Combinatoire des automates

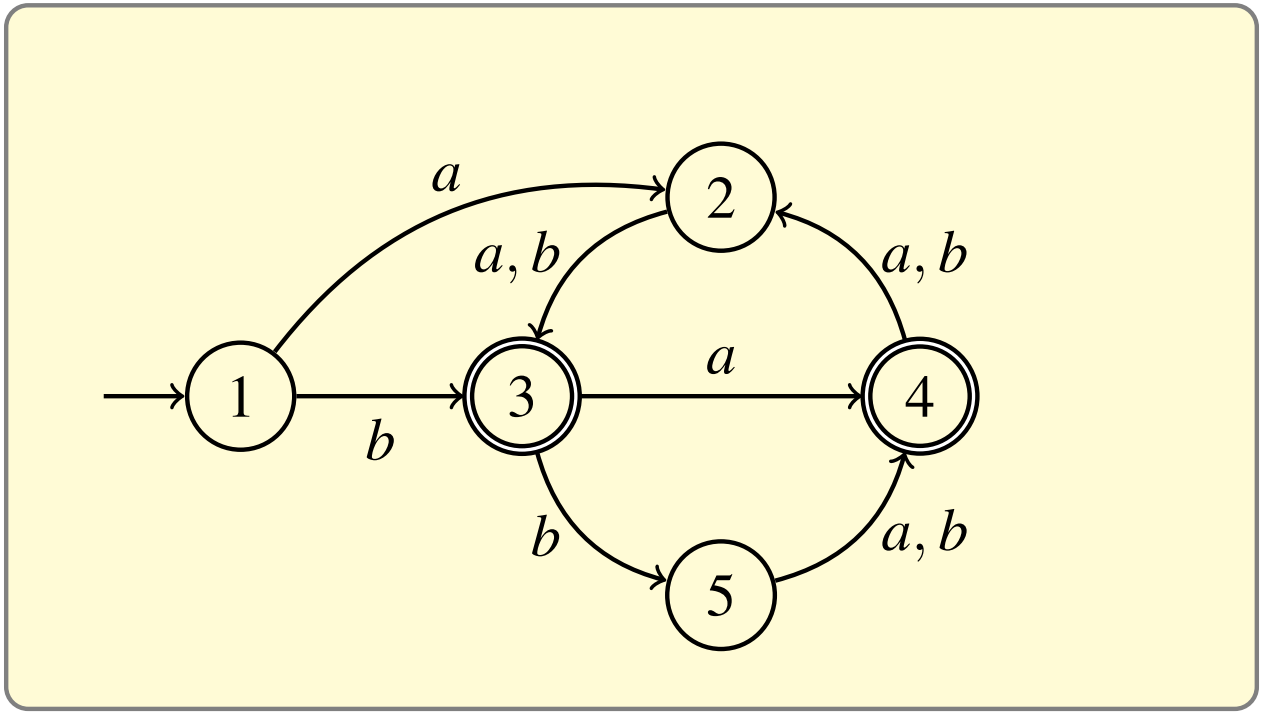
Cyril Nicaud

LIGM, Paris-Est, Marne-la-Vallée

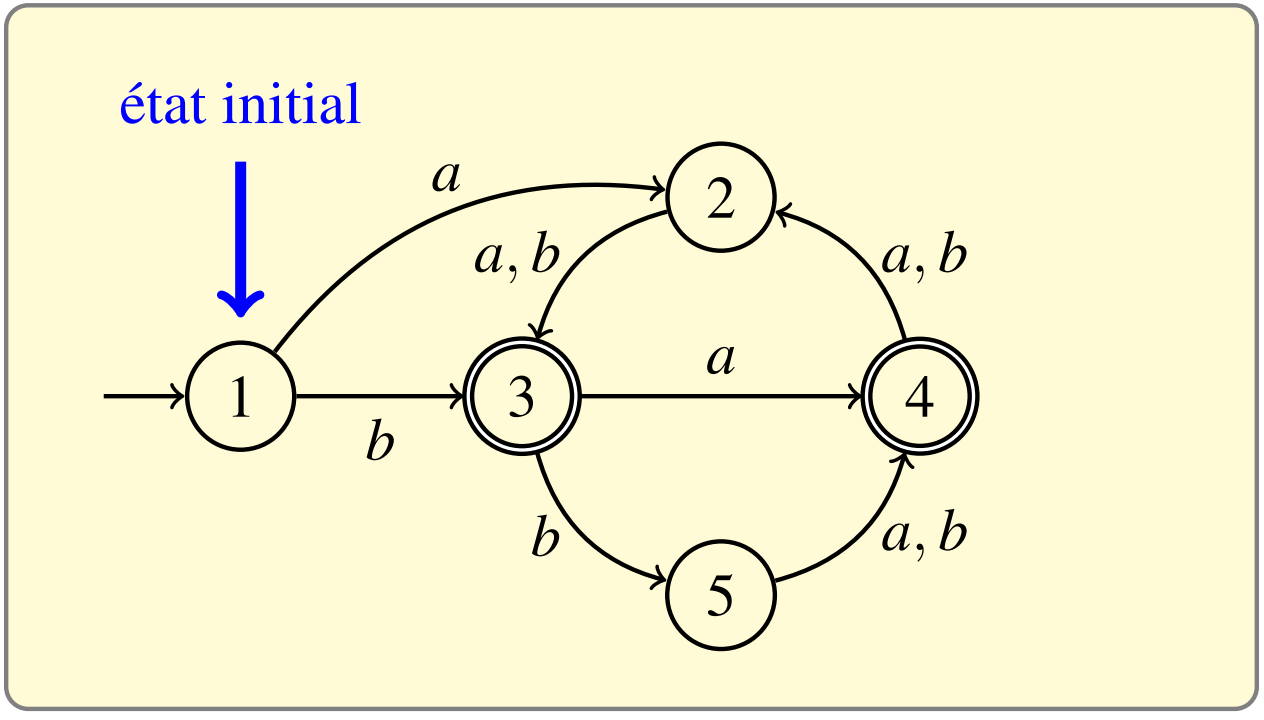
IHP le 7 avril 2011

Partie I : Définitions

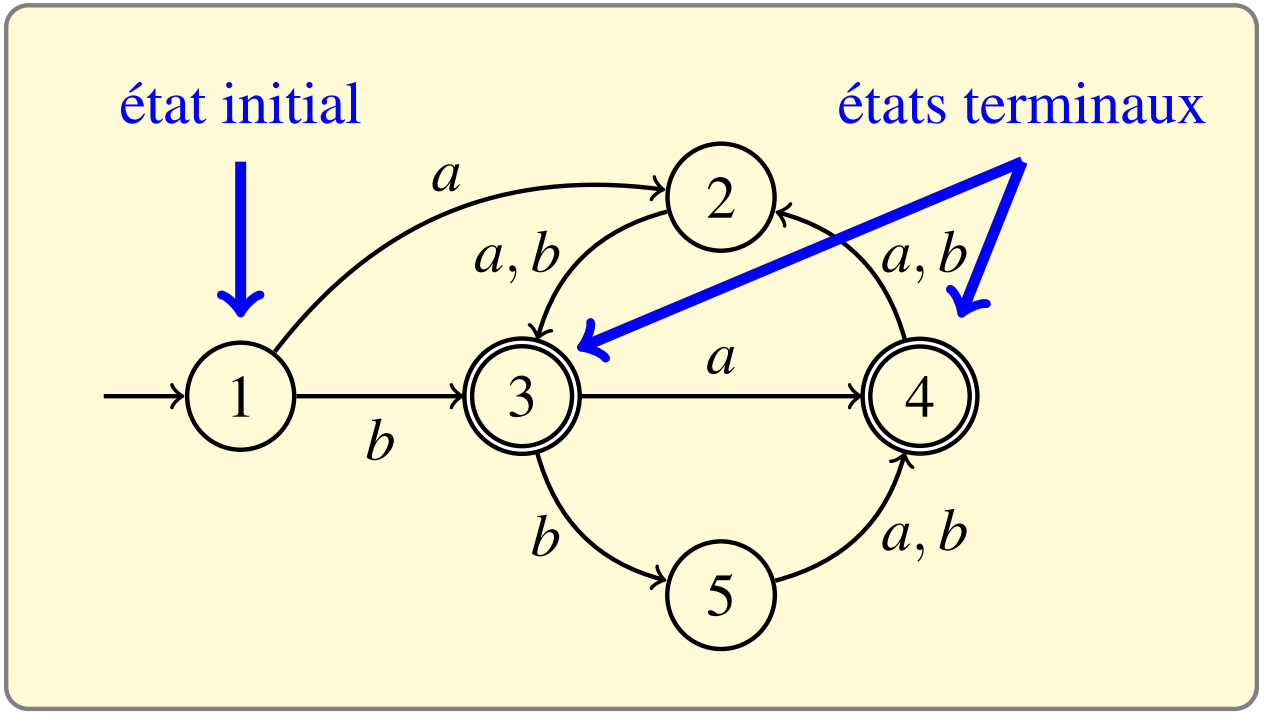
Automate déterministe et complet



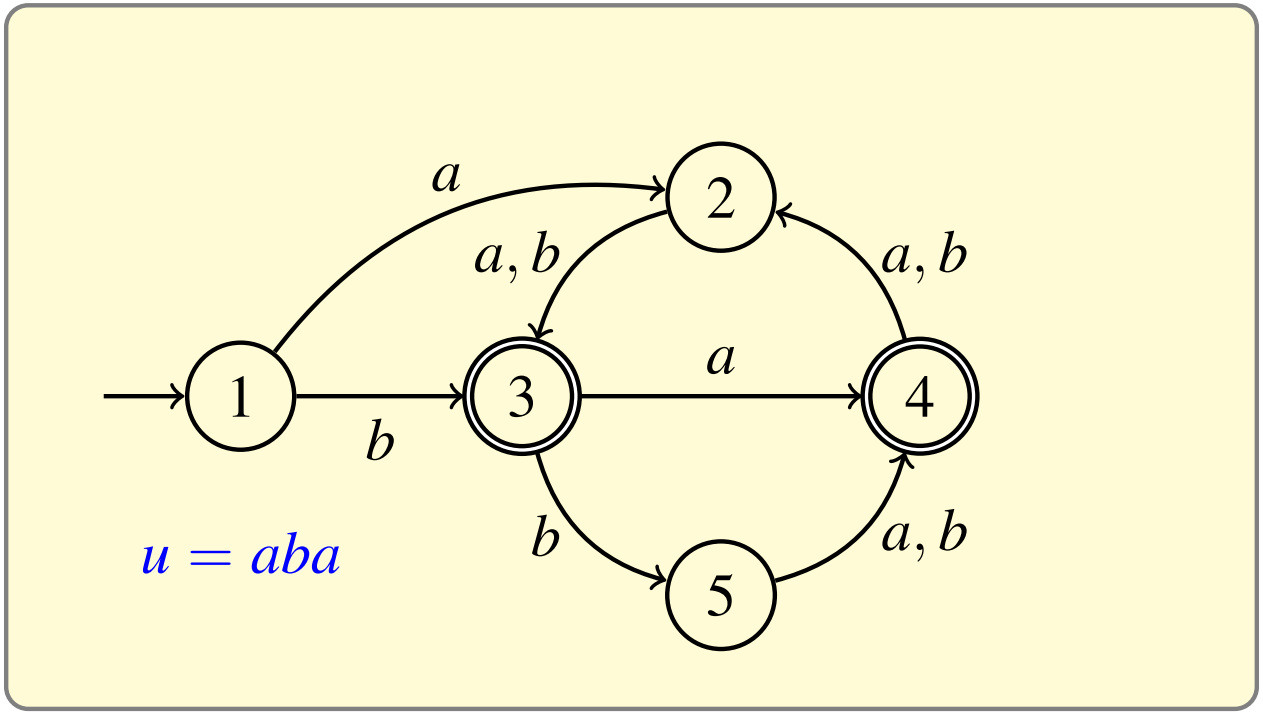
Automate déterministe et complet



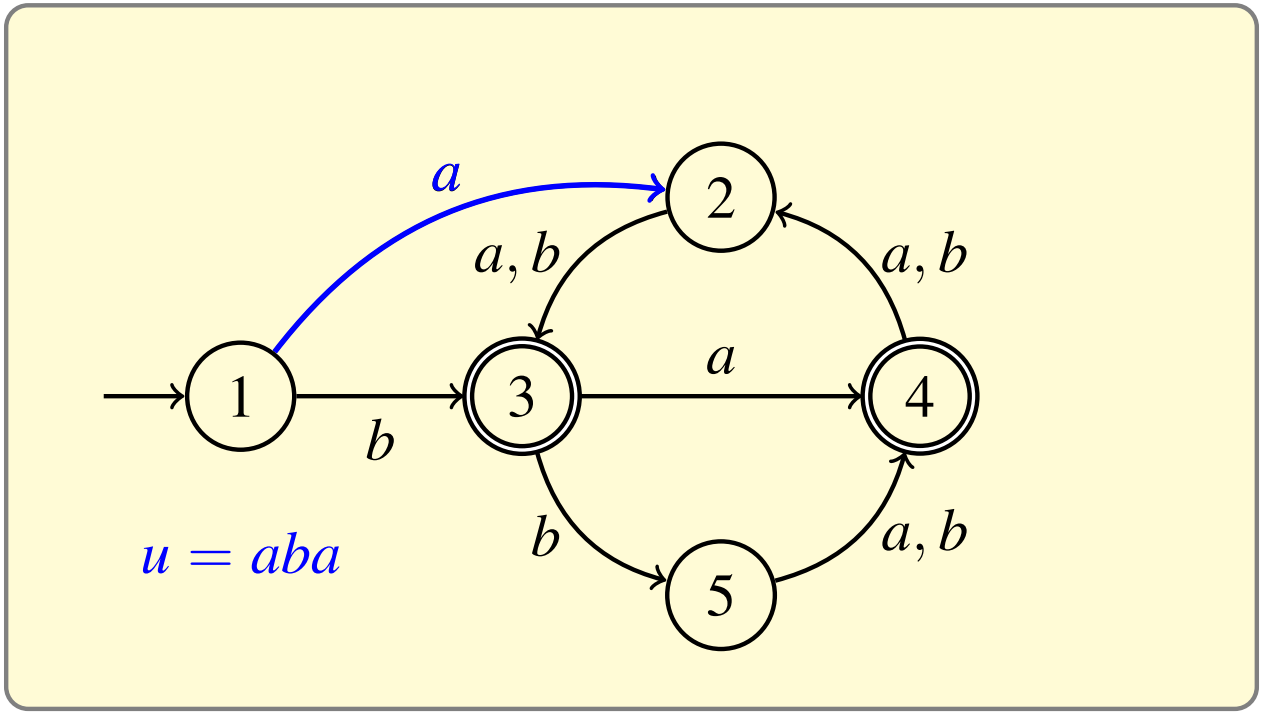
Automate déterministe et complet



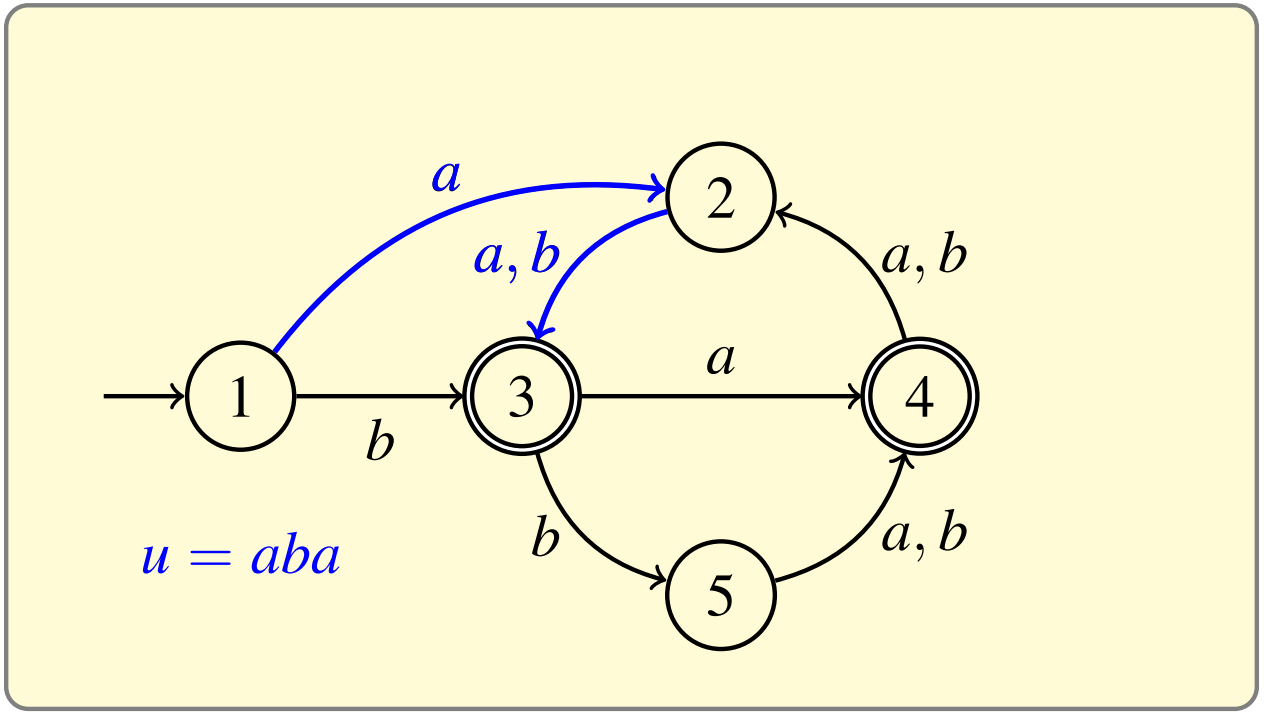
Automate déterministe et complet



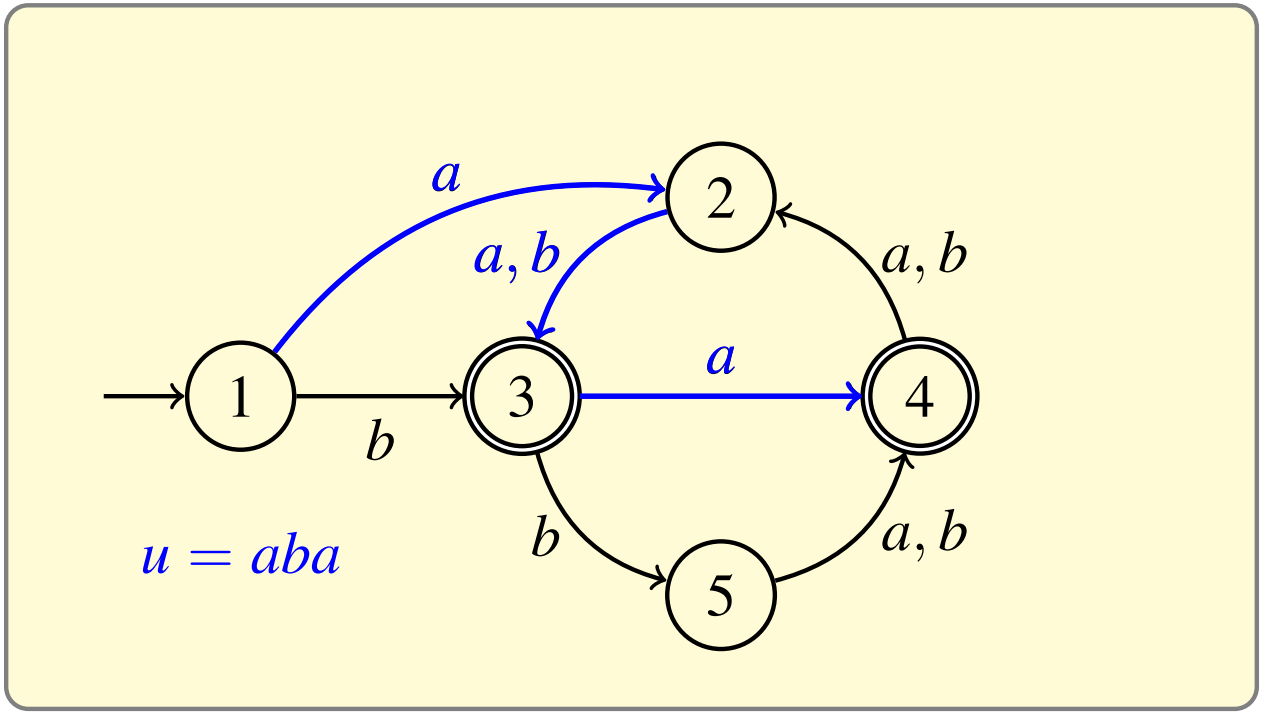
Automate déterministe et complet



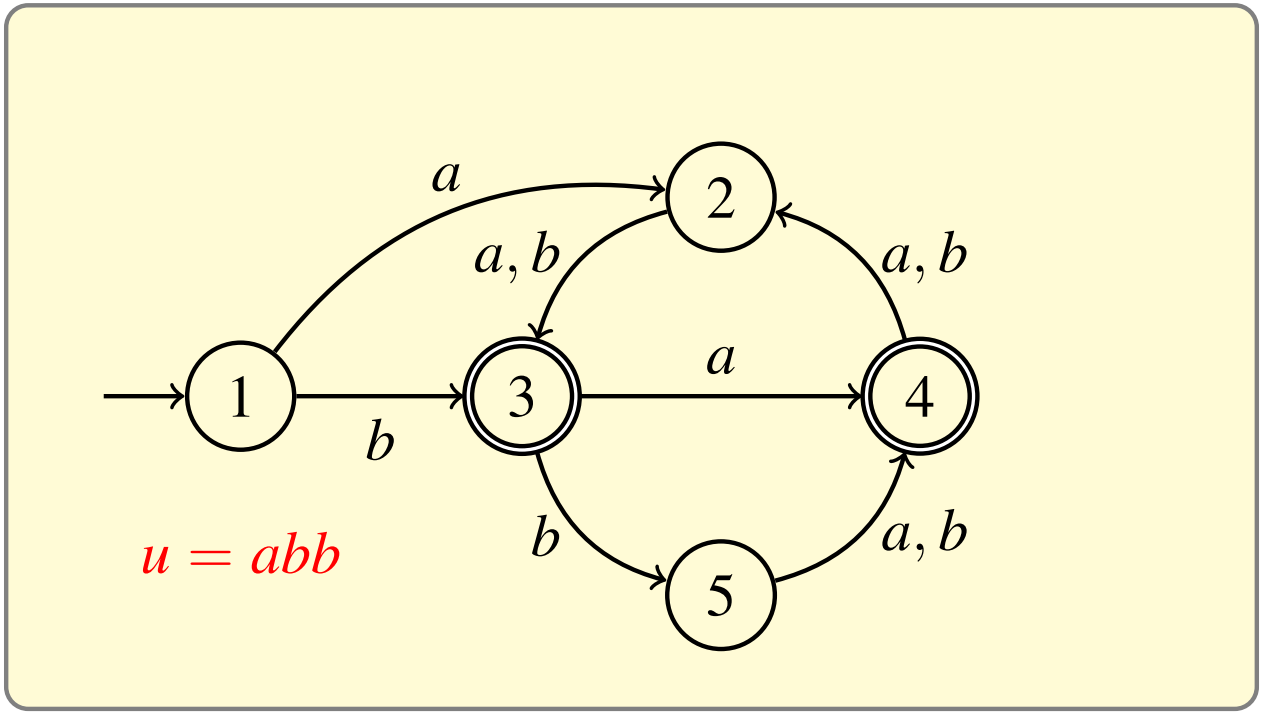
Automate déterministe et complet



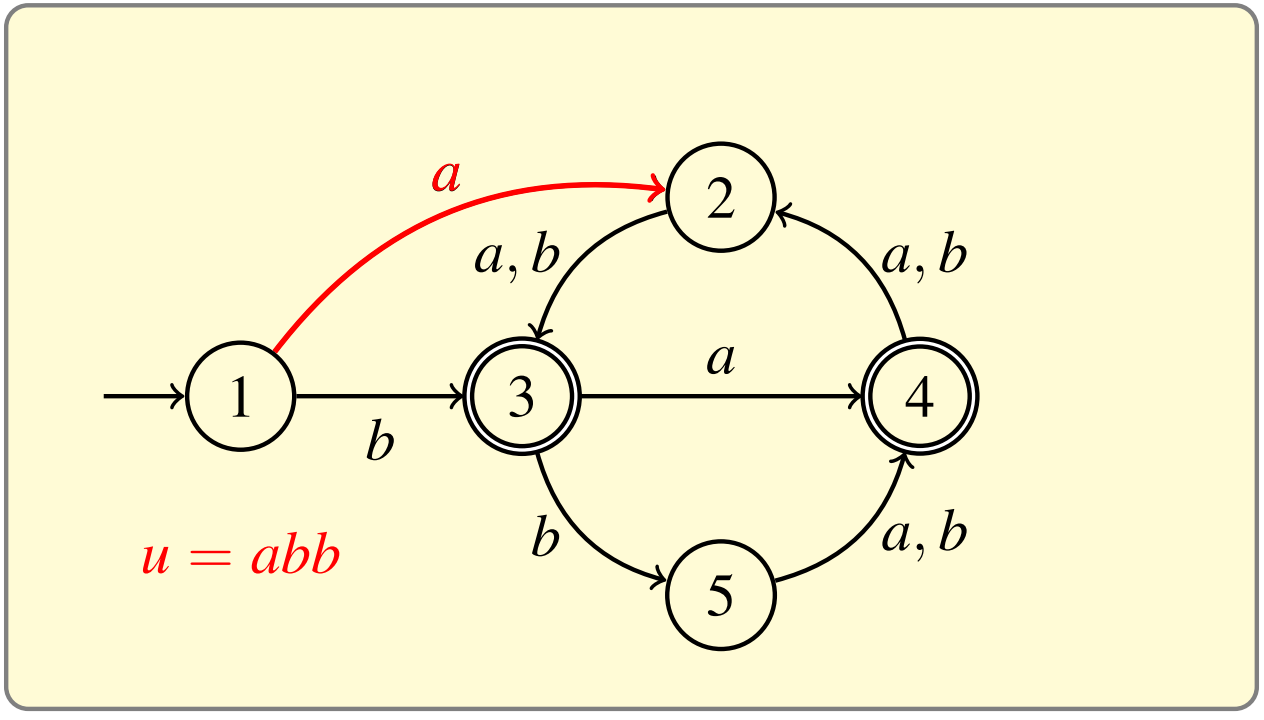
Automate déterministe et complet



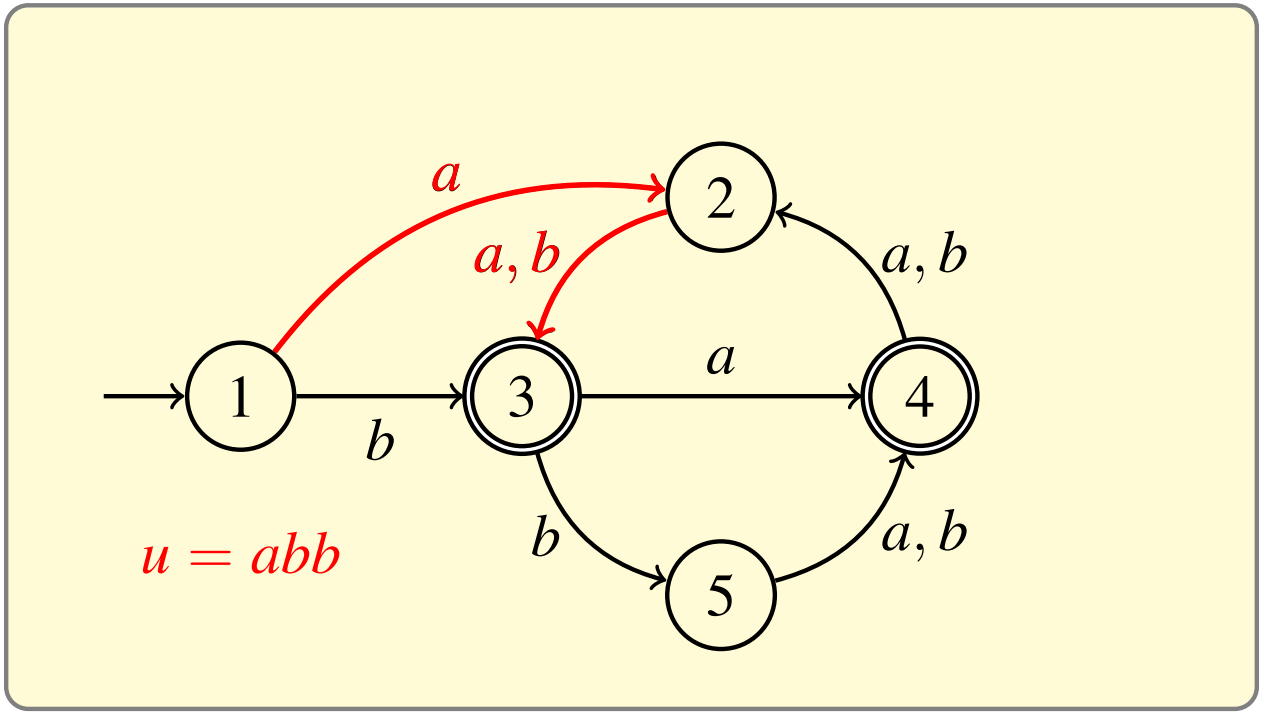
Automate déterministe et complet



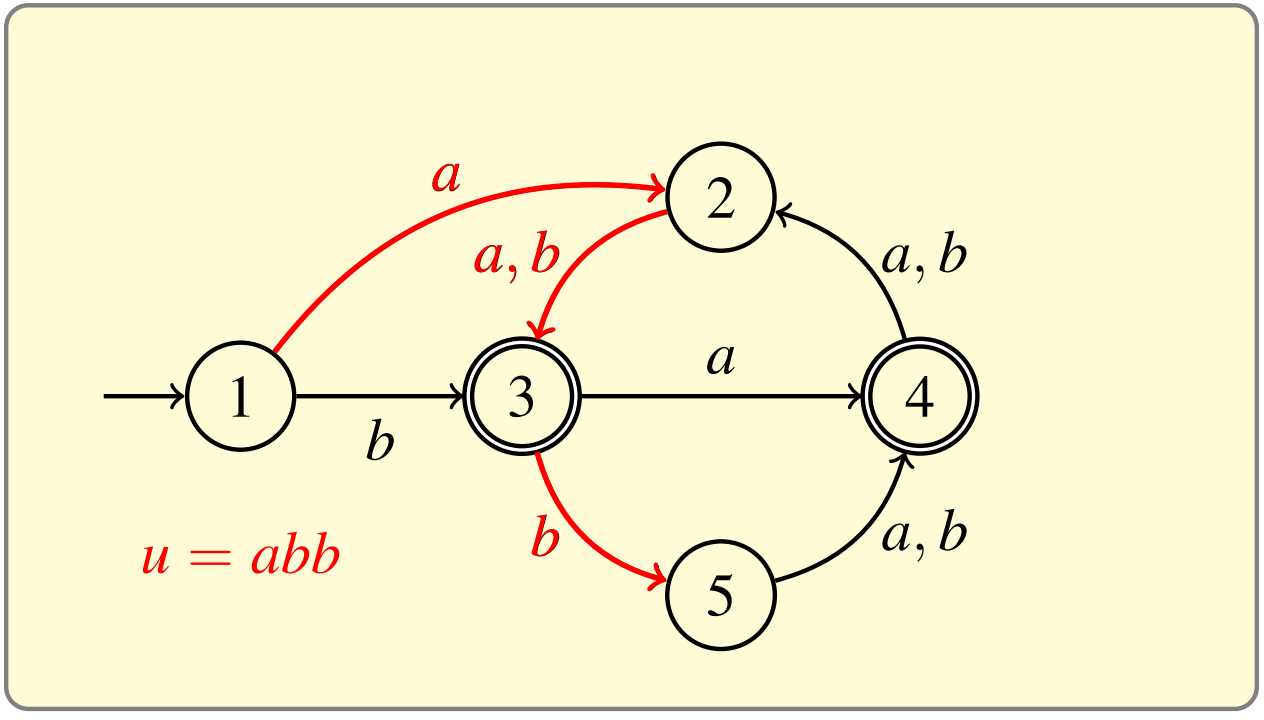
Automate déterministe et complet



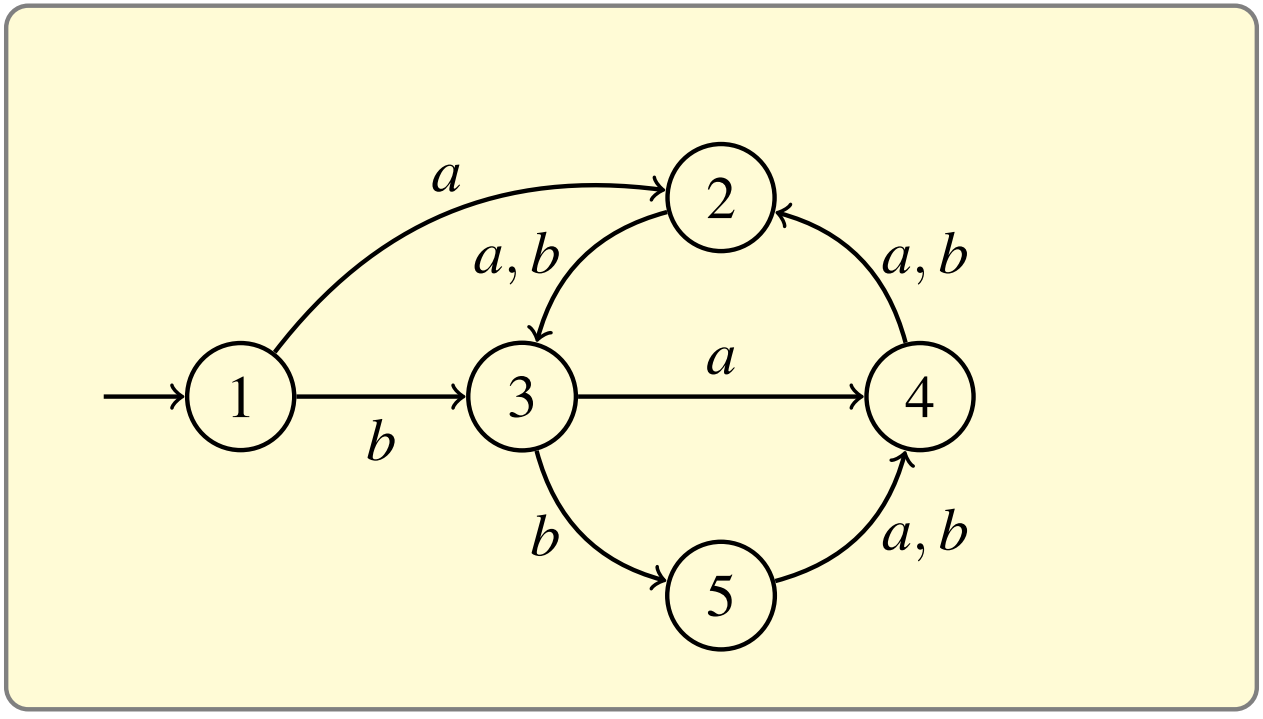
Automate déterministe et complet



Automate déterministe et complet

