

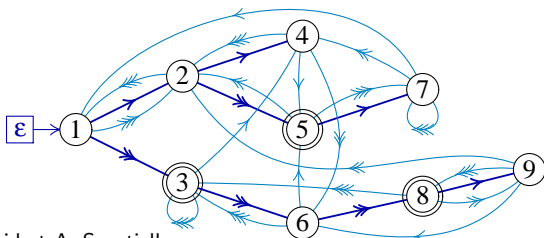
# Énumération asymptotique des automates minimaux

**Frédérique Bassino**

(Laboratoire d'Informatique de Paris Nord)

travail en collaboration avec

Julien David et Andrea Sportiello



F. Bassino, J. David et A. Sportiello  
*Asymptotic enumeration of Minimal Automata*  
arXiv:1109.5683, STACS 2012

Journées Combinatoires de Bordeaux 2012

# Objectifs généraux

- ▶ Etudier de manière quantitative les propriétés des langages réguliers
- ▶ Utiliser des automates minimaux pour représenter les langages réguliers.
- ▶ Enumérer, simuler ces structures
- ▶ Analyser en moyenne des algorithmes manipulant ces structures  
(par exemple l'algorithme de minimisation de Moore)

- ▶ Etudier de manière quantitative les propriétés des langages réguliers
- ▶ Utiliser des automates minimaux pour représenter les langages réguliers.
- ▶ Enumérer, simuler ces structures
- ▶ Analyser en moyenne des algorithmes manipulant ces structures  
(par exemple l'algorithme de minimisation de Moore)

# Pourquoi s'intéresser aux automates minimaux ?

- ▶ Pour chaque langage régulier  $\mathcal{L}$ , il existe **un unique** automate déterministe complet de taille minimale qui le reconnaît : son **automate minimal**.
- ▶ La taille de l'automate minimal d'un langage peut être interprétée comme une mesure de la complexité en espace d'un langage régulier.

*Dans la suite, on va :*

- ▶ Compter (asymptotiquement pour  $n$  grand) les automates minimaux de taille  $n$ .
- ▶ Caractériser l'ensemble des automates minimaux, parmi tous les automates déterministes finis.  
*(propriétés statistiques, probabilités d'occurrence de motifs).*

*Conséquences :*

- ▶ Génération aléatoire des automates minimaux,
- ▶ Analyse en moyenne d'algorithmes sur les langages réguliers.

# Pourquoi s'intéresser aux automates minimaux ?

- ▶ Pour chaque langage régulier  $\mathcal{L}$ , il existe **un unique** automate déterministe complet de taille minimale qui le reconnaît : son **automate minimal**.
- ▶ La taille de l'automate minimal d'un langage peut être interprétée comme une mesure de la complexité en espace d'un langage régulier.

*Dans la suite, on va :*

- ▶ Compter (asymptotiquement pour  $n$  grand) les automates minimaux de taille  $n$ .
- ▶ Caractériser l'ensemble des automates minimaux, parmi tous les automates déterministes finis.  
*(propriétés statistiques, probabilités d'occurrence de motifs).*

*Conséquences :*

- ▶ Génération aléatoire des automates minimaux,
- ▶ Analyse en moyenne d'algorithmes sur les langages réguliers.

# Pourquoi s'intéresser aux automates minimaux ?

- ▶ Pour chaque langage régulier  $\mathcal{L}$ , il existe **un unique** automate déterministe complet de taille minimale qui le reconnaît : son **automate minimal**.
- ▶ La taille de l'automate minimal d'un langage peut être interprétée comme une mesure de la complexité en espace d'un langage régulier.

*Dans la suite, on va :*

- ▶ Compter (asymptotiquement pour  $n$  grand) les automates minimaux de taille  $n$ .
- ▶ Caractériser l'ensemble des automates minimaux, parmi tous les automates déterministes finis.  
*(propriétés statistiques, probabilités d'occurrence de motifs).*

*Conséquences :*

- ▶ Génération aléatoire des automates minimaux,
- ▶ Analyse en moyenne d'algorithmes sur les langages réguliers.

# Pourquoi s'intéresser aux automates minimaux ?

- ▶ Pour chaque langage régulier  $\mathcal{L}$ , il existe **un unique** automate déterministe complet de taille minimale qui le reconnaît : son **automate minimal**.
- ▶ La taille de l'automate minimal d'un langage peut être interprétée comme une mesure de la complexité en espace d'un langage régulier.

*Dans la suite, on va :*

- ▶ Compter (asymptotiquement pour  $n$  grand) les automates minimaux de taille  $n$ .
- ▶ Caractériser l'ensemble des automates minimaux, parmi tous les automates déterministes finis.  
*(propriétés statistiques, probabilités d'occurrence de motifs).*

*Conséquences :*

- ▶ Génération aléatoire des automates minimaux,
- ▶ Analyse en moyenne d'algorithmes sur les langages réguliers.

- ▶  $\Sigma$  un alphabet fini:  $\Sigma = \{a, b, c\}$
- ▶  $\Sigma^*$  l'ensemble des mots sur  $\Sigma$ :  $\Sigma^* = \{\epsilon, a, b, c, aa, ab, \dots\}$
- ▶ Un langage régulier : langage obtenu à partir de sous-ensembles finis de  $\Sigma^*$  et d'un nombre fini d'opérations rationnelles (union  $\cup$ , concaténation  $\cdot$  et étoile  $*$ ) où

$$X^* = \{x_1x_2 \cdots x_n \mid n \geq 0, x_i \in X\}$$



# Expressions rationnelles

Les langages finis peuvent être représentés par des expressions rationnelles écrites sur

- ▶ un alphabet fini  $\Sigma \cup \{\epsilon\}$
- ▶ avec les opérateurs binaires : union  $+$  et concaténation  $\cdot$
- ▶ et l'opérateur unaire  $*$ .

Expression rationnelle :  $a^*b + a(c + d^*)$

- ▶ Cette représentation des langages n'a pas de forme canonique
- ▶ Tester si une expression représente  $\Sigma^*$  est PSPACE-complet.

Un **automate fini**  $A$  est

- ▶ un graphe orienté
- ▶ dont les transitions sont étiquetées sur un alphabet fini
- ▶ avec un ensemble  $\mathcal{I}$  d'états initiaux
- ▶ et un ensemble  $\mathcal{T}$  d'états terminaux.
  
- ▶ Le **langage reconnu** par un automate fini est l'ensemble des étiquette des chemins d'un état initial à un état final.

