Analyse en moyenne d'algorithmes en théorie des langages

Cyril Nicaud

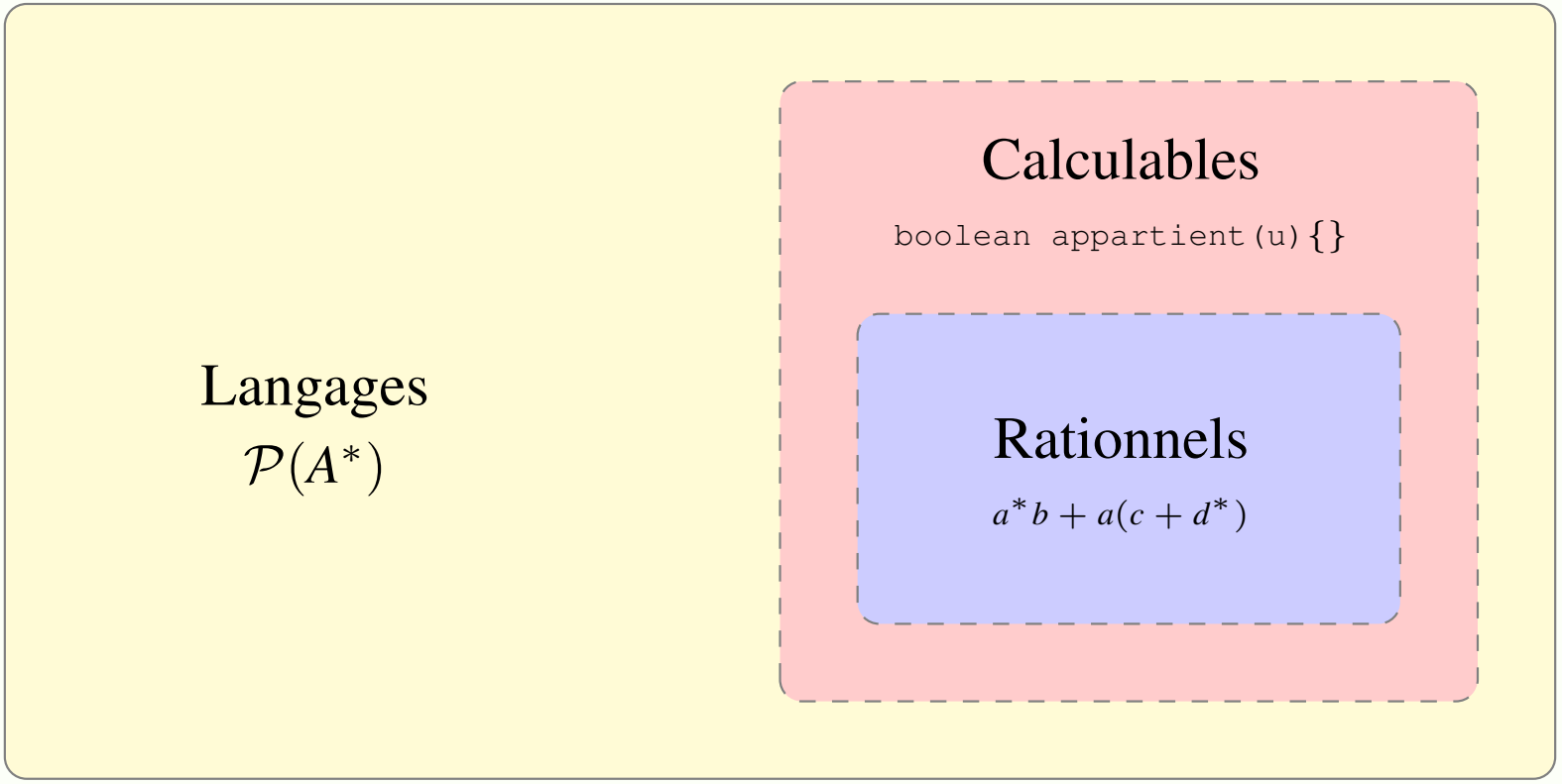
LIGM, Paris-Est

25 janvier 2011

Partie I : Introduction

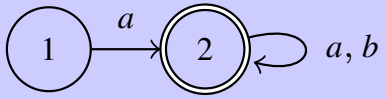
- A un alphabet fini : $A = \{a, b, c\}$
- A^* l'ensemble des mots sur A : $A^* = \{\varepsilon, a, b, c, aa, ab, \ldots\}$

- A un alphabet fini : $A = \{a, b, c\}$
- A^* l'ensemble des mots sur A : $A^* = \{\varepsilon, a, b, c, aa, ab, \ldots\}$

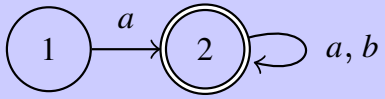


Expressions rationnelles : $a^*b + a(c + d^*)$

Expressions rationnelles : $a^*b + a(c + d^*)$

Automates finis : 

Expressions rationnelles : $a^*b + a(c + d^*)$

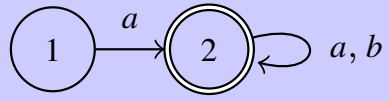
Automates finis : 

Formules logiques : $\forall x\ (a(x) \rightarrow (\exists y\ (y = x + 1) \wedge b(y)))$

Monoïdes : $L = \varphi^{-1}(P),\ \varphi : A^* \longrightarrow M$

$\ldots$

Expressions rationnelles : $a^*b + a(c + d^*)$

Automates finis : 

- On va s'intéresser à des algorithmes qui manipulent des expressions rationnelles ou des automates finis.

- Besoin d'une notion de taille : on note $\mathcal{C}_n$ l'ensemble des entrées de taille n.
- Pire des cas :

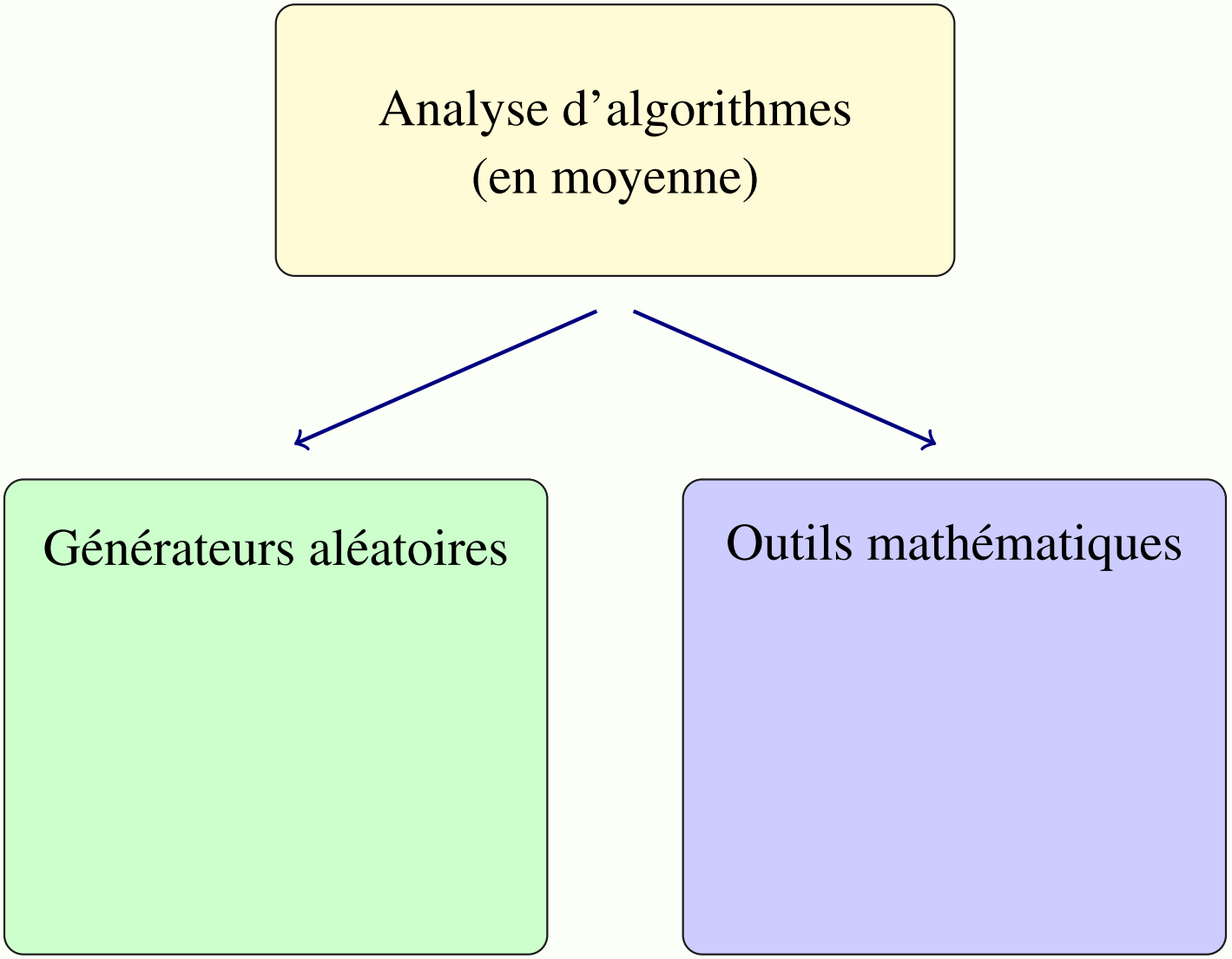
$$n \mapsto \max \{Temps(C) \mid C \in \mathcal{C}_n\}$$

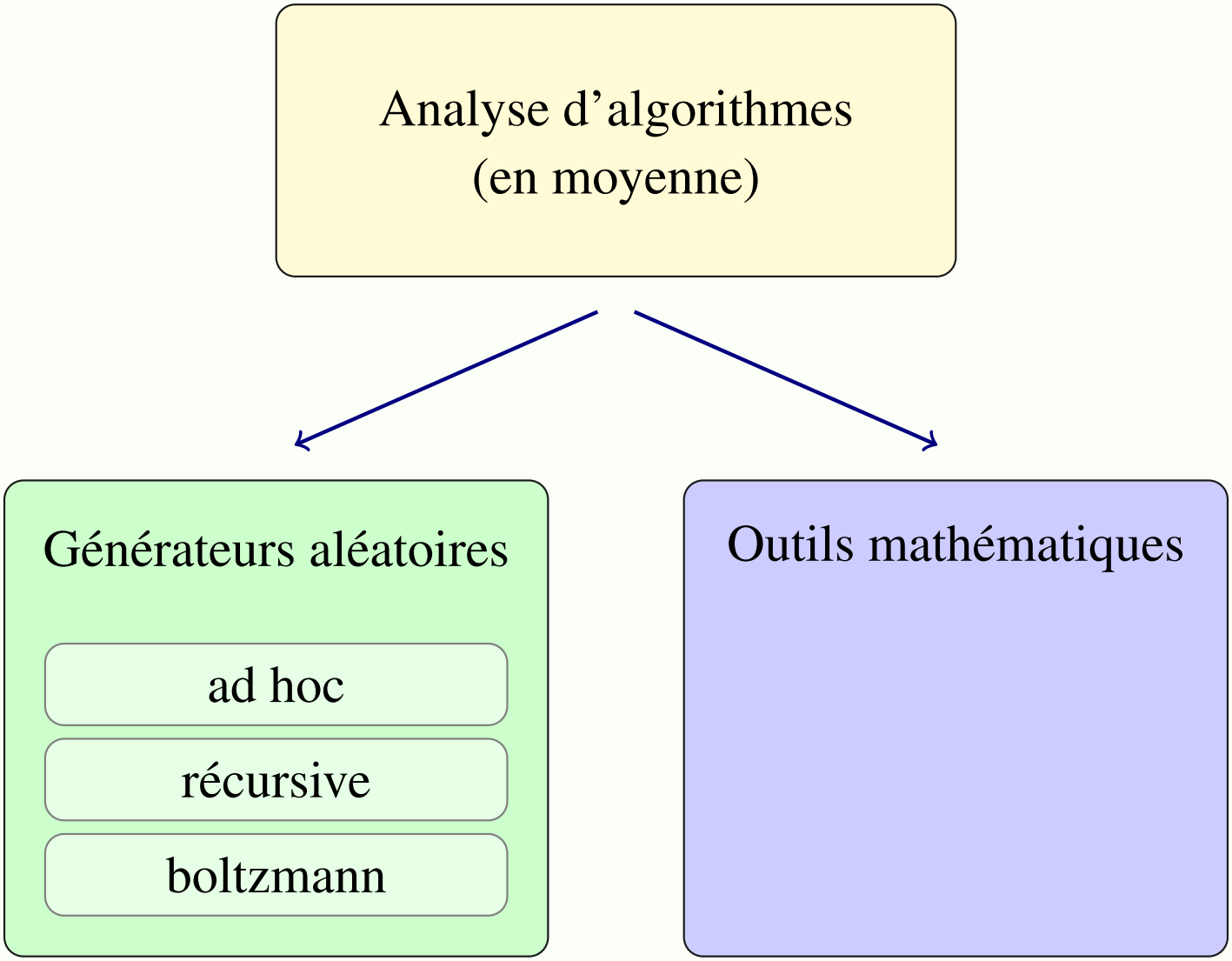
- En moyenne :