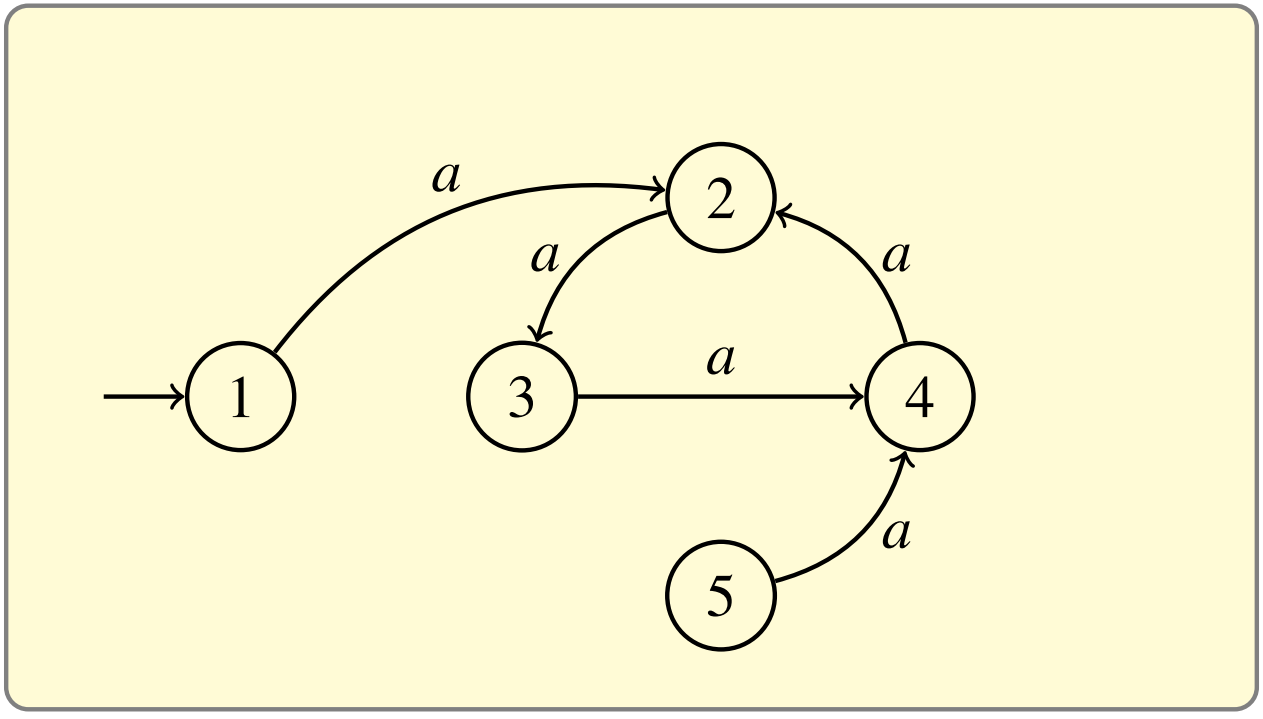


Automate



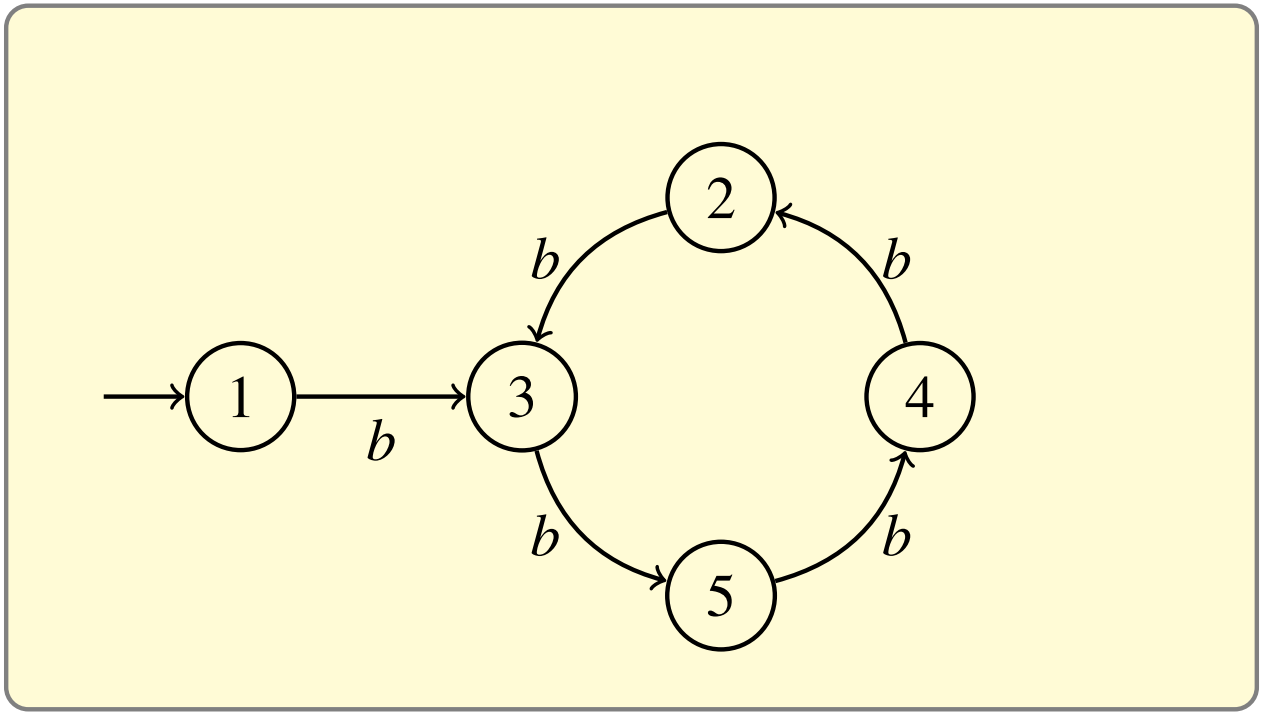
- On oublie les états terminaux pour cet exposé
- On veut que les états soient accessibles

Automate



- Chaque lettre est une application de $[n]$ dans $[n]$
- n est le nombre d'états (paramètre), k le nombre de lettres (fixé)

Automate



- Chaque lettre est une application de $[n]$ dans $[n]$
- n est le nombre d'états (paramètre), k le nombre de lettres (fixé)

Automate

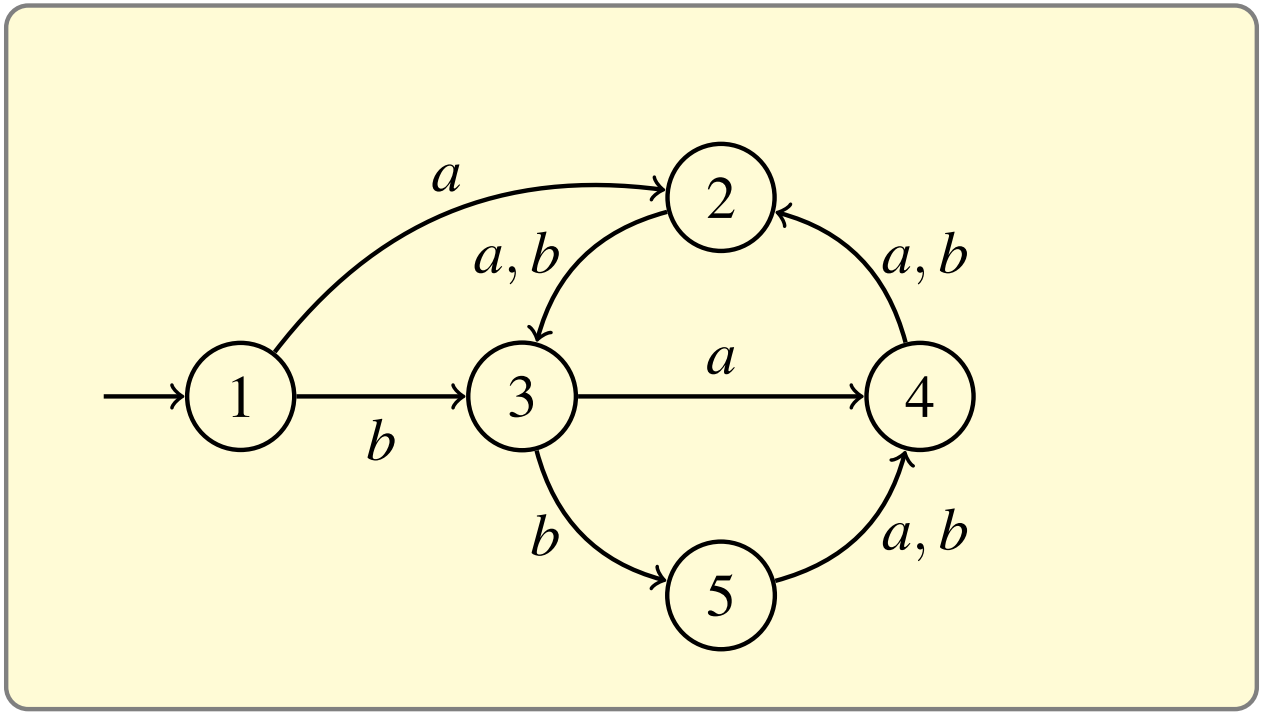
- La taille de l'alphabet k est fixée
- Il y a n états
- Chaque lettre est une application de $[n]$ dans $[n]$
- L'état initial est 1
- L'automate est accessible

Automate

- La taille de l'alphabet k est fixée
- Il y a n états
- Chaque lettre est une application de $[n]$ dans $[n]$
- L'état initial est 1
- L'automate est accessible

- Enumération (asymptotique)
- Génération aléatoire

Remarque : pas de symétrie



- Il n'y a pas d'automorphisme non trivial
- Il y a $(n-1)!$ façons d'étiqueter un automate