Automate fini:

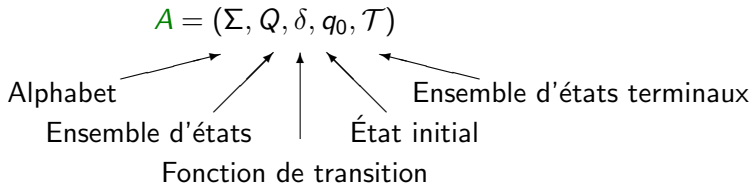


D'après le théorème de Kleene, les langages reconnus par un automate sont les langages réguliers.

# Les automates : précisions

Un automate  $A$  est **déterministe et complet** s'il n'a qu'un seul état initial et que pour tout état  $q$  et pour toute lettre  $\ell$ , il existe exactement une transition d'étiquette  $\ell$  partant de  $q$ .

$A$  est identifié à un quintuplet



*Notations.*

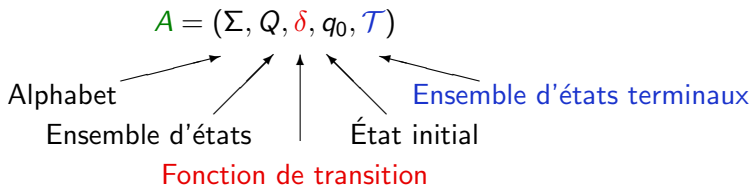
Taille de l'alphabet :  $|\Sigma| = k$ . Nombre d'états :  $|Q| = n$ .

Pour l'énumération, on fait un choix canonique,  $\Sigma \equiv [k]$ ,  $Q \equiv [n]$  et  $q_0 = 1$ . Si on définit un ordre total sur  $Q$ , le couple  $(\delta, \mathcal{T})$  permet de reconstruire l'automate  $A$ .

# Les automates : précisions

Un automate  $A$  est **déterministe et complet** s'il n'a qu'un seul état initial et que pour tout état  $q$  et pour toute lettre  $\ell$ , il existe exactement une transition d'étiquette  $\ell$  partant de  $q$ .

$A$  est identifié à un quintuplet

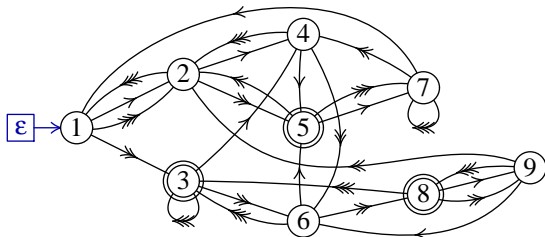


*Notations.*

Taille de l'alphabet :  $|\Sigma| = k$ . Nombre d'états :  $|Q| = n$ .

Pour l'énumération, on fait un choix canonique,  $\Sigma \equiv [k]$ ,  $Q \equiv [n]$  et  $q_0 = 1$ . Si on définit un ordre total sur  $Q$ , le couple  $(\delta, \mathcal{T})$  permet de reconstruire l'automate  $A$ .

# Les automates : accessibilité

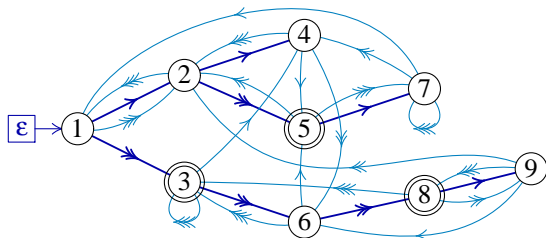


Un automate est **accessible** si, pour tout état  $i$ , il existe un mot  $w$  tel que  $\delta(q_0, w) = i$ .

Dans ce cas, un ordre canonique sur  $Q$  est donné par le parcours en largeur (selon l'ordre de  $\Sigma$ ) de l'automate à partir de l'état initial  $q_0$ .

L'arbre de recherche associé couvre alors tous les états de l'automate

# Les automates : accessibilité



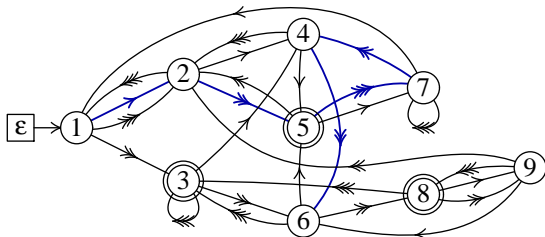
Un automate est **accessible** si, pour tout état  $i$ , il existe un mot  $w$  tel que  $\delta(q_0, w) = i$ .

Dans ce cas, un ordre canonique sur  $Q$  est donné par le parcours en largeur (selon l'ordre de  $\Sigma$ ) de l'automate à partir de l'état initial  $q_0$ .

L'arbre de recherche associé couvre alors tous les états de l'automate



# Les automates : accessibilité



Un automate est **accessible** si, pour tout état  $i$ , il existe un mot  $w$  tel que  $\delta(q_0, w) = i$ .

Dans ce cas, un ordre canonique sur  $Q$  est donné par le parcours en largeur (selon l'ordre de  $\Sigma$ ) de l'automate à partir de l'état initial  $q_0$ .

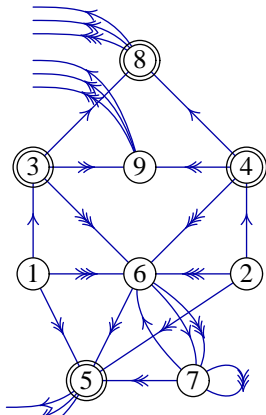
L'arbre de recherche associé couvre alors tous les états de l'automate



# Les automates : Minimalité

Deux états  $i$  et  $j$  sont les **équivalents** - au sens de Myhill–Nerode - ssi, pour tout mot  $w$ , ou bien  $\delta(i, w), \delta(j, w) \in \mathcal{T}$  ou bien  $\delta(i, w), \delta(j, w) \notin \mathcal{T}$ .

On peut vérifier cette propriété par un parcours en largeur simultané à partir des deux états.