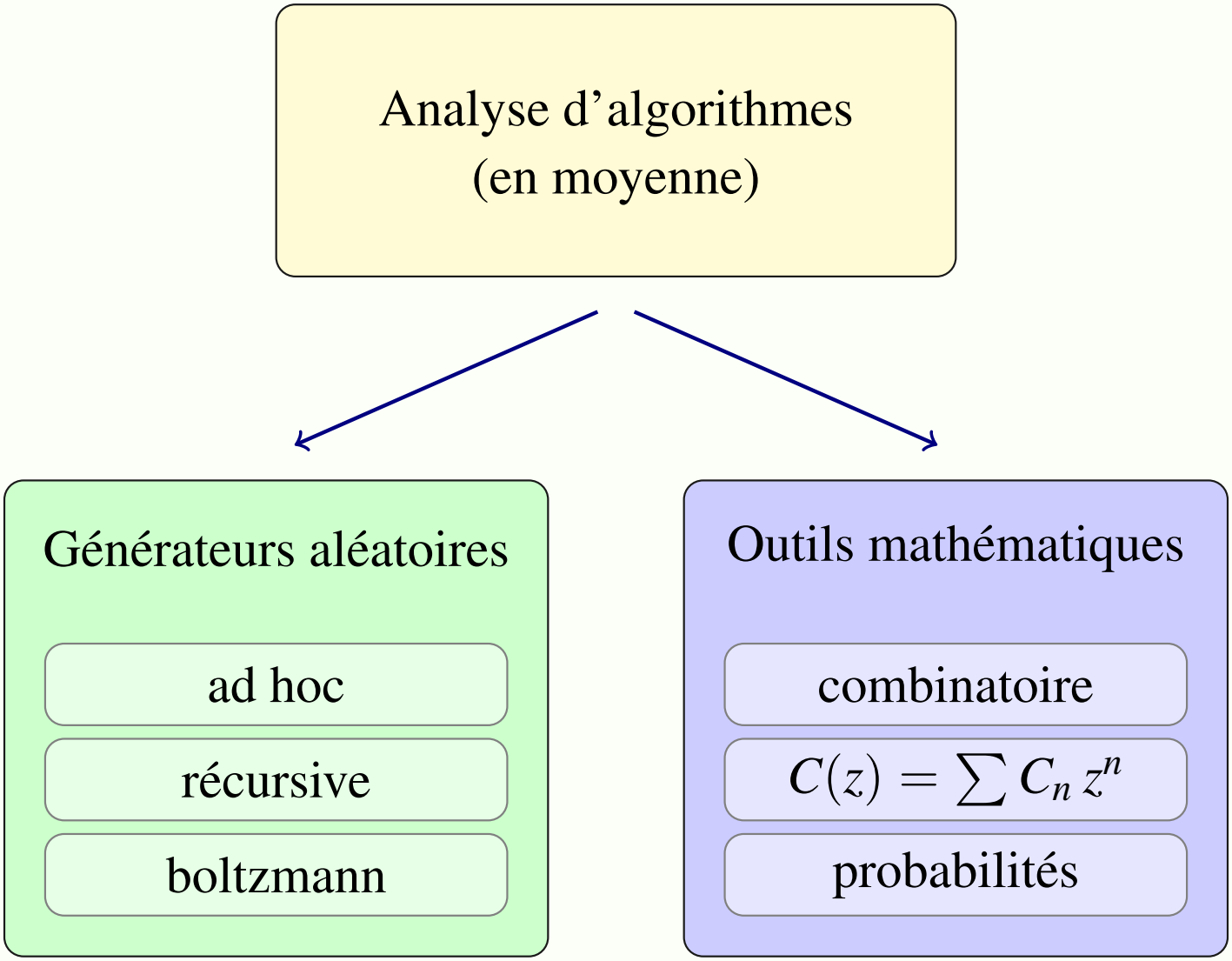
$$n \mapsto \sum_{C \in \mathcal{C}_n} p_n(C) \cdot Temps(C)$$

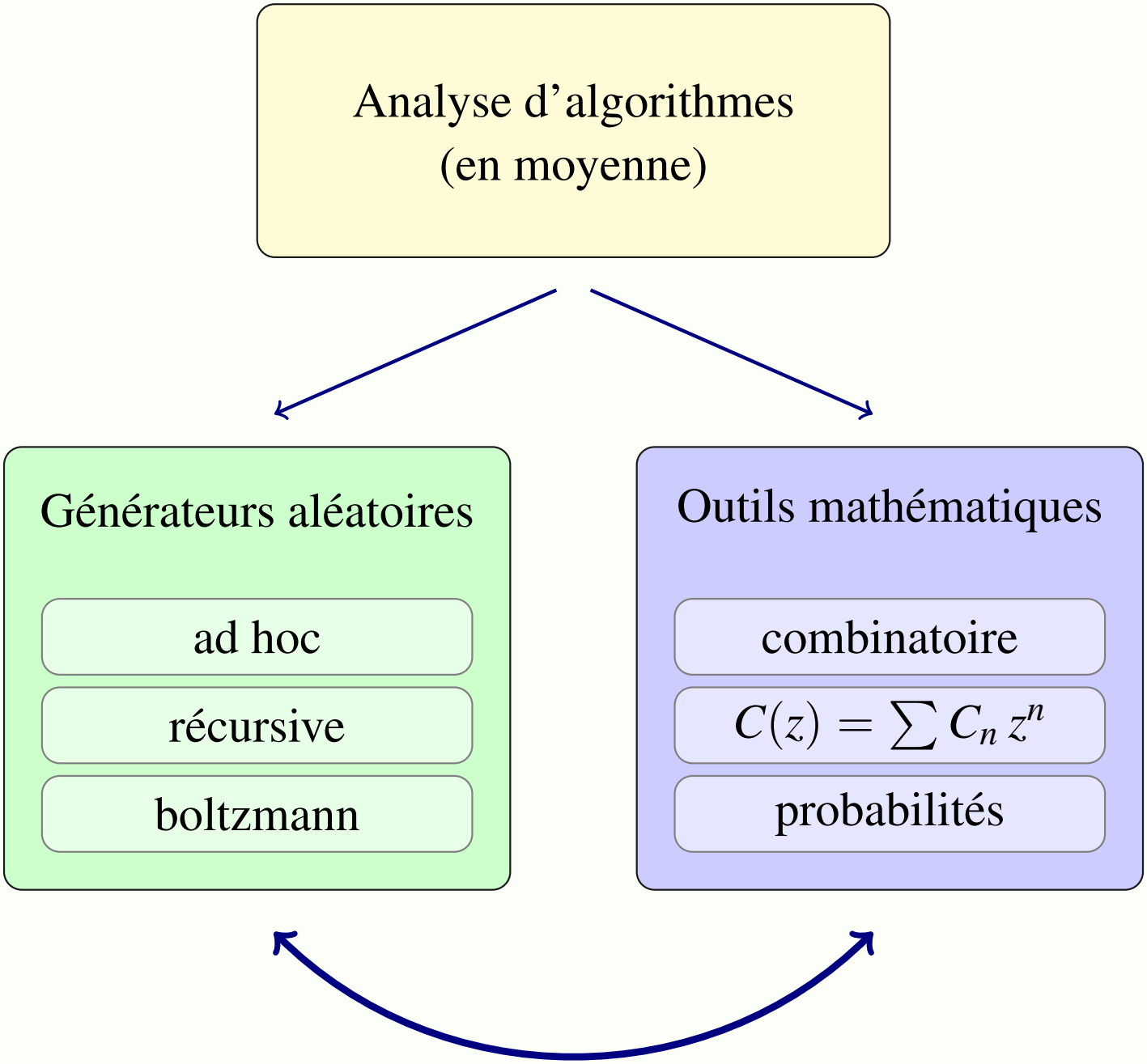
- En moyenne, pour la distribution uniforme :

$$n \mapsto \frac{1}{|\mathcal{C}_n|} \sum_{C \in \mathcal{C}_n} Temps(C)$$


Analyse d'algorithmes
(en moyenne)
Générateurs aléatoires
Outils mathématiques


Analyse d'algorithmes
(en moyenne)
Générateurs aléatoires
ad hoc
récursive
boltzmann
Outils mathématiques


Analyse d'algorithmes
(en moyenne)
Générateurs aléatoires
ad hoc
récursive
boltzmann
Outils mathématiques
combinatoire
$C(z) = \sum C_n z^n$
probabilités


Analyse d'algorithmes
(en moyenne)
Générateurs aléatoires
ad hoc
récursive
boltzmann
Outils mathématiques
combinatoire
$C(z) = \sum C_n z^n$
probabilités

a_n nombres ou probabilités

a_n nombres ou probabilités

$\downarrow$

$A(z) = \sum_{n \geq 0} a_n z^n$

série génératrice

a_n nombres ou probabilités

$\downarrow$

$A(z) = \sum_{n \geq 0} a_n z^n$

série génératrice

$\downarrow$

$A(z)$ fonction de $\mathbb{C} \longrightarrow \mathbb{C}$

a_n nombres ou probabilités

$\downarrow$

$A(z) = \sum_{n \geq 0} a_n z^n$

série génératrice

$\downarrow$

$A(z)$ fonction de $\mathbb{C} \longrightarrow \mathbb{C}$

$\downarrow$

$$\frac{A(z)}{z^{n+1}} = \frac{a_0}{z^{n+1}} + \frac{a_1}{z^n} + \cdots + \frac{a_n}{z} + \sum_{i \geq n+1} a_i z^{i-n-1}$$

a_n nombres ou probabilités

$\downarrow$

$A(z) = \sum_{n\geq 0} a_n z^n$

série génératrice

$\downarrow$

$A(z)$ fonction de $\mathbb{C} \longrightarrow \mathbb{C}$

$\downarrow$

$$\frac{A(z)}{z^{n+1}} = \frac{a_0}{z^{n+1}} + \frac{a_1}{z^n} + \cdots + \frac{a_n}{z} + \sum_{i\geq n+1} a_i z^{i-n-1}$$

$\downarrow$

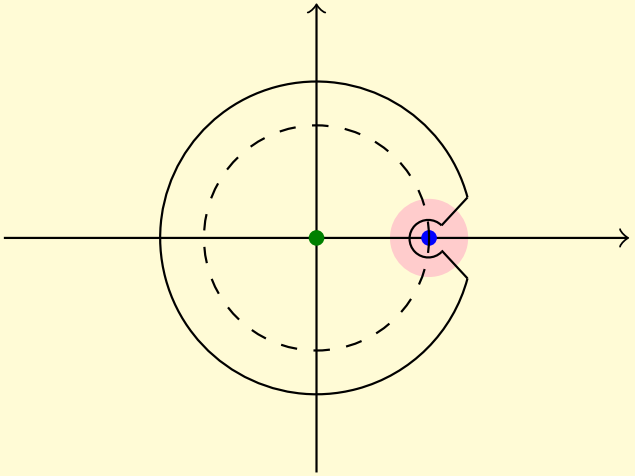
$$a_n = \frac{1}{2i\pi} \int_\gamma \frac{A(z)}{z^{n+1}} dz$$

$$a_n = \frac{1}{2i\pi} \int_\gamma \frac{A(z)}{z^{n+1}} dz$$

$$A(z) \sim (1-z)^{-\alpha}, \; z \to 1, \alpha \notin \{0, -1, -2, \cdots\}$$

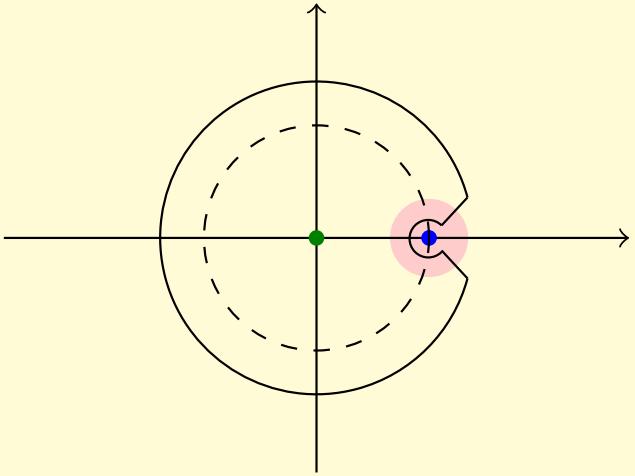
$$a_n = \frac{1}{2i\pi} \int_\gamma \frac{A(z)}{z^{n+1}} dz$$

$$A(z) \sim (1-z)^{-\alpha}, \; z \to 1, \alpha \notin \{0, -1, -2, \cdots\}$$



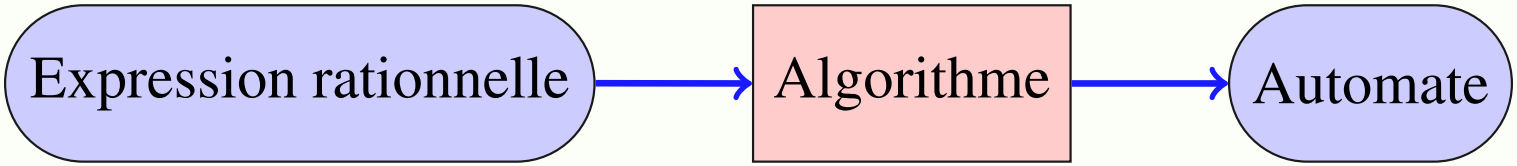
$$a_n = \frac{1}{2i\pi} \int_\gamma \frac{A(z)}{z^{n+1}} dz$$

