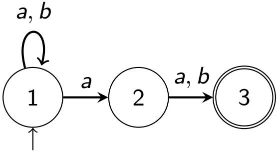
One Drop of Non-Determinism in a Random Deterministic Automaton

Arnaud Carayol, Philipe Duchon, Florent Koechlin, Cyril Nicaud
LORIA, Inria, CNRS, Nancy, France

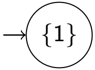
STACS 2023
March 2023, 8th

Determinising an automaton: the Powerset construction

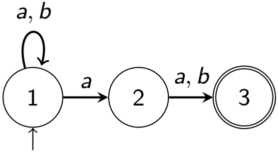


$\mathcal{L}(\mathcal{A}) = \Sigma^* a \Sigma$

$Q_{det} \subseteq 2^Q$

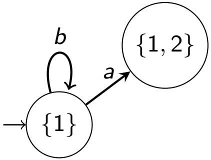


Determinising an automaton: the Powerset construction

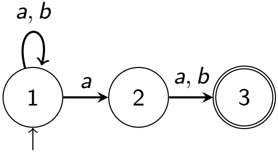


$\mathcal{L}(\mathcal{A}) = \Sigma^* a \Sigma$

$Q_{det} \subseteq 2^Q$

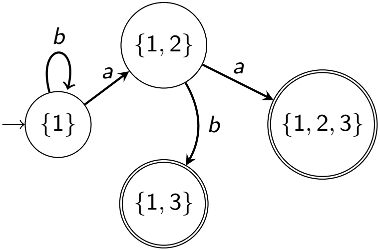


Determinising an automaton: the Powerset construction

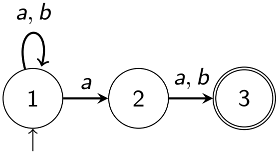


$\mathcal{L}(\mathcal{A}) = \Sigma^* a \Sigma$

$Q_{det} \subseteq 2^Q$

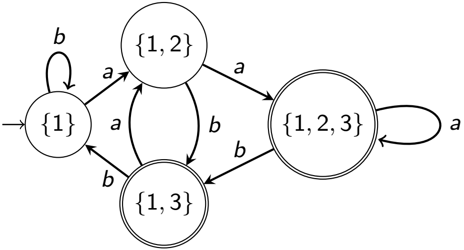


Determinising an automaton: the Powerset construction



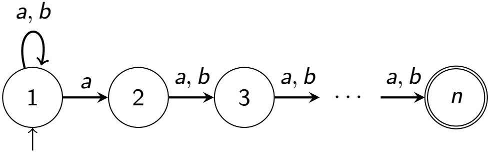
$\mathcal{L}(\mathcal{A}) = \Sigma^* a \Sigma$

$Q_{det} \subseteq 2^Q$



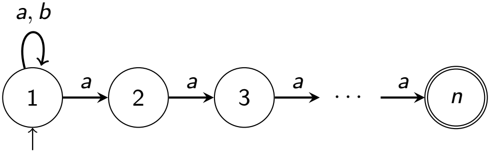
Question

The powerset construction yields an exponential blow-up on the number of states. It is **tight**.



recognizing $\Sigma^* a \Sigma^{n-1}$. Its minimal automaton has 2^{n-1} states.

Question



Question: How likely (in a probabilistic setting) is this state complexity explosion for non deterministic automata?

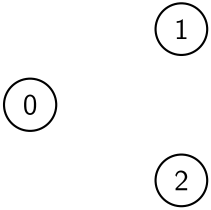
Difficulty: no canonical distribution for random non deterministic automata

- uniform distribution: degenerated
- each transition added with probability p: degenerated until critical probability $p = 1/n$

Random deterministic automaton

Let $|Q| = [n]$ the set of states, $k = |\Sigma|$ the size of the alphabet.
A **uniform random deterministic automaton** on Q is obtained as follows:

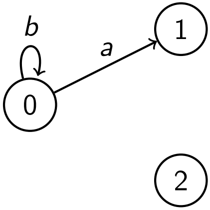
- Draw uniformly the destination of $q \xrightarrow{a} \ldots$ for all q and a.