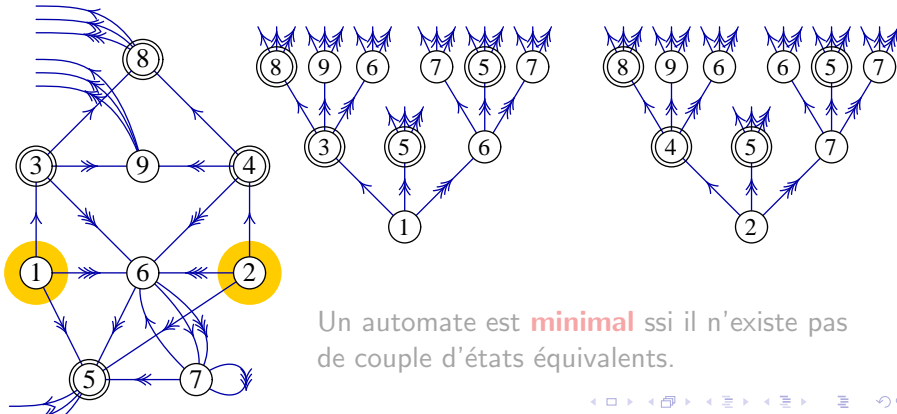


Un automate est **minimal** ssi il n'existe pas de couple d'états équivalents.

# Les automates : Minimalité

Deux états  $i$  et  $j$  sont les **équivalents** - au sens de Myhill–Nerode - ssi, pour tout mot  $w$ , ou bien  $\delta(i, w), \delta(j, w) \in \mathcal{T}$  ou bien  $\delta(i, w), \delta(j, w) \notin \mathcal{T}$ .

On peut vérifier cette propriété par un parcours en largeur simultané à partir des deux états.

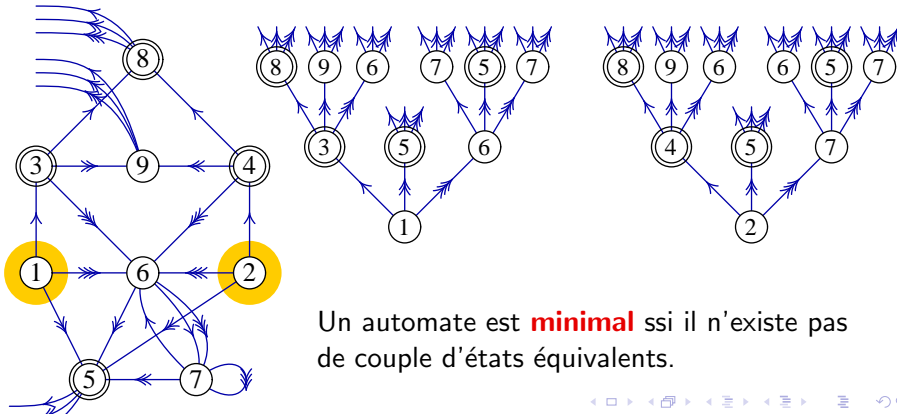


Un automate est **minimal** ssi il n'existe pas de couple d'états équivalents.

# Les automates : Minimalité

Deux états  $i$  et  $j$  sont les **équivalents** - au sens de Myhill–Nerode - ssi, pour tout mot  $w$ , ou bien  $\delta(i, w), \delta(j, w) \in \mathcal{T}$  ou bien  $\delta(i, w), \delta(j, w) \notin \mathcal{T}$ .

On peut vérifier cette propriété par un parcours en largeur simultané à partir des deux états.

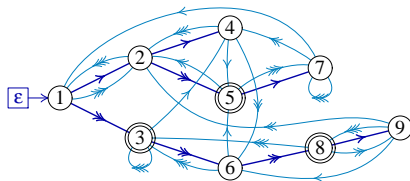


Un automate est **minimal** ssi il n'existe pas de couple d'états équivalents.

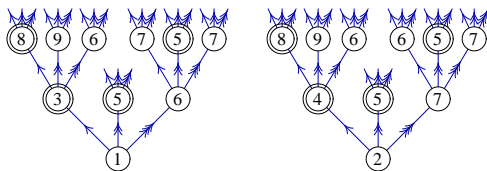
# Les automates : accessibilité et minimalité

Un automate est...

**accessible**, ssi les arbres de recherche issu de l'état initial couvre tous les états de l'automate



**minimal**, ssi il n'existe pas couple d'arbres de recherche certifiant que deux états sont équivalents

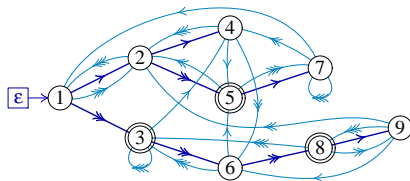


Il s'agit de contraintes **non-locales** et difficiles à exprimer d'un point de vue combinatoire.

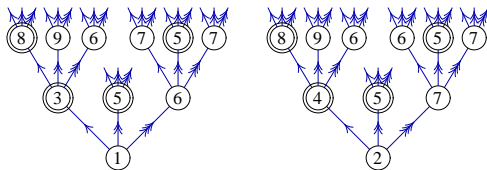
# Les automates : accessibilité et minimalité

Un automate est...

**accessible**, ssi les arbres de recherche issu de l'état initial couvre tous les états de l'automate



**minimal**, ssi il n'existe pas couple d'arbres de recherche certifiant que deux états sont équivalents

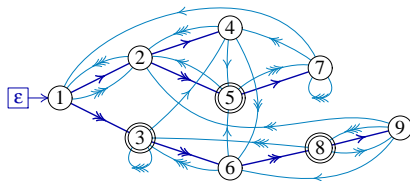


Il s'agit de contraintes **non-locales** et difficiles à exprimer d'un point de vue combinatoire.

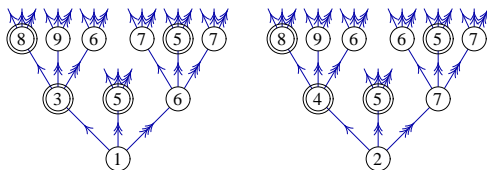
# Les automates : accessibilité et minimalité

Un automate est...

**accessible**, ssi les arbres de recherche issu de l'état initial couvre tous les états de l'automate



**minimal**, ssi il n'existe pas couple d'arbres de recherche certifiant que deux états sont équivalents



Il s'agit de contraintes **non-locales** et difficiles à exprimer d'un point de vue combinatoire.

# Enumérer les automates unaires ( $k = 1$ )

Les automates unaires (déterministes et complets) vérifient

$$i = 1 \cdot a^{i-1}$$

et sont donc caractérisés par le choix de l'état  $c$  tel que  $c = n \cdot a$ .

*Nicaud 1999*

- ▶ Il y a  $n2^n$  automates unaires canoniquement étiquetés
- ▶ Asymptotiquement la moitié des automates unaires (déterministes, complets) accessibles sont minimaux.