$$A(z) \sim (1-z)^{-\alpha}, \; z \to 1, \alpha \notin \{0, -1, -2, \cdots\}$$

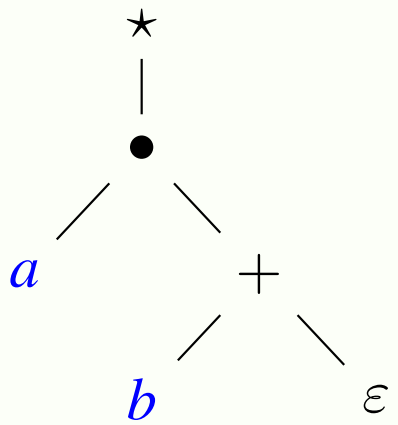


$$a_n \sim \frac{n^{\alpha-1}}{\Gamma(\alpha)}$$

Partie II : automate de Glushkov

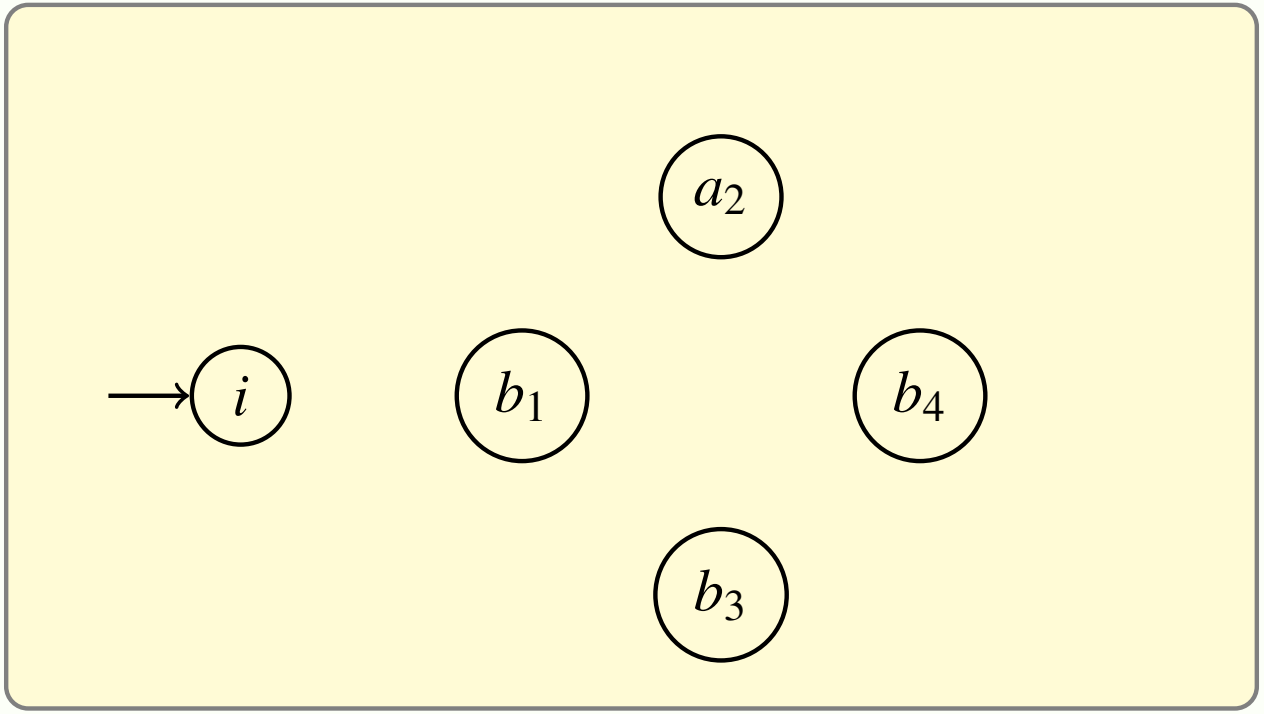


On voit une expression rationnelle comme un arbre :



- Arbre de $(a \cdot (b + \varepsilon))^*$
- La taille de l'expression est 6

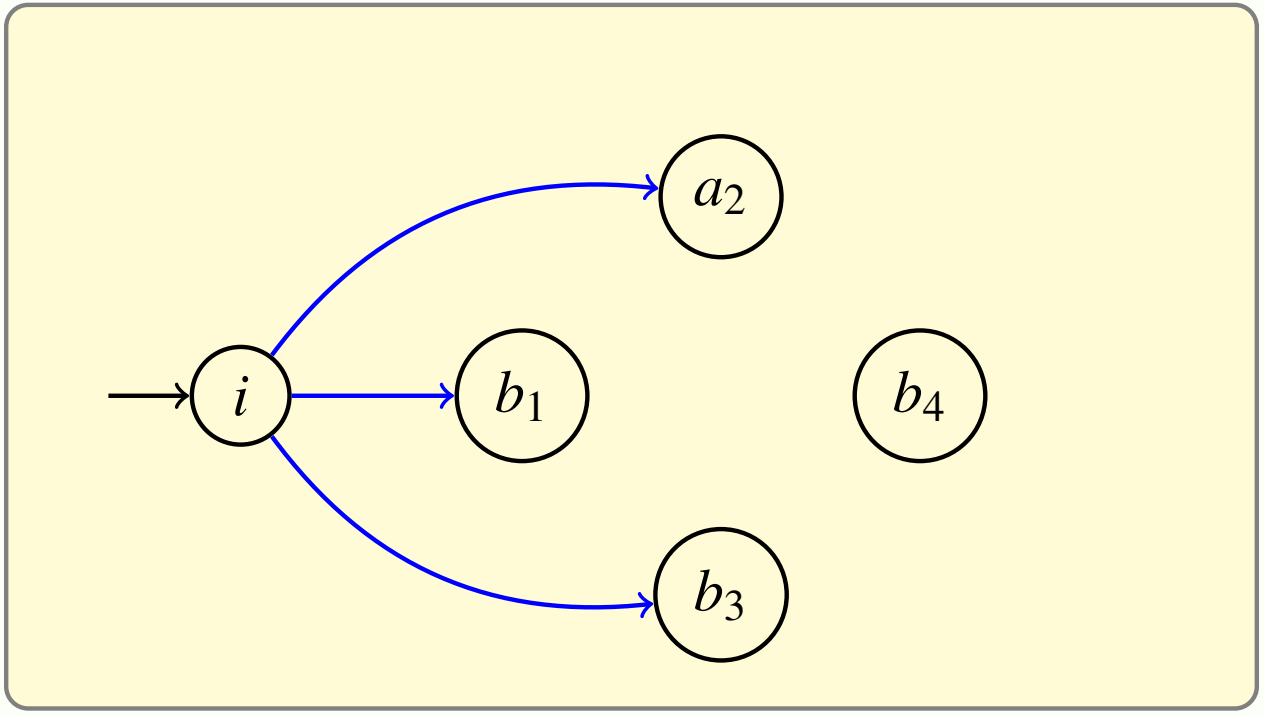
$E = b^* \cdot (a + b \cdot b)^*$, on distingue les lettres : $\tilde{E} = b_1^* \cdot (a_2 + b_3 \cdot b_4)^*$



Ensembles

- $First(\tilde{E}) = \{\alpha \mid \alpha \text{ commence un mot de } \tilde{L}\}$
- $Last(\tilde{E}) = \{\alpha \mid \alpha \text{ termine un mot de } \tilde{L}\}$
- $Follow(\tilde{E}) = \{\alpha \to \beta \mid \alpha\beta \text{ est facteur d'un mot de } \tilde{L}\}$

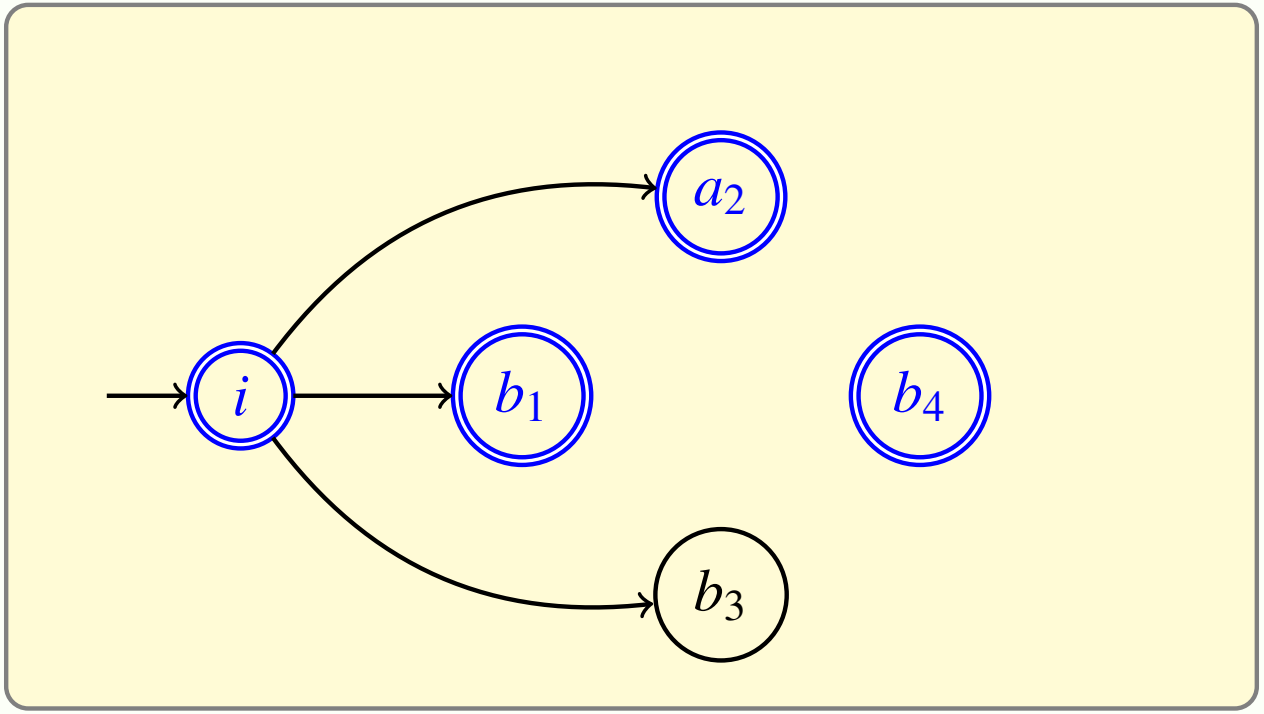
$E = b^* \cdot (a + b \cdot b)^*$, on distingue les lettres : $\tilde{E} = b_1^* \cdot (a_2 + b_3 \cdot b_4)^*$



Ensembles

- $First(\tilde{E}) = \{\alpha \mid \alpha \text{ commence un mot de } \tilde{L}\}$
- $Last(\tilde{E}) = \{\alpha \mid \alpha \text{ termine un mot de } \tilde{L}\}$
- $Follow(\tilde{E}) = \{\alpha \to \beta \mid \alpha\beta \text{ est facteur d'un mot de } \tilde{L}\}$

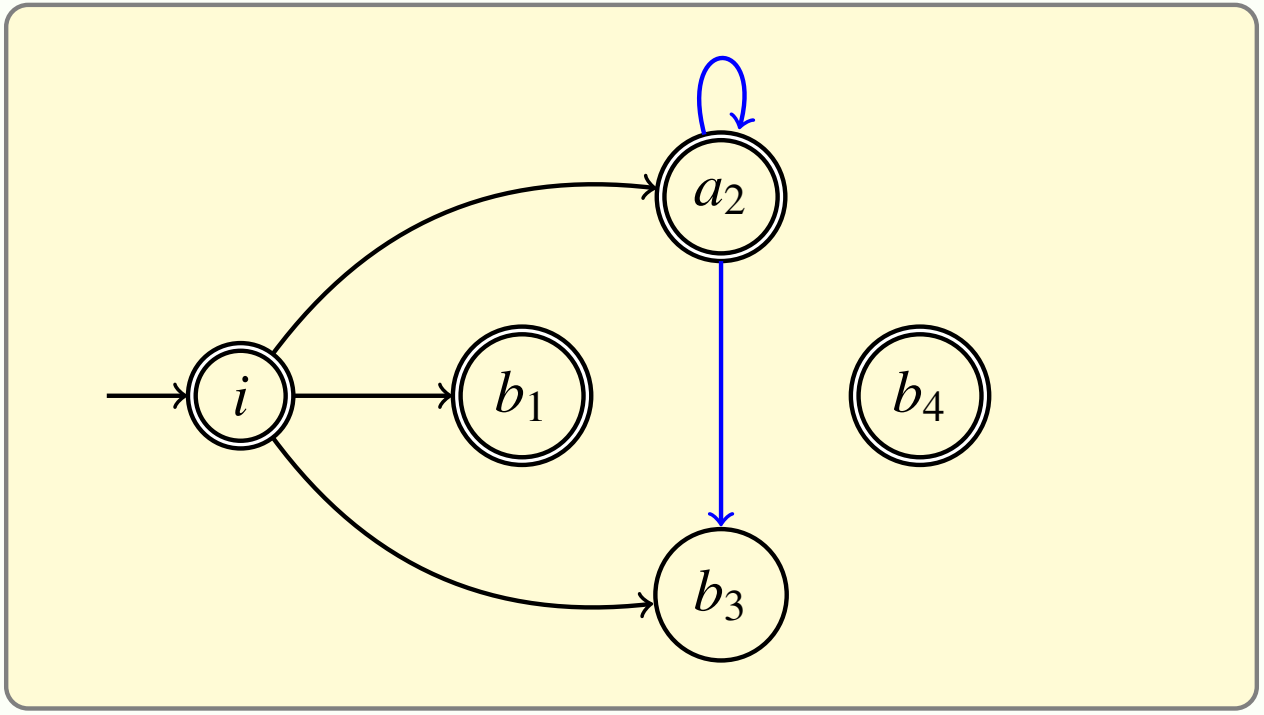
$E = b^* \cdot (a + b \cdot b)^*$, on distingue les lettres : $\tilde{E} = b_1^* \cdot (a_2 + b_3 \cdot b_4)^*$



Ensembles

- $First(\tilde{E}) = \{\alpha \mid \alpha \text{ commence un mot de } \tilde{L}\}$
- $Last(\tilde{E}) = \{\alpha \mid \alpha \text{ termine un mot de } \tilde{L}\}$
- $Follow(\tilde{E}) = \{\alpha \to \beta \mid \alpha\beta \text{ est facteur d'un mot de } \tilde{L}\}$

