

# Random Generation Models for NFAs

J.-M. Champarnaud, G. Hansel, T. Paranthoën, and D. Ziadi  
LIFAR, Université de Rouen, 76821 Mont-Saint-Aignan, France

## Abstract

The aim of this study is the random generation of non-deterministic automata. We focus our attention on the random generation processed with bitstreams for which we present a probabilistic analysis. Let  $m$  be the size of the alphabet. We show that the DFAs obtained by subset construction from  $n$ -state NFAs based on equiprobable bitstreams have a probability of being of size  $m + 2$  that tends to 1 when  $n$  tends to infinity. This property gives an asymptotical explanation to van Zijl's experimental results concerning the succinctness of NFAs. We also determine the probability that a state is reachable from an equiprobably chosen DFA state. We show that the distribution of the subsets that appear during the subset construction is an equiprobable one in the case of bitstreams generated with the probability  $2 - 2^{\frac{n-1}{n}}$ . This result is related to the conjecture of Leslie, Raymond and Wood, which says that the number of states of the DFA is maximum when the density of the NFA is approximately equal to  $2/n$ . Finally we extend this probabilistic study to the case of  $\star$ -NFAs defined by van Zijl. **Keywords:** Random generation, non deterministic finite automaton,  $\star$ -NFA, subset construction.

## 1 Introduction

Random generation of structures allows us to test algorithms' performance and to illustrate theoretical results. A good knowledge of the structures' space that we want to generate is indispensable to design a generation algorithm. Nicaud has studied the equiprobable random generation of deterministic  $n$ -state automata [7].

Our aim is to make a similar study concerning the random generation of NFAs. We consider in this paper the random NFA generation

method based on random bitstreams. Van Zijl [8] has used this method with equiprobable bitstreams in order to compare the succinctness of various representations of regular languages. We develop a probabilistic analysis of nondeterministic transition tables produced by this method, which highlights some properties of the associated NFAs. We especially focus on the number of states of the DFAs obtained by subset construction.

In the first part, we follow the subset construction based on reachability and we consider the case of equiprobable bitstreams. We determine the probability that the image of a given subset of size  $i$  by a given symbol is of size  $k$ . We deduce the probability that the image of the initial state set by a word of length  $t$  is of size  $k$ . We show that, for equiprobable bitstreams, the DFAs generated by the subset construction are asymptotically of size  $m + 2$ , where  $m$  is the alphabet size. Moreover, in this case, the random NFAs are asymptotically accessible. This study provides an asymptotical justification of the experimental results obtained by van Zijl [8].

In the second part, we follow the brute force subset construction. We determine the probability that a state is reachable from an equiprobably chosen DFA state. Consequently, the distribution of the subsets occurring during the subset construction is an equiprobable one in the case of bitstreams generated with the probability  $2 - 2^{\frac{n-1}{n}}$ . We explain how this result is connected to the conjecture of Leslie, Raymond and Wood [4], which says that the number of states of the DFA is maximum when the density of the NFA is approximately equal to  $2/n$ .

We also extend this probabilistic study to the case of  $\star$ -NFAs defined by van Zijl. The experimental results described in [8] are based on an equiprobable bitstream generation; they show that  $\cup$ -NFAs and  $\cap$ -NFAs are very rarely succinct whereas  $\oplus$ -NFAs are very often succinct. We explain how our analysis is related to these results.

The following section gathers some definitions and conventions. Section 3 describes the bitstream generation method and presents the studies of van Zijl and of Leslie *et al.* Sections 4 and 5 develop the probabilistic analysis of this method. Section 4 is devoted to equiprobable bitstreams whereas Section 5 addresses non-equiprobable bitstreams.

## 2 Definitions and notation

An automaton is a 5-tuple  $\mathcal{A} = \langle Q, \Sigma, \delta, I, F \rangle$  where  $Q$  is a finite set of *states*,  $\Sigma$  is the *alphabet* on which the automaton is defined,  $\delta$  is the *transition function* ( $\delta : Q \times \Sigma \rightarrow 2^Q$ ) that associates a subset of  $Q$  to each element of  $Q \times \Sigma$ ,  $I$  is a non-empty subset of  $Q$  whose elements are the *initial states*, and  $F$  is a subset of  $Q$  whose elements are the *final states*.

An automaton  $\mathcal{A}$  is *unary* if its alphabet is restricted to one symbol.

An automaton is *accessible* if and only if for any state  $q \in Q$  there is a path from one of the initial states to this state. An automaton is *co-accessible* if and only if for any state  $q \in Q$  there is a path from this state and one of the final states. An automaton that is both accessible and co-accessible is *trim*.

An automaton  $\mathcal{A}$  is a *deterministic automaton* if it has a unique initial state and if for all states and for all symbols there is at most one transition outgoing from this state. A deterministic (resp. non-deterministic) automaton is called a *DFA* (resp. an *NFA*). For each NFA  $\mathcal{A}$  that recognizes a language  $L$ , we know how to build an equivalent DFA (i.e. a deterministic automaton that recognizes the same language). The algorithm used to perform this conversion is the *subset construction* [1, 11]. It produces the so-called *subset automaton*. Since the subset automaton of an  $n$ -state NFA can have  $2^n$  states [5], we generally think that the NFAs are (much) smaller than the DFAs.