Partie II : Première tentative

Génération aléatoire sans accessibilité

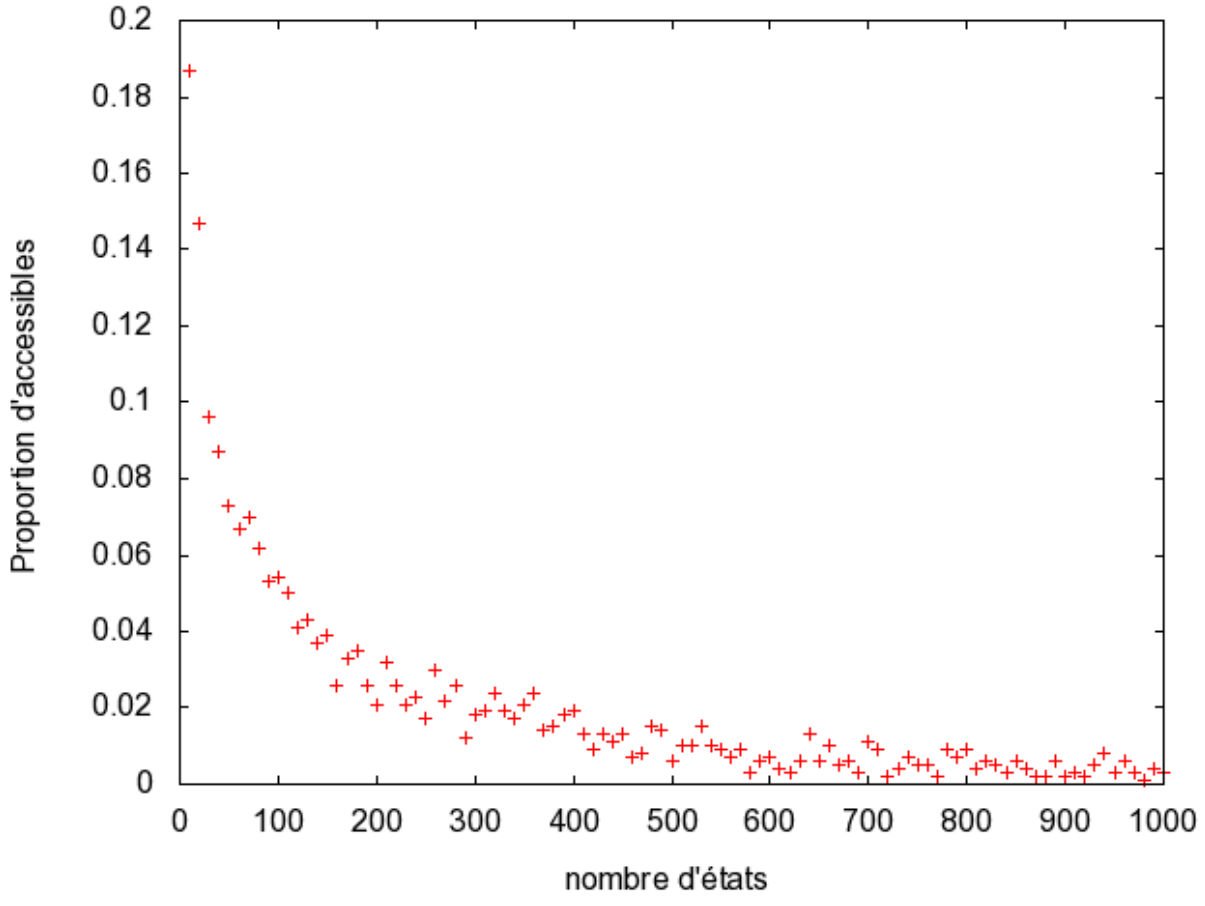
RandomMap(n)

for $i \in [n]$ **do**
  $R[i] = \text{Uniforme}([n])$
return R

RandomAutomaton(n, k)

for $a \in A$ **do**
  $\mathcal{A}[a] =$ **RandomMap**(n)
return $\mathcal{A}$

Proportion d'accessibles



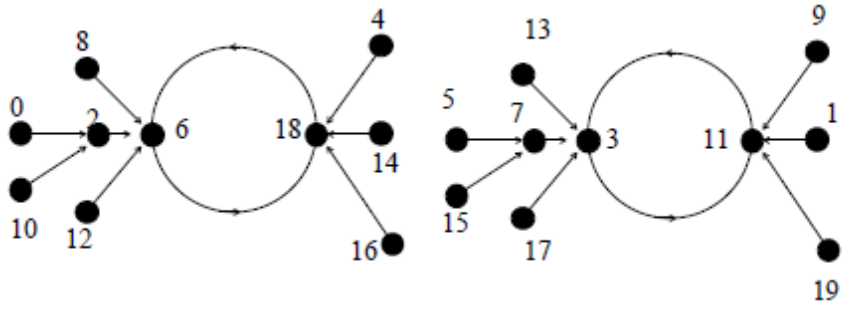
Random Mapping Statistics

Advances in Cryptology, Proc. Eurocrypt'89 :

RANDOM MAPPING STATISTICS

Philippe Flajolet
INRIA Rocquencourt,
F–78150 Le Chesnay (France)

Andrew M. Odlyzko
AT&T Bell Laboratories,
Murray Hill, NJ 07974 (USA)



Random Mapping Statistics

- Random Mapping : ensemble de cycles d'arbres de Cayley
- Arbres de Cayley : $T(z) = z \cdot \exp(T(z))$
- Random Mapping :

$$R(z) = \sum_{n \geq 0} \frac{n^n}{n!} z^n = \frac{1}{1 - T(z)}$$

Random Mapping Statistics

- Nombre de feuilles : $T(z, u) = z \cdot \exp(T(z, u)) + (u - 1)z$
- Nombre de feuilles dans un random mapping :

$$R(z, u) = \frac{1}{1 - z \cdot \exp\left(T(z, u)\right)}$$

- Nombre moyen d'éléments sans antécédent équivalent à $\frac{1}{e}n$

Random Mapping Statistics

- Nombre de feuilles : $T(z, u) = z \cdot \exp(T(z, u)) + (u - 1)z$
- Nombre de feuilles dans un random mapping :

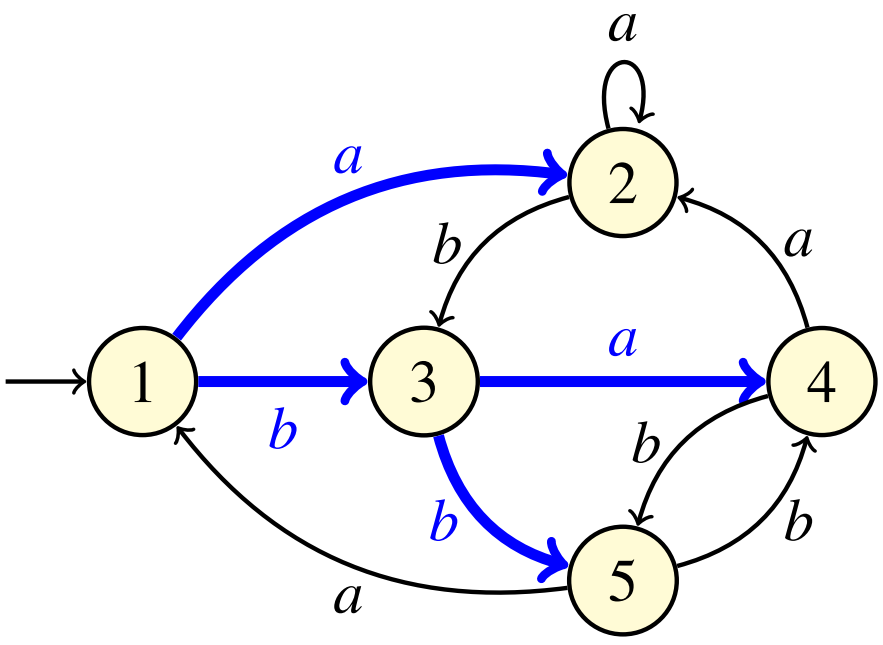
$$R(z, u) = \frac{1}{1 - z \cdot \exp\left(T(z, u)\right)}$$

- Nombre moyen d'éléments sans antécédent équivalent à $\frac{1}{e}n$

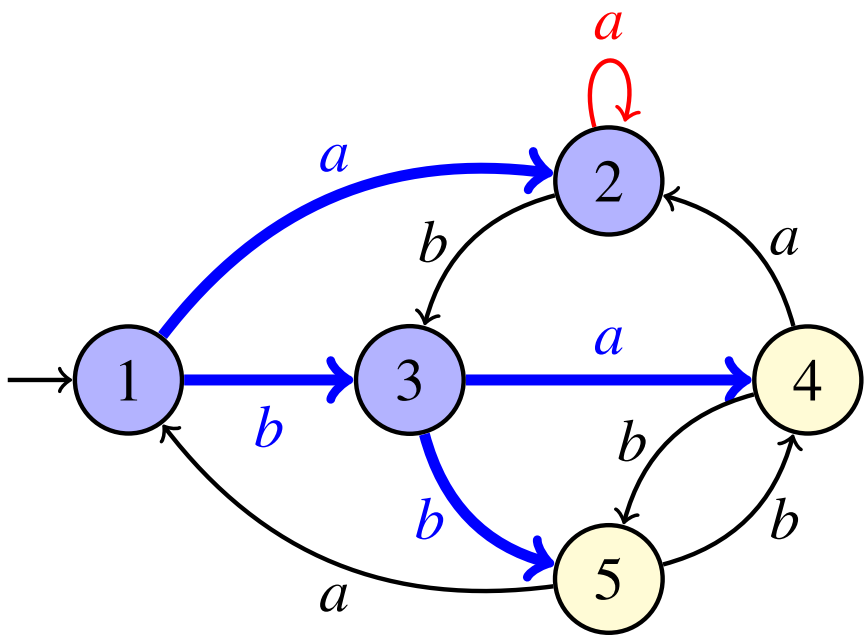
- Dans l'automate, la probabilité qu'un état n'ait pas de transition entrante est de l'ordre $\frac{1}{e^k}$
- Grandes déviations $\Rightarrow$ proportion exponentiellement faible d'accessibles

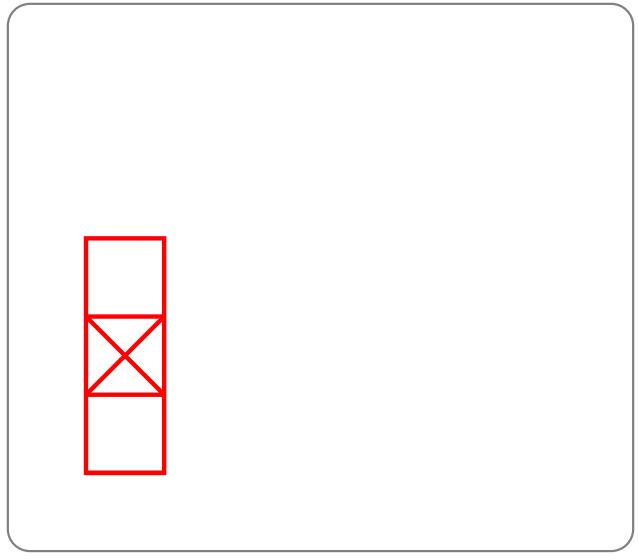
Partie III : Tableaux

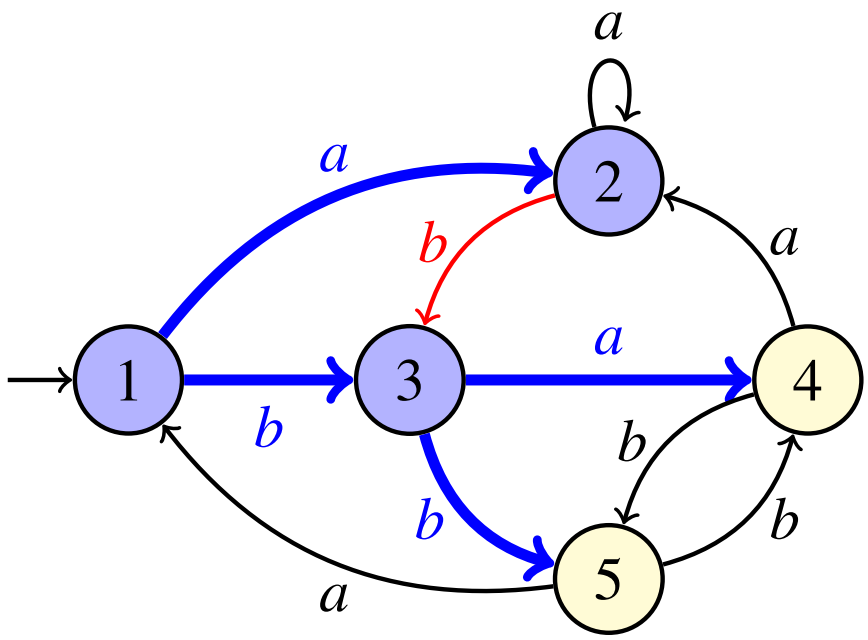
Etiquetage

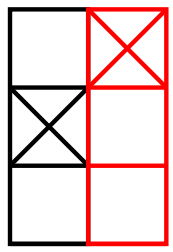


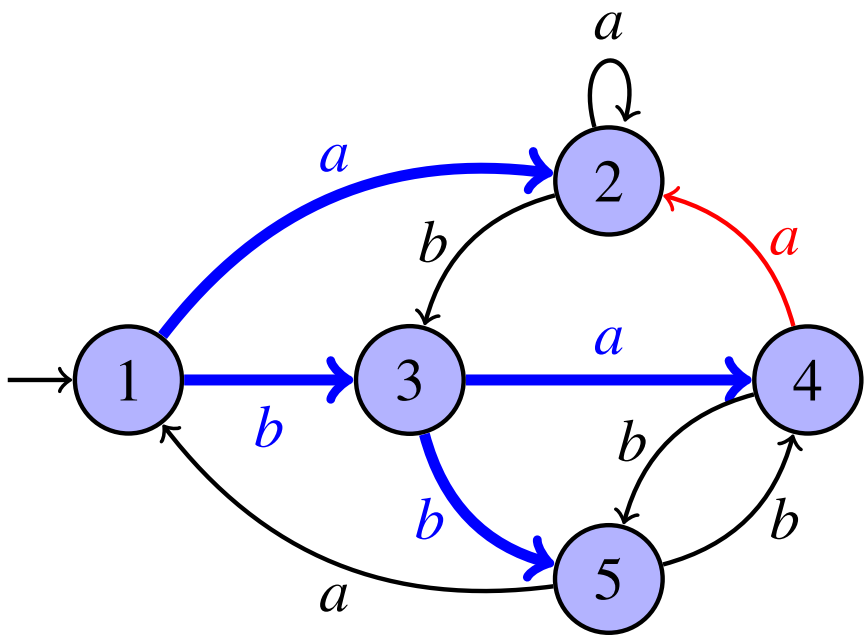
- On ordonne l'alphabet : $a < b$
- On numérote les sommets par ordre de visite lors d'un parcours en largeur

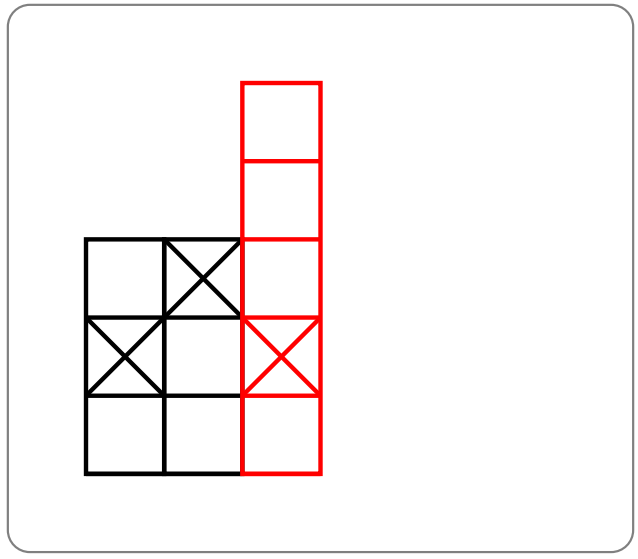

a
a
2
b
a
1
3
a
4
b
b
b
5
a

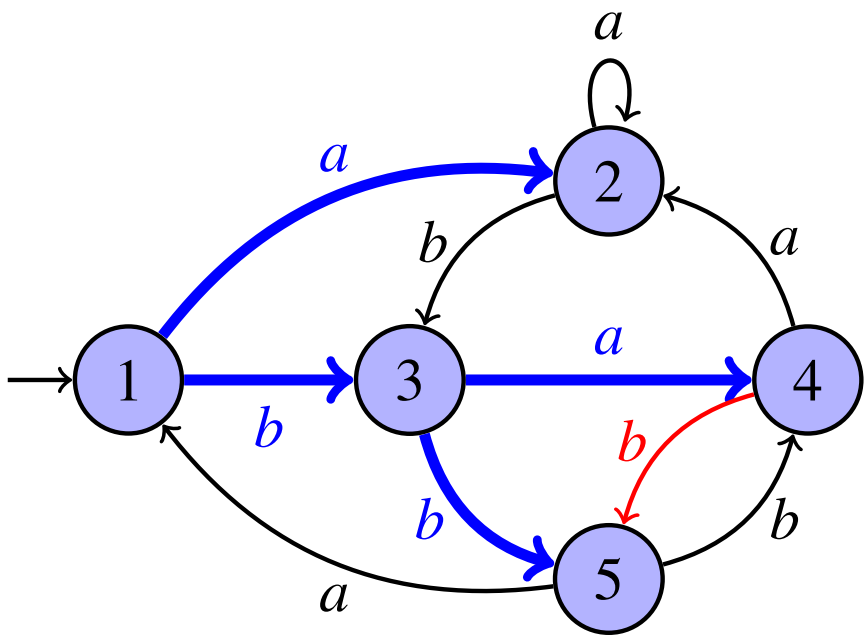


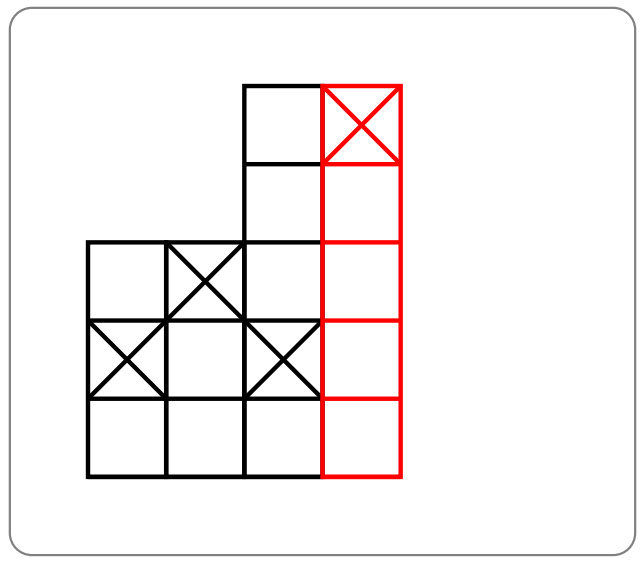

a
a
2
b
a
1
3
a
4
b
b
b
5
a

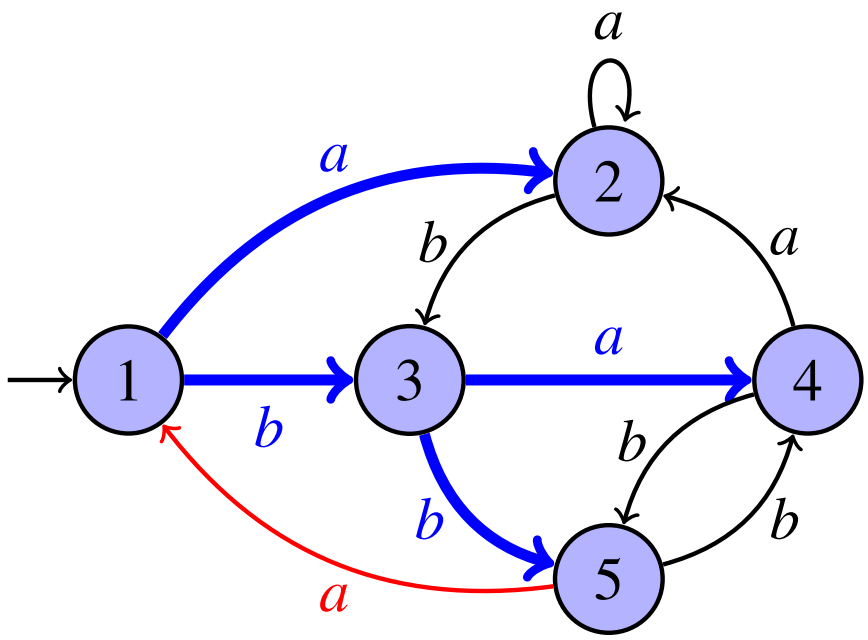


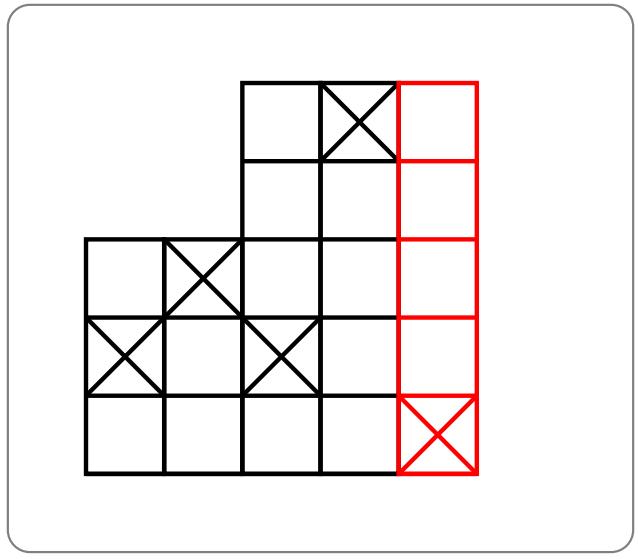

a
a
2
b
a
1
3
a
4
b
b
b
5
a

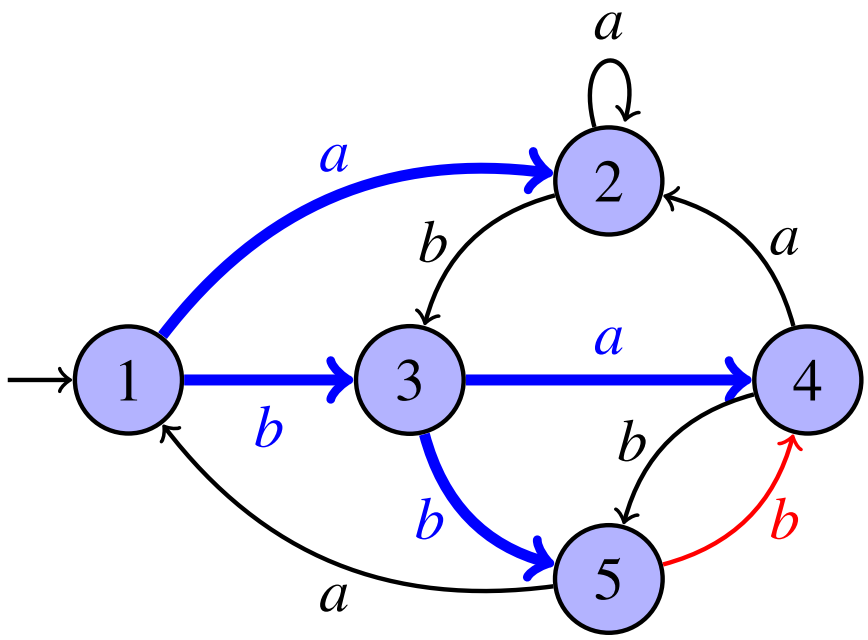


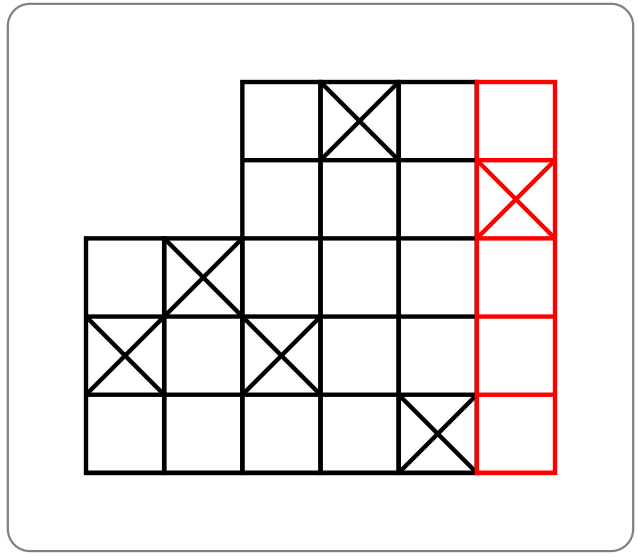

a
a
2
b
a
1
3
a
4
b
b
b
5
a




a
a
2
b
a
1
3
a
4
b
b
b
a
5

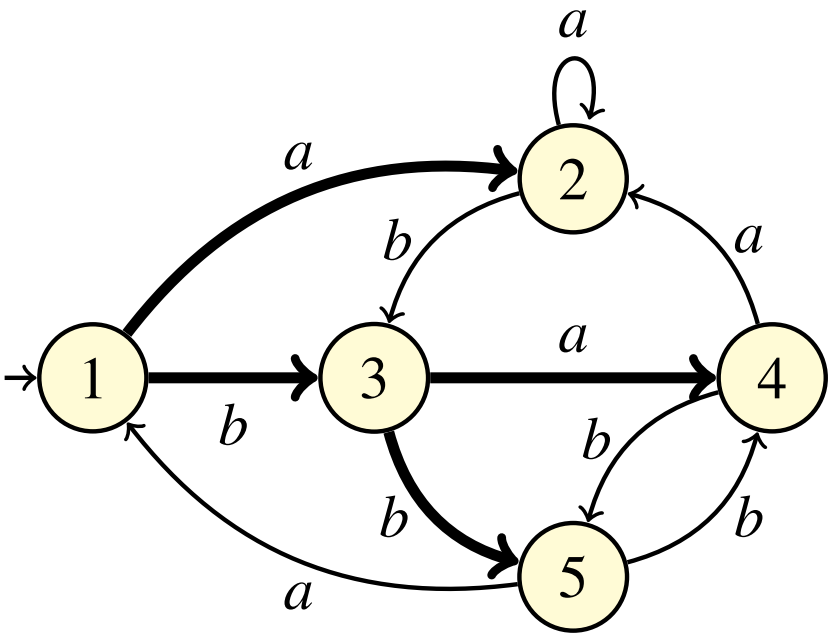



a
a
2
b
a
1
3
a
4
b
b
b
5
a



Automate $\longrightarrow$ tableau

- A un automate à n états sur un alphabet à k lettres on associe un tableau
- Le tableau possède $(k-1)n+1$ colonnes
- Les hauteurs des colonnes sont croissantes
- La hauteur maximale est n