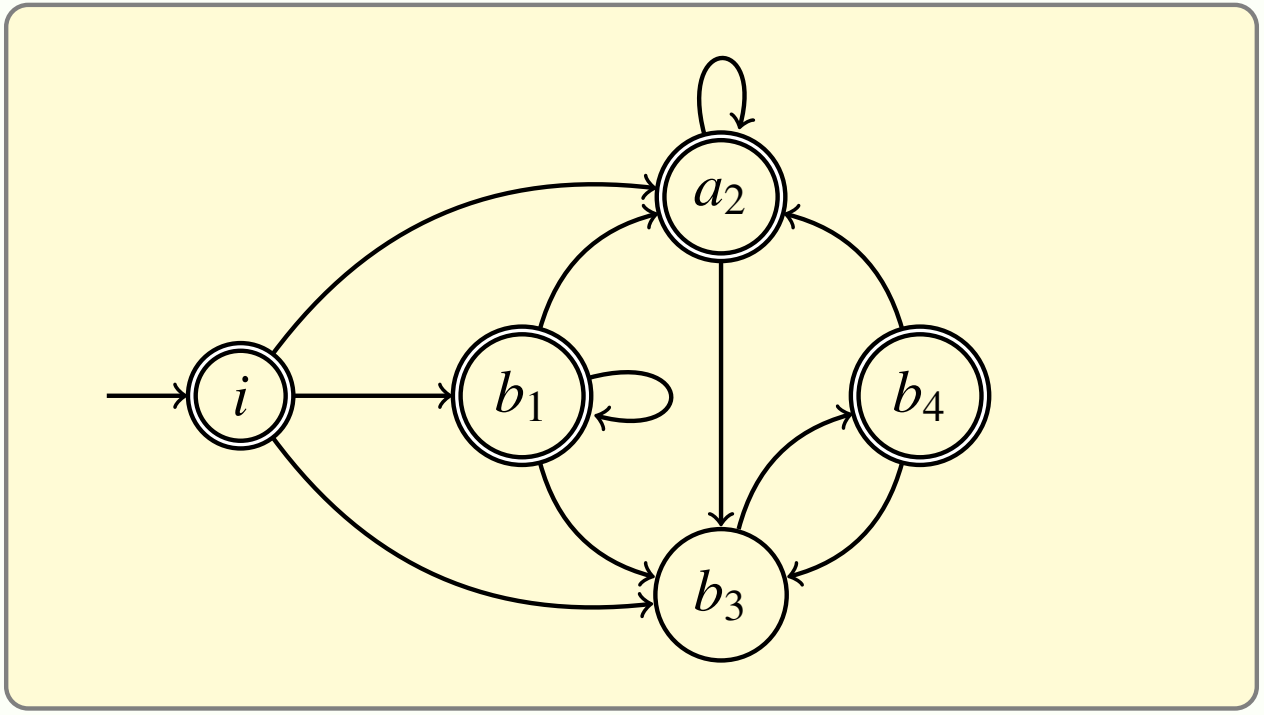
$E = b^* \cdot (a + b \cdot b)^*$, on distingue les lettres : $\tilde{E} = b_1^* \cdot (a_2 + b_3 \cdot b_4)^*$



Ensembles

- $First(\tilde{E}) = \{\alpha \mid \alpha \text{ commence un mot de } \tilde{L}\}$
- $Last(\tilde{E}) = \{\alpha \mid \alpha \text{ termine un mot de } \tilde{L}\}$
- $Follow(\tilde{E}) = \{\alpha \to \beta \mid \alpha\beta \text{ est facteur d'un mot de } \tilde{L}\}$

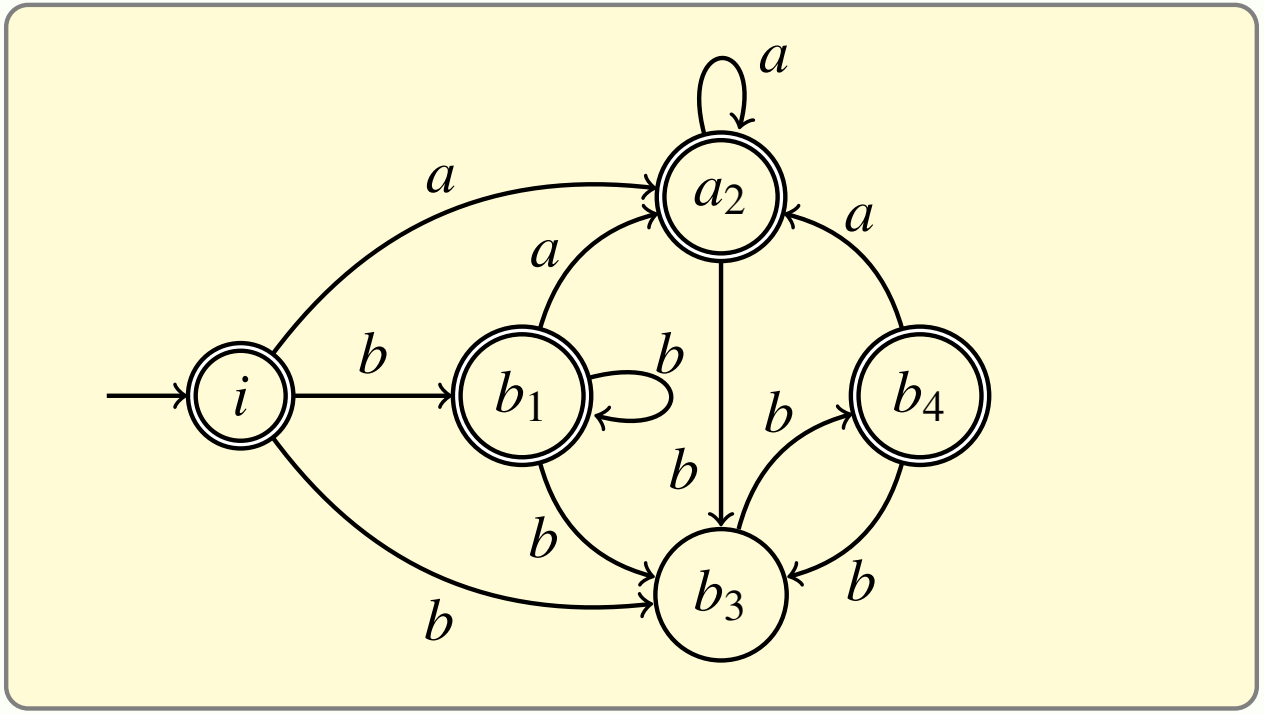
$E = b^* \cdot (a + b \cdot b)^*$, on distingue les lettres : $\tilde{E} = b_1^* \cdot (a_2 + b_3 \cdot b_4)^*$



Ensembles

- $First(\tilde{E}) = \{\alpha \mid \alpha \text{ commence un mot de } \tilde{L}\}$
- $Last(\tilde{E}) = \{\alpha \mid \alpha \text{ termine un mot de } \tilde{L}\}$
- $Follow(\tilde{E}) = \{\alpha \to \beta \mid \alpha\beta \text{ est facteur d'un mot de } \tilde{L}\}$

$E = b^* \cdot (a + b \cdot b)^*$, on distingue les lettres : $\tilde{E} = b_1^* \cdot (a_2 + b_3 \cdot b_4)^*$



Ensembles

- $First(\tilde{E}) = \{\alpha \mid \alpha \text{ commence un mot de } \tilde{L}\}$
- $Last(\tilde{E}) = \{\alpha \mid \alpha \text{ termine un mot de } \tilde{L}\}$
- $Follow(\tilde{E}) = \{\alpha \to \beta \mid \alpha\beta \text{ est facteur d'un mot de } \tilde{L}\}$

Complexités :

- Espace $O(n^2)$ [Glushkov 61] temps $O(n^2)$.

Complexités :

- Espace $O(n^2)$ [Glushkov 61] temps $O(n^2)$.

- Espace $O(n \log^2 n)$ [Hromkovic, Seibert, Wilke 97].

Complexités :

- Espace $O(n^2)$ [Glushkov 61] temps $O(n^2)$.
- Espace $O(n \log^2 n)$ [Hromkovic, Seibert, Wilke 97].
- Temps $O(n \log^2 n)$, [Hagenah, Muscholl 98].

Complexités :

- Espace $O(n^2)$ [Glushkov 61] temps $O(n^2)$.
- Espace $O(n \log^2 n)$ [Hromkovic, Seibert, Wilke 97].
- Temps $O(n \log^2 n)$, [Hagenah, Muscholl 98].
- Espace $\Omega(n \log n)$.

Pire des cas

Le nombre de transitions est $First + Follow$, dans le pire des cas, il y en a $\Theta(n^2)$.

Pire des cas

Le nombre de transitions est $First + Follow$, dans le pire des cas, il y en a $\Theta(n^2)$.

Théorème [N. 2009]

Pour la distribution uniforme sur les expressions, le nombre moyen de transitions est en $\Theta(n)$.

Pire des cas

Le nombre de transitions est $First + Follow$, dans le pire des cas, il y en a $\Theta(n^2)$.

Théorème [N. 2009]

Pour la distribution uniforme sur les expressions, le nombre moyen de transitions est en $\Theta(n)$.

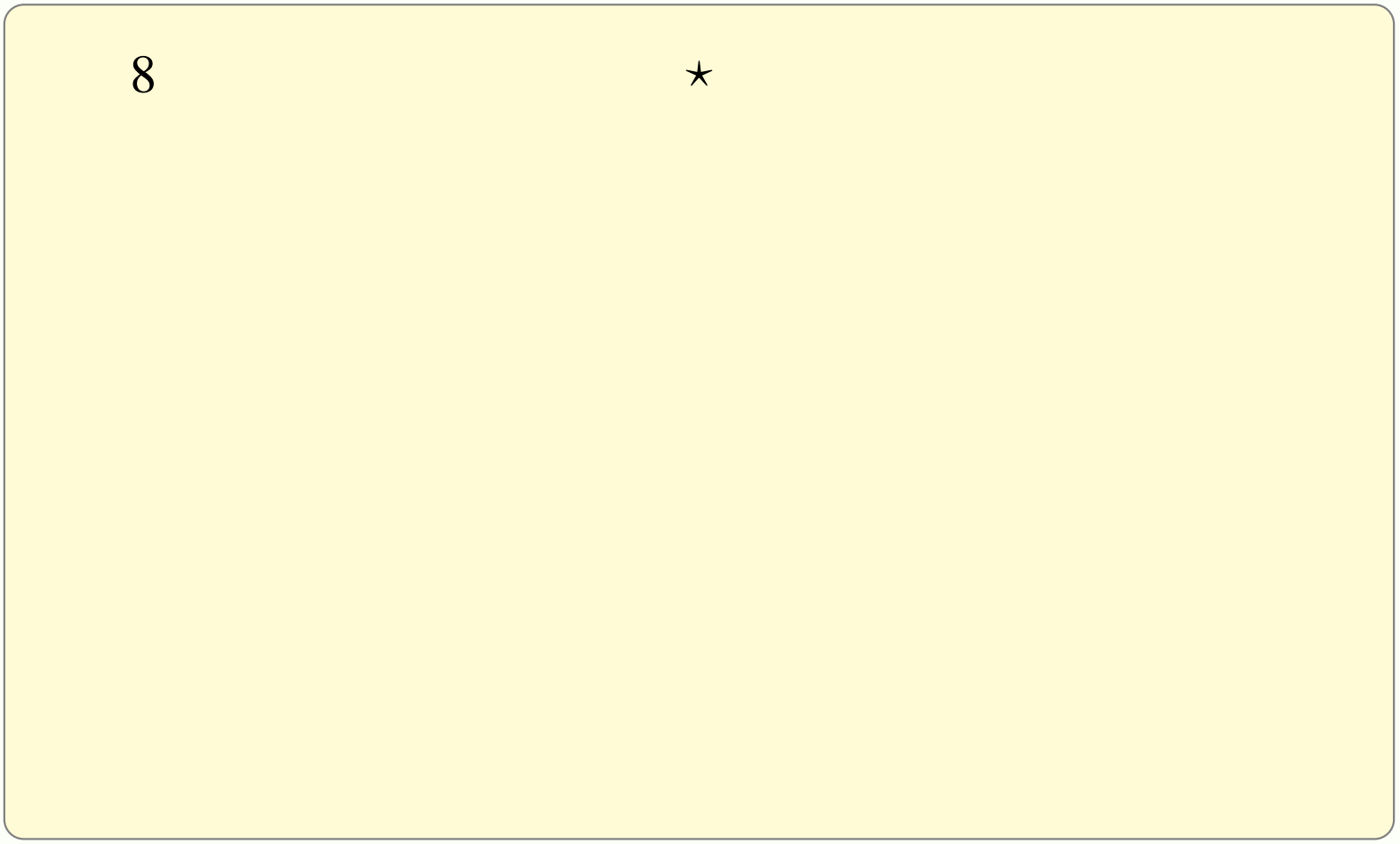
Théorème [N., Pivoteau, Razet 2010]

Pour la distribution “ABR”, le nombre moyen de transitions est en $\Theta(n^2)$.