Following van Zijl [9], we call  $\star$ -NFA an automaton for which the classical union operation is replaced by any associative and commutative binary operation on sets. In the case of the *intersection*, we obtain an  $\cap$ -NFA. The transition function has the form:  $\delta(P, a) = \delta(p_1, a) \cap \delta(p_2, a) \cap \dots \cap \delta(p_k, a)$ , where  $P = \{p_1, p_2, \dots, p_k\}$  is a subset of  $Q$ . A word  $u$  is recognized by an  $\cap$ -NFA if and only if  $F \subseteq \delta(I, u)$ . In the case of the *symmetric difference*, we obtain a  $\oplus$ -NFA. The transition function has the form:  $\delta(P, a) = \delta(p_1, a) \oplus \delta(p_2, a) \oplus \dots \oplus \delta(p_k, a)$ . A word  $u$  is recognized by a  $\oplus$ -NFA if and only if  $|\delta(I, u) \cap F|$  is odd.

There is no agreement about the size of an NFA. The choice of the number of transitions [2] is justified by the memory space required to implement an automaton. Specific complexity properties depend on the sum of the number of states and of the number of transitions [3]. However, most of the studies use the number of states to measure the size of an NFA, and we will adopt this convention.

### 3 NFA generation by bitstreams

The method used by van Zijl [10] to randomly build a nondeterministic automaton is the following:

- the alphabet is  $\Sigma = \{1, \dots, m\}$  and the set of states is  $Q = \{1, \dots, n\}$ ,
- an equiprobable bitstream of size  $mn^2$  is generated; it describes the transition function  $\delta$ ; the occurrence of a non-zero bit at position  $(l-1)n^2 + (i-1)n + j$  denotes the existence of a transition from state  $i$  to state  $j$  and labeled by  $l$ ,
- there is a unique initial state (state 1),
- the set of final states is randomly chosen, each state having an equal chance of being final or not.

This method provides an NFA  $\langle Q, \Sigma, \delta, I, F \rangle$  that is not necessarily accessible nor co-accessible. It is used by van Zijl to measure the succinctness of  $n$ -state NFAs. An NFA sequence is built and for each NFA the following operations are performed:

- check whether it is trim or not (non-trim automata are rejected),
- compute the equivalent deterministic minimal automaton.

Thanks to these operations the size distribution of minimal automata equivalent to random NFAs of a given size has been studied [8], leading to results similar to those of Figure 1.

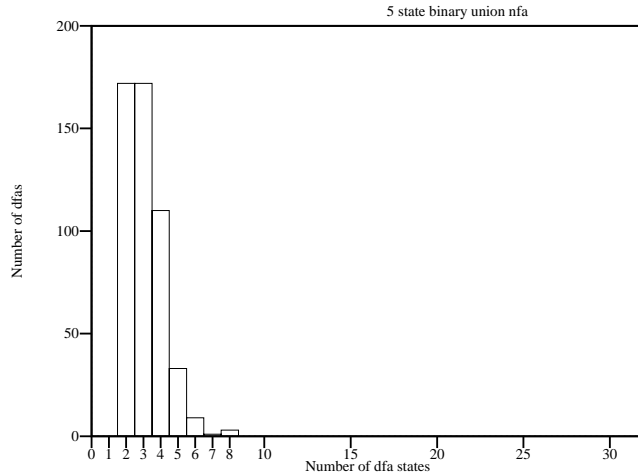


Figure 1: The number of minimal  $n$ -state DFAs for 5-state binary NFAs according to [8].

The work of Leslie *et al.* suggests another approach: their aim is to study the size of the DFAs obtained by determinization. The NFAs are built in the following way:

- a connected structure is randomly generated in order to generate an accessible automaton,
- a unique initial state is randomly chosen,
- the transitions are randomly chosen: if a transition appears twice, it is rejected and another one is chosen.

The *density* of an NFA with  $e$  transitions is  $da = \frac{e}{mn^2}$  and its *deterministic density* is  $dd = \frac{e}{mn}$ . Leslie *et al.* carry out an experimental study on the size of the DFAs w.r.t. the density of the random NFAs. Their results point out an optimal deterministic density equal to 2. Therefore, they suggest the following conjecture:

**Conjecture 1** [4] For a given NFA, we can compute the expected number of states and transitions in the corresponding DFA, produced by subset construction, from the deterministic density of the NFA. In addition, this functional relationship gives rise to a Poisson-like curve with its peak approximately at a deterministic density of 2.

## 4 Probabilistic analysis: equiprobable case

We consider here an NFA  $\mathcal{A}$  of size  $n$  built on an alphabet  $\Sigma$  of size  $m$ , associated with an equiprobable bitstream of  $mn^2$  bits. We suppose that the initial state is unique.

Let  $\mathcal{D}$  be the deterministic automaton equivalent to  $\mathcal{A}$  generated by the subset construction. We point out that the probability that the size of  $\mathcal{D}$  is equal to  $m + 2$  tends to 1 when  $n$  tends to infinity. We take our inspiration from reachability subset construction. The idea is that the average size of the image of the initial state by a letter is equal to  $n/2$  and that the size of the image by a word of length  $t$  grows very quickly to  $n$  [6]. For a given subset  $X$  of size  $i$  and a given symbol  $a$  we compute the probability that the image  $\delta(X, a)$  is of size  $k$ . From this we deduce the probability that the size of the image of the set of initial states by a given word of length  $t$  is equal to  $k$ .