# Enumérer les automates unaires ( $k = 1$ )

Les automates unaires (déterministes et complets) vérifient

$$i = 1 \cdot a^{i-1}$$

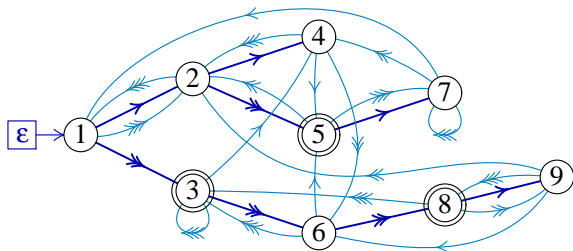
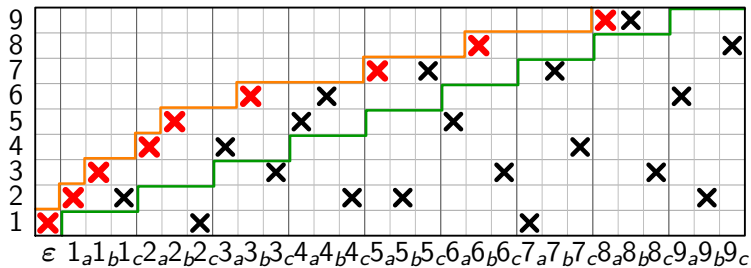
et sont donc caractérisés par le choix de l'état  $c$  tel que  $c = n \cdot a$ .

*Nicaud 1999*

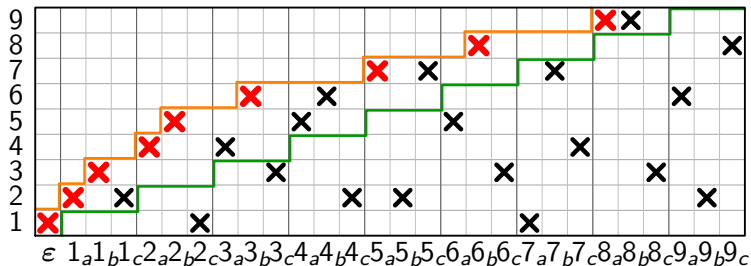
- ▶ Il y a  $n2^n$  automates unaires canoniquement étiquetés
- ▶ Asymptotiquement la moitié des automates unaires (déterministes, complets) accessibles sont minimaux.



# Automates et tableaux

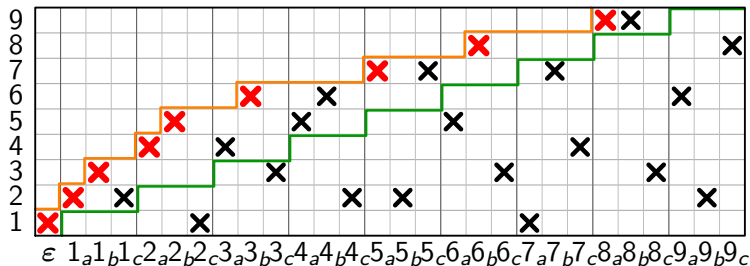


# Automates et tableaux



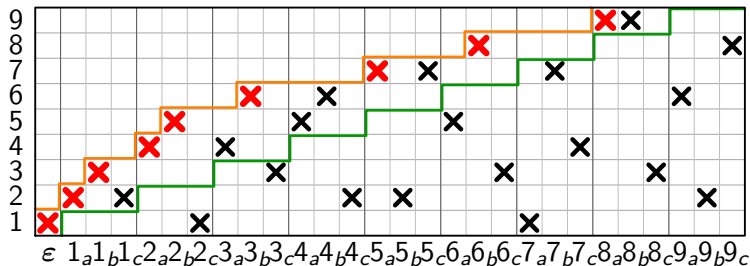
- Une croix par colonne

# Automates et tableaux



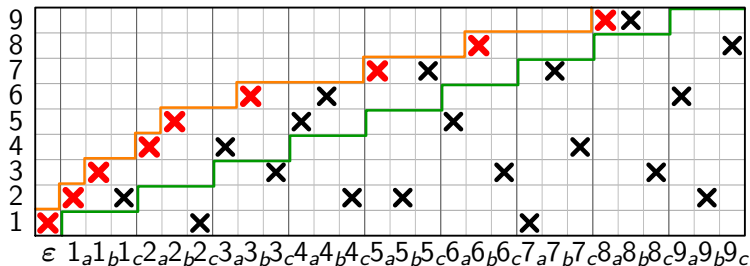
- Une croix par colonne
- Au moins une croix par ligne – la plus à gauche est rouge

# Automates et tableaux



- Une croix par colonne
- Au moins une croix par ligne – la plus à gauche est rouge
- Les croix rouges (*squelette*) forment une séquence monotone

# Automates et tableaux



- Une croix par colonne
- Au moins une croix par ligne – la plus à gauche est rouge
- Les croix rouges (*squelette*) forment une séquence monotone
- Enfin, il faut que les croix rouges restent à la gauche de la diagonale (*contrainte  $k$ -Dyck*)

# Enumérer les automates accessibles

Soit  $\mathcal{D}_{n,k}$  l'ensemble des structures de transition (déterministes complètes) accessibles, sur un alphabet de taille  $|\Sigma| = k \geq 2$ , avec  $|Q| = n$  états.

*Korshunov, 1978; Bassino et Nicaud, 2007; Lebensztayn, 2010.*

$$|\mathcal{D}_{n,k}| = \omega_k \left\{ \begin{matrix} kn+1 \\ n \end{matrix} \right\} (1 + \mathcal{O}(n^{-1})) ,$$

où  $\left\{ \begin{matrix} n \\ m \end{matrix} \right\}$  est le *nombre de Stirling de seconde espèce*, et  $\omega_k$  satisfait  $-k\omega_k = \ln(1 - \omega_k)$ .

Qu'est ce qui se cache derrière ce théorème ?

- Le nombre de tableaux  $n \times m$  est exactement  $\left\{ \begin{matrix} n \\ m \end{matrix} \right\}$  (par interprétation graphique des nombres de Stirling de seconde espèce).
- Un tableau n'est pas  $k$ -Dyck, s'il passe sous la diagonale : cela se produit avec une probabilité d'ordre  $1 - \omega_k$ .

# Enumérer les automates accessibles

Soit  $\mathcal{D}_{n,k}$  l'ensemble des structures de transition (déterministes complètes) accessibles, sur un alphabet de taille  $|\Sigma| = k \geq 2$ , avec  $|Q| = n$  états.

*Korshunov, 1978; Bassino et Nicaud, 2007; Lebensztayn, 2010.*

$$|\mathcal{D}_{n,k}| = \omega_k \left\{ \begin{matrix} kn+1 \\ n \end{matrix} \right\} (1 + \mathcal{O}(n^{-1})) ,$$

où  $\left\{ \begin{matrix} n \\ m \end{matrix} \right\}$  est le *nombre de Stirling de seconde espèce*, et  $\omega_k$  satisfait  $-k\omega_k = \ln(1 - \omega_k)$ .