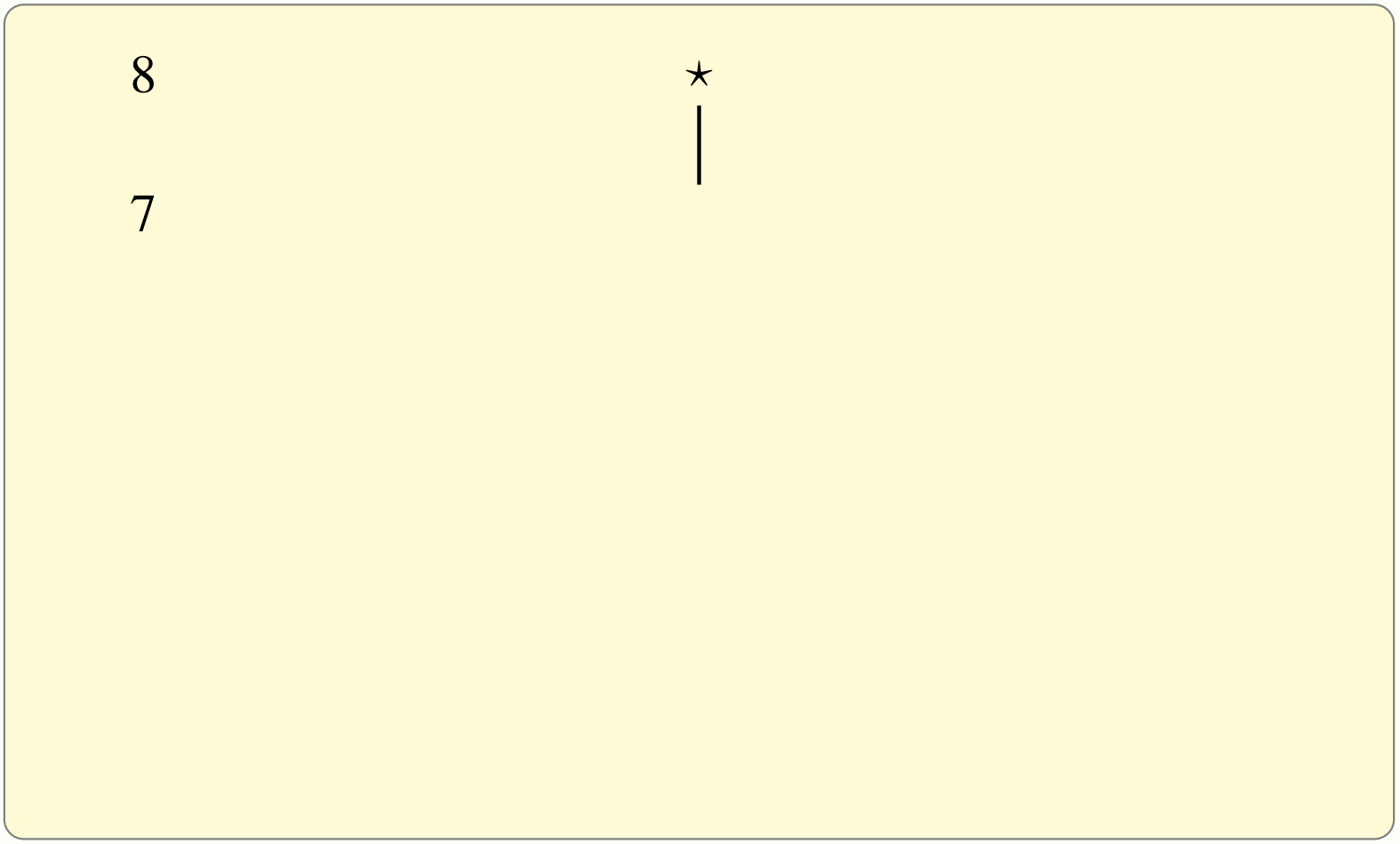
Une feuille a probabilité p_ε d'être le mot vide. Un nœud interne, probabilité $p_\star$ d'être une étoile.

8

Une feuille a probabilité p_ε d'être le mot vide. Un nœud interne, probabilité $p_\star$ d'être une étoile.



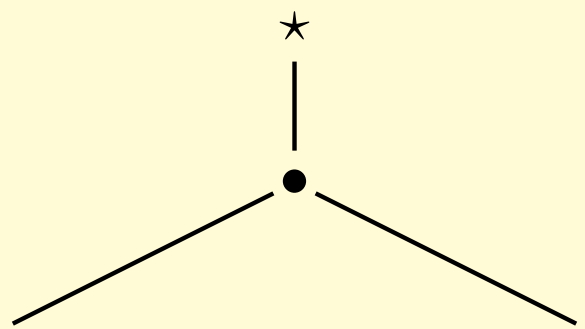
Une feuille a probabilité p_ε d'être le mot vide. Un nœud interne, probabilité $p_\star$ d'être une étoile.



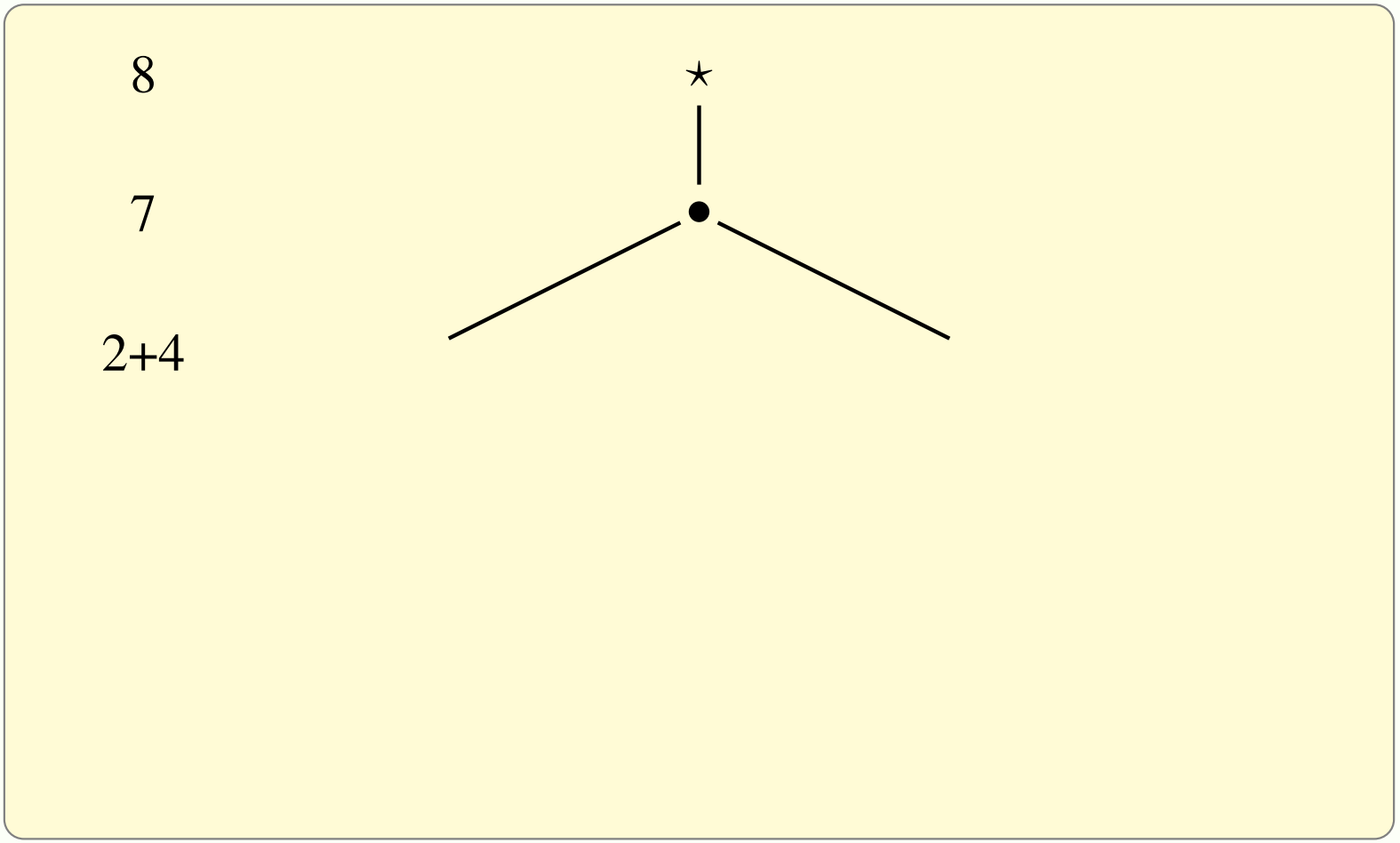
Une feuille a probabilité p_ε d'être le mot vide. Un nœud interne, probabilité $p_\star$ d'être une étoile.

8

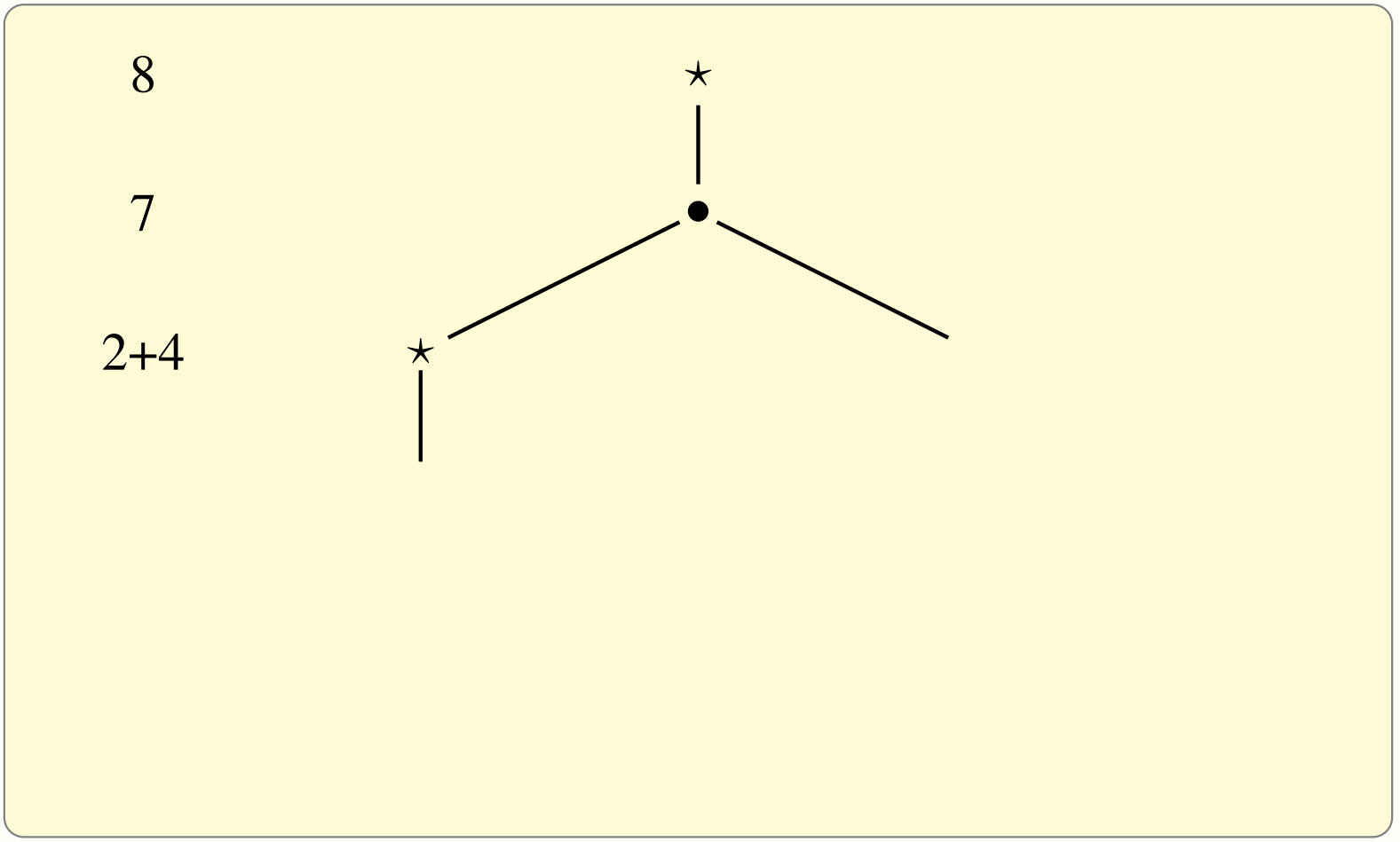
7



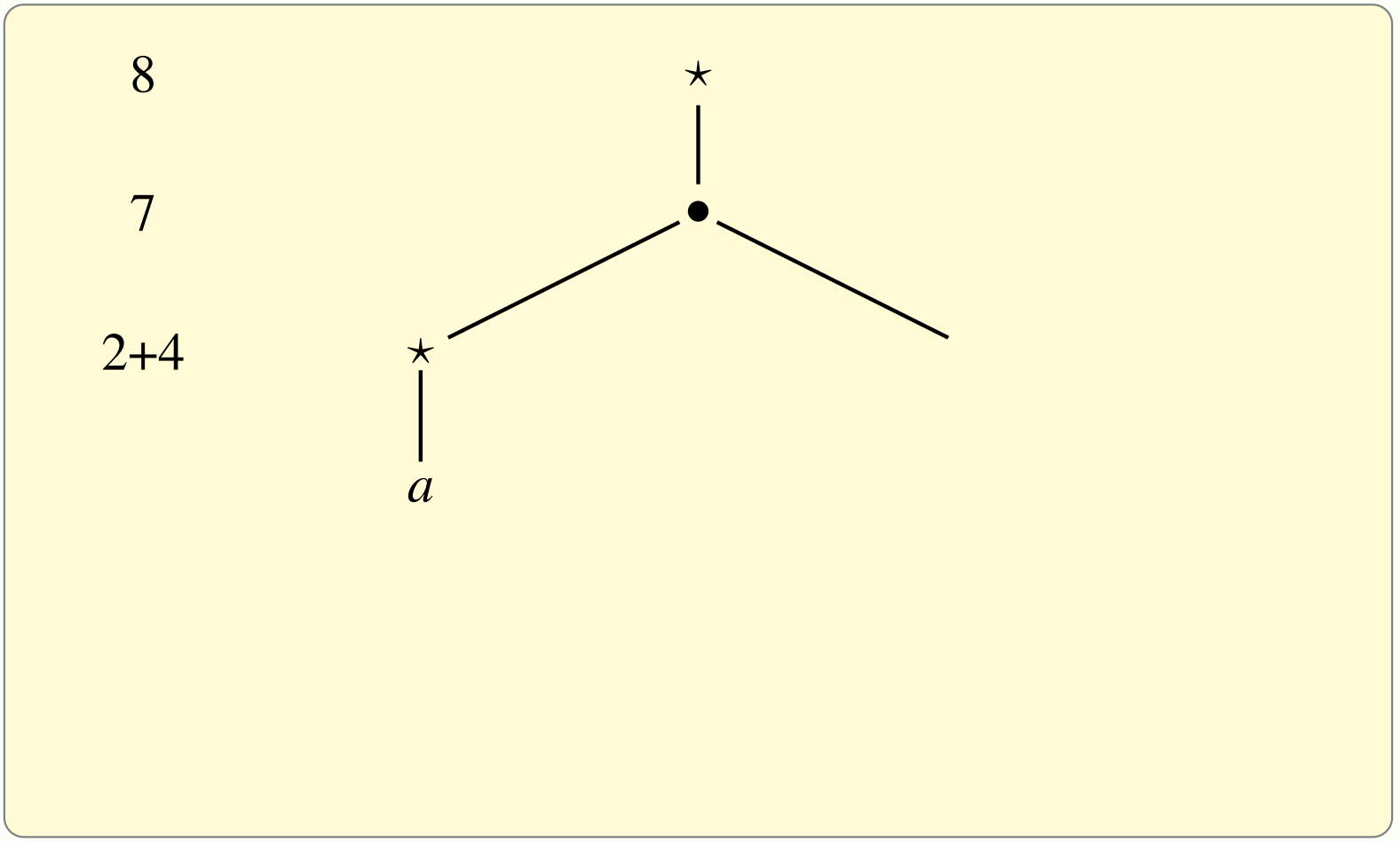
Une feuille a probabilité p_ε d'être le mot vide. Un nœud interne, probabilité $p_\star$ d'être une étoile.



