l, l')$ by removing these pairs of loops. Putting all results together we can prove that if l and l' do not satisfy $\mathcal{P}$ then, for $l \leq l'$, $|G(l, l')| \geq 2^{l+l'} - 2^{\alpha l}2^{l'}$ for some real $\alpha \in]0, 1[$. But

$$\sum_{l=1}^{n} \sum_{l'=1}^{n'} 2^{\alpha l} 2^{l'} 2^{-l} 2^{-l'} (l \vee l') \leq C n'^2$$

for some constant C. Hence by bounding the number of l, l' satisfying $\mathcal{P}$ we can prove Theorem 4.

Using the same kind of methods we can also prove that the result of Theorem 4 still holds if we consider the average on minimal automata only:

Theorem 6. *The following result holds:*

$$\frac{1}{|\mathcal{MU}_n|\,|\mathcal{MU}_{n'}|} \langle || \times ||, \mathcal{MU}_n \times \mathcal{MU}_{n'} \rangle \asymp \frac{3\zeta(3)}{2\pi^2} n n'$$

Thus for the intersection the average and worst cases only differ by a multiplicative constant. The theorem also shows that the naive algorithm which constructs the product automaton cannot be substantially improved.

5 Star operation

The purpose of this section is to prove that the average state complexity of the star operation is bounded. S. Yu, Q. Zhuang and K. Salomaa have already proved that in the worst case it is quadratic in the number of states [YZS94]:

Theorem 7. *For every regular language L of complexity n, the complexity of L^* is bounded by $(n-1)^2+1$. Furthermore, this upper bound is reached for every $n \geq 1$.*

We establish the following result, where $\| \, ^*\|$ is the function from $\mathcal{U}_n$ into $\mathbb{N}$ such that the image of an n-state automaton $\mathcal{A}$ is $\|L(\mathcal{A})^*\|$.

Theorem 8. *There exists a constant $C_* \in \mathbb{R}^+$ such that for every $n \geq 2$,*

$$\frac{1}{|\mathcal{U}_n|}\langle\| \, ^*\|, \mathcal{U}_n\rangle \leq C_*$$

In the proof of the theorem we encode automata by words on the alphabet $\{0,1\}$. Removing a negligible subset of $\mathcal{U}_n$ containg all the automata such that not to consecutive states are both final, we reduce the problem to a problem of combinatorics on words, which is sufficient to prove the theorem. Remark that the bound found in the proof is approximatively 50, whereas an experimental computation gives a bound lower than 6.

This result shows that the average behavior of the star operation is very different from its worst case behavior, since the first one is bounded whereas the second one has a quadratic growth. Moreover we can use this result to obtain an algorithm that constructs the minimal automaton of the star of a given regular language that has an average complexity in $O(1)$ whereas the classical construction is in $O(n^2)$.

6 Concatenation product

The purpose of this section is to prove that the concatenation product is polynomialy bounded.

S. Yu, Q. Zhuang and K. Salomaa gave the following result:

Theorem 9. *[YZS94] For every regular languages L and L' such that $\|L\| = n$ and $\|L'\| = n'$, $\|LL'\| \leq nn'$.*

They also proved that the bound is reached if $n \wedge n' = 1$.

With this result we can obtain a equivalent to the worst case complexity of the product of two languages (once more we use number theory results [BH96,Dav74,Hux72]):

Proposition 2. *The complexity in the worst case of the concatenation product of two unary automata with respectively n and n' states is equivalent to nn'.*

Now we are now going to show that in the average case, the asymptotic complexity of the concatenation product is bounded, provided the growth of n' is bounded by a polynomial in n. The image of $(\mathcal{A}, \mathcal{A}') \in \mathcal{U}_n \times \mathcal{U}_{n'}$ by $\|\,.\,\|$ is the integer $\|L(\mathcal{A})L(\mathcal{A}')\|$.