### 4.1 Study of the classical NFAs

We consider here that  $\mathcal{A}$  is a  $\cup$ -NFA.

**Proposition 4.1** Let  $\mathcal{A} = \langle Q, \Sigma, \delta, I, F \rangle$  be an  $n$ -state NFA associated with an equiprobable bitstream. Let  $X$  be a subset of  $Q$  of size  $i$  and  $a$  be a symbol of  $\Sigma$ . The probability  $P(i \rightarrow k)$  that the image  $\delta(X, a)$  of  $X$  by  $a$  has a size equal to  $k$  is given by:

$$P(0 \rightarrow 0) = 1 \text{ and } P(0 \rightarrow k) = 0, \forall k \neq 0$$

$$P(i \rightarrow k) = \frac{\binom{n}{k} (2^i - 1)^k}{2^{in}}, \forall i \neq 0$$

#### Proof

Let  $Z$  be the random variable in  $\{0, \dots, n\}$  that is equal to the size of  $\delta(X, a)$ . Let  $Z_j$ ,  $1 \leq j \leq n$ , be the random variable in  $\{0, 1\}$  that is equal to 1 if the state  $j$  belongs to  $\delta(X, a)$  and 0 otherwise. As the size of  $X$  is equal to  $i$ , we have:  $P(Z_j = 0) = 1/2^i$  and  $P(Z_j = 1) = 1 - 1/2^i$ . On the other hand, the variables  $Z_j$  are independent. Since  $Z = \sum_{j=1}^{j=n} Z_j$ ,  $Z$  is a binomial of parameters  $n$  and  $1 - 1/2^i$ . So we have:  $P(Z = k) = \binom{n}{k} (1 - 1/2^i)^k (1/2^i)^{n-k}$ . Hence the result. □

In the case of a unary alphabet, the probability  $P(k, t)$  of reaching a subset  $Y$  of size  $k$  by  $t$  successive transitions from the set  $I$  of initial states can be calculated as follows. Let us consider the matrix of probabilities  $M$  built as below:

$$M = \begin{pmatrix} P(0 \rightarrow 0) & \cdots & \cdots & \cdots & P(0 \rightarrow n) \\ \vdots & \ddots & & & \vdots \\ \vdots & & P(i \rightarrow k) & & \vdots \\ \vdots & & & \ddots & \vdots \\ P(n \rightarrow 0) & \cdots & \cdots & \cdots & P(n \rightarrow n) \end{pmatrix}$$

The coefficient  $M_{i,k}$  represents the probability of obtaining a subset of size  $k$  by a transition from a subset of size  $i$ .

For example, for an automaton of size 3 we have the following matrix:

$$M = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{8} & \frac{3}{8} & \frac{3}{8} & \frac{1}{8} \\ \frac{1}{64} & \frac{9}{64} & \frac{27}{64} & \frac{27}{64} \\ \frac{1}{512} & \frac{21}{512} & \frac{147}{512} & \frac{343}{512} \end{pmatrix}$$

The probability  $P(k, t)$  can be computed according to the formula:

$$P(k, t) = \sum_{i=0}^n P(i, t-1) \times P(i \rightarrow k)$$

Hence the proposition:

**Proposition 4.2** We consider the case of a unary alphabet. Let  $V_t$  be the vector of size  $n$  whose  $k^{th}$  element corresponds to the probability  $P(k, t)$  of obtaining a subset of size  $k$  by  $t$  successive transitions from the set of initial states. We will take  $V_0 = (0, 1, 0, \dots, 0)$ , which corresponds to the choice of a unique initial state. We have:

$$V_t = V_0 M^t$$

The following table, realized with Maple, shows how the probability  $P(n, 2)$  grows w.r.t. the size  $n$  of the NFA. It indicates for example that  $P(n, 2)$  is greater than 0.9 when  $n$  is greater than 17.

$n >$	17	28	37	46	54	100
$P(n, 2) >$	$1 - 10^{-1}$	$1 - 10^{-2}$	$1 - 10^{-3}$	$1 - 10^{-4}$	$1 - 10^{-5}$	$1 - 10^{-10}$

It clearly shows that  $P(n, 2)$  tends to 1 when  $n$  tends to infinity. As a result, we have:

**Proposition 4.3** Let  $\mathcal{A}$  be a unary  $n$ -state NFA associated with an equiprobable bitstream. The probability that the deterministic automaton obtained by subset construction from  $\mathcal{A}$  has 3 states tends to 1 when  $n$  tends to infinity.

We can notice that the deterministic automaton obtained by the subset construction has a size of 2 if we take an equiprobable set of initial states. Moreover, this result can easily be generalized to the case of an arbitrary alphabet size.

**Proposition 4.4** Let  $\mathcal{A}$  be an  $n$ -state NFA associated with an equiprobable bitstream. Then, the probability that the deterministic automaton obtained from the subset construction of  $\mathcal{A}$  has  $m + 2$  states tends to 1 when  $n$  tends to infinity.

Notice that the deterministic automaton obtained by the subset construction has  $m + 1$  states if we take an equiprobable set of initial states.