Random deterministic automaton

Let $|Q| = [n]$ the set of states, $k = |\Sigma|$ the size of the alphabet.
A **uniform random deterministic automaton** on Q is obtained as follows:

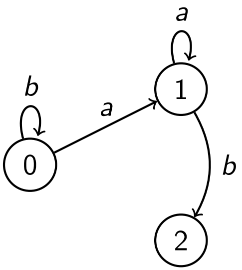
- Draw uniformly the destination of $q \xrightarrow{a} \ldots$ for all q and a.



Random deterministic automaton

Let $|Q| = [n]$ the set of states, $k = |\Sigma|$ the size of the alphabet.
A **uniform random deterministic automaton** on Q is obtained as follows:

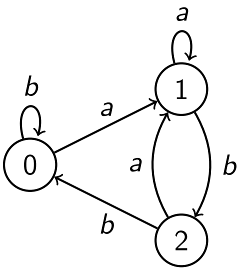
- Draw uniformly the destination of $q \xrightarrow{a} \ldots$ for all q and a.



Random deterministic automaton

Let $|Q| = [n]$ the set of states, $k = |\Sigma|$ the size of the alphabet.
A **uniform random deterministic automaton** on Q is obtained as follows:

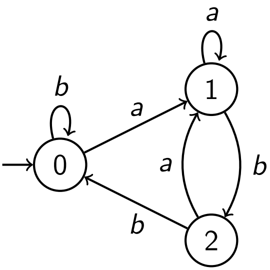
- Draw uniformly the destination of $q \xrightarrow{a} \ldots$ for all q and a.



Random deterministic automaton

Let $|Q| = [n]$ the set of states, $k = |\Sigma|$ the size of the alphabet.
A **uniform random deterministic automaton** on Q is obtained as follows:

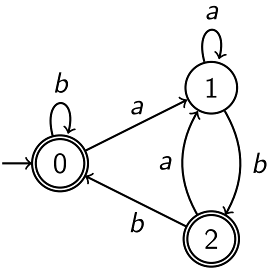
- Draw uniformly the destination of $q \xrightarrow{a} \ldots$ for all q and a.
- Draw uniformly the initial state q_0



Random deterministic automaton

Let $|Q| = [n]$ the set of states, $k = |\Sigma|$ the size of the alphabet.
A **uniform random deterministic automaton** on Q is obtained as follows:

- Draw uniformly the destination of $q \xrightarrow{a} \ldots$ for all q and a.
- Draw uniformly the initial state q_0
- Every state is final with probability f.

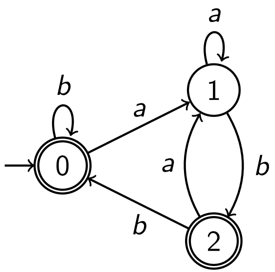


Random deterministic automaton

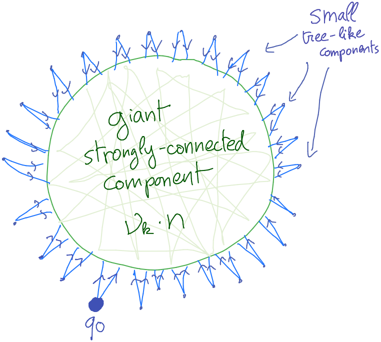
Let $|Q| = [n]$ the set of states, $k = |\Sigma|$ the size of the alphabet.
A **uniform random deterministic automaton** on Q is obtained as follows:

- Draw uniformly the destination of $q \xrightarrow{a} \ldots$ for all q and a.
- Draw uniformly the initial state q_0
- Every state is final with probability f.

$\rightarrow$ on average $f \cdot n$ states are final
$\rightarrow$ we can have f_n as long as f_n and $1 - f_n \geq \Omega(\frac{1}{\sqrt{n}})$.



Typical shape of a random deterministic automaton

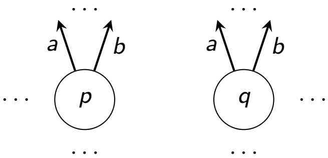


With high probability as $n \to \infty$

- a constant proportion of states are accessible from the initial state q_0 and lie into a giant strongly connected component of size $\approx \nu_k n$ [Grusho]
- typical distances are of order $\log(n)$ [Addario-Berry Balle Perarnau]
- cycles outside the s.c. component are small [Cai Devroye]
- minimal automaton has size $\nu_k n$ [Bassino, David, Sportiello]

Almost deterministic random automata

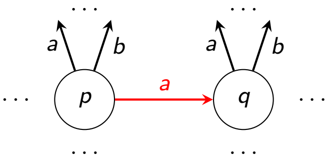
An **almost deterministic** automaton is an automaton obtained from a complete deterministic automaton by adding one extra transition randomly, labelled by a fixed letter $a \in \Sigma$.