Qu'est ce qui se cache derrière ce théorème ?

- Le nombre de tableaux  $n \times m$  est exactement  $\left\{ \begin{matrix} n \\ m \end{matrix} \right\}$  (par interprétation graphique des nombres de Stirling de seconde espèce).
- Un tableau n'est pas  $k$ -Dyck, s'il passe sous la diagonale : cela se produit avec une probabilité d'ordre  $1 - \omega_k$ .

# Enumérer les automates accessibles

Soit  $\mathcal{D}_{n,k}$  l'ensemble des structures de transition (déterministes complètes) accessibles, sur un alphabet de taille  $|\Sigma| = k \geq 2$ , avec  $|Q| = n$  états.

*Korshunov, 1978; Bassino et Nicaud, 2007; Lebensztayn, 2010.*

$$|\mathcal{D}_{n,k}| = \omega_k \left\{ \begin{matrix} kn+1 \\ n \end{matrix} \right\} (1 + \mathcal{O}(n^{-1})) ,$$

où  $\left\{ \begin{matrix} n \\ m \end{matrix} \right\}$  est le *nombre de Stirling de seconde espèce*, et  $\omega_k$  satisfait  $-k\omega_k = \ln(1 - \omega_k)$ .

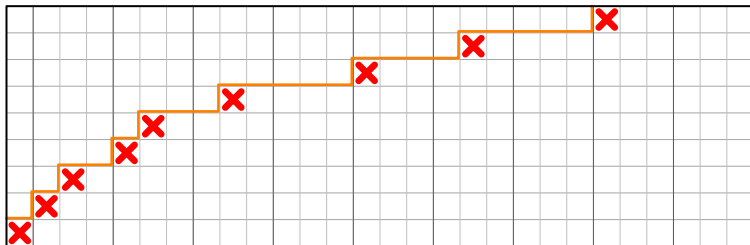
Qu'est ce qui se cache derrière ce théorème ?

- Le nombre de tableaux  $n \times m$  est exactement  $\left\{ \begin{matrix} n \\ m \end{matrix} \right\}$  (par interprétation graphique des nombres de Stirling de seconde espèce).
- Un tableau n'est pas  $k$ -Dyck, s'il passe sous la diagonale : cela se produit avec une probabilité d'ordre  $1 - \omega_k$ .



# Relations entre squelette et croix noires

Q: Combien de tableaux  $T$  peuvent compléter un squelette donné ?



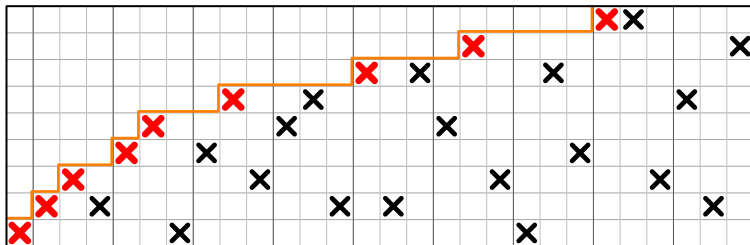
R: Code le squelette par la séquence  $\mathbf{c} = \{c_y\}_{1 \leq y \leq n}$ . C'est

$$\mu(\mathbf{c}) = \begin{cases} \prod_{y=1}^n y^{c_y} & \sum_y c_y = (k-1)n + 1 \\ 0 & \text{autrement} \end{cases}$$

Corollaire :  $\sum_T \phi(\mathbf{c}(T)) = \sum_{\mathbf{c}} \mu(\mathbf{c}) \phi(\mathbf{c})$ .

# Relations entre squelette et croix noires

Q: Combien de tableaux  $T$  peuvent compléter un squelette donné ?



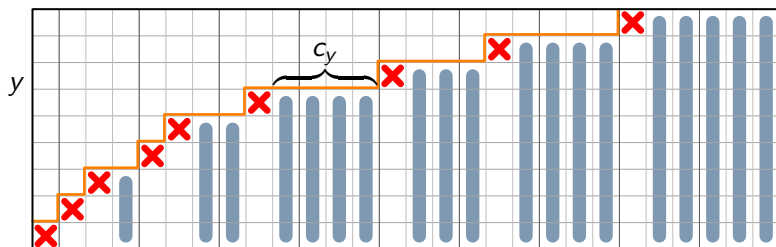
R: Code le squelette par la séquence  $\mathbf{c} = \{c_y\}_{1 \leq y \leq n}$ . C'est

$$\mu(\mathbf{c}) = \begin{cases} \prod_{y=1}^n y^{c_y} & \sum_y c_y = (k-1)n + 1 \\ 0 & \text{autrement} \end{cases}$$

Corollaire :  $\sum_T \phi(\mathbf{c}(T)) = \sum_{\mathbf{c}} \mu(\mathbf{c}) \phi(\mathbf{c})$ .

# Relations entre squelette et croix noires

**Q:** Combien de tableaux  $T$  peuvent compléter un squelette donné ?



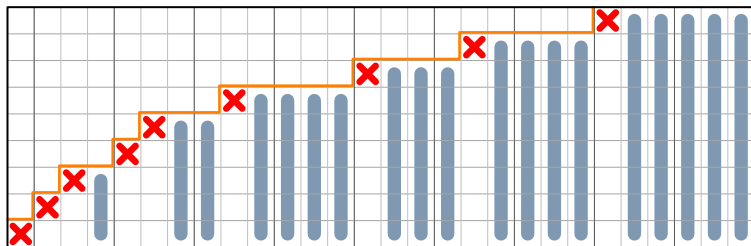
**R:** Code le squelette par la séquence  $\mathbf{c} = \{c_y\}_{1 \leq y \leq n}$ . C'est

$$\mu(\mathbf{c}) = \begin{cases} \prod_{y=1}^n y^{c_y} & \sum_y c_y = (k-1)n + 1 \\ 0 & \text{autrement} \end{cases}$$

**Corollaire :**  $\sum_T \phi(\mathbf{c}(T)) = \sum_{\mathbf{c}} \mu(\mathbf{c}) \phi(\mathbf{c})$ .

# Relations entre squelette et croix noires

**Q:** Combien de tableaux  $T$  peuvent compléter un squelette donné ?



**R:** Code le squelette par la séquence  $\mathbf{c} = \{c_y\}_{1 \leq y \leq n}$ . C'est

$$\mu(\mathbf{c}) = \begin{cases} \prod_{y=1}^n y^{c_y} & \sum_y c_y = (k-1)n + 1 \\ 0 & \text{autrement} \end{cases}$$

**Corollaire :**  $\sum_T \phi(\mathbf{c}(T)) = \sum_{\mathbf{c}} \mu(\mathbf{c}) \phi(\mathbf{c})$ .

# Enumérer les automates accessibles

Soit  $\mathcal{D}_{n,k}$  l'ensemble des structures de transition (déterministes complètes) accessibles, sur un alphabet de taille  $|\Sigma| = k \geq 2$ , avec  $|Q| = n$  états.