Theorem 10. *There exist $C \in \mathbb{R}^+$ such that*

$$\frac{1}{|\mathcal{U}_n \times \mathcal{U}_{n'}|}\langle \|\,.\,\|, \mathcal{U}_n \times \mathcal{U}_{n'}\rangle \underset{P}{\preccurlyeq} C$$

Once again we establish the proof by removing negligible subsets of $\mathcal{U}_n \times \mathcal{U}_{n'}$. We also encode automata with words and use combinatorics on words. We procede in three steps, for $n \leq n' \leq P(n)$ for some polynomial P:

- We first remark that almost all pairs of automata $(\mathcal{A}, \mathcal{A}')$ are such that $L(\mathcal{A})L(\mathcal{A}')$ recognizes every word of length between $[n/2]$ and $[3n/2]$. This step is quite technical and uses basic combinatorial tools.
- Almost all pairs satisfying the first condition are such that $L(\mathcal{A})L(\mathcal{A}')$ is cofinite. To prove this we consider two cases; namely the loop of $\mathcal{A}$ contains more than $\sqrt{n}$ states or not.
- Finaly we precisely compute the size of the minimal automaton of $L(\mathcal{A})L(\mathcal{A}')$ for pairs of automata satisfying the two previous conditions

Remark that the condition $n \leq n' \leq P(n)$ is certainly necessary to obtain a bounded average complexity, but is not very restrictive in practice.

7 Conclusion

Putting all things together we can summarize our results as follows:

Operation	Worst case	Average case
Minimization	anything in $\{1, \cdots, n\}$	$\sim n$
Star operation	$(n-1)^2+1$	$\leq C_*$
Concatenation product	$\asymp nn'$	$\preccurlyeq_P C$
Intersection	$\asymp nn'$	$\asymp \frac{3\zeta(3)}{2\pi^2}nn'$

8 Acknowledgements

I would like to thank Jean-Eric Pin for reading and commenting numerous versions of this paper, Gerald Tenenbaum for his help on analytic number theory problems, and Arnaud Delobelle for his assistance.

References

[All99] J.P. Allouche. Transcendence of formal power series with rational coefficients. *Th. Comp. Sc.*, 1999. to appear.

[BH96] R.C. Baker and G. Harman. The difference between consecutive primes. *Proc. Lond. Math. Soc.*, III ser. 72:261–280, 1996.

[Dav74] H. Davenport. *Multiplicative number theory. Second edition.* Graduate Texts in Mathematics. Springer Verlag, 1974.

[Del99] A. Delobelle. Primitive circular words and irreducible polynomials. 1999. to appear.

[Eil74] S. Eilenberg. *Automata, Languages and Machines*, volume A. Academic Press, 1974.

[HU79] J. Hopcroft and J.D. Ullman. *Introduction to Automata Theory, Language and Computation.* Addison Wesley, 1979.

[Hux72] M.N. Huxley. *The distribution of prime numbers, large sieves and zero-density theorems.* Clarendon press, 1972.

[Knu68] D. E. Knuth. *The Art of Computer Programming*, volume 1: fundamental algorithms. Addison-Wesley, Reading, MA, 1968.

[Knu69] D. E. Knuth. *The Art of Computer Programming*, volume 2: seminumerical algorithms. Addison-Wesley, Reading, MA, 1969.

[Knu73] D. E. Knuth. *The Art of Computer Programming*, volume 3: sorting and searching. Addison-Wesley, Reading, MA, 1973.

[SF96] R. Sedgewick and P. Flajolet. *An Introduction to the Analysis of Algorithms.* Addison-Wesley Publishing Company, 1996.

[Ten96] G. Tenenbaum. *Introduction à la théorie analytique et probabiliste des nombres, Cours spécialisés 1.* Société Mathématique de France, 1996.

[Ten97] G. Tenenbaum. Private communication, may 1997.

[Yu97] S. Yu. Regular languages. In G. Rozenberg and A. Salomaa, editors, *Handbook of language theory*, volume 1, chapter 2, pages 96–105. Springer Verlag, 1997.

[YZS94] S. Yu, Q. Zhuang, and K. Salomaa. The state complexities of some basic operations on regular languages. *Th. Comp. Sc.*, 125:315–328, 1994.