$\Rightarrow$

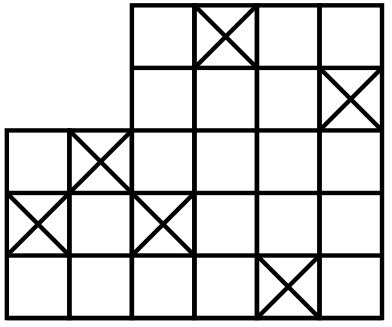


Tableau $\longrightarrow$ automate

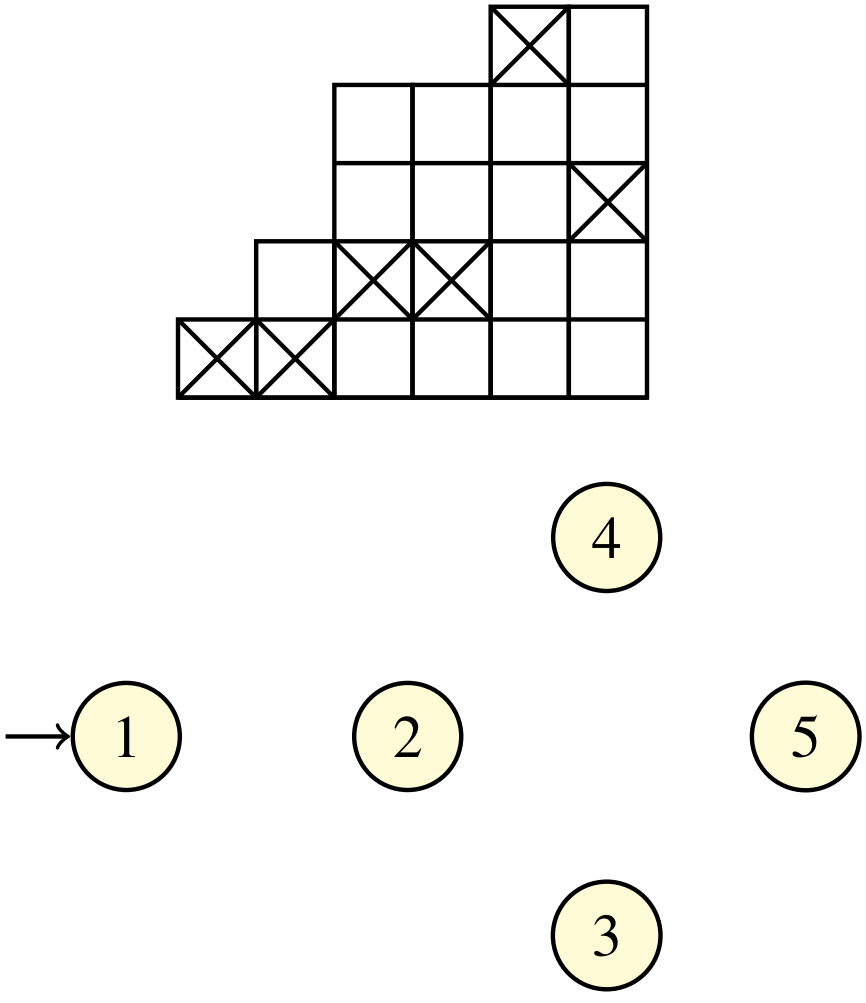


Tableau $\longrightarrow$ automate

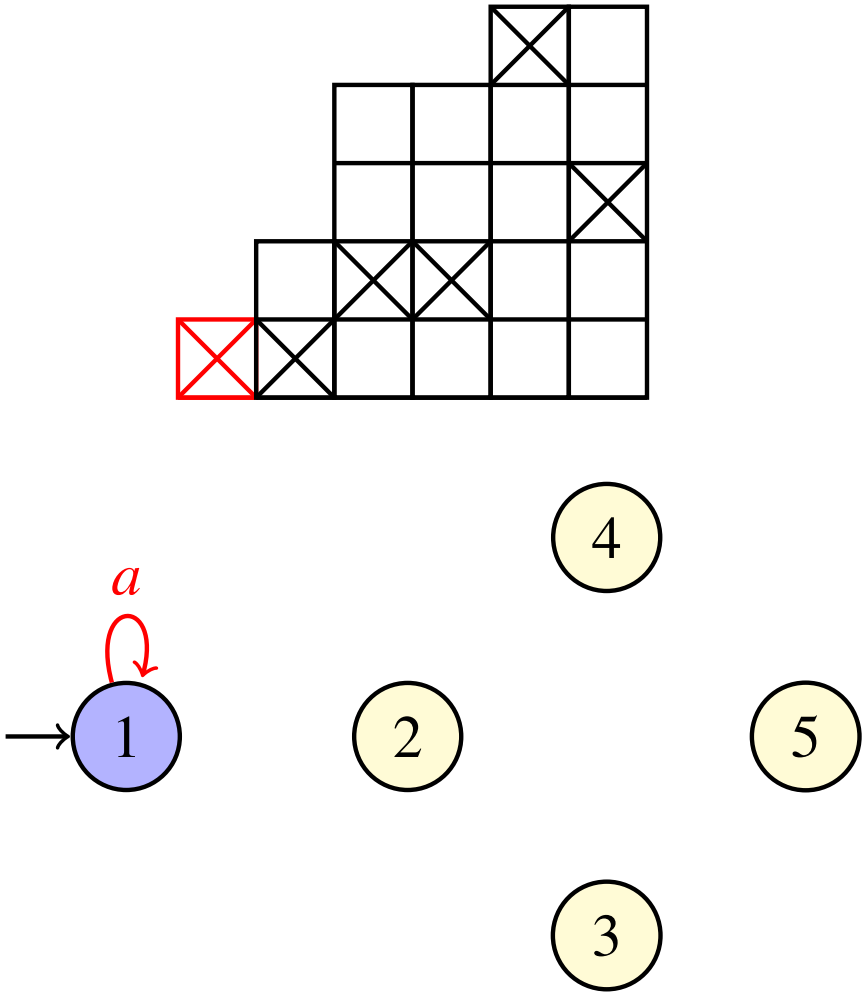
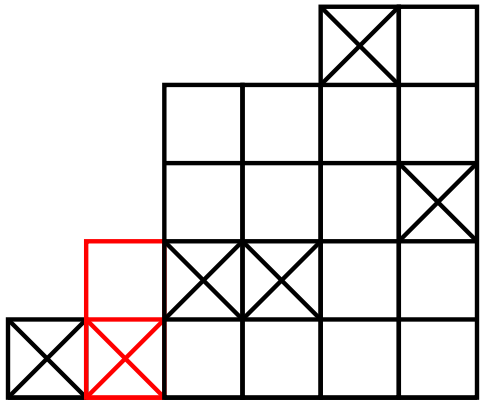


Tableau $\longrightarrow$ automate



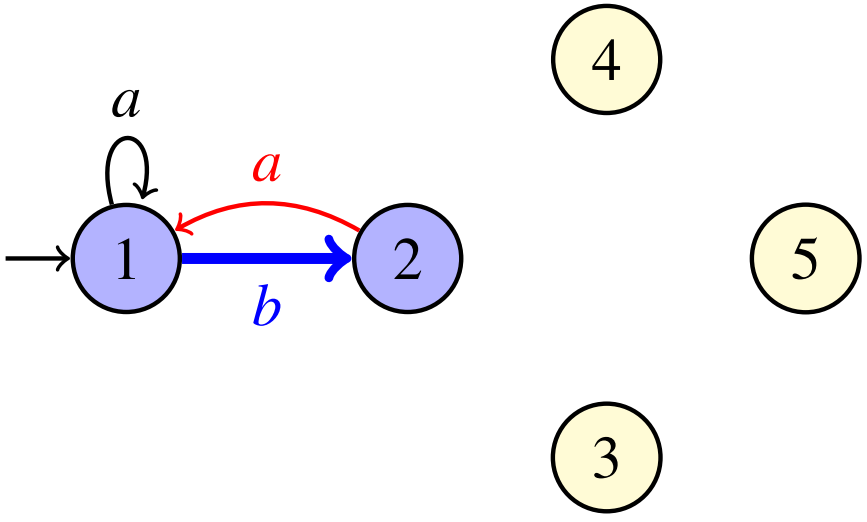
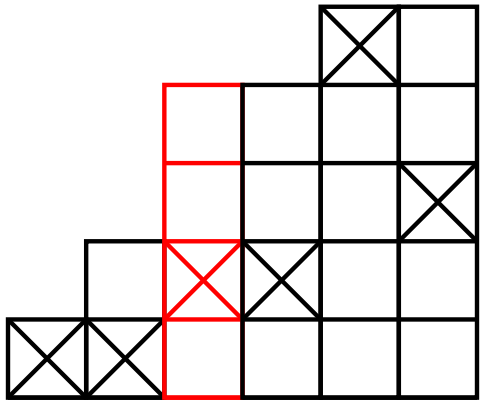


Tableau $\longrightarrow$ automate



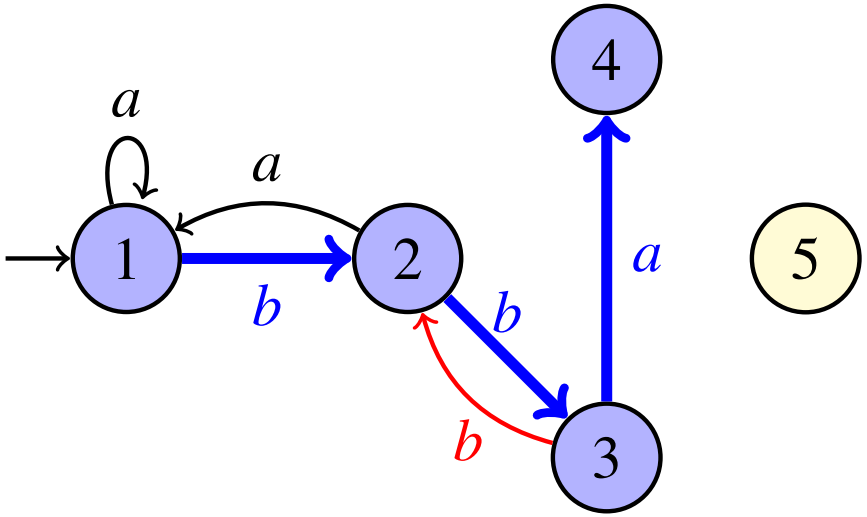
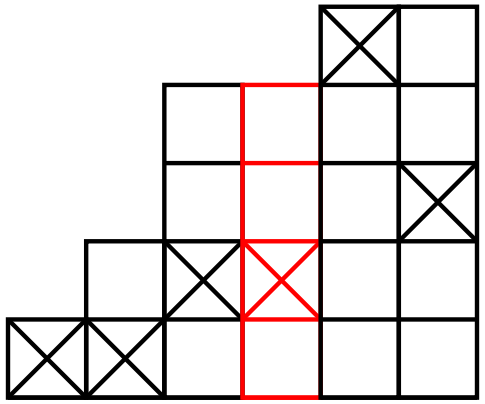


Tableau $\longrightarrow$ automate



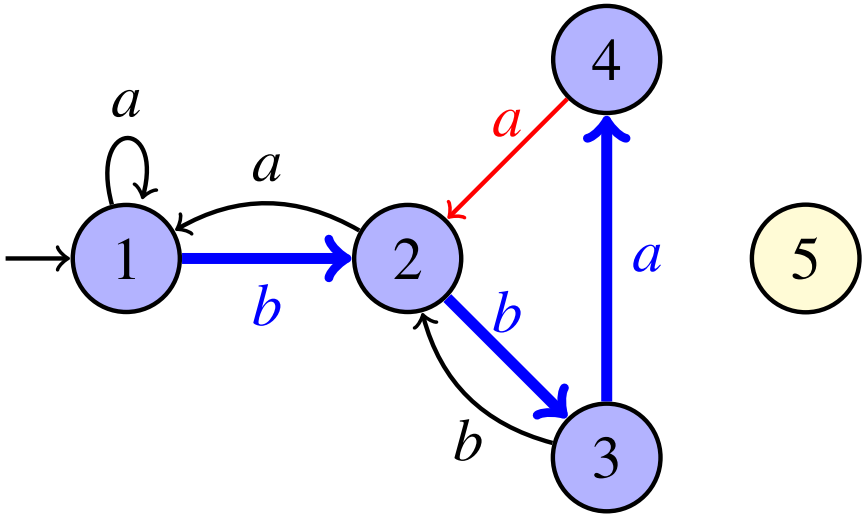
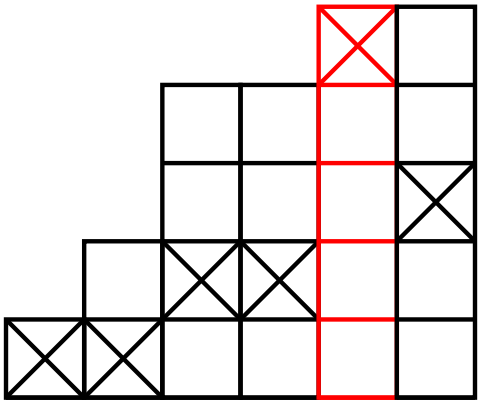


Tableau $\longrightarrow$ automate



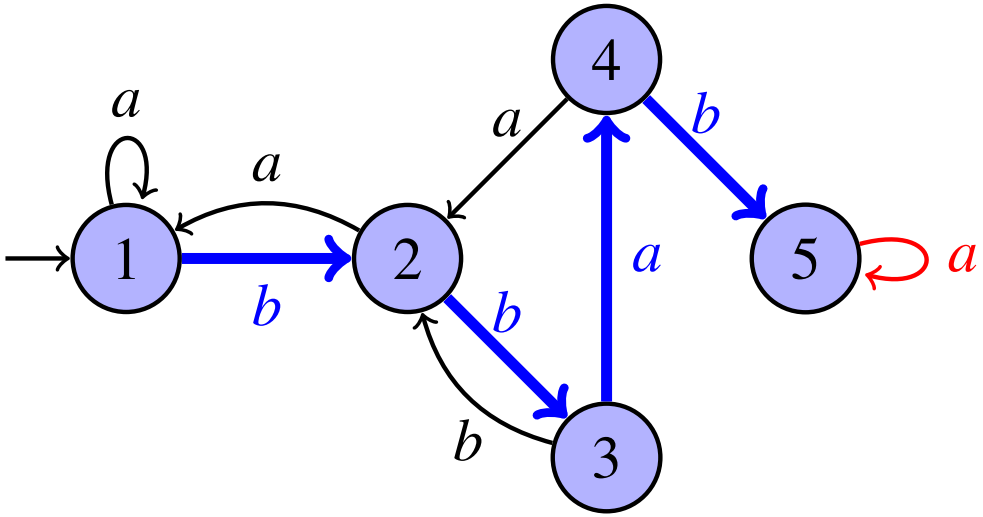
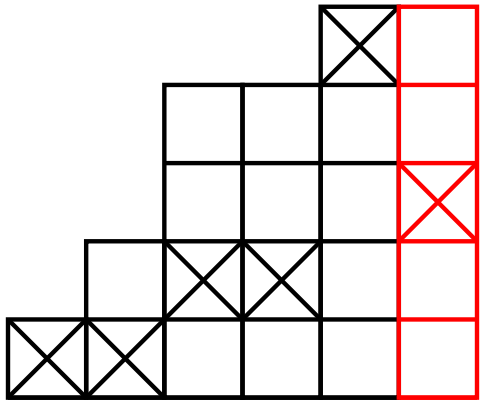
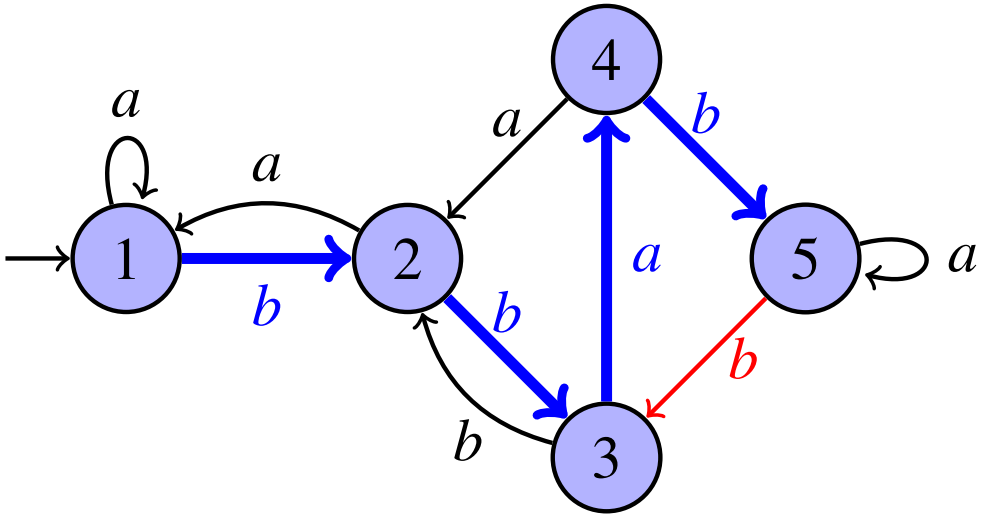
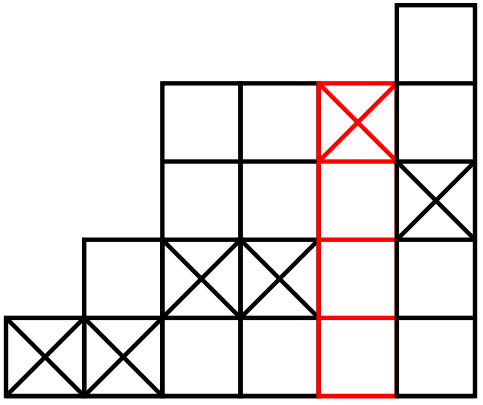


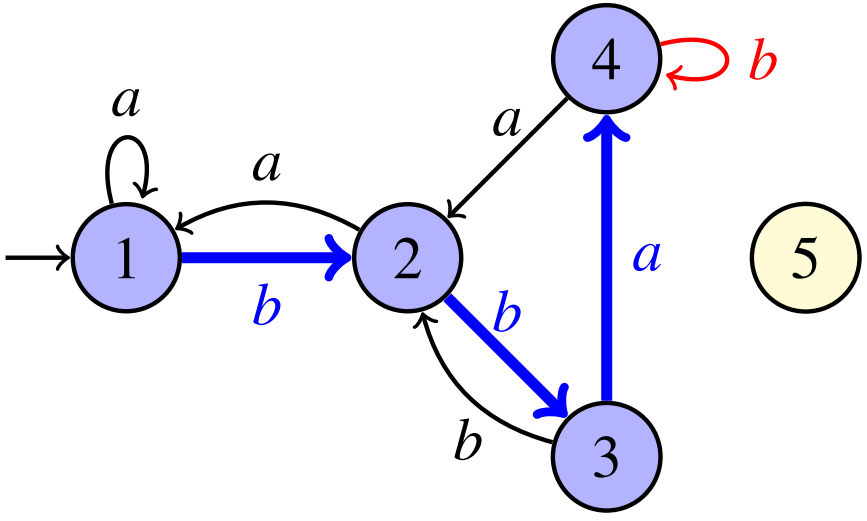
Tableau $\longrightarrow$ automate





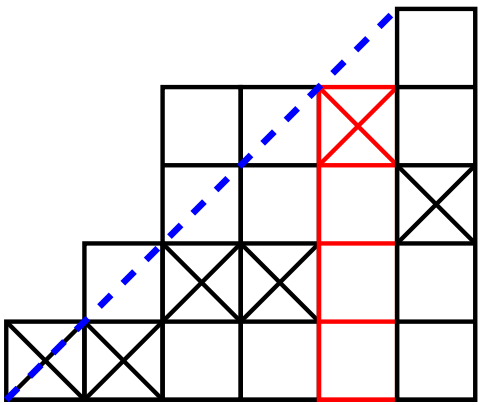
Condition d'accessibilité

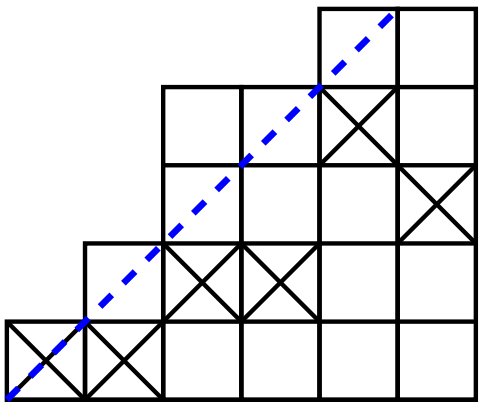




Tableaux admissibles

- Un tableau est admissible quand ses colonnes (sauf la dernière) sont au dessus de la diagonale