Informal question: How bad can the addition of a single non-deterministic choice be for state complexity?

Almost deterministic random automata

An **almost deterministic** automaton is an automaton obtained from a complete deterministic automaton by adding one extra transition randomly, labelled by a fixed letter $a \in \Sigma$.



Informal question: How bad can the addition of a single non-deterministic choice be for state complexity?

Our main result

Theorem: Let Σ with at least two letters a and b, and $\mathcal{A}$ a uniform almost deterministic n-state automaton, where final states have probability f_n, satisfying $f_n, 1 - f_n > \Omega(1/\sqrt{n})$.

Then with visible probability[1], $\mathcal{L}(\mathcal{A})$ has a **super-polynomial** state complexity.

For every $d \in \mathbb{N}$ and for n big enough, $\mathcal{L}(\mathcal{A})$ has a state complexity bigger than n^d with probability $c_d > 0$.

[1] $\mathbb{P}(X) \geq c$ for some $c > 0$ independent of n

Our main result

Theorem: Let Σ with at least two letters a and b, and $\mathcal{A}$ a uniform almost deterministic n-state automaton, where final states have probability f_n, satisfying $f_n, 1 - f_n > \Omega(1/\sqrt{n})$.

Then with visible probability[1], $\mathcal{L}(\mathcal{A})$ has a **super-polynomial** state complexity.

> For every $d \in \mathbb{N}$ and for n big enough, $\mathcal{L}(\mathcal{A})$ has a state complexity bigger than n^d with probability $c_d > 0$.

Corollary: The expected state complexity of $\mathcal{L}(\mathcal{A})$ grows faster that any polynomial in n.

Proof: $\mathbb{E}_n[S] \geq n^d \cdot \mathbb{P}(S \geq n^d) \geq c_d \cdot n^d$.

[1] $\mathbb{P}(X) \geq c$ for some $c > 0$ independent of n

Idea of the proof

We fix $d \in \mathbb{N}$.

We exhibit a substructure that has visible probability and n^d non-equivalent states

Proof technique:

- we build such a structure by drawing random transitions on the fly, when needed
- at each step, we ensure that the built structure can be extended into the final target substructure
- while not consuming too much random transitions to keep some for the next steps

Birthday-like theorems

Classical setting: if we draw uniformly elements of $[n]$, there is a collision after $\approx \sqrt{n}$ draws on average.

We use also some variants:

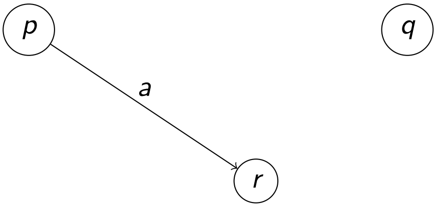
- If we draw $\Theta(\sqrt{n})$ elements of $[n]$, there is no collision with visible probability
- If we draw $\Theta(\sqrt{n})$ elements of $[n]$, we hit a given subset of size $\Theta(\sqrt{n})$ with visible probability before a self collision occurs

even if we forbid in the process to draw some subset of size $\sqrt{n}$.

Construction of the substructure

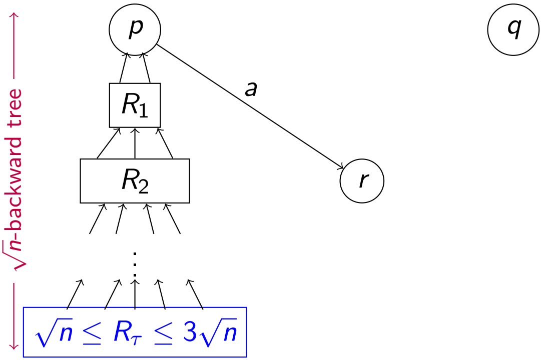
Step 1: cycle around p of logarithmic length starting with a