Moreover, a corollary of Proposition 4.4 is that the probability that  $\mathcal{A}$  is accessible tends to 1 when  $n$  tends to infinity. Furthermore, if the final states are randomly chosen, as the reversed automaton holds all the properties defined previously, the automaton is co-accessible.

## 4.2 Study of the $\star$ -NFAs

A probabilistic analysis can also be carried out concerning  $\cap$ -NFAs and  $\oplus$ -NFAs associated with an equiprobable bitstream.

**Proposition 4.5** Let  $\mathcal{A}$  be an  $n$ -state  $\cap$ -NFA associated with an equiprobable bitstream. If the initial state is unique, the probability that the deterministic automaton obtained by subset construction has  $m + 2$  states tends to 1 when  $n$  tends to infinity. Otherwise, the probability that the deterministic automaton obtained by subset construction has  $m + 1$  states tends to 1 when  $n$  tends to infinity.

### Proof

We use the same reasoning as for the  $\cup$ -NFAs, the role of the values 1 and 0 being inverted.

□

**Proposition 4.6** Let  $\mathcal{A}$  be an  $n$ -state  $\oplus$ -NFA associated with an equiprobable bitstream. The probability  $P(i \rightarrow k)$  of reaching a subset  $Y$  of  $Q$  of size  $k$  by a transition from a non-empty subset  $X$  of size  $i$  is independent of  $i$ . We have:

$$P(i \rightarrow k) = \frac{\binom{n}{k}}{2^n}$$

**Proof**

Let  $O$  (resp.  $E$ ) be the set of the odd (resp. even) integers of  $\{0, \dots, n\}$ . A state  $y$  belongs (resp. does not belong) to  $y$  if there exists an odd (resp. even) number of states of  $X$  reaching  $y$ . As a result, we have:

$$P(i \rightarrow k) = \frac{\binom{n}{k} \times A^k \times B^{n-k}}{2^{in}}$$

where  $A$  and  $B$  are defined as follows:

$$A = \sum_{k \in E} \binom{i}{k} = 2^{i-1} = \sum_{k \in O} \binom{i}{k} = B$$

Hence the result.

□

To conclude this analysis for the case of equiprobable bitstreams, let us notice that our results (based on subset automata) are in accordance with van Zijl's experimental observations (based on minimal DFAs). Indeed Proposition 4.4 (resp. Proposition 4.5) obviously means that  $\cup$ -NFAs (resp.  $\cap$ -NFAs) generated by an equiprobable bitstream are asymptotically never succinct. On the contrary, by Proposition 4.6 the subsets produced by determinization of a  $\oplus$ -NFA are equiprobably distributed. Therefore equiprobable bitstreams lead to a correct random model to study the succinctness of  $\oplus$ -NFAs.

## 5 Probabilistic study: non-equiprobable case

The approach is inspired by the exhaustive construction of the subset automaton. We compute the probability that the image of a subset equiprobably chosen among the  $2^n$  possible subsets of  $Q$  contains

a given state. It allows us to show that the subsets produced by determinization are equiprobably distributed in the case where the bitstream is generated with the probability  $\frac{1}{x(n)} = 2 - 2^{\frac{n-1}{n}}$ .

## 5.1 Study of the classical NFAs

**Proposition 5.1** Let  $\mathcal{A} = \langle Q, \Sigma, \delta, I, F \rangle$  be an  $n$ -state NFA associated with a  $\frac{1}{x}$ -probability bitstream. Let  $q \in Q$ ,  $a \in \Sigma$ . Let  $X$  be an equiprobably chosen subset of  $Q$ . The probability  $P_{\cup}(x, n)$  such that  $q$  belongs to  $\delta(X, a)$  is equal to:

$$P_{\cup}(x, n) = \frac{(2x)^n - (2x - 1)^n}{(2x)^n}$$

### Proof

Let  $V$  be the vector of  $\{0, 1\}^n$  associated with  $X$ . Since the subset  $X$  is chosen equiprobably among the  $2^n$  subsets of  $Q$ , the probability that  $V[i]$  is equal to 1 is equal to  $1/2$ , for all  $i$  in  $\{1, \dots, n\}$ . Let  $V'$  be the vector of  $\{0, 1\}^n$  associated with  $\delta^{-1}(q, a)$ . By hypothesis, the probability that  $V'[i]$  is equal to 1 is  $1/x$ , for all  $i$  in  $\{1, \dots, n\}$ . Consequently, the probability that  $(V[i], V'[i])$  is different from  $(1, 1)$  is equal to  $1 - \frac{1}{2x}$ , for all  $i$  in  $\{1, \dots, n\}$ .

As  $\mathcal{A}$  is a  $\cup$ -NFA, we have:  $q \in \delta(X, a) \Leftrightarrow \exists y \in X \mid q \in \delta(y, a)$ . Thus,  $q \in \delta(X, a) \Leftrightarrow \exists i \in \{1, \dots, n\} \mid (V[i], V'[i]) = (1, 1)$ . Then the probability that  $q$  does not belong to  $\delta(X, a)$  is equal to:  $P' = (1 - \frac{1}{2x})^n$ . Hence the result.