*Korshunov, 1978; Bassino et Nicaud, 2007; Lebensztayn, 2010.*

$$|\mathcal{D}_{n,k}| = \omega_k \left\{ \begin{matrix} kn+1 \\ n \end{matrix} \right\} (1 + \mathcal{O}(n^{-1})) ,$$

où  $\left\{ \begin{matrix} n \\ m \end{matrix} \right\}$  est le *nombre de Stirling de seconde espèce*, et  $\omega_k$  satisfait  $-k\omega_k = \ln(1 - \omega_k)$ .

*Bassino, Nicaud 2007*

En utilisant un générateur de Boltzmann, on peut tirer uniformément au hasard des automates déterministe, complets et accessibles à  $n$  états en temps moyen  $\Theta(n^{3/2})$ .

# Enumérer les automates accessibles

Soit  $\mathcal{D}_{n,k}$  l'ensemble des structures de transition (déterministes complètes) accessibles, sur un alphabet de taille  $|\Sigma| = k \geq 2$ , avec  $|Q| = n$  états.

*Korshunov, 1978; Bassino et Nicaud, 2007; Lebensztayn, 2010.*

$$|\mathcal{D}_{n,k}| = \omega_k \left\{ \begin{matrix} kn+1 \\ n \end{matrix} \right\} (1 + \mathcal{O}(n^{-1})) ,$$

où  $\left\{ \begin{matrix} n \\ m \end{matrix} \right\}$  est le *nombre de Stirling de seconde espèce*, et  $\omega_k$  satisfait  $-k\omega_k = \ln(1 - \omega_k)$ .

*Bassino, Nicaud 2007*

En utilisant un générateur de Boltzmann, on peut tirer uniformément au hasard des automates déterministe, complets et accessibles à  $n$  états en temps moyen  $\Theta(n^{3/2})$ .

## Résultats expérimentaux

Taille	10	100	500	1 000	5 000	10 000
$k = 2$ (%)	81.33	85.06	85.32	85.09	85.32	85,24
$k = 3$ (%)	96.92	99.78	99.99	99.98	99.99	99.99
$k = 4$ (%)	99.53	99.99	99.99	99.99	99.99	99.99

- ▶ Résultats obtenus avec la bibliothèque C++ REGAL développée par J. David
- ▶ 20 000 automates de chaque taille ont été engendrés.
- ▶ Générateur par rejet dont l'efficacité n'était pas prouvée théoriquement

# Proportion d'automates minimaux parmi les accessibles

On munit l'ensemble des automates accessibles  $A = (\delta, \mathcal{T})$  de la mesure  $\mu(\delta, \mathcal{T}) = \mu_{\text{uniforme}}(\delta) \times \mu_{\text{Bernoulli}}^{(b)}(\mathcal{T})$ .

*Bassino, David et Sportiello, 2011*

Quand  $n$  tend vers  $+\infty$ , la probabilité qu'un automate accessible  $A$  soit minimal tend vers

$$\exp\left(-\frac{1-2b(1-b)}{2} \omega_k^k n^{-k+2}\right).$$

En prenant  $b = 1/2$  (uniforme), on obtient, pour  $k \geq 2$  fixé, que le nombre d'automates minimaux à  $n$  états est asymptotiquement égal à

$$\exp\left(-\frac{1}{4} \omega_k^k n^{-k+2}\right) 2^n \omega_k \left\{ \begin{matrix} kn+1 \\ n \end{matrix} \right\} (1 + \mathcal{O}(n^{-1})).$$



# Proportion d'automates minimaux parmi les accessibles

On munit l'ensemble des automates accessibles  $A = (\delta, \mathcal{T})$  de la mesure  $\mu(\delta, \mathcal{T}) = \mu_{\text{uniforme}}(\delta) \times \mu_{\text{Bernoulli}}^{(b)}(\mathcal{T})$ .

*Bassino, David et Sportiello, 2011*

Quand  $n$  tend vers  $+\infty$ , la probabilité qu'un automate accessible  $A$  soit minimal tend vers

$$\exp\left(-\frac{1-2b(1-b)}{2} \omega_k^k n^{-k+2}\right).$$

En prenant  $b = 1/2$  (uniforme), on obtient, pour  $k \geq 2$  fixé, que le nombre d'automates minimaux à  $n$  états est asymptotiquement égal à

$$\exp\left(-\frac{1}{4} \omega_k^k n^{-k+2}\right) 2^n \omega_k \left\{ \begin{matrix} kn+1 \\ n \end{matrix} \right\} (1 + \mathcal{O}(n^{-1})).$$

# Proportion d'automates minimaux parmi les accessibles

On munit l'ensemble des automates accessibles  $A = (\delta, \mathcal{T})$  de la mesure  $\mu(\delta, \mathcal{T}) = \mu_{\text{uniforme}}(\delta) \times \mu_{\text{Bernoulli}}^{(b)}(\mathcal{T})$ .

*Bassino, David et Sportiello, 2011*

Quand  $n$  tend vers  $+\infty$ , la probabilité qu'un automate accessible  $A$  soit minimal tend vers

$$\exp\left(-\frac{1-2b(1-b)}{2} \omega_k^k n^{-k+2}\right).$$

En prenant  $b = 1/2$  (uniforme), on obtient, pour  $k \geq 2$  fixé, que le nombre d'automates minimaux à  $n$  états est asymptotiquement égal à

$$\exp\left(-\frac{1}{4} \omega_k^k n^{-k+2}\right) 2^n \omega_k \left\{ \begin{matrix} kn+1 \\ n \end{matrix} \right\} (1 + \mathcal{O}(n^{-1})).$$

# Énumération des minimaux

*Bassino, David et Sportiello, 2011*

Pour  $n$  grand, la probabilité que  $A$  soit minimal converge vers

$$\exp\left(-\frac{1-2b(1-b)}{2} \omega_k^k n^{-k+2}\right).$$

Certains motifs, dans les graphes orientés correspondant aux structures de transition, ont un rôle central dans la preuve.