Construction of the substructure



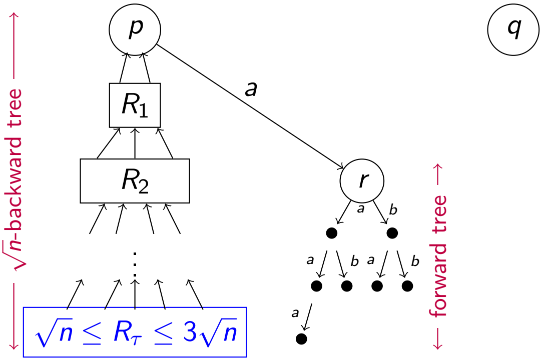
Step 1: cycle around p of logarithmic length starting with a

Construction of the substructure



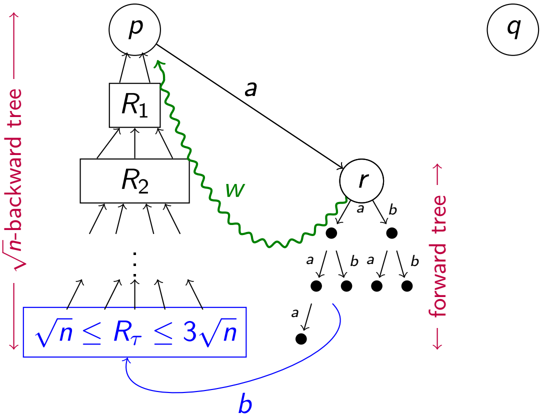
Step 1: cycle around p of logarithmic length starting with a

Construction of the substructure



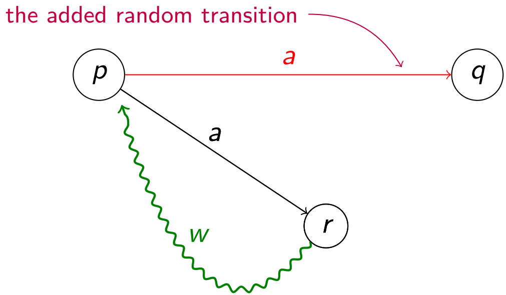
Step 1: cycle around p of logarithmic length starting with a

Construction of the substructure



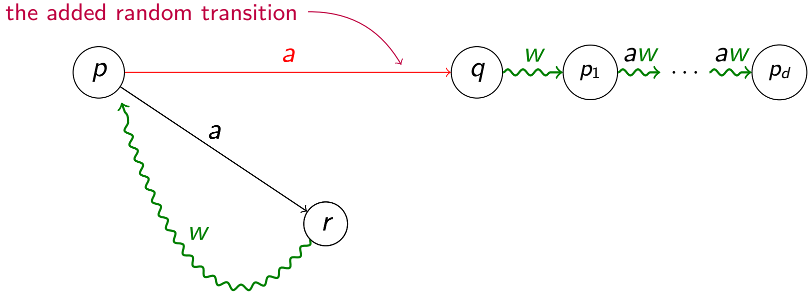
Step 1: cycle around p of logarithmic length starting with a

Construction of the substructure



Step 2: add extra transition

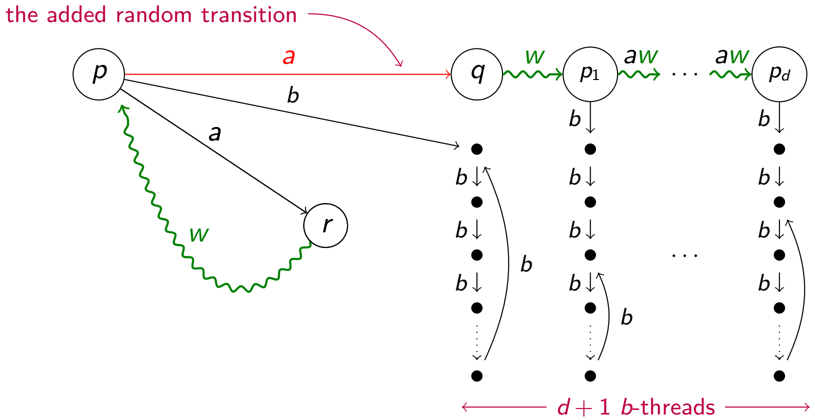
Construction of the substructure



Step 3: generate transitions reading $(aw)^{d-1}$

$\delta_{(aw)^d}(\{p\}) = \{p, p_1, \ldots, p_d\}$

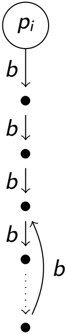
Construction of the substructure



Step 4: b threads of length $\approx \sqrt{n}$ with collision creating cycles of length ℓ_i