□

When the bitstream is equiprobable, we obtain:  $P_{\cup}(2, n) = 1 - (3/4)^n$ . Clearly, the probability that  $V'[i]$  is equal to 1, for  $i$  in  $\{1, \dots, n\}$ , is thus different from  $1/2$ . We conclude that the subsets of  $\delta(X, a)$  cannot be equiprobably generated by an equiprobable bitstream during the subset construction. The resolution of the equation

$$\frac{(2x)^n - (2x - 1)^n}{(2x)^n} = \frac{1}{2}$$

gives the following solution:

$$x(n) = \frac{1}{2 - 2^{\frac{n-1}{n}}}$$

Using bitstreams of probability  $\frac{1}{x(n)}$  allows us to maintain that at each step of the subset construction each subset of  $Q$  has an equal chance of appearing. Intuitively it means that for a fixed  $n$ , the probability  $\frac{1}{x(n)}$  allows us to maximize the average number of DFA's states. Let us recall that this property occurs with  $\oplus$ -NFAs generated by equiprobable bitstreams (Proposition 4.6 and Proposition 5.3). Van Zijl's results concerning equiprobable bitstreams show that  $\oplus$ -NFAs are more succinct than  $\cup$ -NFAs and  $\cap$ -NFAs. Our experimental results bring to light that  $\cup$ -NFAs coming from a bitstream with a probability equal to  $\frac{1}{x(n)}$  are more succinct than  $\cup$ -NFAs coming from an equiprobable bitstream.

Moreover the probability  $\frac{1}{x(n)}$  increases the chance to get a DFA having a given (large) size. This is coherent with the experimental results of Leslie *et al.* that show that DFAs with maximal size come from NFAs generated with a density equal to  $\frac{2}{n}$ . The gap between  $\frac{1}{x(n)}$  and  $\frac{2}{n}$  can be partly explained by the specific parameters of the experimental protocol used by Leslie *et al.*: a unique initial state, an alphabet of size greater than 10 and the use of a connected structure to build the random NFA.

## 5.2 Experimental results

First, we have carried out several tests to study the size distribution of the DFAs built by the subset construction applied to the randomly generated NFAs. Notice that our approach is sensibly different from van Zijl's one, which focuses on the size distribution of the minimal DFAs equivalent to the randomly generated NFAs. On the contrary, our study is close to Leslie's one [4].

Figures 2, 3 and 4 illustrate the results obtained from the tests. For the first two figures the abscissa represents the number of states of the DFA obtained by the subset construction applied to a random NFA; the ordinate represents the percentage of DFAs of a fixed size generated by subset construction.

For Figure 4 the abscissa represents the size of the subsets produced by the subset construction applied to random 50-state NFAs; the ordinate represents the percentage of subsets of a given size generated during the subset construction.

Concerning the figures 2.[a-f], a sample of 100 000 NFAs of size 5 has been used. Figures 2.a and 2.b show the gain obtained by using the probability  $\frac{1}{x(n)}$  (Figure 2.b) rather than the probability  $1/2$  (Figure

2.a) to produce the bitstream. The alphabet size is equal to 2 and the set of initial states is equiprobably and randomly chosen.

Figures 2.c and 2.d allow the comparison between the two possible choices for the set of initial states (a set of initial states equiprobably chosen or a unique initial state). These two tests are based on equiprobable bitstreams.

Figures 2.e and 2.f point out the influence of the alphabet size on the size distribution of the DFAs. The bitstream is generated with a probability equal to  $\frac{1}{x(n)}$ . The alphabet size is equal to 2 in Figure 2.e and equal to 4 in Figure 2.f.

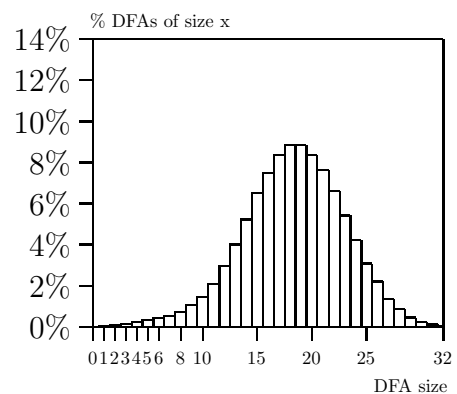
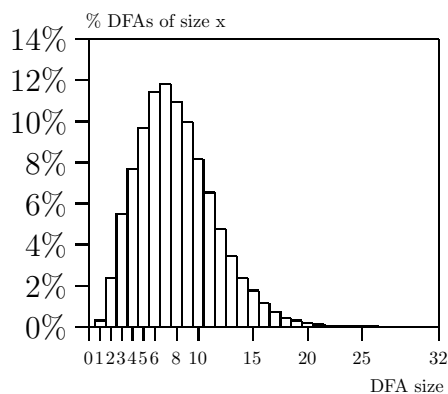
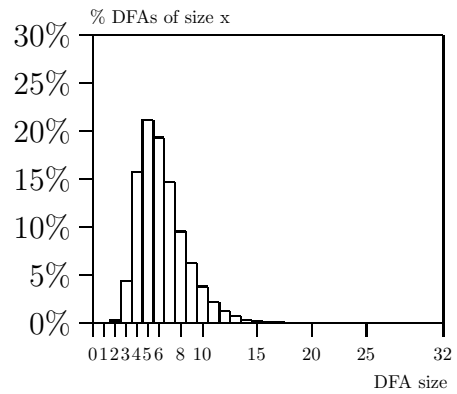
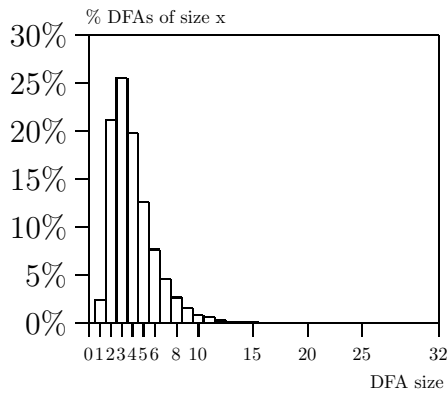
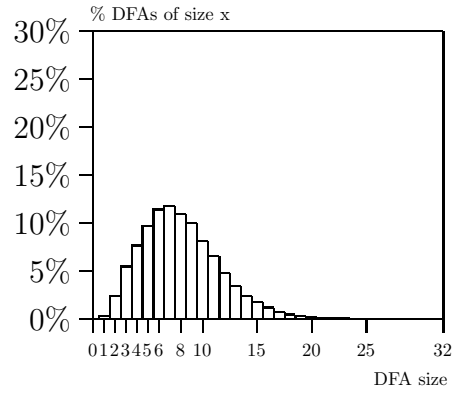
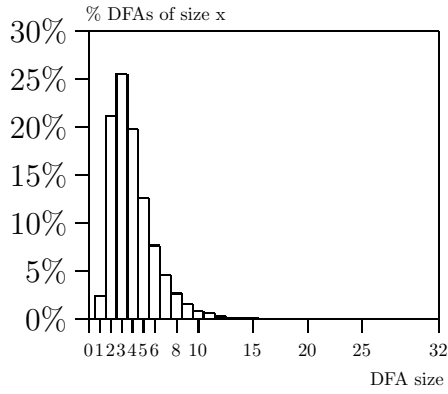