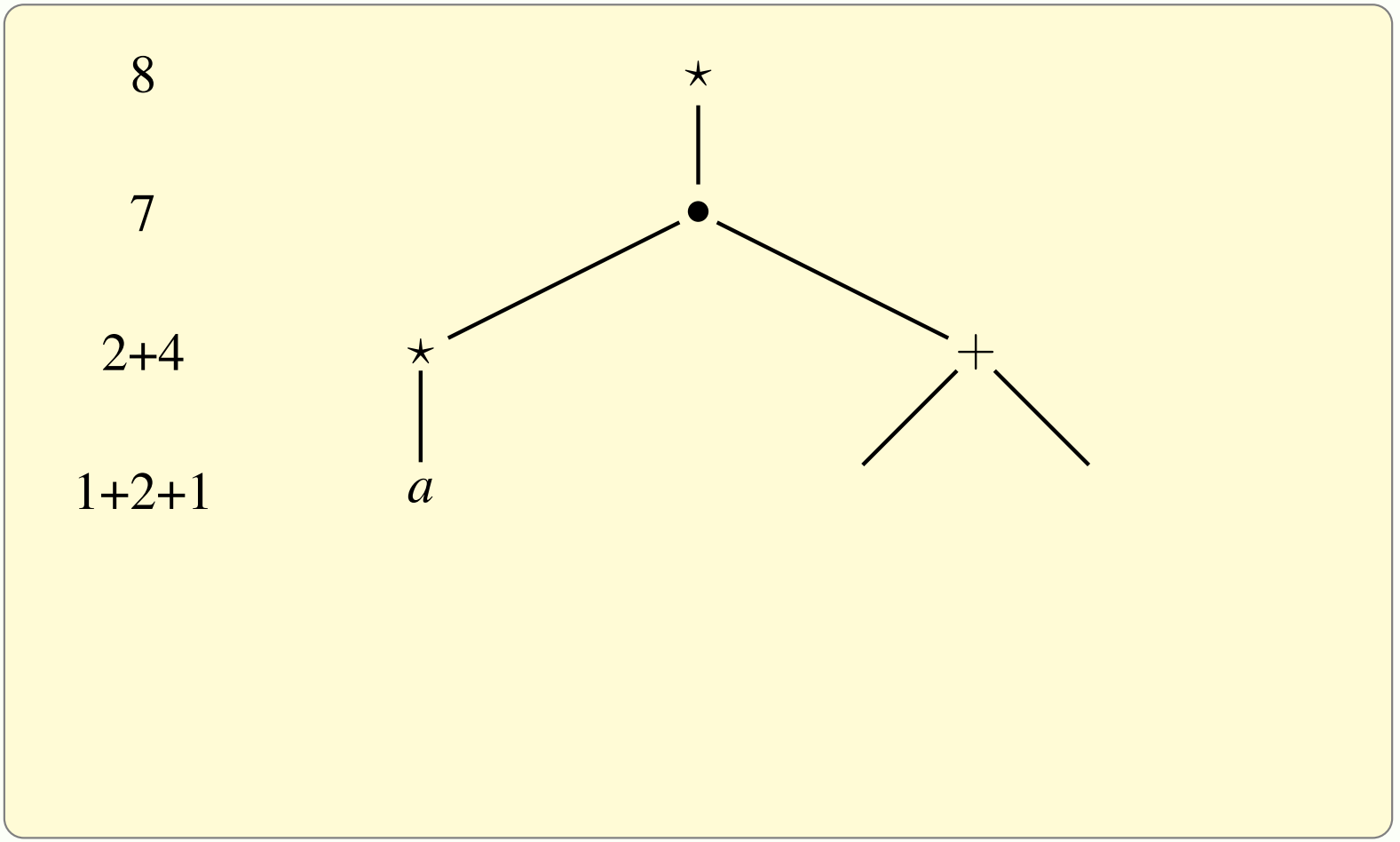
Une feuille a probabilité p_ε d'être le mot vide. Un nœud interne, probabilité $p_\star$ d'être une étoile.



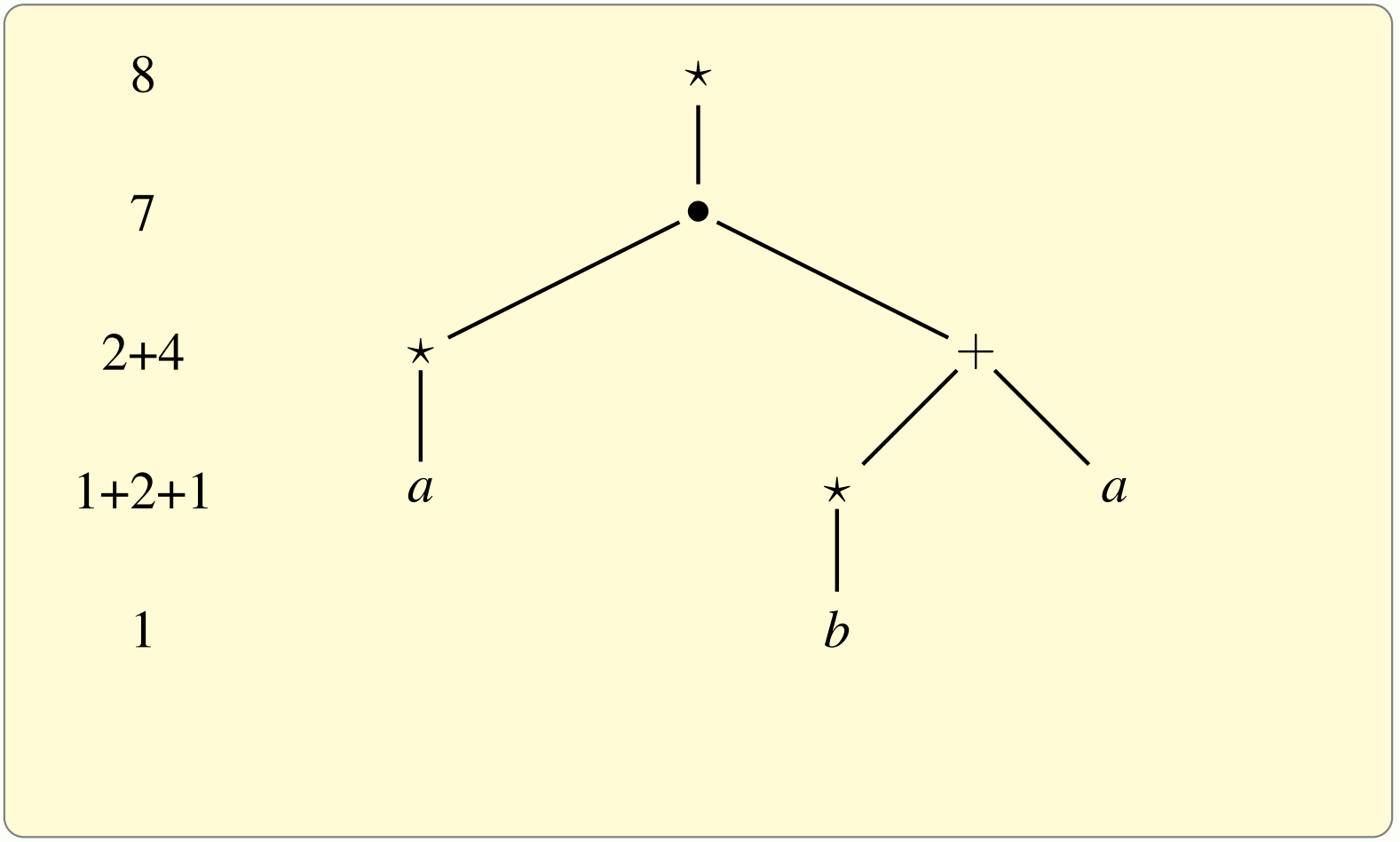
Une feuille a probabilité p_ε d'être le mot vide. Un nœud interne, probabilité $p_\star$ d'être une étoile.

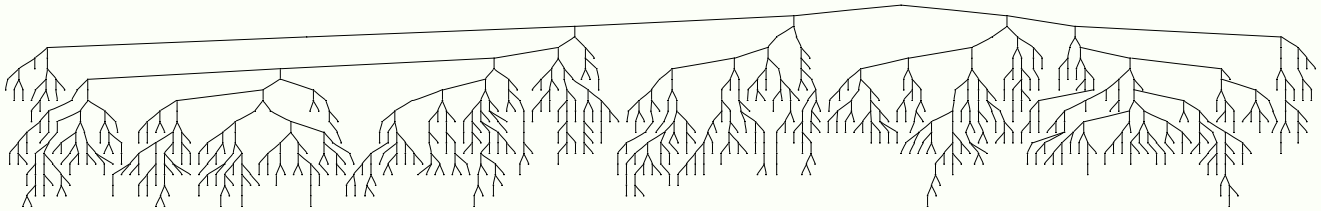


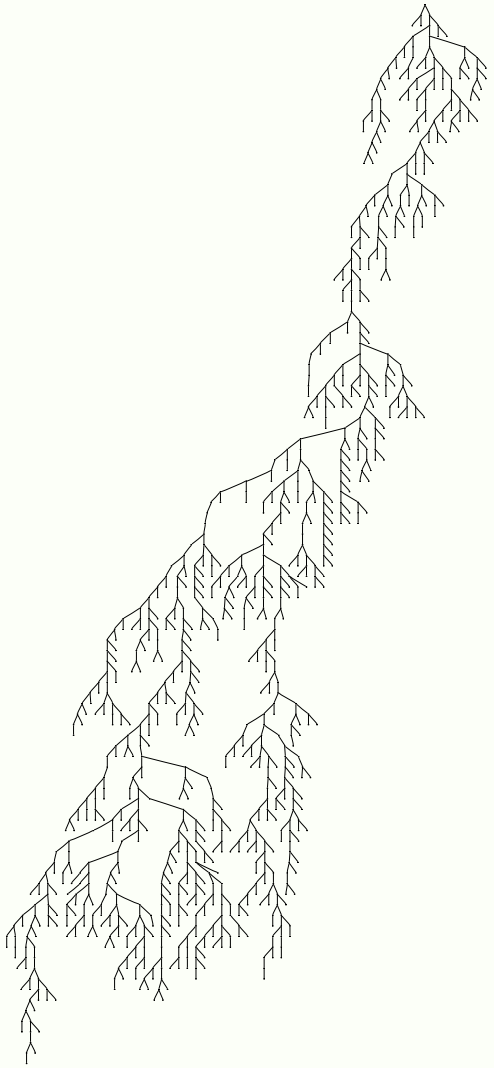
Une feuille a probabilité p_ε d'être le mot vide. Un nœud interne, probabilité $p_\star$ d'être une étoile.



Une feuille a probabilité p_ε d'être le mot vide. Un nœud interne, probabilité $p_\star$ d'être une étoile.







- En haut, distribution ABR (1000 nœuds)
- A gauche, distribution uniforme (1021 nœuds)

Uniforme

ABR

Uniforme

Hauteur $\Theta(\sqrt{n})$

ABR

Hauteur $\Theta(\log n)$

Uniforme

Hauteur $\Theta(\sqrt{n})$

$$T(z) = (k+1)z + zT(z) + 2zT^2(z)$$

ABR

Hauteur $\Theta(\log n)$

$$zR'(z) + \lambda z = \frac{(1-p_\star)z^2 - 2z + 2}{1-z} R(z)$$

Uniforme

Hauteur $\Theta(\sqrt{n})$

$$T(z) = (k+1)z + zT(z) + 2zT^2(z)$$

$$T(z) = \frac{1-z-\sqrt{1-2z-(7+8k)z^2}}{4z}$$

ABR

Hauteur $\Theta(\log n)$

$$zR'(z) + \lambda z = \frac{(1-p_\star)z^2-2z+2}{1-z}R(z)$$

$$R(z) = \lambda(1 - z\int_0^1 g)ze^{(1-p_\star)z}(1-z)^{p_\star-1}$$

Uniforme

Hauteur $\Theta(\sqrt{n})$

$$T(z) = (k+1)z + zT(z) + 2zT^2(z)$$

$$T(z) = \frac{1-z-\sqrt{1-2z-(7+8k)z^2}}{4z}$$

Théorèmes de transfert

ABR

Hauteur $\Theta(\log n)$

$$zR'(z) + \lambda z = \frac{(1-p_\star)z^2-2z+2}{1-z}R(z)$$

$$R(z) = \lambda(1 - z\int_0^1 g)ze^{(1-p_\star)z}(1-z)^{p_\star-1}$$

Théorèmes de transfert

Uniforme

Hauteur $\Theta(\sqrt{n})$

$$T(z) = (k+1)z + zT(z) + 2zT^2(z)$$

$$T(z) = \frac{1-z-\sqrt{1-2z-(7+8k)z^2}}{4z}$$

Théorèmes de transfert

$\text{Proba}(\varepsilon \notin L) = \Theta(1)$

Taille de First $= \Theta(1)$

Nbr transitions $= \Theta(n)$

ABR

Hauteur $\Theta(\log n)$

$$zR'(z) + \lambda z = \frac{(1-p_\star)z^2 - 2z + 2}{1-z} R(z)$$

$$R(z) = \lambda(1 - z\int_0^1 g)ze^{(1-p_\star)z}(1-z)^{p_\star - 1}$$

Théorèmes de transfert

$\text{Proba}(\varepsilon \notin L) = o(1)$

Taille de First $= \Theta(n)$

Nbr transitions $= \Theta(n^2)$

- On a utilisé de la combinatoire analytique pour obtenir les résultats.