From $\{p, \ldots, p_d\}$, we can reach a **b-cycle Λ of length $\mathrm{lcm}(\ell_i)$**

b-threads properties



b threads of length $\approx \sqrt{n}$ with collisions creating cycles of length ℓ_i.

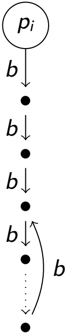
From $\{p, \ldots, p_d\}$, we can reach a **b-cycle Λ of length** $\mathrm{lcm}(\ell_i)$

With visible probability, $\ell_i \approx c\sqrt{n}$ which leads to

$$\mathrm{lcm}(\ell_i) \approx n^{(d+1)/2}$$

Conclusion: With visible probability, in the powerset automaton, we can reach from $\{p, \ldots, p_d\}$ a b-cycle Λ of length $n^{(d+1)/2}$.

b-threads properties



b threads of length $\approx \sqrt{n}$ with collisions creating cycles of length ℓ_i.

From $\{p, \ldots, p_d\}$, we can reach a **b-cycle Λ of length** $\mathrm{lcm}(\ell_i)$

With visible probability, $\ell_i \approx c\sqrt{n}$ which leads to

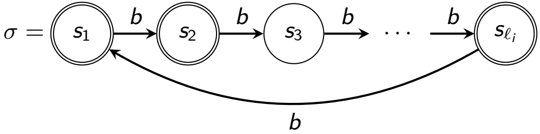
$$\mathrm{lcm}(\ell_i) \approx n^{(d+1)/2}$$

Conclusion: With visible probability, in the powerset automaton, we can reach from $\{p, \ldots, p_d\}$ a b-cycle Λ of length $n^{(d+1)/2}$.

We still need to show that with visible probability, this cycle Λ does not collapse after minimisation.

Adding final states

Adding final states: Every state is final with probability f.



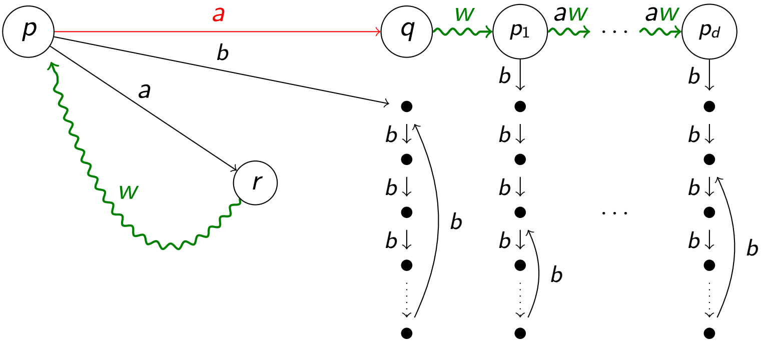
$w_\sigma = 110\ldots1$

Proposition: If w_σ is primitive, then the states of σ are pairwise non-equivalent.

Proposition: If every ℓ_i are pairwise coprime and every w_{σ_i} are primitive, then w_Λ is primitive.
It happens with visible probability [Tóth], [De Felice, Nicaud].

Conclusion: The states of Λ are pairwise non-equivalent in the powerset automaton with visible probability.

Finally



With visible probability:

- $\{p\}$ is accessible from $\{q_0\}$ as $|w| \approx \log(n)$
- hence the b-cycle Λ with $n^{(d+1)/2}$ non-pairwise equivalent states is accessible in the powerset automaton
- hence the minimal automaton for $\mathcal{L}(\mathcal{A})$ has at least $n^{(d+1)/2}$ states.

Conclusion

Main result:
the state complexity of a random almost deterministic automaton is greater than n^d with probability at least $c_d > 0$.

Experiments suggest this probability should be actually rather high.

Further work: avoid many final states by having a single final state. The result should still hold but our proof does not apply directly.