$M$        $i$        $j$

$M^{(3)}$        $i$        $j$        $h$

$l_1$        $l_2 \dots l_k$

$l_1$        $l_2 \dots l_k$

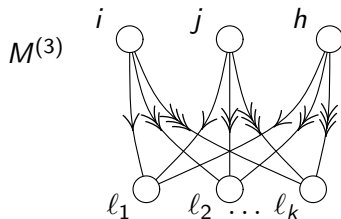
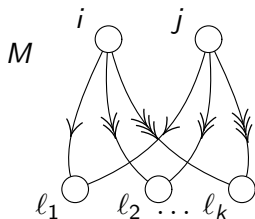
# Énumération des minimaux

*Bassino, David et Sportiello, 2011*

Pour  $n$  grand, la probabilité que  $A$  soit minimal converge vers

$$\exp\left(-\frac{1-2b(1-b)}{2} \omega_k^k n^{-k+2}\right).$$

Certains **motifs**, dans les graphes orientés correspondant aux structures de transition, ont un rôle central dans la preuve.



# Probabilités qui interviennent dans la preuve

On va définir les quantités

- ▶  $P_{\text{rare}}$  : probabilité que  $A$  ne contienne pas de motifs  $M$ , et pourtant l'automate n'est pas minimal ;
- ▶  $P_{\text{confl}}$  : probabilité que  $A$  contienne des motifs  $M^{(3)}$  ;
- ▶  $P(r)$  : probabilité que  $A$  ne contienne pas des motifs  $M^{(3)}$ , et qu'il contienne exactement  $r$  motifs  $M$ .

Notons que, pour l'ensemble Bernoulli des états terminaux,

$$\mathbb{P}[A \text{ n'est pas minimal}] = \sum_{r \geq 0} (1 - (2b(1-b)))^r P(r) + \mathcal{O}(P_{\text{rare}}, P_{\text{confl}})$$

Si on prouve que  $P_{\text{rare}}, P_{\text{confl}} = o(1 - P(0))$

et que  $P(r) = \text{Poisson}_{\rho}(r)(1 + o(1))$  un certain  $\rho$ ,

$$\mathbb{P}[A \text{ n'est pas minimal}] = (1 - \exp(-\rho(1 - 2b(1 - b))))(1 + o(1)).$$

# Probabilités qui interviennent dans la preuve

On va définir les quantités

- ▶  $P_{\text{rare}}$  : probabilité que  $A$  ne contienne pas de motifs  $M$ , et pourtant l'automate n'est pas minimal ;
- ▶  $P_{\text{confl}}$  : probabilité que  $A$  contienne des motifs  $M^{(3)}$  ;
- ▶  $P(r)$  : probabilité que  $A$  ne contienne pas des motifs  $M^{(3)}$ , et qu'il contienne exactement  $r$  motifs  $M$ .

Notons que, pour l'ensemble Bernoulli des états terminaux,

$$\mathbb{P}[A \text{ n'est pas minimal}] = \sum_{r \geq 0} (1 - (2b(1-b)))^r P(r) + \mathcal{O}(P_{\text{rare}}, P_{\text{confl}})$$

Si on prouve que  $P_{\text{rare}}, P_{\text{confl}} = o(1 - P(0))$

et que  $P(r) = \text{Pois}_\rho(r)(1 + o(1))$  un certain  $\rho$ ,

$$\mathbb{P}[A \text{ n'est pas minimal}] = (1 - \exp(-\rho(1 - 2b(1-b))))(1 + o(1)).$$

📖 N. Alon et J. Spencer, *The probabilistic method*, sect. 8.3

Soit  $\{\mathcal{E}_j\}$  une collection d'évènements. On cherche  $p(r) = \mathbb{P}[r \text{ des } \mathcal{E}_j \text{ sont vrais}]$ . Soit

$$S^{(r)} = \sum_{(i_1, \dots, i_r)} \mathbb{P}[\mathcal{E}_{i_1} \wedge \dots \wedge \mathcal{E}_{i_r}]$$

## Crible de Brun

S'il existe une constante  $\mu$  tel que, pour  $n$  grand, et chaque  $r$ ,

$$S^{(r)} \rightarrow \frac{1}{r!} \mu^r$$

alors

$$p(r) \rightarrow \exp(-\mu) \frac{1}{r!} \mu^r .$$

Par le crible de Brun, il nous suffira de montrer que

- ▶ **Estimation des motifs** :

$$S^{(r)} \rightarrow \rho^r / r!, \text{ pour } \rho = \frac{1}{2} \omega_k^k n^{-k+2},$$

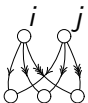
et

$$P_{\text{confl}} \rightarrow \frac{1}{6} \omega_k^{2k} n^{-2k+3}$$

- ▶ **Estimation de la “boîte noire”** :

$$P_{\text{rare}} = o(n^{-k+2}).$$



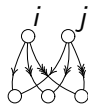
**Q** : Quelle est la probabilité d'avoir un motif  (avec  $i < j$ ), conditionné à un squelette  $\mathbf{c}$ ?

**R** : Zéro, s'il y a des croix rouges dans les colonnes de  $j$ ,  $y^{-k}$  si ce n'est pas le cas.

$$\mathbb{P}(M_{ij}|\mathbf{c}) \mu(\mathbf{c}) \equiv \mu(\mathbf{c}') \quad \mathbf{c}' = \mathbf{c} \setminus \text{colonnes de } j$$

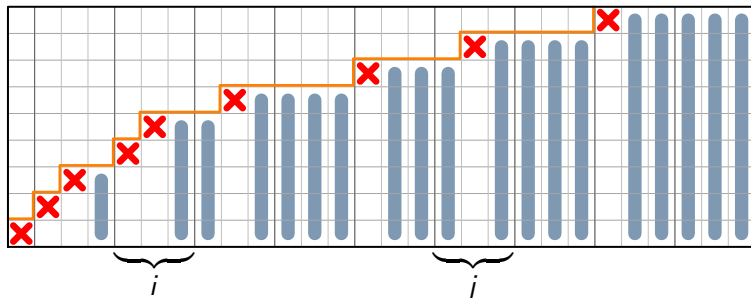
# Au sujet de $P(r)$ et $P_{\text{confl}}$