Figure 2: Tests on 5-state NFAs.

Figures 3 and 4 illustrate the pertinence of the probability  $\frac{1}{x(n)}$ . Figure 3 shows the distribution of the DFAs size, for NFAs of size 20, 50 and 100. The alphabet is of size 2. A sample of 100 000 NFAs has been used for each graph. The DFAs sizes have been gathered by brackets of size 10.

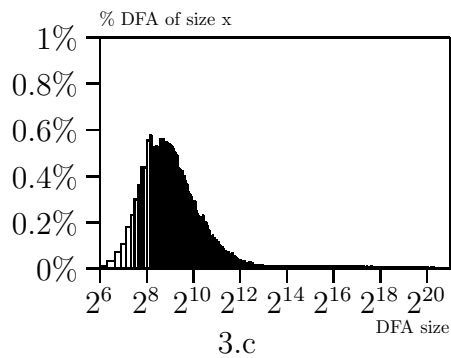
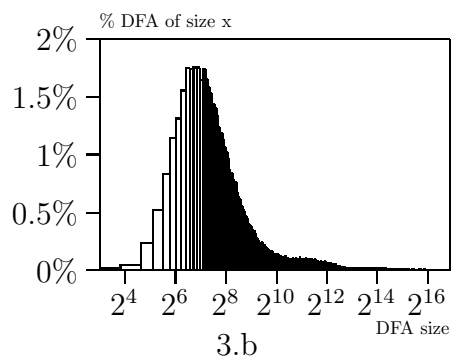
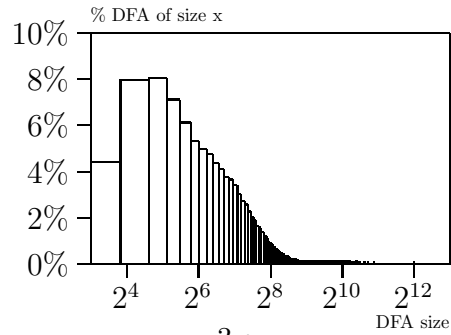


Figure 3: Size distribution of DFAs obtained by subset construction applied on 20-state NFAs (a), 50-state NFAs (b) and 100-state NFAs (c), with the probability  $\frac{1}{x(n)}$ .

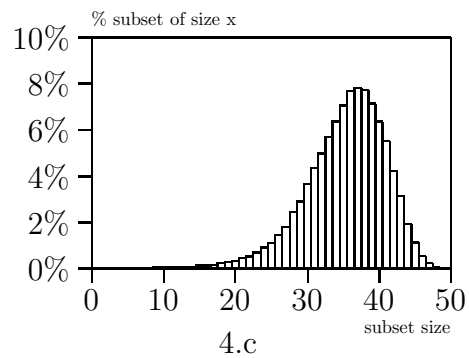
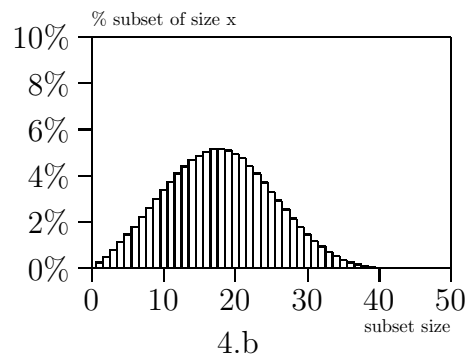
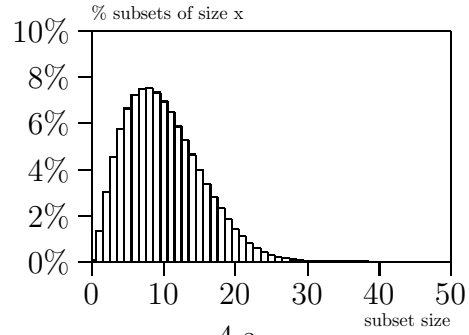


Figure 4: Size distribution of the subsets obtained during subset construction applied on 50-state NFAs. (a):  $\frac{1}{x(50)} - 0.01$  (b):  $\frac{1}{x(50)}$  (c):  $\frac{1}{x(50)} + 0.01$ .