- On a utilisé de la combinatoire analytique pour obtenir les résultats.
- Les deux distributions donnent des résultats différents.

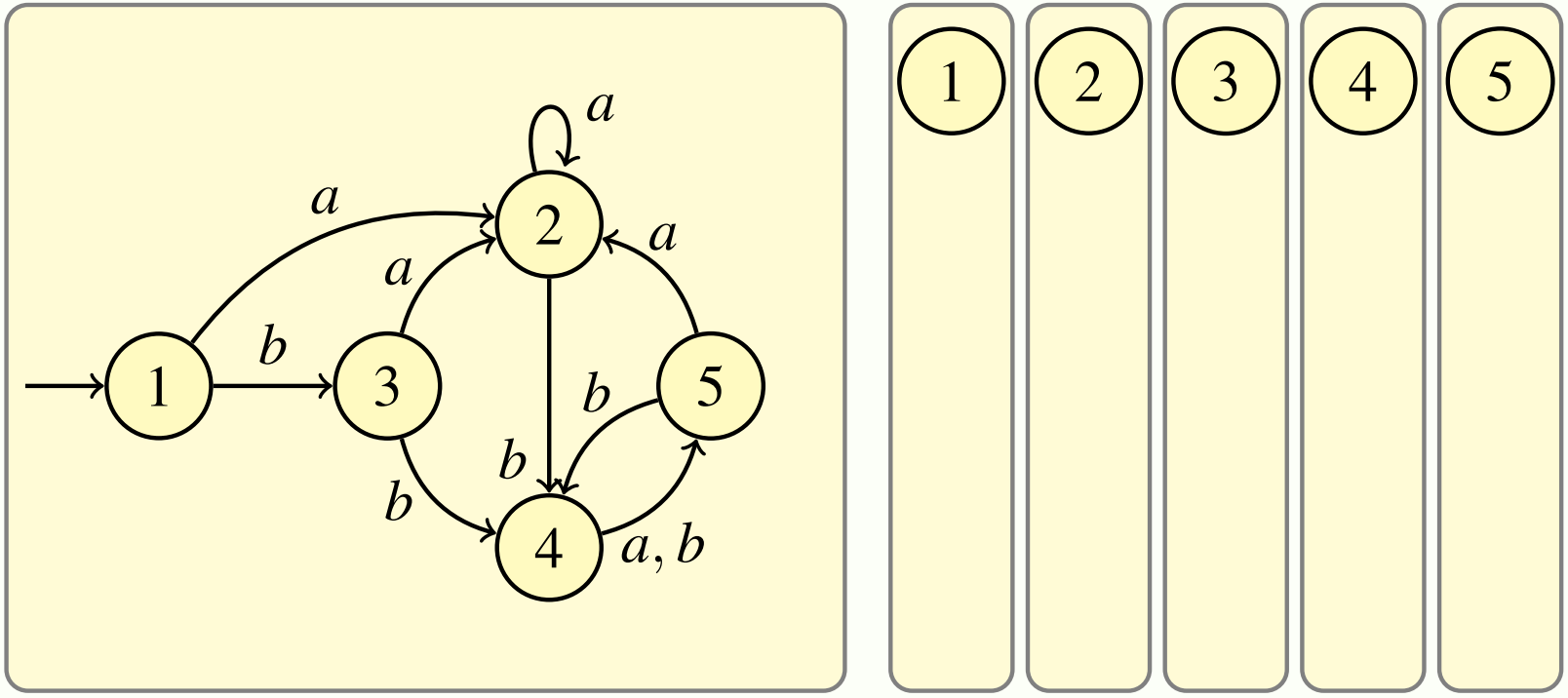
- On a utilisé de la combinatoire analytique pour obtenir les résultats.
- Les deux distributions donnent des résultats différents.
- Si on change $\star$ en $+$ dans la distribution ABR, on pense retrouver une nombre moyen de transitions linéaire.

- On a utilisé de la combinatoire analytique pour obtenir les résultats.
- Les deux distributions donnent des résultats différents.
- Si on change $\star$ en $+$ dans la distribution ABR, on pense retrouver une nombre moyen de transitions linéaire.

- Peut-on obtenir un résultat plus général ?

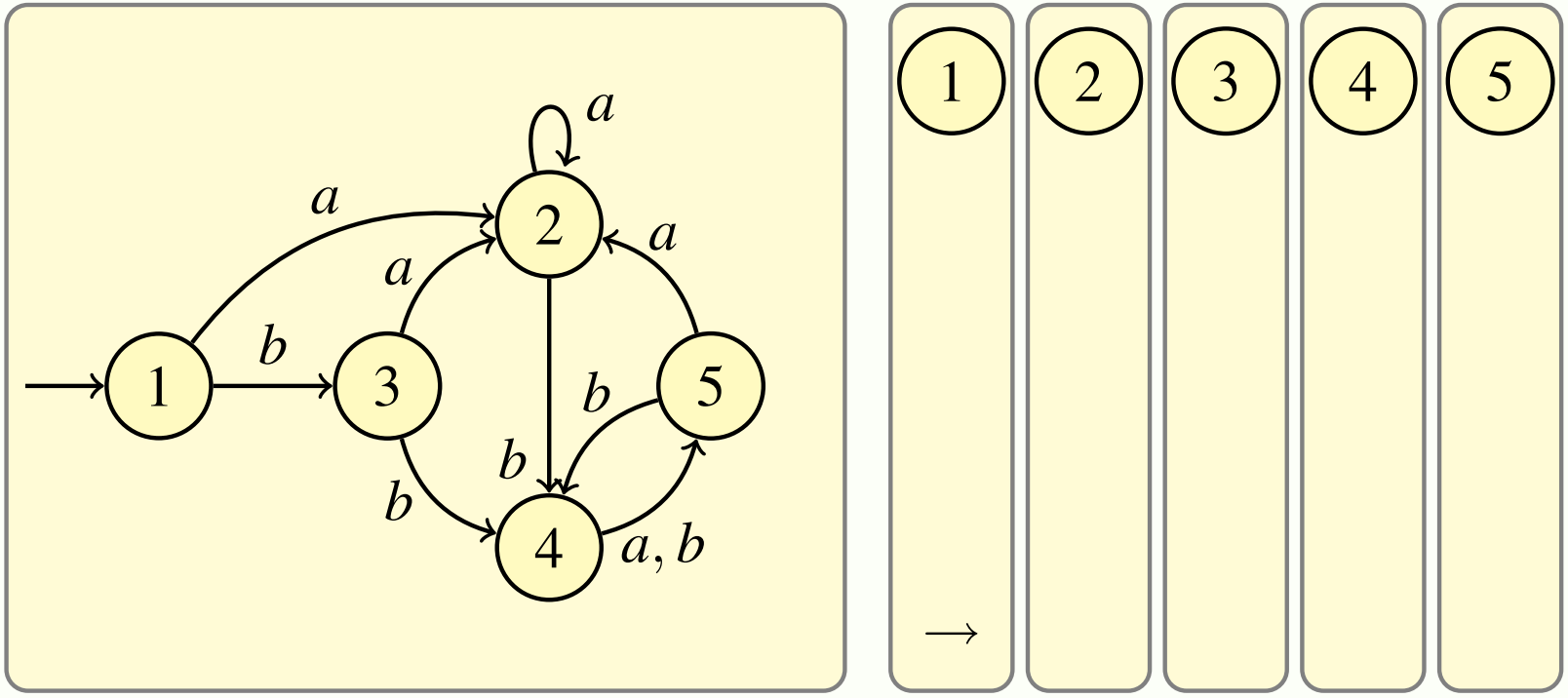
Partie III : Génération aléatoire d'automates

Automate déterministe accessible et complet.



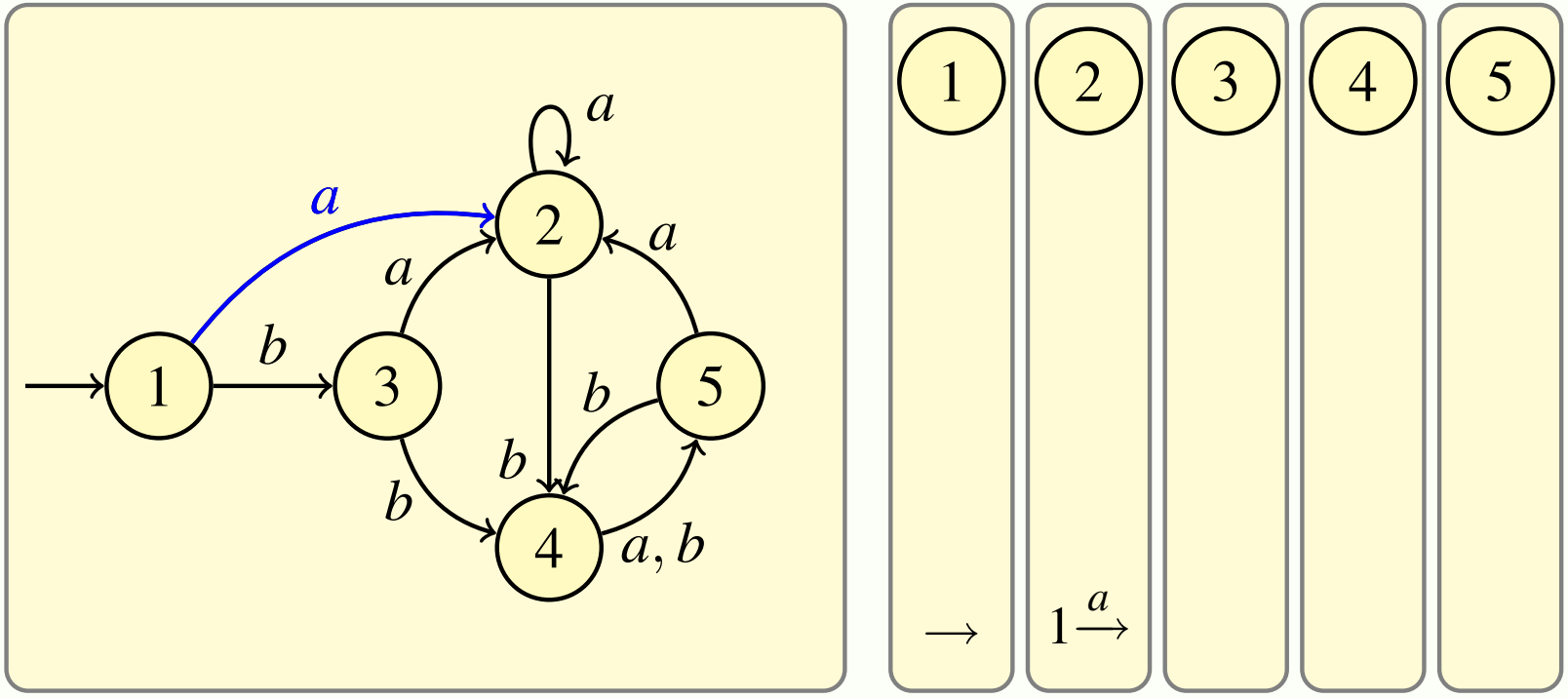
Construction d'une partition des $kn + 1$ flèches en n parts.

Automate déterministe accessible et complet.



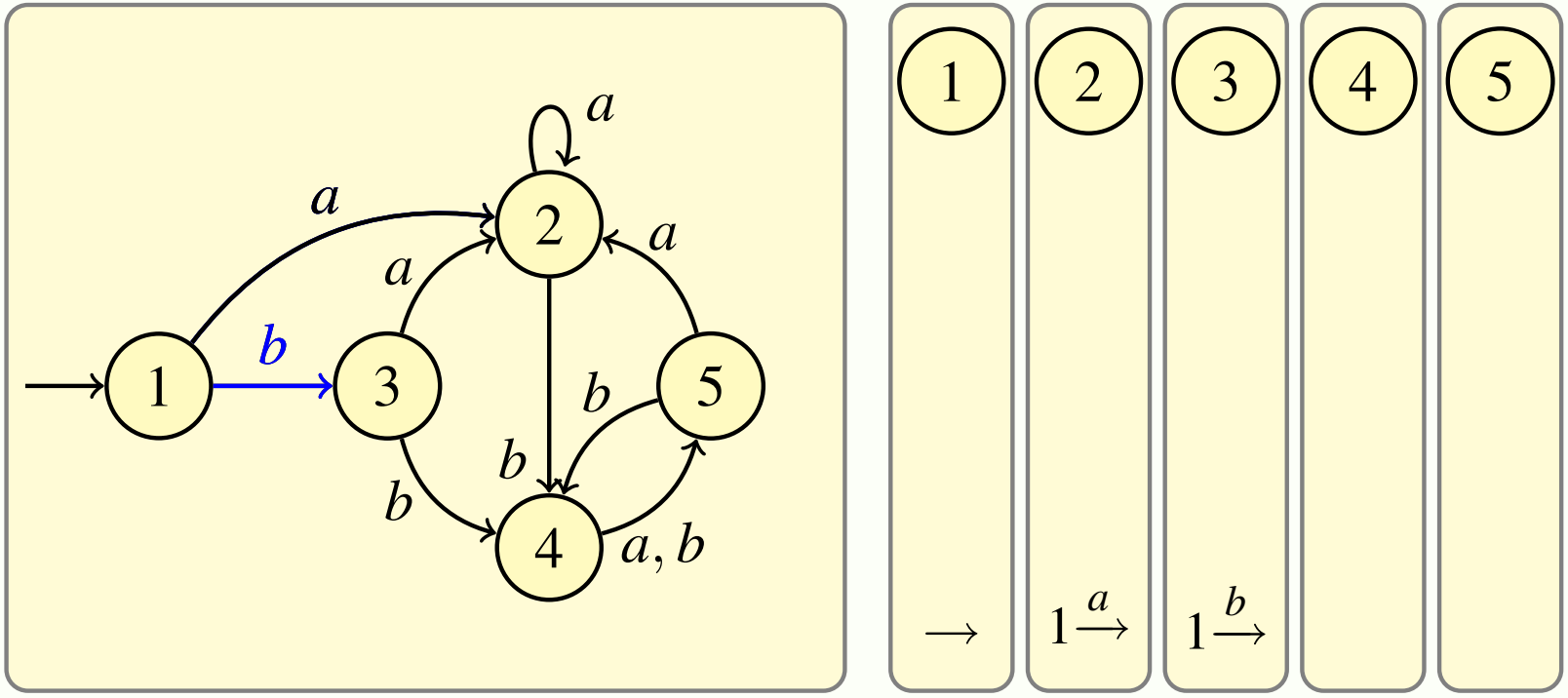
Construction d'une partition des $kn + 1$ flèches en n parts.