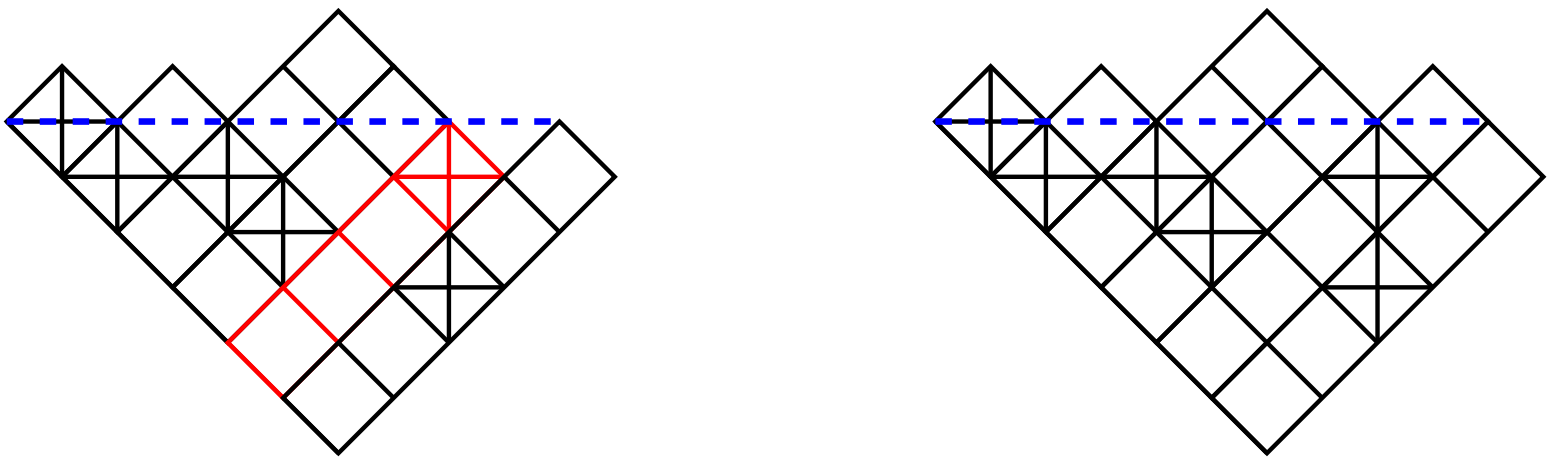




- Les tableaux admissibles sont en bijection avec les automates [N. 00, Champarnaud Paranthoën 05]

Tableaux admissibles

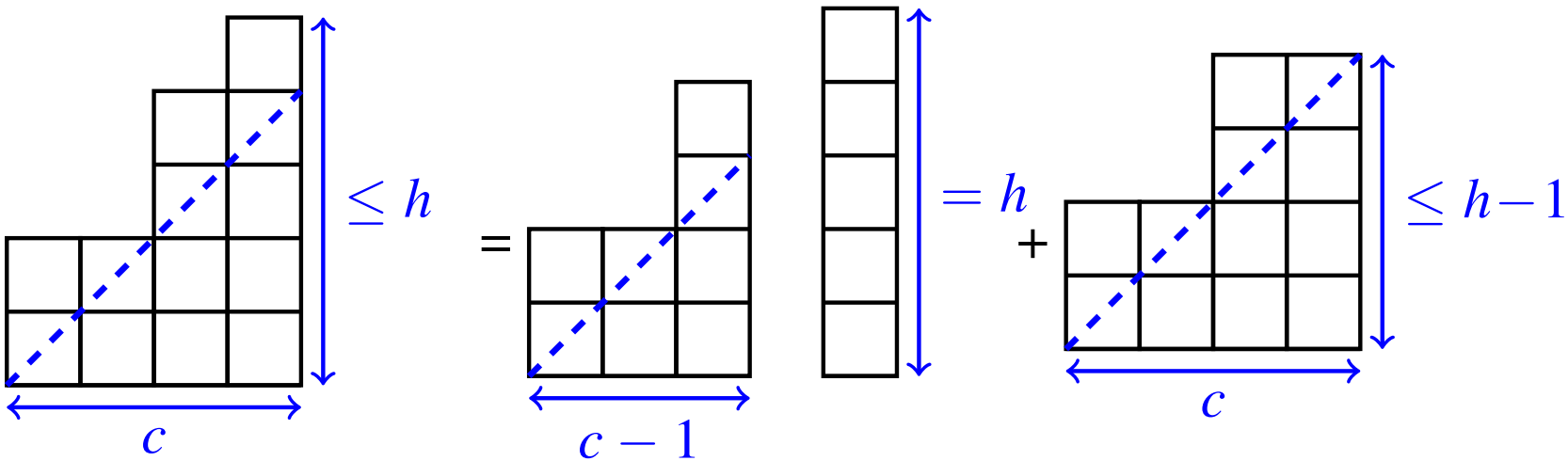
- Un tableau est admissible quand ses colonnes (sauf la dernière) sont au dessus de la diagonale



- Les tableaux admissibles sont en bijection avec les automates [N. 00, Champarnaud Paranthoën 05]

Génération aléatoire : méthode récursive

- On enlève la dernière colonne qui est toujours de hauteur n
- On a une décomposition récursive des tableaux avec c colonnes de hauteur au plus h, au-dessus de la diagonale



$$T_{h,c} = h\,T_{h,c-1} + T_{h-1,c}$$

Génération aléatoire : méthode récursive

Théorème [N. 00, Champarnaud Paranthoën 05]

Après un précalcul en $\Theta(n^2)$, on peut générer aléatoirement des automates en temps linéaire.

Génération aléatoire : méthode récursive

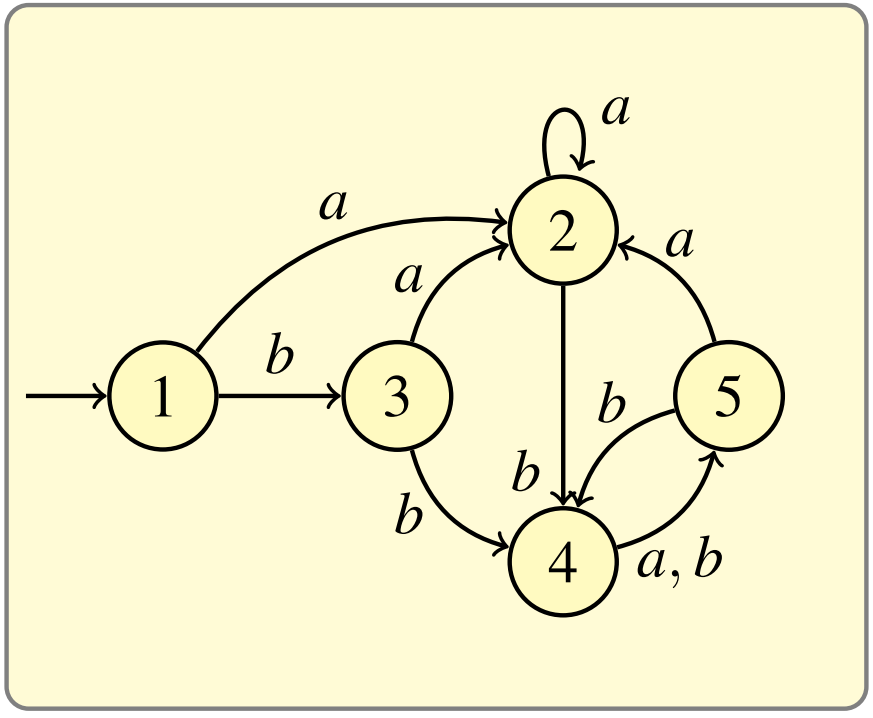
Théorème [N. 00, Champarnaud Paranthoën 05]

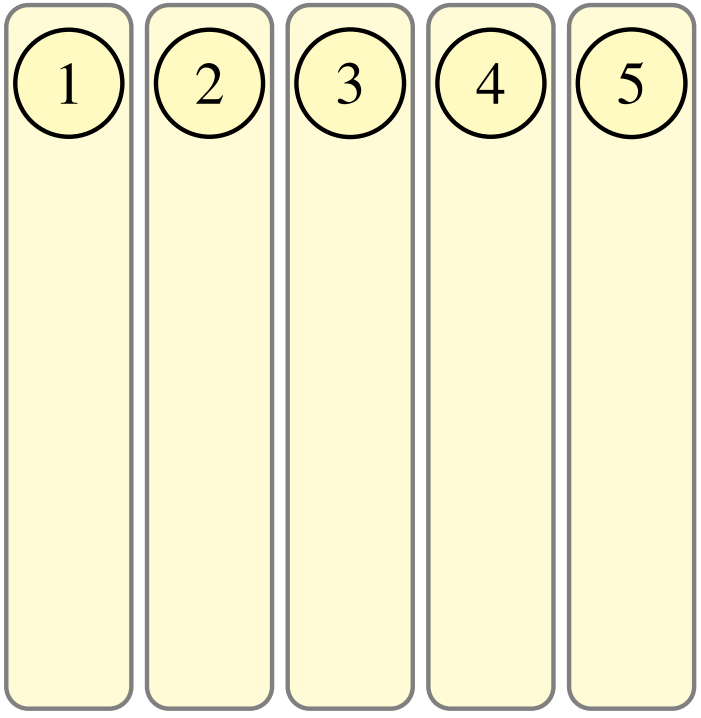
Après un précalcul en $\Theta(n^2)$, on peut générer aléatoirement des automates en temps linéaire.

- On a besoin d'entiers très grands, ou alors il faut approximer par des flottants.