Figure 4 enlightens the influence of small variations around the value  $\frac{1}{x(n)}$  on the repartition of the size of subsets. With a smaller value

$(\frac{1}{x(n)} - 0.01)$  the size of subsets is smaller (Figure 4.a). With a larger value  $(\frac{1}{x(n)} + 0.01)$  the size of subsets is larger (Figure 4.a). For each one of these graphs, the size of the alphabet is 2, and a sample of 1000 NFAs has been used.

### 5.3 Study of the $\star$ -NFAs

A probabilistic analysis can also be carried out concerning the  $\cap$ -NFAs and the  $\oplus$ -NFAs associated with non equiprobable bitstreams.

**Proposition 5.2** Let  $\mathcal{A}$  be an  $\cap$ -NFAs of size  $n$  associated with a bitstream with a  $1/x$ -probability. Let  $q \in Q$ ,  $a \in \Sigma$ . Let  $X$  be an equiprobably chosen subset of  $Q$ . The probability  $P$  that  $q$  belongs to  $\delta(X, a)$  is equal to:

$$P_{\cap}(x, n) = \left(\frac{2x-1}{2x}\right)^n - \frac{1}{x^n}$$

#### Proof

Let  $V$  (resp.  $V'$ ) be the vector of  $\{0, 1\}^n$  associated with  $X$  (resp. to  $\delta^{-1}(q, a)$ ). For any  $i$  in  $\{1, \dots, n\}$ , the probability that  $V[i]$  (resp.  $V'[i]$ ) is equal to 1 is equal to  $1/2$  (resp.  $1/x$ ). Therefore, the probability that  $(V[i], V'[i])$  is different from  $(1, 0)$  is equal to  $\frac{2x-1}{2x}$ , for any  $i$  in  $\{1, \dots, n\}$ .

Since  $\mathcal{A}$  is an  $\cap$ -NFA, we have  $q \in \delta(X, a) \Leftrightarrow \forall y \in X, q \in \delta(y, a)$ , and consequently:

$$q \in \delta(X, a) \Leftrightarrow \forall i \in \{1, \dots, n\}, (V[i], V'[i]) \neq (1, 0) \text{ and } \exists i \in \{1, \dots, n\} \mid V[i] = 1$$

Therefore the probability that  $q$  belongs to  $\delta(X, a)$  is equal to:

$$P_{\cap}(x, n) = \left(\frac{2x-1}{2x}\right)^n - \frac{1}{x^n}$$

□

When the bitstream is equiprobable, we get  $P_{\cap}(2, n) = \left(\frac{3}{4}\right)^n - \frac{1}{2^n}$ .

**Proposition 5.3** Let  $\mathcal{A}$  be a  $\oplus$ -NFA of size  $n$  associated with a bitstream with a probability of  $1/x$ . Let  $q \in Q$ ,  $a \in \Sigma$ . Let  $X$  be an equiprobably chosen subset of  $Q$ . The probability  $P_{\oplus}(x, n)$  that  $q$  belongs to  $\delta(X, a)$  is equal to:

$$P_{\oplus}(x, n) = \frac{1}{2(x-1)} - \frac{1}{2x^{n-1}(x-1)}$$

**Proof**

Let  $V$  (resp.  $V'$ ) be the vector of  $\{0, 1\}^n$  associated with  $X$  (resp. with  $\delta^{-1}(q, a)$ ). For any  $i$  in  $\{1, \dots, n\}$ , the probability that  $V[i]$  (resp.  $V'[i]$ ) is equal to 1 is equal to  $1/2$  (resp.  $1/x$ ). Therefore, the probability that  $(V[i], V'[i])$  is different from  $(1, 1)$  is equal to  $\frac{2x-1}{2x}$ , and the probability that  $(V[i], V'[i])$  is equal to  $(1, 1)$  is equal to  $\frac{1}{2x}$ , for any  $i$  in  $\{1, \dots, n\}$ .