**Q** : Quelle est la probabilité d'avoir un motif **conditionné** à un squelette  $\mathbf{c}$ ?



(avec  $i < j$ ),

**R** : **Zéro**, s'il y a des croix rouges dans les colonnes de  $j$ ,  
 $y^{-k}$  si ce n'est pas le cas.

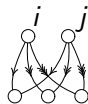


$$\mathbb{P}(M_{ij} | \mathbf{c}) \mu(\mathbf{c}) \equiv \mu(\mathbf{c}')$$

$$\mathbf{c}' = \mathbf{c} \setminus \text{colonnes de } j$$

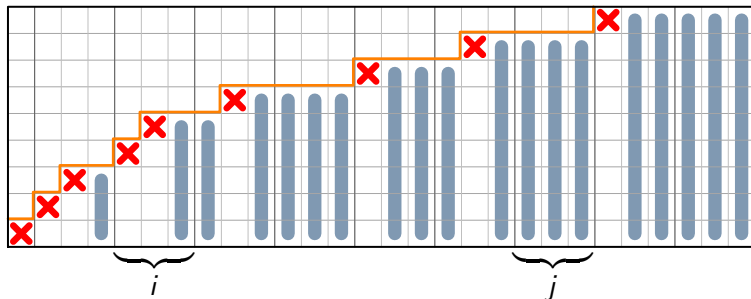
# Au sujet de $P(r)$ et $P_{\text{conf}}$

**Q** : Quelle est la probabilité d'avoir un motif **conditionné** à un squelette  $\mathbf{c}$ ?



(avec  $i < j$ ),

**R** : **Zéro**, s'il y a des croix rouges dans les colonnes de  $j$ ,  
 $y^{-k}$  si ce n'est pas le cas.

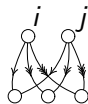


$$\mathbb{P}(M_{ij} | \mathbf{c}) \mu(\mathbf{c}) \equiv \mu(\mathbf{c}')$$

$$\mathbf{c}' = \mathbf{c} \setminus \text{colonnes de } j$$

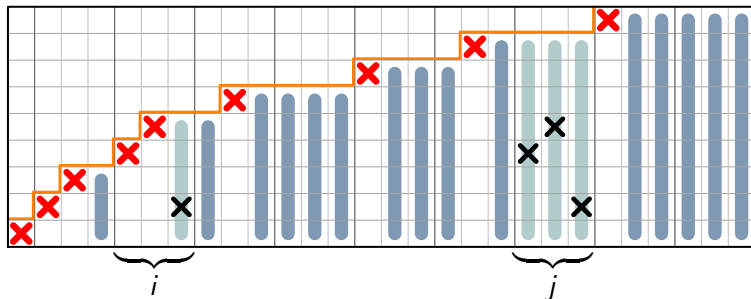
# Au sujet de $P(r)$ et $P_{\text{confl}}$

**Q** : Quelle est la probabilité d'avoir un motif **conditionné** à un squelette  $\mathbf{c}$ ?



(avec  $i < j$ ),

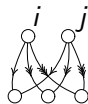
**R** : **Zéro**, s'il y a des croix rouges dans les colonnes de  $j$ ,  
 $y^{-k}$  si ce n'est pas le cas.



$$\mathbb{P}(M_{ij} | \mathbf{c}) \mu(\mathbf{c}) \equiv \mu(\mathbf{c}')$$

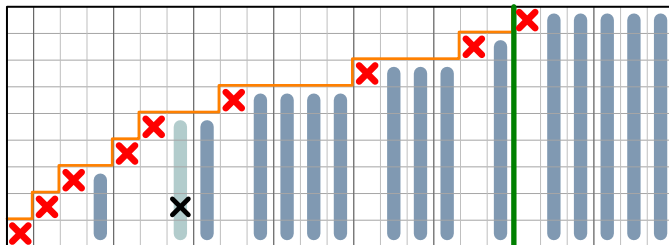
$$\mathbf{c}' = \mathbf{c} \setminus \text{colonnes de } j$$

**Q** : Quelle est la probabilité d'avoir un motif **conditionné** à un squelette  $\mathbf{c}$ ?



(avec  $i < j$ ),

**R** : **Zéro**, s'il y a des croix rouges dans les colonnes de  $j$ ,  $y^{-k}$  si ce n'est pas le cas.



$$\mathbb{P}(M_{ij} | \mathbf{c}) \mu(\mathbf{c}) \equiv \mu(\mathbf{c}') \quad \mathbf{c}' = \mathbf{c} \setminus \text{colonnes de } j$$

*Bassino-David-Sportiello 2012*

On peut engendrer aléatoirement des automates minimaux en temps moyen  $\mathcal{O}(n^{3/2})$ , avec un générateur de Boltzmann.

- ▶ Rejet à partir du générateur aléatoire d'automates accessibles dont la complexité est en  $\mathcal{O}(n^{3/2})$  (Bassino-Nicaud 2007).
- ▶ Comme il y a une proportion constante d'automates déterministes, le rejet moyen est en  $\mathcal{O}(1)$ .

## *Bassino-David-Sportiello 2012*

La complexité moyenne de la minimisation avec l'algorithme de Moore est en  $\mathcal{O}(n \log \log n)$  pour la distribution uniforme sur les automates déterministes complets et accessibles.

- ▶ La complexité de l'algorithme de minimisation de Moore pour la distribution uniforme sur les automates déterministes à  $n$  états est  $\mathcal{O}(n \log \log n)$  (*David 2011*).
- ▶ La borne supérieure vient de la distribution de la taille de la partie accessible d'une  $k$ -application aléatoire (*Carayol-Nicaud 2012*)
- ▶ La borne inférieure de l'algorithme de Moore appliquée aux automates minimaux à  $n$  états est  $\Omega(n \log \log n)$  (*Bassino-David- Nicaud 2009*).

# Analyse en moyenne de la minimisation (Hopcroft)

*Bassino-David-Sportiello 2012*

Pour la distribution uniforme sur les automates déterministes complets et accessibles, il existe une famille d'implantations de l'algorithme de Hopcroft dont la complexité moyenne est  $\mathcal{O}(n \log \log n)$ .

Il existe une famille d'implantations de l'algorithme de minimisation de Hopcroft qui sont toujours plus rapides que l'algorithme de Moore. (*David 2011*).



# Que reste-t-il à faire ?

- ▶ Extension au cas des automates incomplets
- ▶ Extension au cas où  $\mathbb{E}[|\mathcal{T}|] \rightarrow 0$  (par exemple,  $\mathbb{E}[|\mathcal{T}|] \sim n^\gamma$ ,  $0 < \gamma < 1$ ).
- ▶ Calcul de propriétés statistiques plus subtiles, en particulier la distribution de la **profondeur de l'arbre de recherche** (à partir de l'état initial).