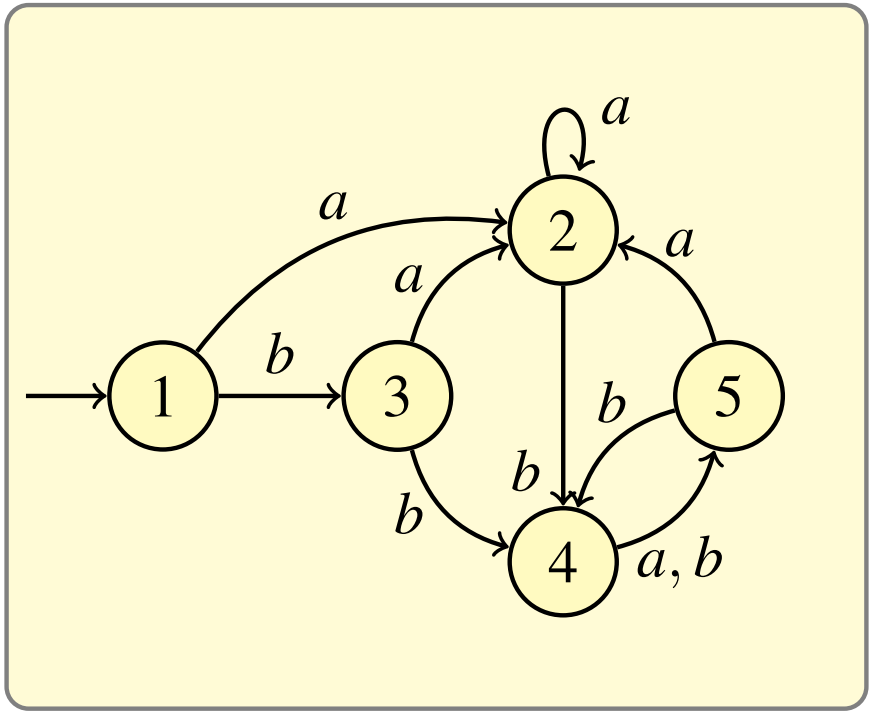
Partie IV : Partitions

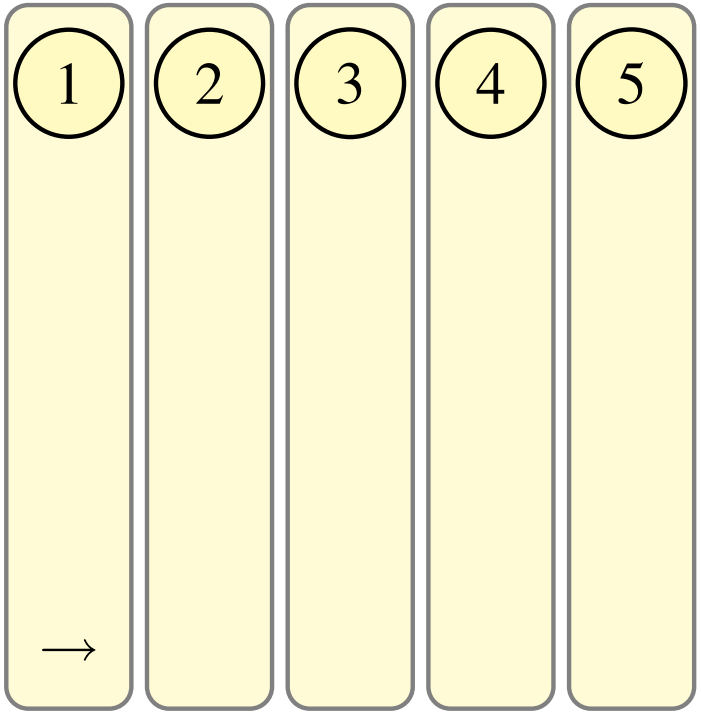
Automate → partition



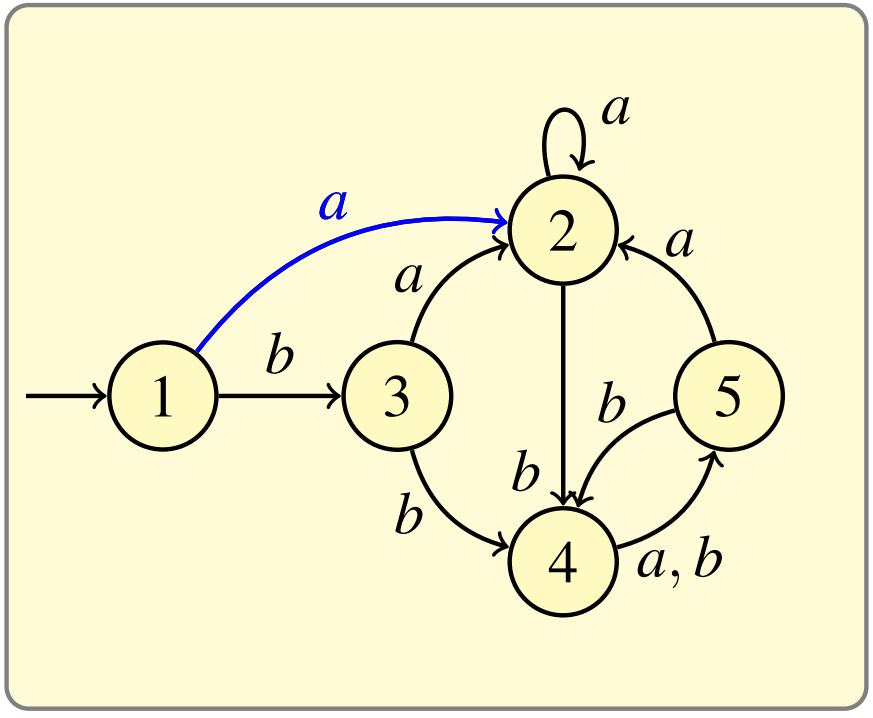


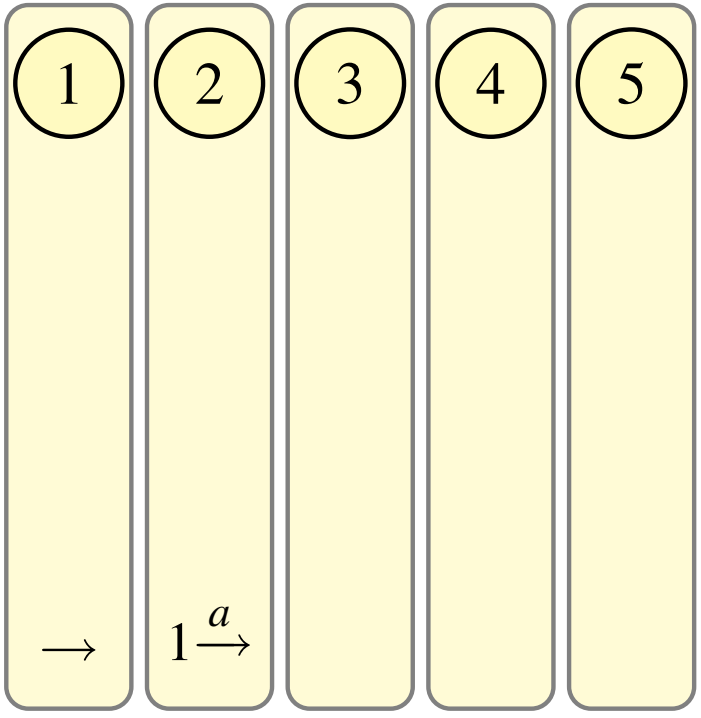
Automate → partition



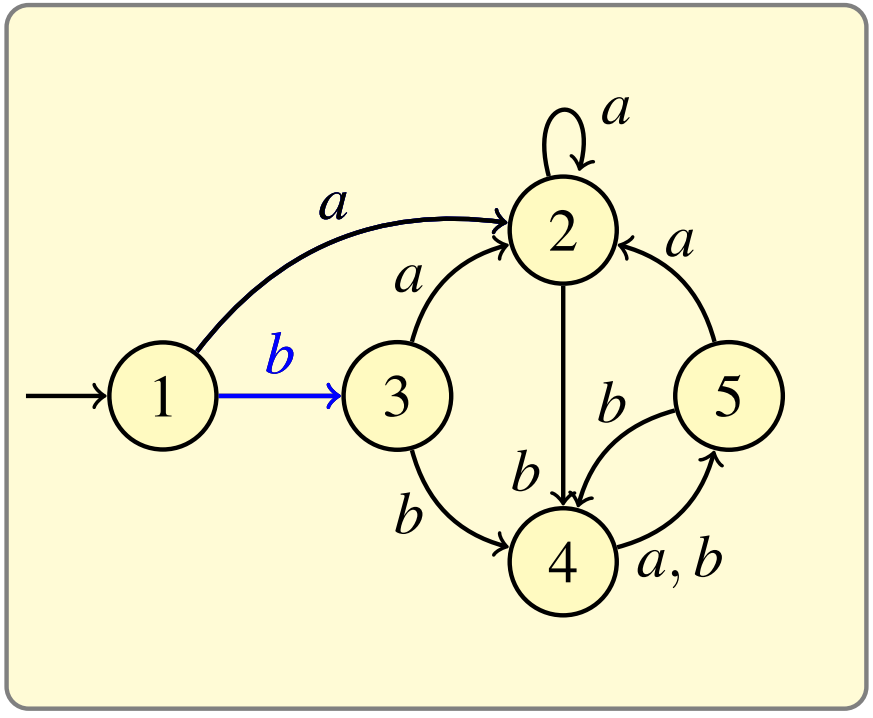


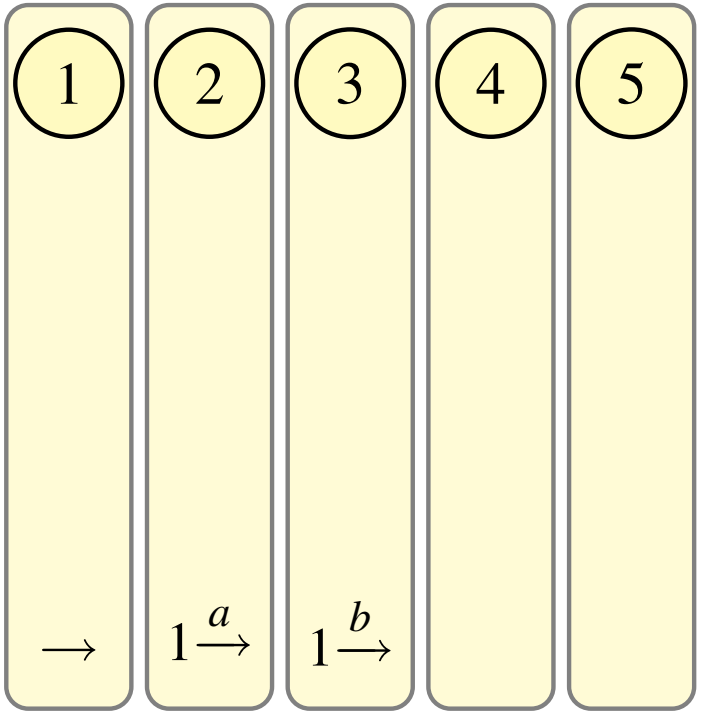
Automate → partition



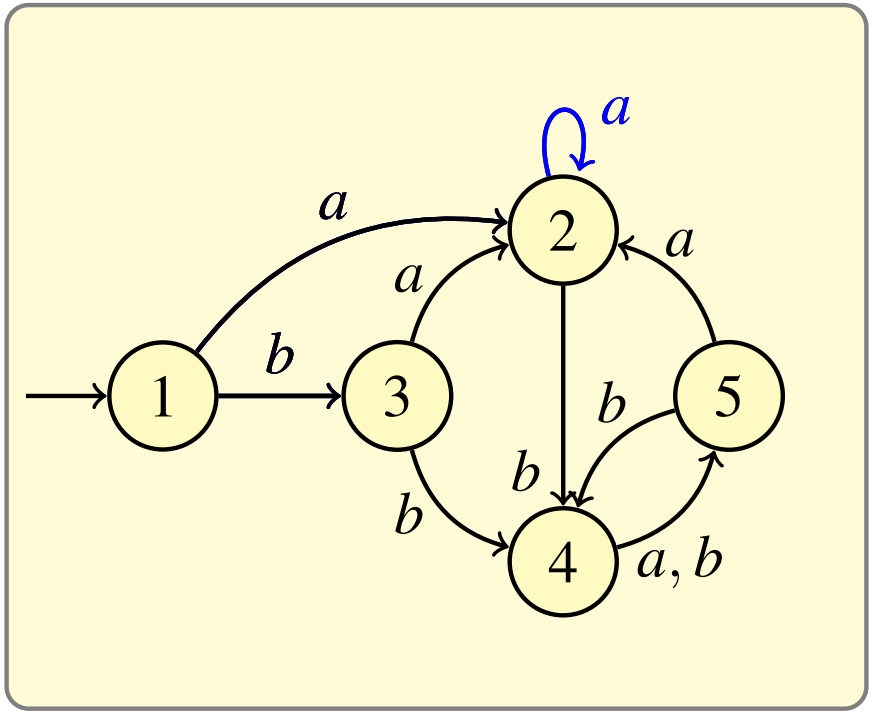


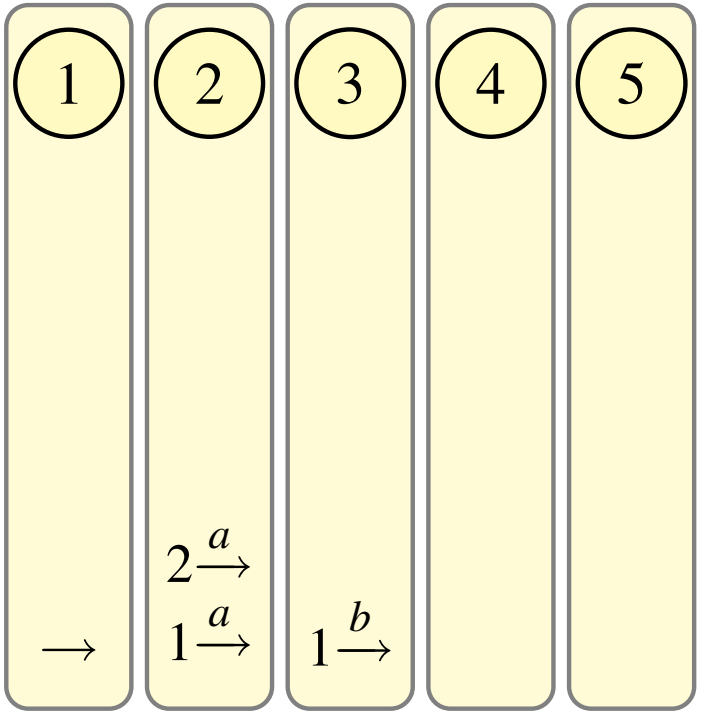
Automate → partition



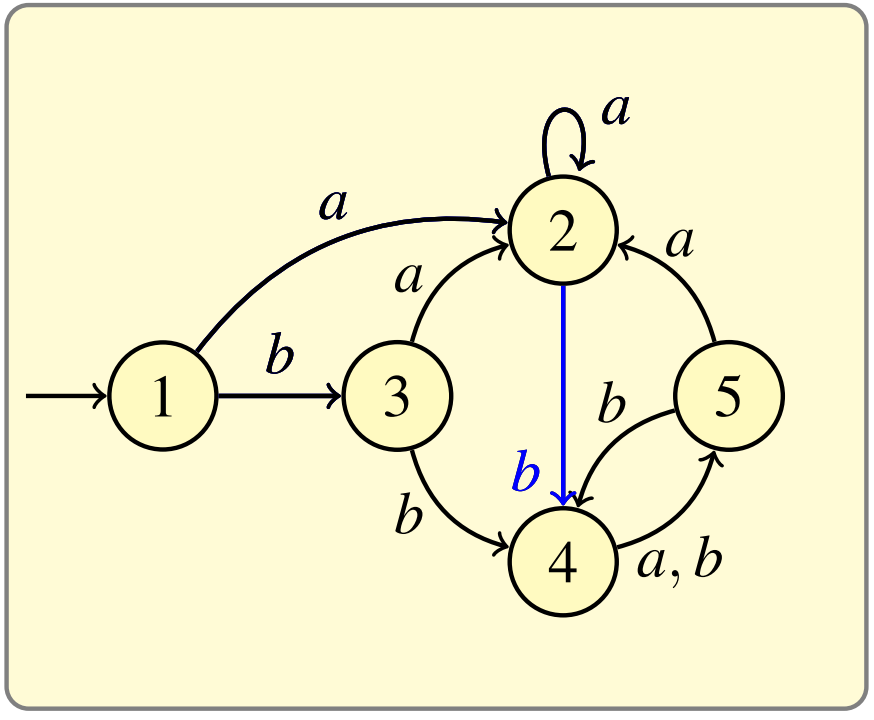


Automate → partition



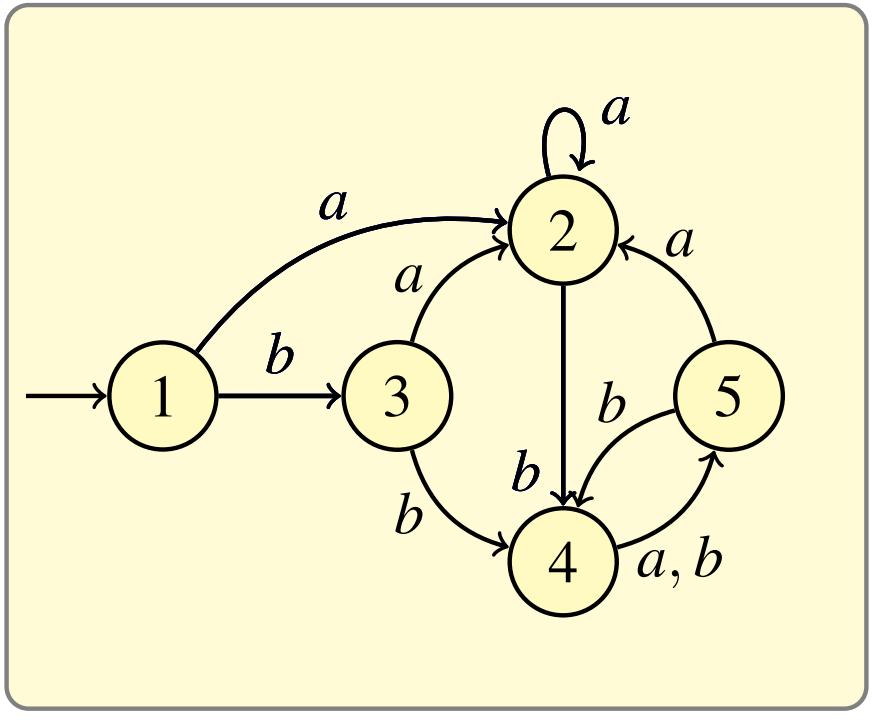


Automate → partition



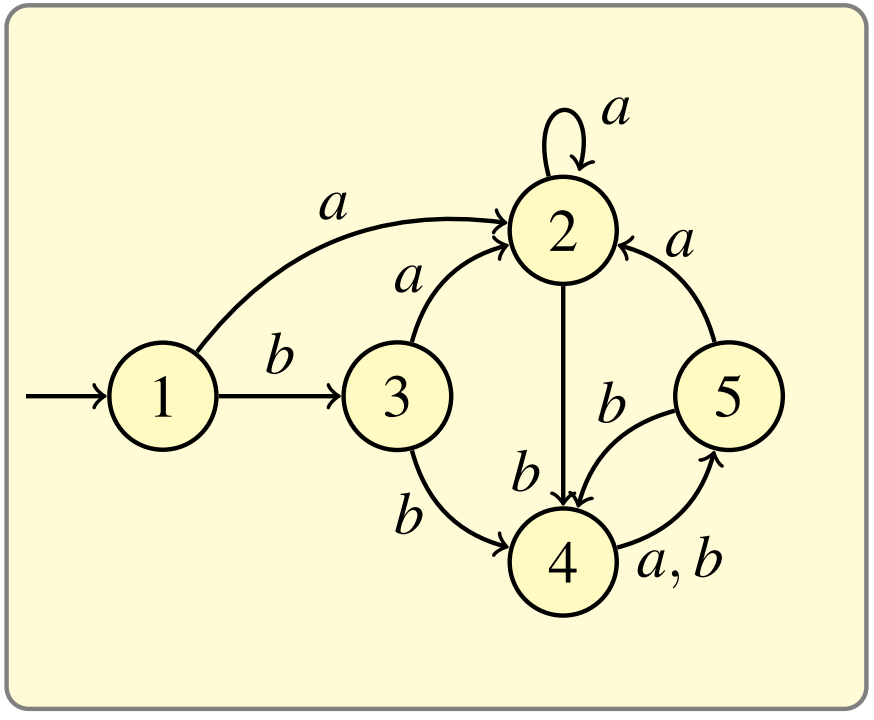
1	2	3	4	5
→	$1 \xrightarrow{a}$ $2 \xrightarrow{a}$	$1 \xrightarrow{b}$	$2 \xrightarrow{b}$	

Automate → partition



1	2	3	4	5
	$5 \xrightarrow{a}$			
	$3 \xrightarrow{a}$		$5 \xrightarrow{b}$	
	$2 \xrightarrow{a}$		$3 \xrightarrow{b}$	$4 \xrightarrow{b}$
$\rightarrow$	$1 \xrightarrow{a}$	$1 \xrightarrow{b}$	$2 \xrightarrow{b}$	$4 \xrightarrow{a}$

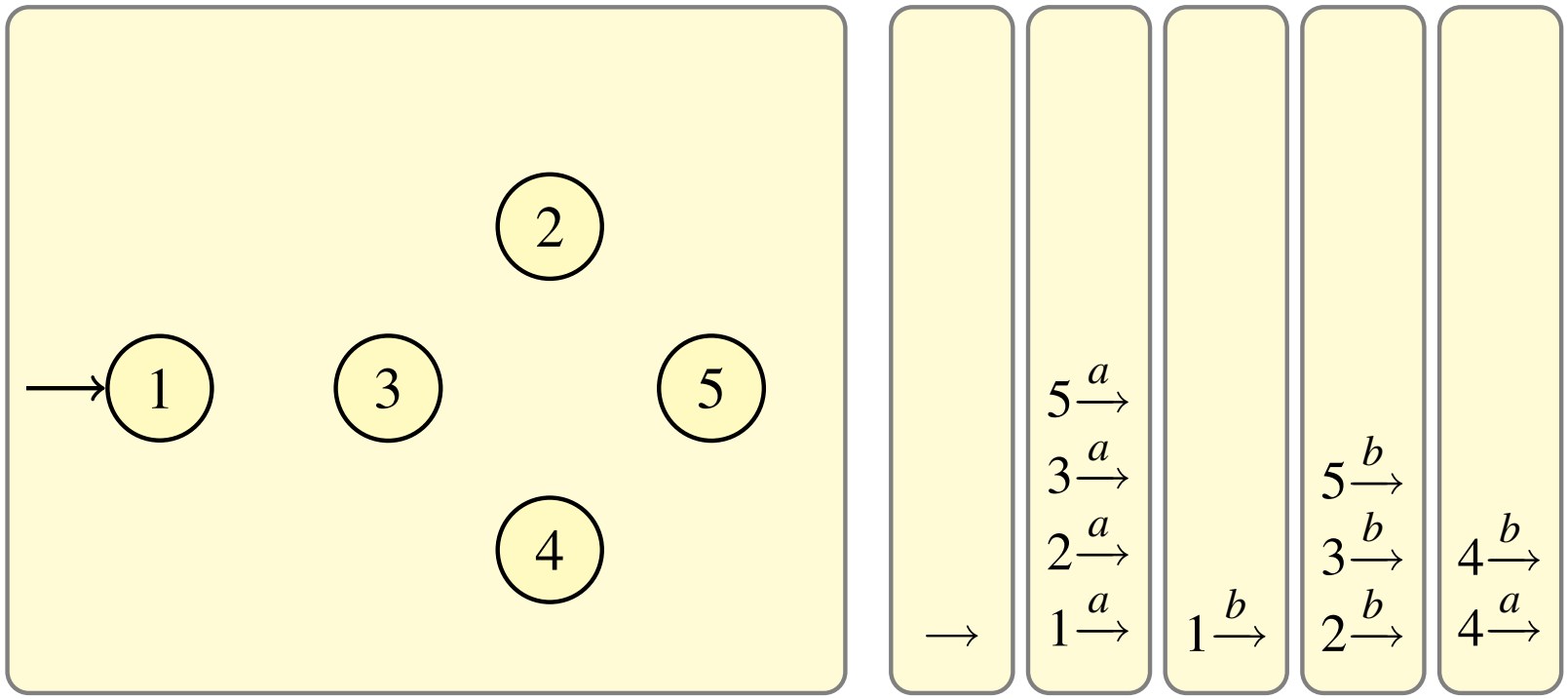
Automate $\longrightarrow$ partition



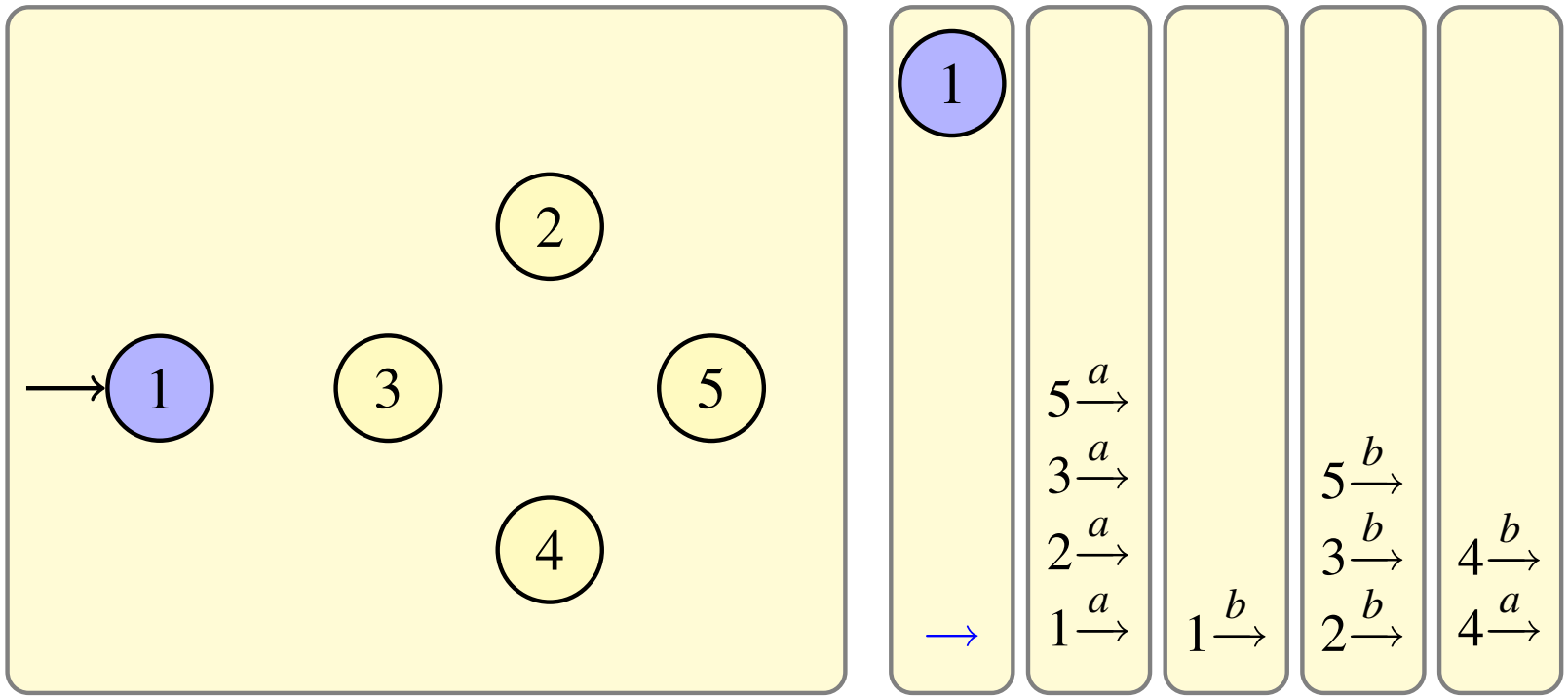
$\longrightarrow$	$5 \xrightarrow{a}$ $3 \xrightarrow{a}$ $2 \xrightarrow{a}$ $1 \xrightarrow{a}$	$1 \xrightarrow{b}$	$5 \xrightarrow{b}$ $3 \xrightarrow{b}$ $2 \xrightarrow{b}$	$4 \xrightarrow{b}$ $4 \xrightarrow{a}$

- On a une partition des $kn + 1$ flèches en n parts.