Since  $\mathcal{A}$  is a  $\oplus$ -NFA, we have:

$$q \in \delta(X, a) \Leftrightarrow |\{y \in X \text{ s.t. } q \in \delta(y, a)\}| \equiv 0 \pmod{2}$$

We assume that  $n = 2p$  and that  $P$  denotes the probability  $P_{\oplus}(x, n)$ . Let  $E$  (resp.  $O$ ) be the set of even (resp. odd) integers of  $\{1, \dots, n\}$ . Let  $E'$  be the set of odd integers ranging between 1 and  $n - 2$ . We have:

$$q \in \delta(X, a) \Leftrightarrow \begin{cases} |\{i \in \{1, \dots, n\} \text{ s.t. } (V[i], V'[i]) = (1, 1)\}| \equiv 1 \pmod{2} \\ |\{i \in \{1, \dots, n\} \text{ s.t. } (V[i], V'[i]) \neq (1, 1)\}| \equiv 1 \pmod{2} \end{cases}$$

Hence:

$$\begin{aligned} P &= \frac{\sum_{k \in O} \binom{n}{k} (2x-1)^k}{(2x)^n} \\ P &= \frac{\sum_{k \in O} \binom{n-1}{k} (2x-1)^k + \sum_{k \in O} \binom{n-1}{k-1} (2x-1)^k}{(2x)^n} \\ P &= \frac{\sum_{k=0}^n \binom{n-1}{k} (2x-1)^k + 2 \sum_{k \in E'} \binom{n-1}{k} (2x-1)^k}{(2x)^n} \\ P &= \frac{1}{2x} + \frac{2 \sum_{k \in E'} \binom{n-1}{k} (2x-1)^k}{(2x)^n} \\ P &= \frac{1}{2x} + \frac{1}{2x^2} + \dots + \frac{1}{2x^{n-1}} \end{aligned}$$

So finally:

$$P_{\oplus}(x, n) = \frac{1}{2(x-1)} - \frac{1}{2x^{n-1}(x-1)}$$

The proof is similar for  $n = 2p + 1$ , with the hypothesis:

$$q \in \delta(X, a) \Leftrightarrow \begin{cases} |\{i \in \{1, \dots, n\} \text{ s.t. } (V[i], V'[i]) = (1, 1)\}| \equiv 1 \pmod{2} \\ |\{i \in \{1, \dots, n\} \text{ s.t. } (V[i], V'[i]) \neq (1, 1)\}| \equiv 0 \pmod{2} \end{cases}$$

□

When the bitstream is equiprobable, we get  $P_{\oplus}(2, n) = 1/2 - 1/2^n$ . In this expression, the term  $-1/2^n$  correlates to the case where  $X$  is empty. So when  $X$  is a non-empty subset, a state  $q$  is reached with a probability equal to  $1/2$ . A  $\oplus$ -NFA generated from an equiprobable bitstream equiprobably produces any subset of  $X$  during the subset construction. This is coherent with the probability  $P = \frac{\binom{n}{k}}{2^n}$  of obtaining a subset of size  $k$  from a subset of size  $i$  (Proposition 4.6).

## 6 Conclusion

The probabilistic analysis of nondeterministic transition tables associated with an equiprobable bitstream gives an asymptotical justification of van Zijl's experimental results. On one hand, a reject algorithm can be used to generate accessible NFAs, and on the other hand the associated DFAs have an asymptotical size of  $m + 2$ . We can notice that the asymptotical behaviour is obtained as soon as the automaton has more than 30 states. Van Zijl's results are based on NFAs of size less than 10 with a unique initial state, so the succinctness is relatively better in her results. Moreover, the analysis of the distribution in size of the DFAs obtained by subset construction seems to be fruitful. The generation of NFAs associated with bitstreams with a probability of  $\frac{1}{x(n)} = 2 - 2^{\frac{n-1}{n}}$  leads to an optimal range. This result is coherent with the conjecture of Leslie *et al.* that says that the number of states of the DFA is maximum when the deterministic density is approximately equal to 2.

## References

- [1] J.E. Hopcroft and J.D. Ullman. *Introduction to Automata Theory, Languages and Computation*. Addison-Wesley, Reading, 1979.
- [2] J. Hromkovič, S. Seibert, and T. Wilke. Translating regular expressions into small  $\epsilon$ -free nondeterministic finite automata. *Journal of Computer and System Sciences*, 62(4):565–588, June 2001.
- [3] L. Ilie and S. Yu. Constructing NFAs by optimal use of positions in regular expressions. In *CPM'2002, Lecture Notes in Computer Science*, pages 279–288, 2002.

- [4] T. Leslie. Efficient approaches to subset construction. Master's thesis, University of Waterloo, Ontario, Canada, 1995. Supervised by D. Raymond and D. Wood.
- [5] A.R. Meyer and M.J. Fisher. Economy of description by automata grammars and formal systems. *FOCS*, 12:188–191, 1971.
- [6] T. Paranthoën. Génération aléatoire d'automates non-déterministes, et application aux systèmes multi-agents. Master's thesis, Université de Rouen, France, 2001.
- [7] C. Nicaud. *Etude du comportement en moyenne des automates finis et des langages rationnels*. PhD thesis, Université Paris 7, 2000.
- [8] L. van Zijl. *Generalized Nondeterminism and the Succinct Representation of Regular languages*. PhD thesis, University of Stellenbosch, 1997.
- [9] L. van Zijl. The quantification of succinctly representable regular languages. Technical report, Stellenbosch University, 1999.
- [10] L. van Zijl *et al.* [www.cs.sun.ac.za/~lynette/MERLin/MerlinManRev.ps](http://www.cs.sun.ac.za/~lynette/MERLin/MerlinManRev.ps), Merlin 1.1 help. Technical report.
- [11] S. Yu. Regular languages. *in: G. Rozenberg, A. Salomaa (Eds.), Handbook of Formal Languages, Word, Grammar*, I:41–110, 1997.