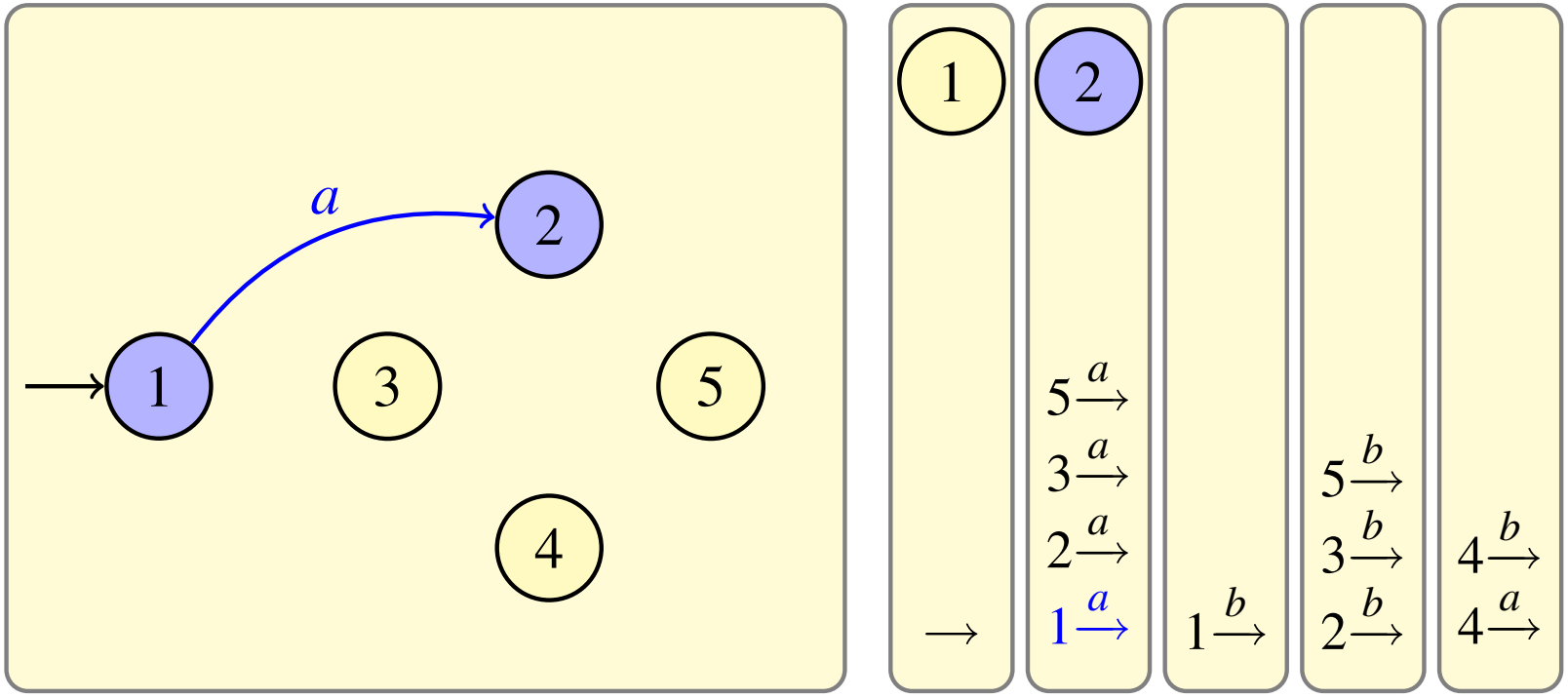
Partition → automate



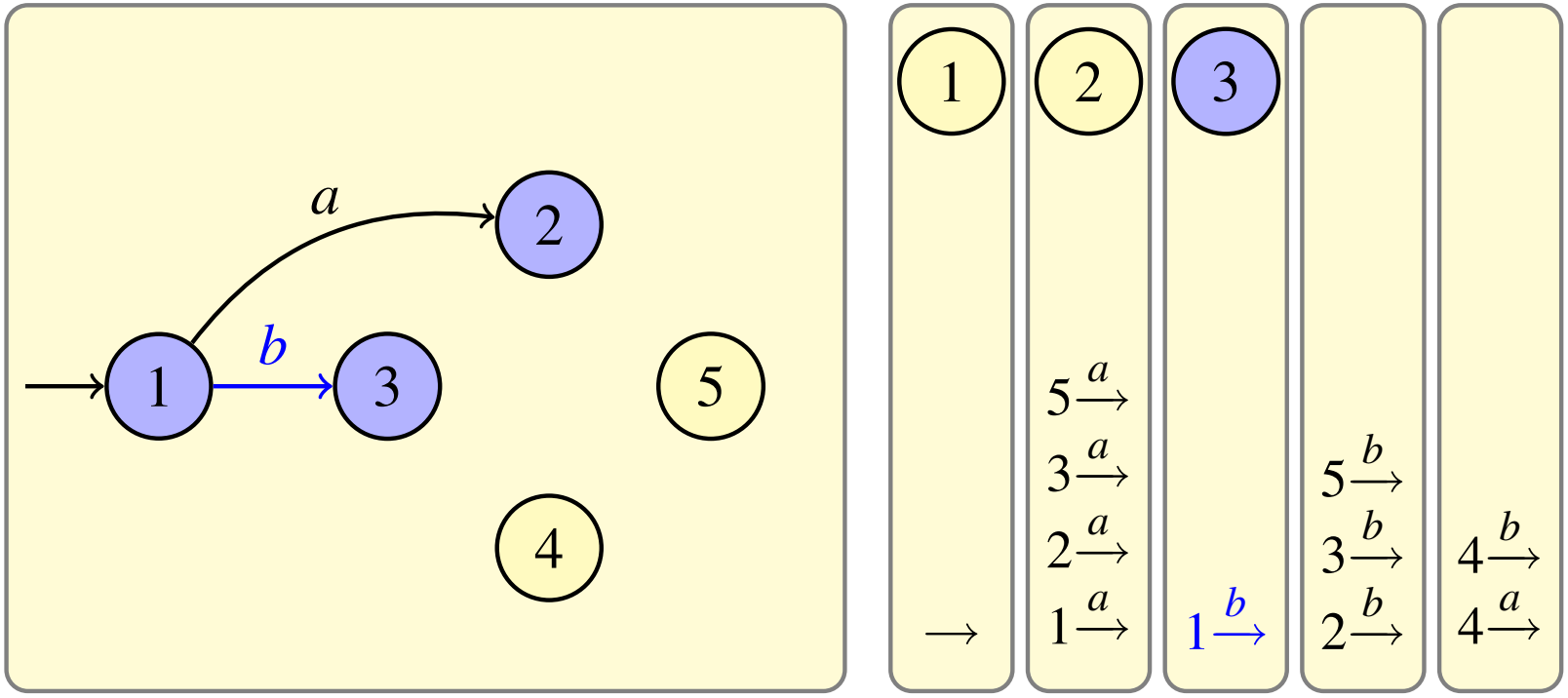
Partition → automate



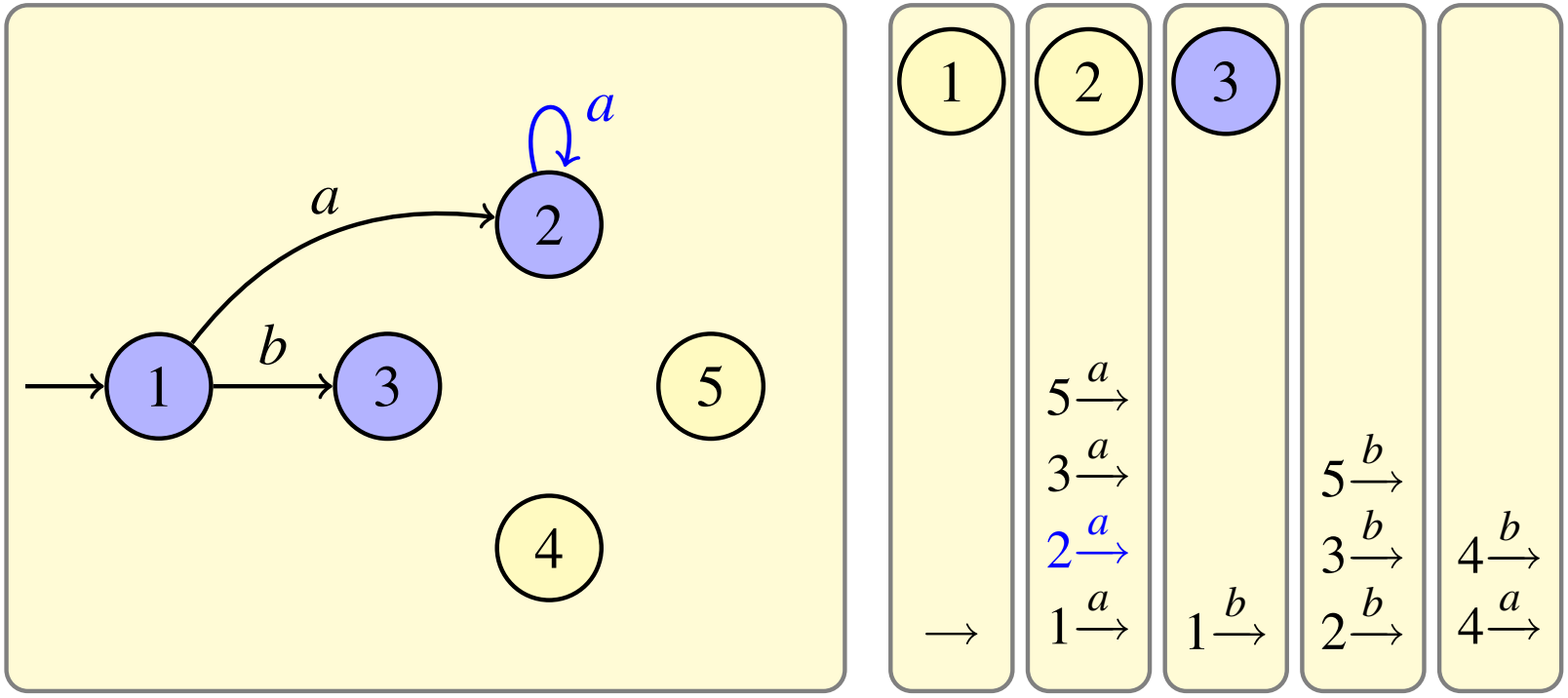
Partition → automate



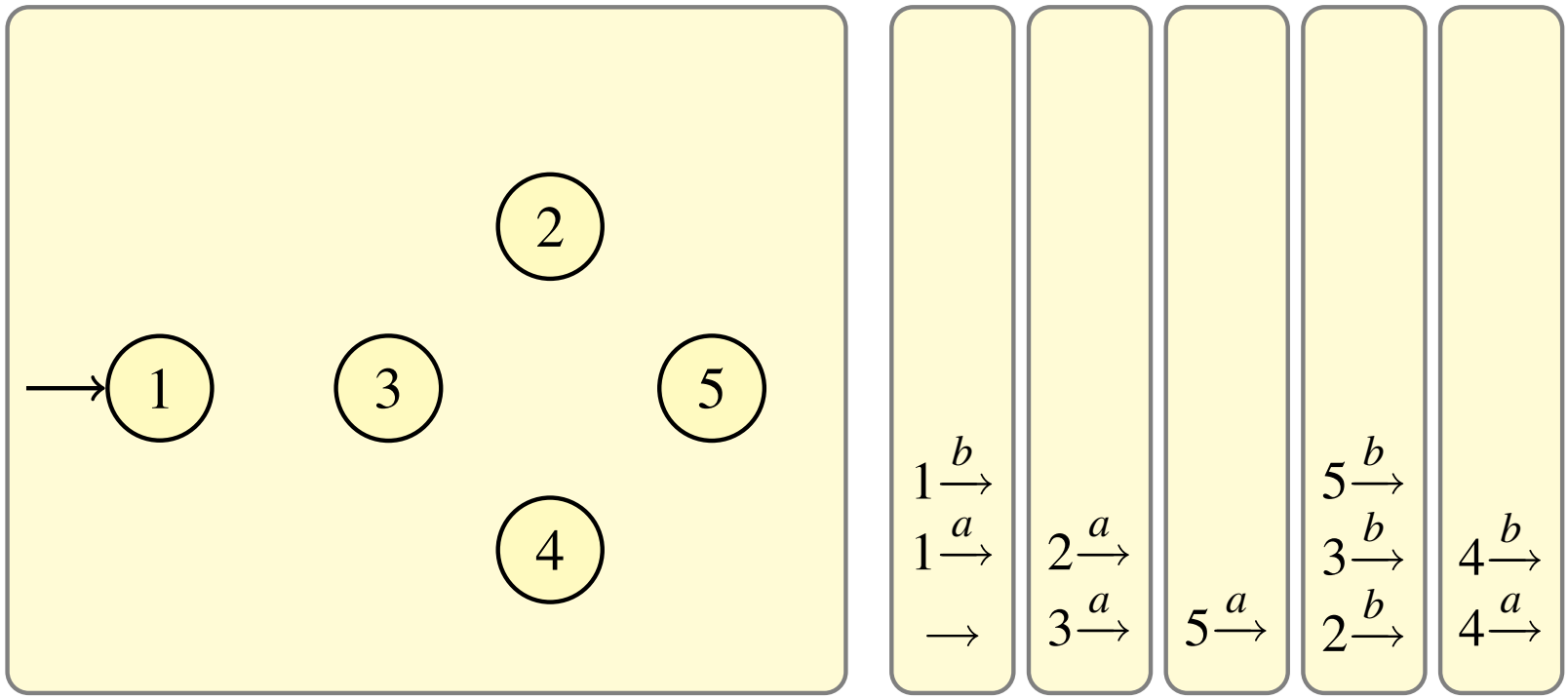
Partition → automate



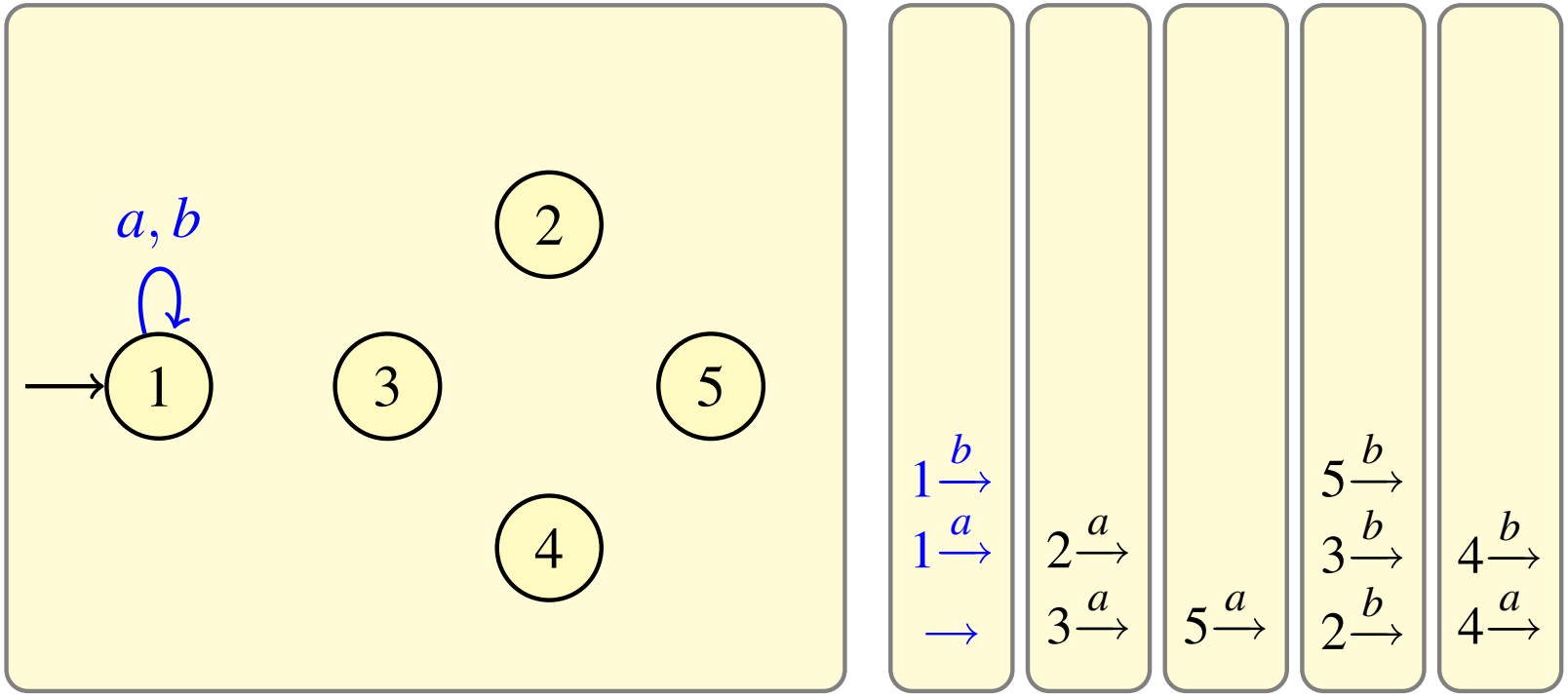
Partition → automate



Partition → automate



Partition → automate



Partitions admissibles

- Une partition est admissible quand elle correspond à un automate accessible
- On peut caractériser facilement les partitions admissibles (similaire aux tableaux)

Partitions admissibles

- Une partition est admissible quand elle correspond à un automate accessible
- On peut caractériser facilement les partitions admissibles (similaire aux tableaux)

- On a une majoration du nombre d'automates par $\left\{ {kn+1 \atop n} \right\}$
- $\left\{ {x \atop y} \right\}$ est le nombre de façons de partitionner un ensemble à x éléments en y parts (nombres de Stirling de deuxième espèce)

Génération aléatoire de partitions

- Comment générer aléatoirement des partitions admissibles ?
- Comment générer aléatoirement des partitions de $kn + 1$ éléments en n parts ?

Génération aléatoire de partitions

- Comment générer aléatoirement des partitions admissibles ?
- Comment générer aléatoirement des partitions de $kn + 1$ éléments en n parts ?

Combinatorics, Probability and Computing (2004) 13, 577–625.
DOI: 10.1017/S0963548304006315 Printed in the United Kingdom

Boltzmann Samplers for the Random Generation of Combinatorial Structures

PHILIPPE DUCHON,[1] PHILIPPE FLAJOLET,[2]
GUY LOUCHARD[3] and GILLES SCHAEFFER[4]

Génération aléatoire de partitions

- On génère chacune des n parts indépendamment
- La taille de chaque part suit une loi de Poisson non nulle de paramètre ζ_k
- ζ_k est choisi de façon à ce que la taille moyenne de l'objet soit kn
- (On traite à part la dernière transition, comme pour les tableaux)

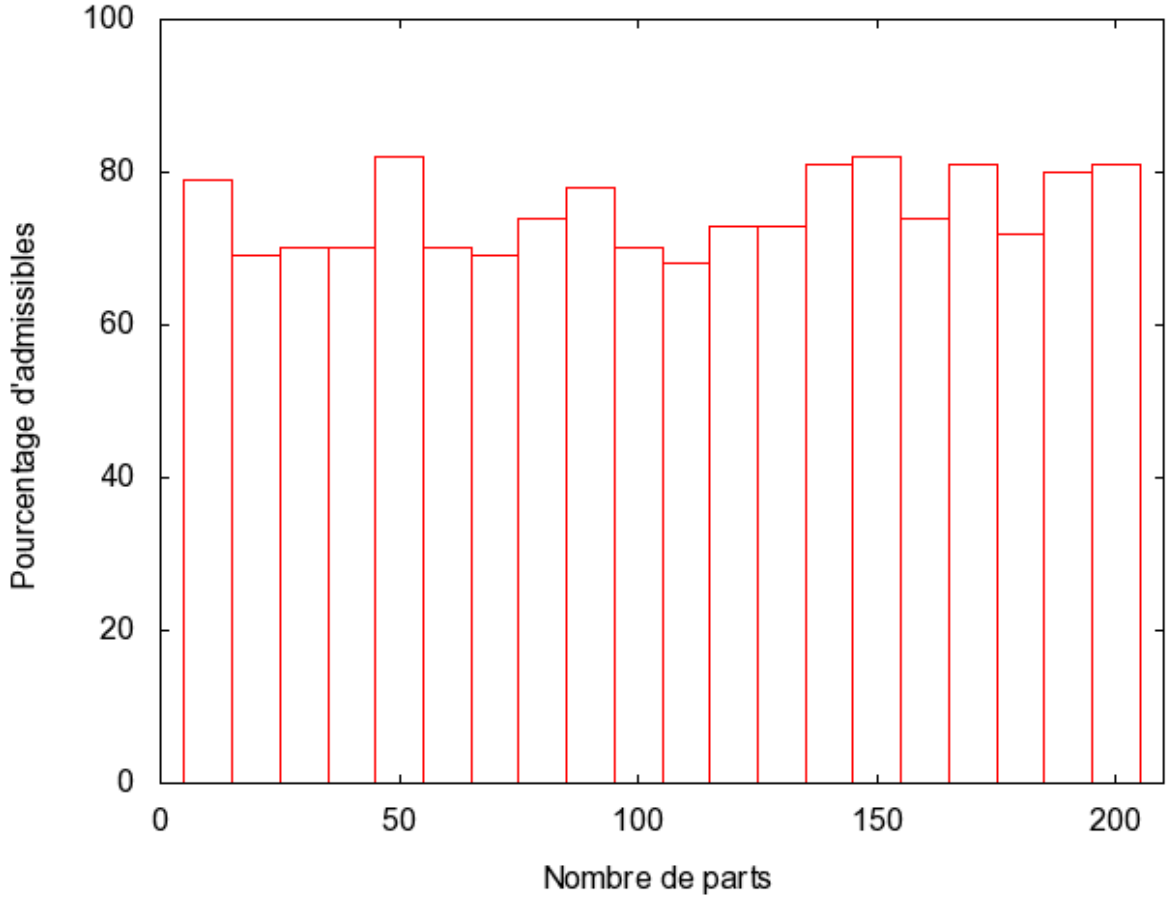
Génération aléatoire de partitions

- On génère chacune des n parts indépendamment
- La taille de chaque part suit une loi de Poisson non nulle de paramètre ζ_k
- ζ_k est choisi de façon à ce que la taille moyenne de l'objet soit kn
- (On traite à part la dernière transition, comme pour les tableaux)
- On génère un objet de taille à peu près kn
- La probabilité qu'il soit de taille exactement kn est en $\Theta(\frac{1}{\sqrt{n}})$
- $\zeta_k = k + W_0(-ke^{-k})$, et

$$W_0(z) = \sum_{n=1}^{\infty} \frac{(-n)^{n-1}}{n!} z^n,$$

W_0 est la branche principale, analytique en 0 de l'inverse $x \mapsto xe^x$.

Test d’admissibilité



Asymptotique des Stirling de deuxième espèce

- Une proportion importante de partitions semblent être admissibles
- L'asymptotique des $\left\{ {kn \atop n} \right\}$ nous intéresse donc

Asymptotique des Stirling de deuxième espèce

- Une proportion importante de partitions semblent être admissibles
- L'asymptotique des $\left\{ {kn \atop n} \right\}$ nous intéresse donc

Ann. Math. Statist. 32 (1961) :

AN ASYMPTOTIC FORMULA FOR THE DIFFERENCES OF THE POWERS AT ZERO

By I. J. Good

Admiralty Research Laboratory, Teddington, England

Asymptotique des Stirling de deuxième espèce

- Une proportion importante de partitions semblent être admissibles
- L'asymptotique des $\left\{ {kn \atop n} \right\}$ nous intéresse donc

Ann. Math. Statist. 32 (1961) :

AN ASYMPTOTIC FORMULA FOR THE DIFFERENCES OF THE POWERS AT ZERO

By I. J. Good

Admiralty Research Laboratory, Teddington, England

- Par un théorème du col, il obtient

$$\left\{ {kn+1 \atop n} \right\} \sim \alpha_k \beta_k^n n^{(k-1)n+\frac{1}{2}}$$

Asymptotique des automates

Théorème [Bassino N. 07]

Asymptotiquement, le nombre d'automates est en $\Theta(\left\{ {kn+1 \atop n} \right\})$.

- On a travaillé sur les tableaux

Asymptotique des automates

Théorème [Bassino N. 07]

Asymptotiquement, le nombre d'automates est en $\Theta(\left\{ {kn+1 \atop n} \right\})$.

- On a travaillé sur les tableaux

Asymptotique des automates

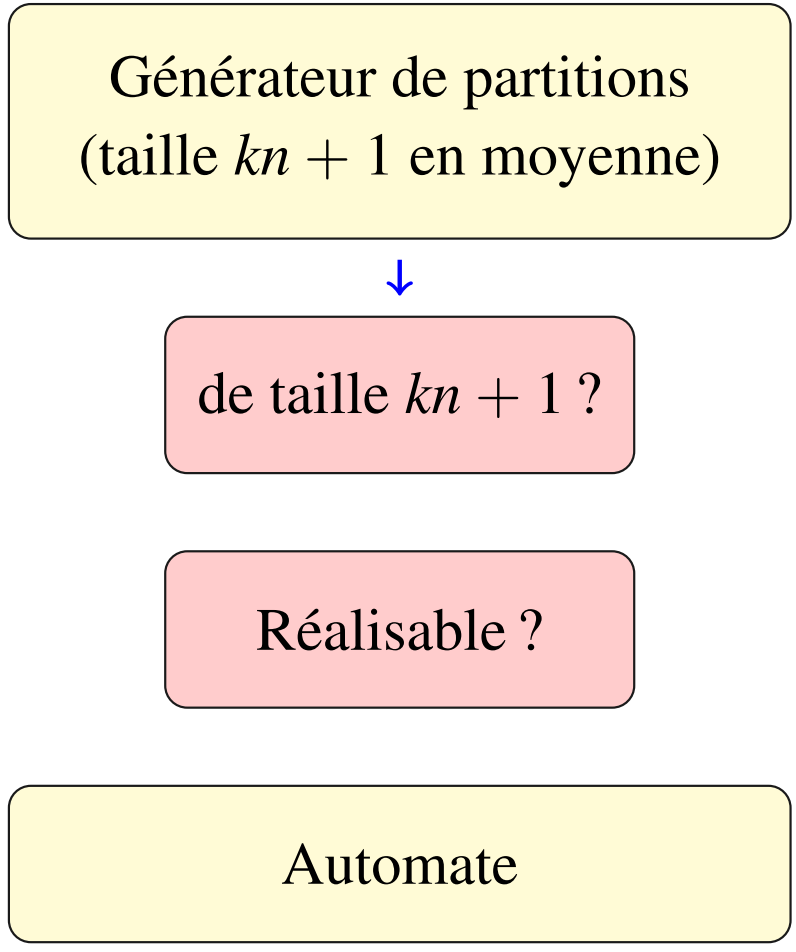
Théorème [Bassino N. 07]

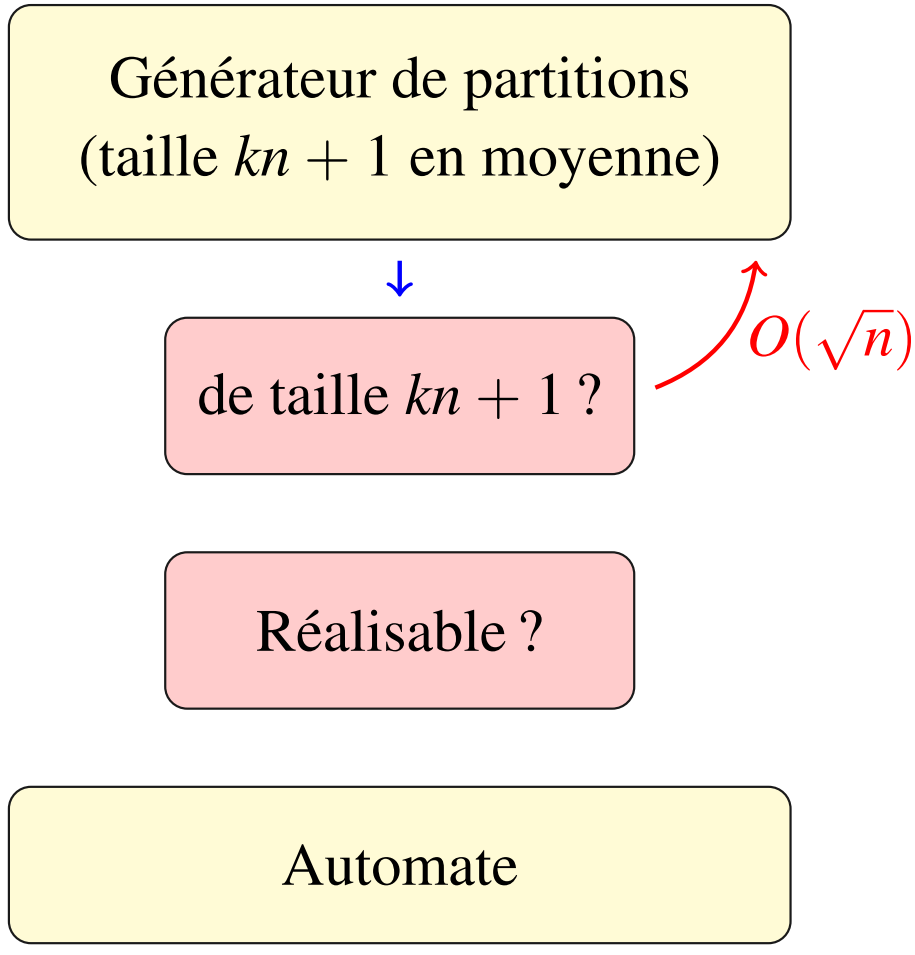
Asymptotiquement, le nombre d'automates est en $\Theta(\left\{ {kn+1 \atop n} \right\})$.

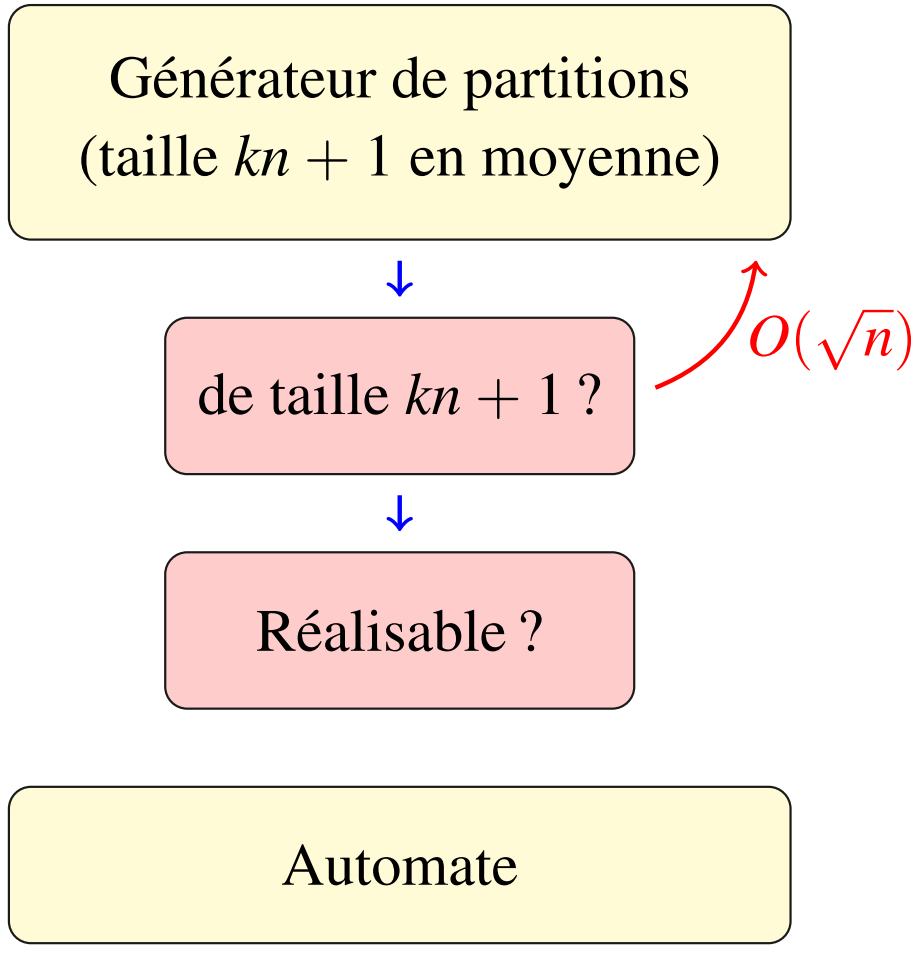
- On a travaillé sur les tableaux

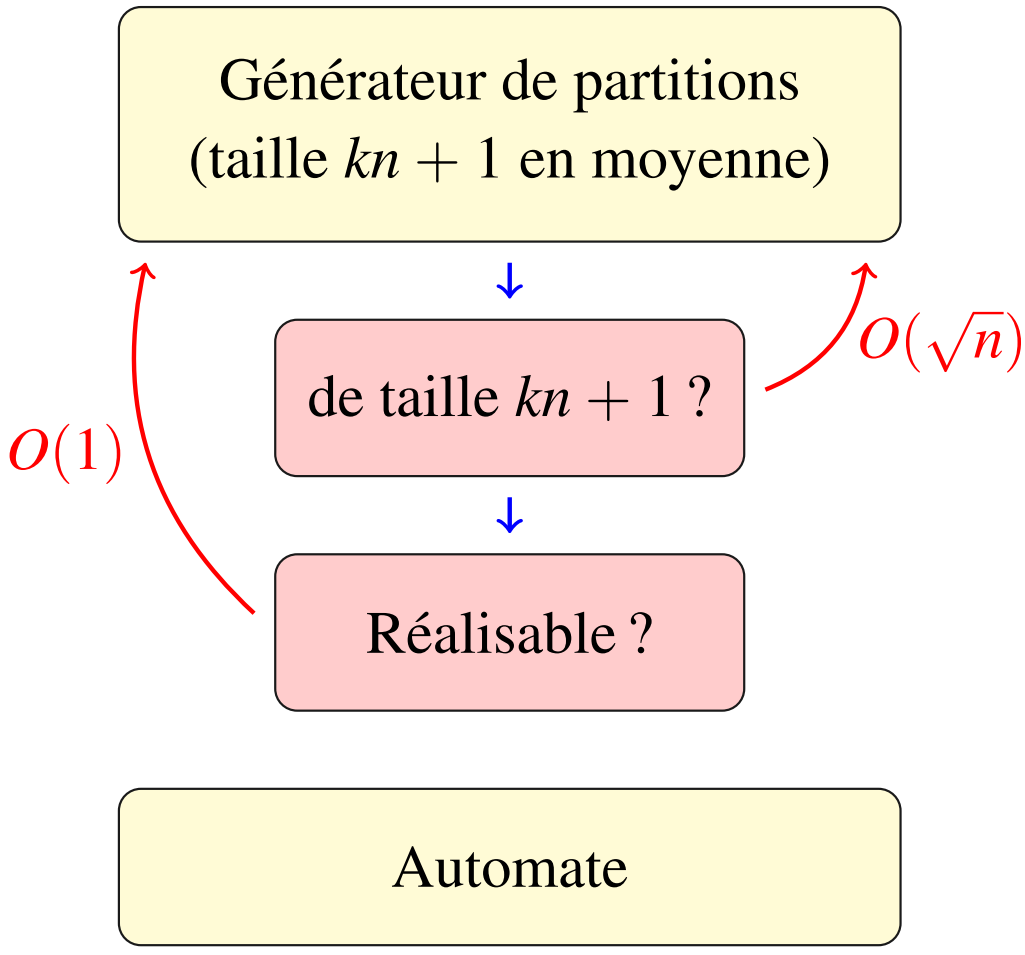
Théorème [Korshunov 78]

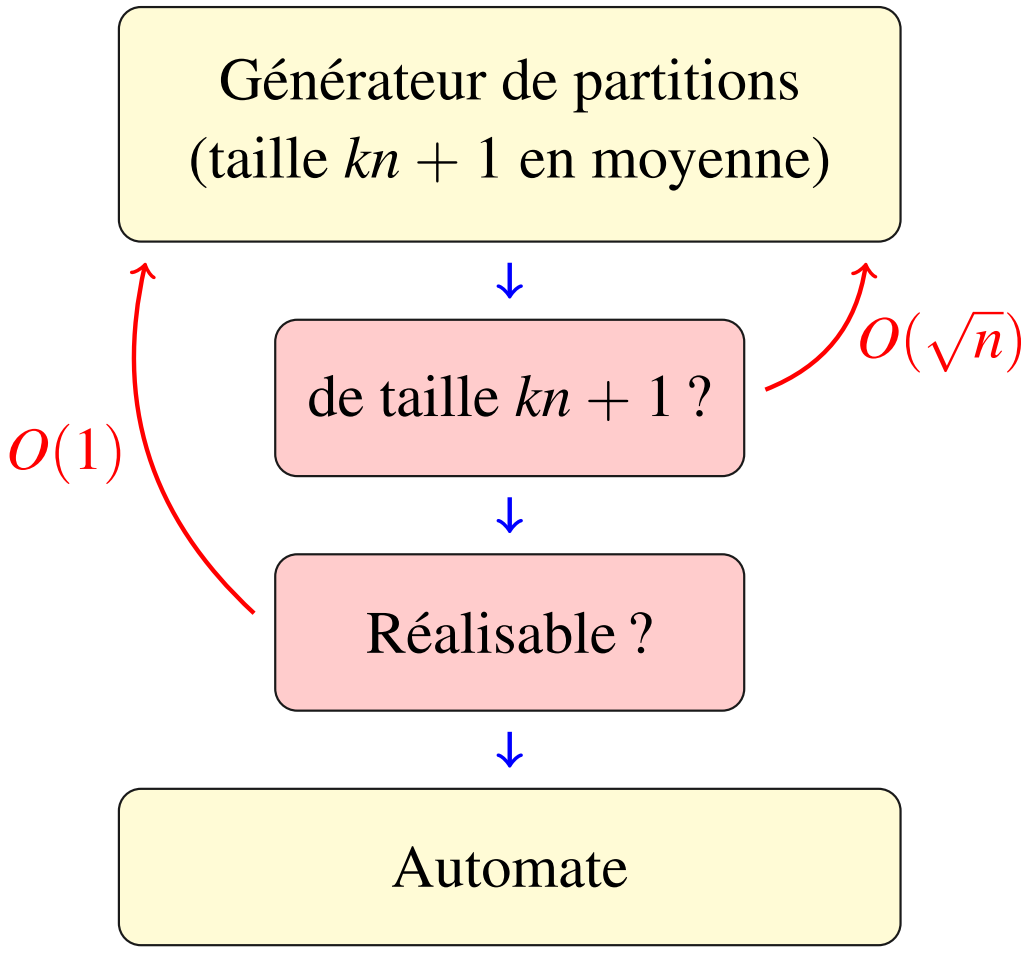
Asymptotiquement, le nombre d'automates est équivalent à $E_k \beta_k^n n^{(k-1)n+\frac{1}{2}}$, avec une expression de E_k avec des sommes infinies convergentes. (Simplifiée dans [Lebensztayn 10]).


Générateur de partitions
(taille $kn+1$ en moyenne)
de taille $kn+1$?
Réalisable ?
Automate

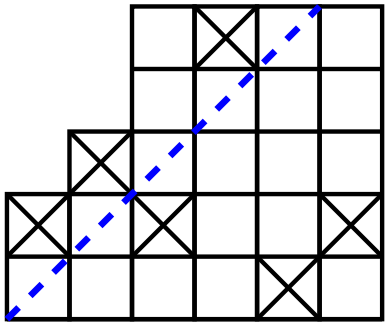

Générateur de partitions
(taille $kn+1$ en moyenne)
de taille $kn+1$?
$O(\sqrt{n})$
Réalisable ?
Automate

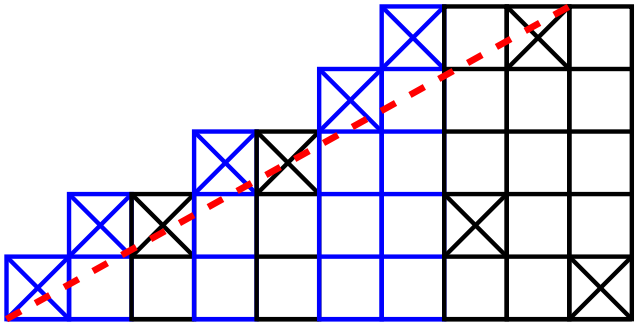

Générateur de partitions
(taille $kn+1$ en moyenne)
de taille $kn+1$?
$O(\sqrt{n})$
Réalisable ?
Automate


Générateur de partitions
(taille $kn+1$ en moyenne)
de taille $kn+1$?
Réalisable ?
Automate
$O(\sqrt{n})$
$O(1)$


Générateur de partitions
(taille $kn+1$ en moyenne)
de taille $kn+1$?
Réalisable ?
Automate
$O(\sqrt{n})$
$O(1)$

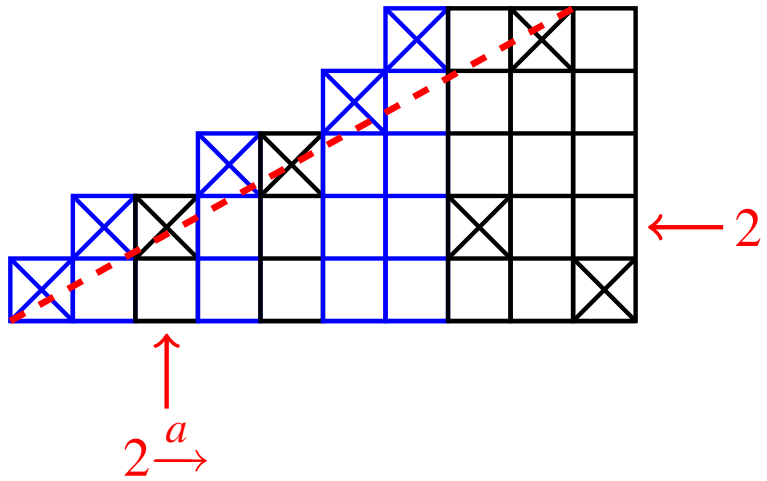
Lien entre tableaux et partitions





- On rajoute une colonne par taille possible, que l'on place le plus à gauche possible
- Cela correspond aux arcs de l'arbre couvrant
- La condition diagonale s'incline (droite $y = \frac{1}{k}x$)

Lien entre tableaux et partitions



- Chaque colonne correspond à une transition, dans l'ordre
- La croix correspond à la part
- $\rightarrow$ est dans la part 1

Lien entre tableaux et partitions

- Sur les tableaux on avait obtenu, avec en plus des conditions diagonales :

$$T_{h,c} = h\,T_{h,c-1} + T_{h-1,c}$$

- Les nombres de Stirling de deuxième espèce satisfont

$$\left\{ {x \atop y} \right\} = y \left\{ {x-1 \atop y} \right\} + \left\{ {x-1 \atop y-1} \right\}$$

Lien entre tableaux et partitions

- Sur les tableaux on avait obtenu, avec en plus des conditions diagonales :

$$T_{h,c} = h\,T_{h,c-1} + T_{h-1,c}$$

- Les nombres de Stirling de deuxième espèce satisfont

$$\left\{ {x \atop y} \right\} = y \left\{ {x-1 \atop y} \right\} + \left\{ {x-1 \atop y-1} \right\}$$

- On en déduit que

$$\left\{ {h+c \atop h} \right\} = h \left\{ {h+c-1 \atop h} \right\} + \left\{ {h-1+c \atop h-1} \right\}$$

- $T_{h,c}$ et $\left\{ {h+c \atop h} \right\}$ vérifient la même relation, avec des conditions initiales différentes.
- Si on enlève la condition diagonale, il y a égalité

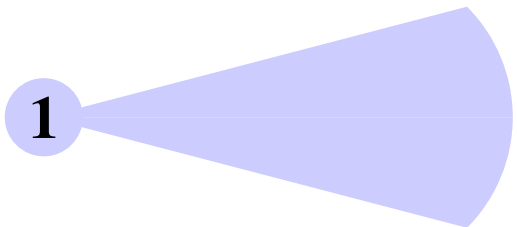
Partie V : Génération par l'intérieur

Idée initiale

- On tire au sort un k-uplet de random mapping
- Ce n'est presque jamais accessible

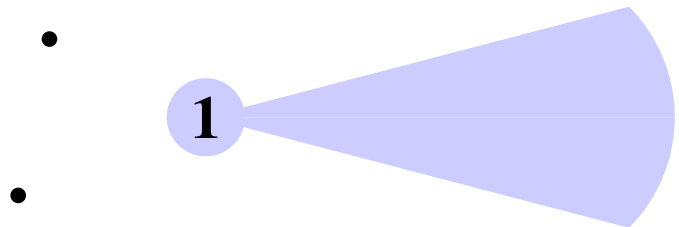
Idée initiale

- On tire au sort un k-uplet de random mapping
- Ce n'est presque jamais accessible
- Et si on regarde la partie accessible depuis 1 ?
- Si elle est de taille i on a



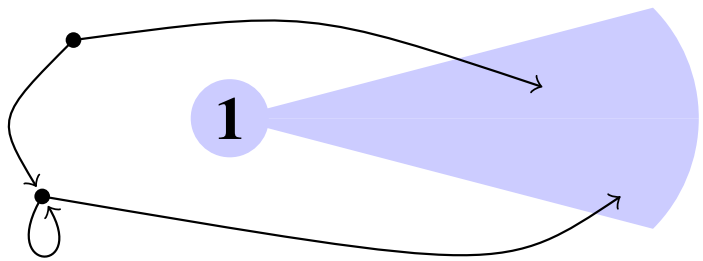
Idée initiale

- On tire au sort un k-uplet de random mapping
- Ce n'est presque jamais accessible
- Et si on regarde la partie accessible depuis 1 ?
- Si elle est de taille i on a

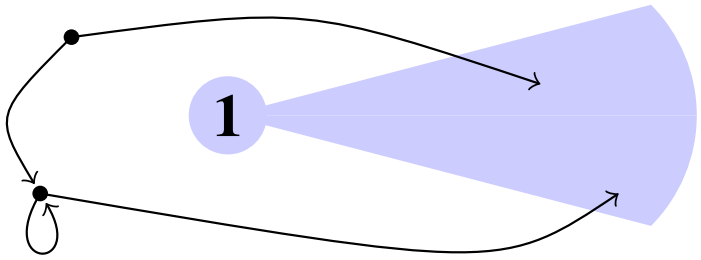


Idée initiale

- On tire au sort un k-uplet de random mapping
- Ce n'est presque jamais accessible
- Et si on regarde la partie accessible depuis 1 ?
- Si elle est de taille i on a



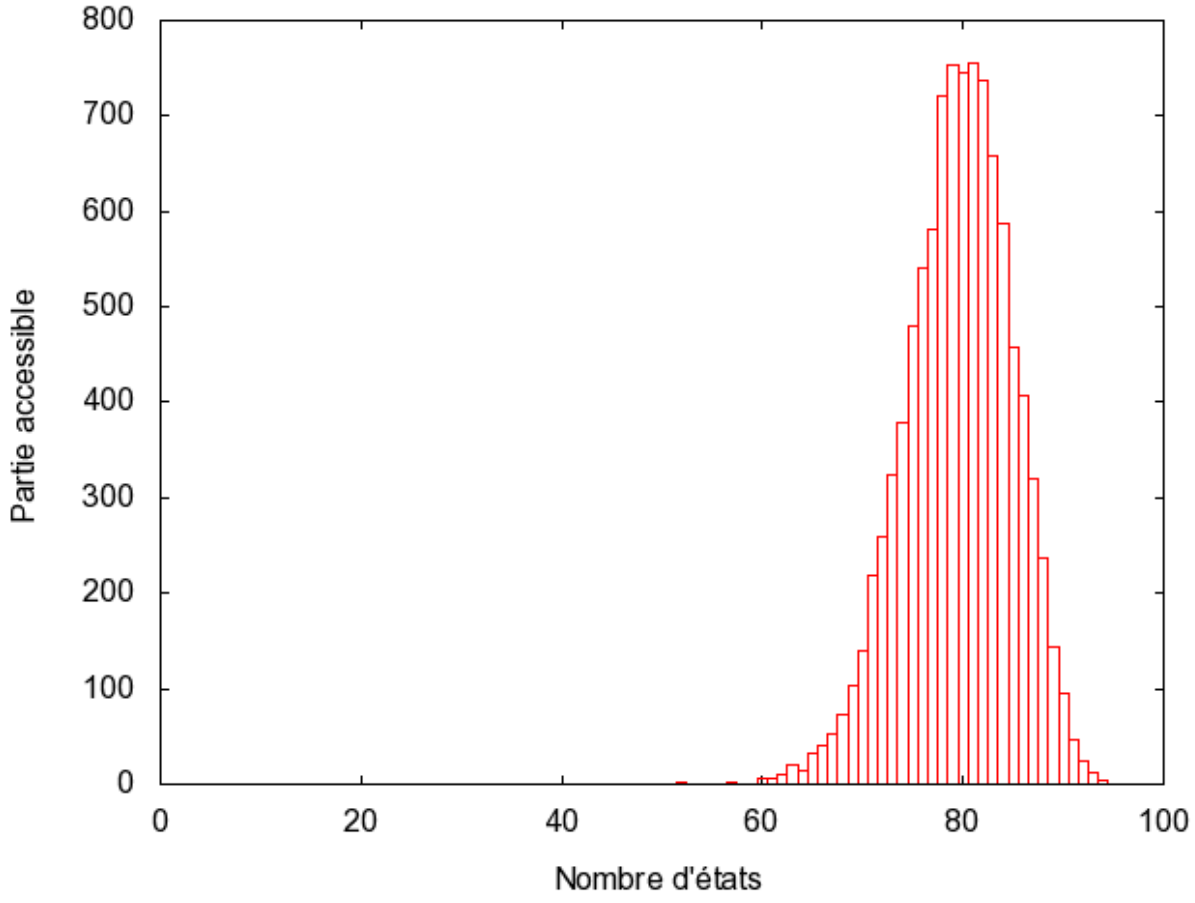
Formule de dénombrement



$$\underbrace{n^{kn}}_{k \text{ random mapping}} = \sum_{i=1}^{n} \underbrace{\binom{n-1}{i-1}}_{\text{accessibles}} \underbrace{(i-1)!A_i}_{\text{auto étiquetés}} \underbrace{n^{k(n-i)}}_{\text{autres transitions}}$$

- A i fixé c'est équiprobable ! (ne dépend pas de la forme de l'automate)

Expérimentons



Formule de dénombrement

$$\sum_{i=1}^{n} \frac{(n-1)!}{(n-i)!} A_i n^{k(n-i)} = n^{kn}$$

Formule de dénombrement

$$\sum_{i=1}^{n} \frac{(n-1)!}{(n-i)!} A_i n^{k(n-i)} = n^{kn}$$

$$\sum_{i=1}^{n} \frac{(n-1)!}{(n-i)!} A_i n^{-ik} = 1$$

Formule de dénombrement

$$\sum_{i=1}^{n} \frac{(n-1)!}{(n-i)!} A_i n^{k(n-i)} = n^{kn}$$

$$\sum_{i=1}^{n} \frac{(n-1)!}{(n-i)!} A_i n^{-ik} = 1$$

- On utilise l'approximation par les $\left\{ {kn+1 \atop n} \right\}$, la formule de Stirling, etc

$$\sum_{i=n/e}^{n-1} \frac{(n-1)!}{(n-i)!} A_i n^{-ik} \to 1$$

Formule de dénombrement

$$\sum_{i=1}^{n} \frac{(n-1)!}{(n-i)!} A_i n^{k(n-i)} = n^{kn}$$

$$\sum_{i=1}^{n} \frac{(n-1)!}{(n-i)!} A_i n^{-ik} = 1$$

- On utilise l'approximation par les $\left\{ {kn+1 \atop n} \right\}$, la formule de Stirling, etc

$$\sum_{i=n/e}^{n-1} \frac{(n-1)!}{(n-i)!} A_i n^{-ik} \to 1$$

$$\Theta(\sqrt{n}) \leq \sum_{i=n/e}^{n-1} g\left(\frac{i}{n}\right) f\left(\frac{i}{n}\right)^n \leq \Theta(n\sqrt{n})$$

- On trouve α tel que $f(\alpha) = 1$
- La taille de la partie accessible est concentrée autour de αn

Génération par l'intérieur

Théorème [Carayol N. en cours]

La méthode de génération par l'intérieur a une complexité moyenne de $\Theta(n^{3/2})$ en taille exacte et est linéaire en taille approchée.

- Il faut générer un k-uplet d'applications de taille $\frac{1}{\alpha}n$
- Pour un alphabet de taille 2 on trouve $\alpha \approx 0.7968$
- Il faut donc utiliser une taille $N \approx 1.255n$

Conclusion

- On a vu trois méthodes de génération aléatoire pour les automates
- C'était l'occasion d'étudier leur combinatoire

Conclusion

- On a vu trois méthodes de génération aléatoire pour les automates
- C'était l'occasion d'étudier leur combinatoire

- Tableaux : méthode récursive, avec précalcul en $\Theta(n^2)$ puis génération en $\Theta(n)$
- Partitions : Boltzmann, génération en temps moyen $\Theta(n^{3/2})$
- Par l'intérieur : temps moyen $\Theta(n^{3/2})$ en exact, linéaire en approché

Conclusion

- On a vu trois méthodes de génération aléatoire pour les automates
- C'était l'occasion d'étudier leur combinatoire

- Tableaux : méthode récursive, avec précalcul en $\Theta(n^2)$ puis génération en $\Theta(n)$
- Partitions : Boltzmann, génération en temps moyen $\Theta(n^{3/2})$
- Par l'intérieur : temps moyen $\Theta(n^{3/2})$ en exact, linéaire en approché

- On peut surement améliorer le générateur de partitions
- Travaux en cours de F. Bassino, J. David et A. Sportiello

FIN