Automate déterministe accessible et complet.



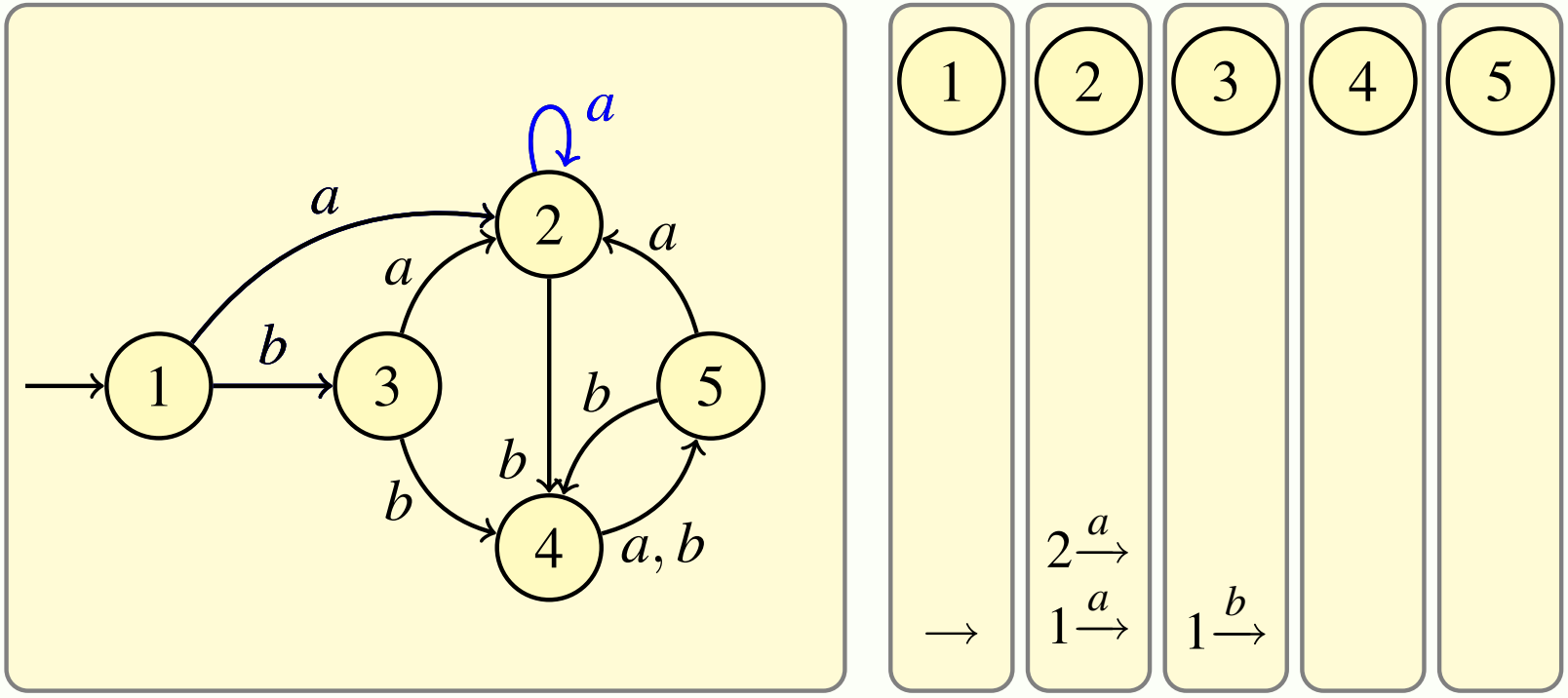
Construction d'une partition des $kn + 1$ flèches en n parts.

Automate déterministe accessible et complet.



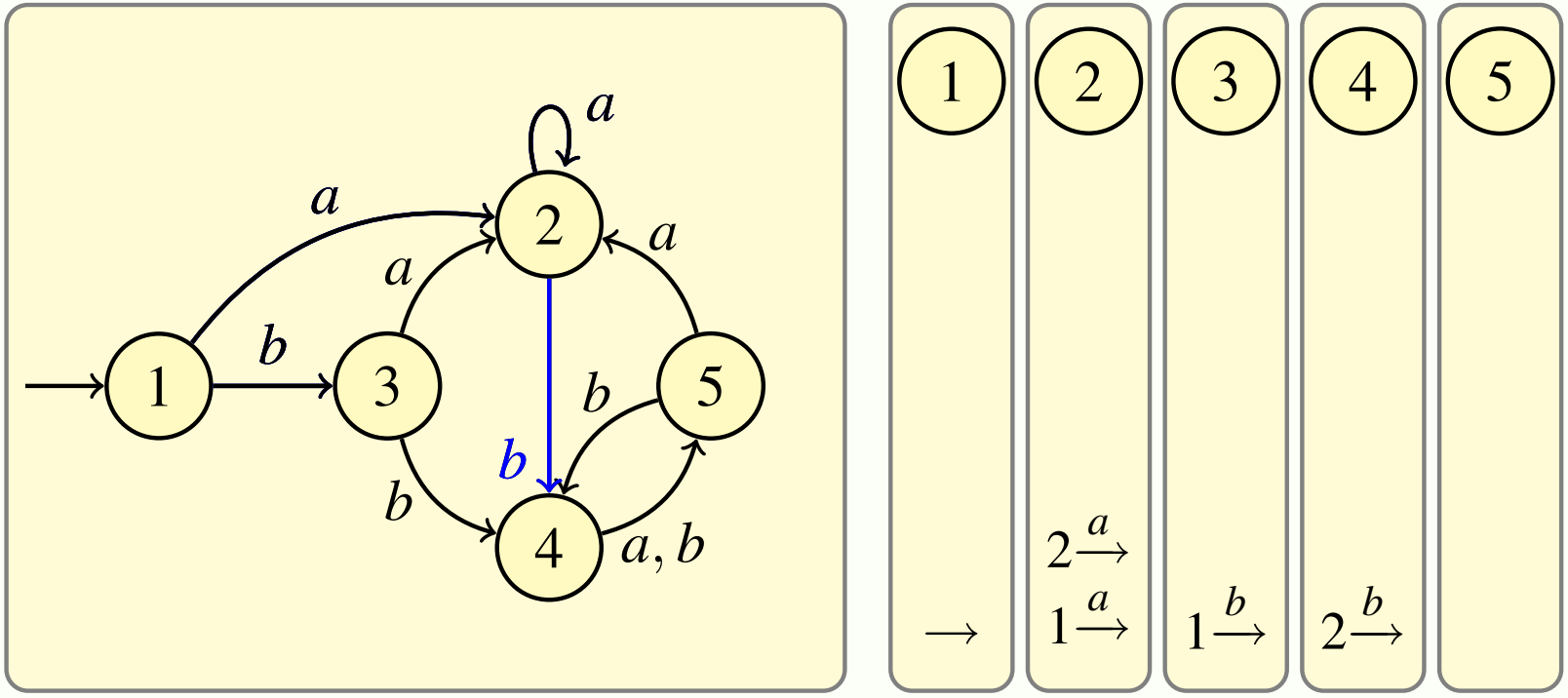
Construction d'une partition des $kn + 1$ flèches en n parts.

Automate déterministe accessible et complet.



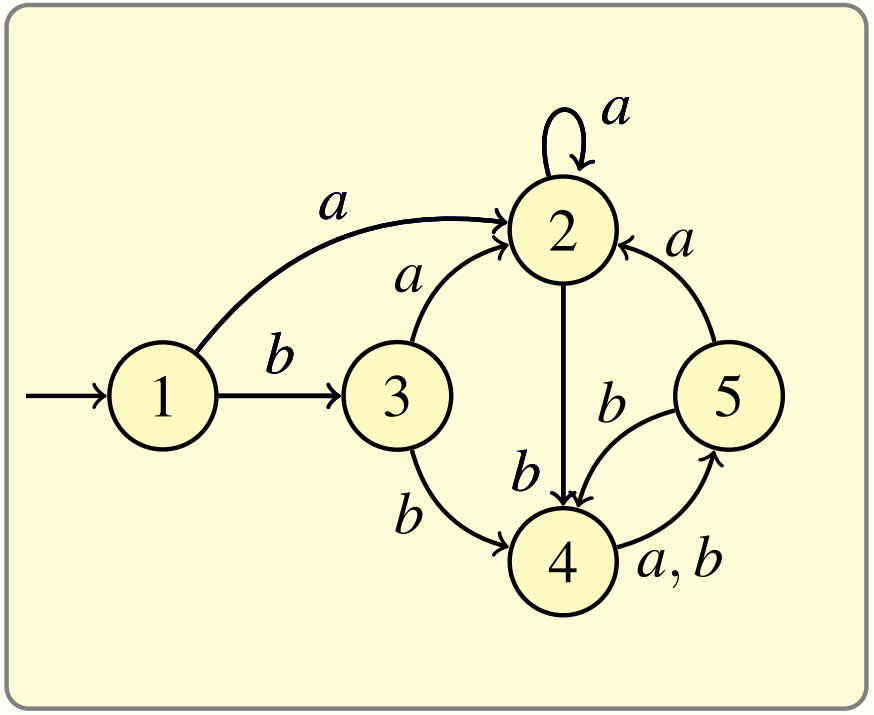
Construction d'une partition des $kn + 1$ flèches en n parts.

Automate déterministe accessible et complet.



Construction d'une partition des $kn + 1$ flèches en n parts.

Automate déterministe accessible et complet.



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

Construction d'une partition des $kn + 1$ flèches en n parts.

Définition

Une partition des $kn + 1$ flèches en n part est dite réalisable quand on peut l'obtenir à partir d'un automate accessible.

Définition

Une partition des $kn+1$ flèches en n part est dite réalisable quand on peut l'obtenir à partir d'un automate accessible.

Théorème [Bassino, N. 07 ; Korshunov 78]

La proportion de partitions réalisables est en $\Theta(1)$.

Définition

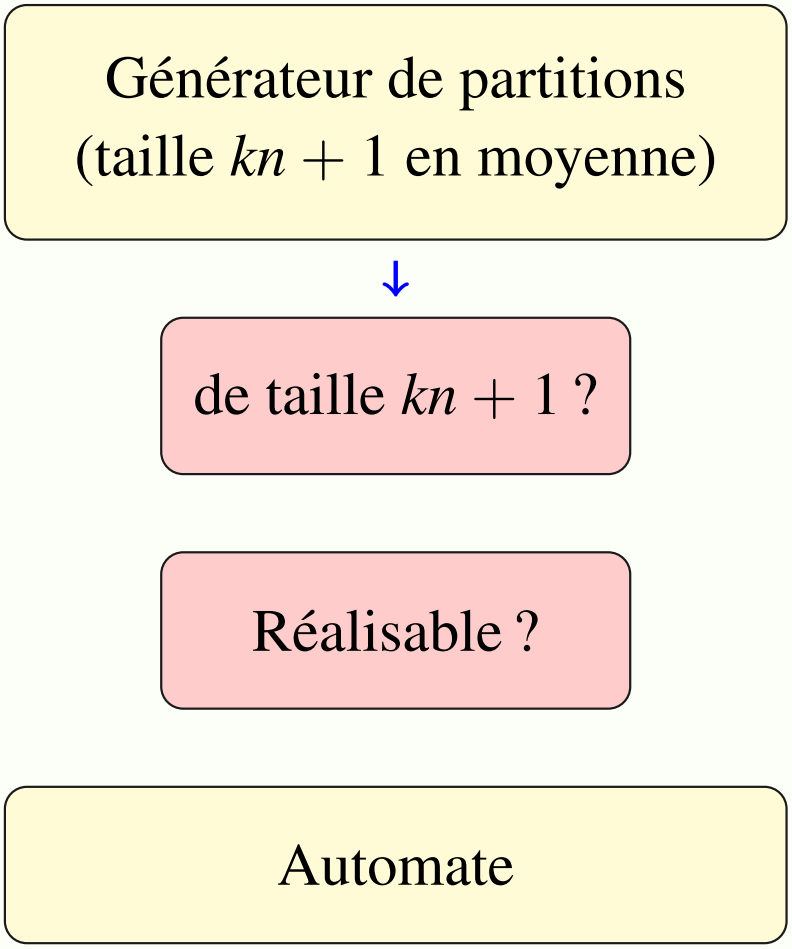
Une partition des $kn+1$ flèches en n part est dite réalisable quand on peut l'obtenir à partir d'un automate accessible.

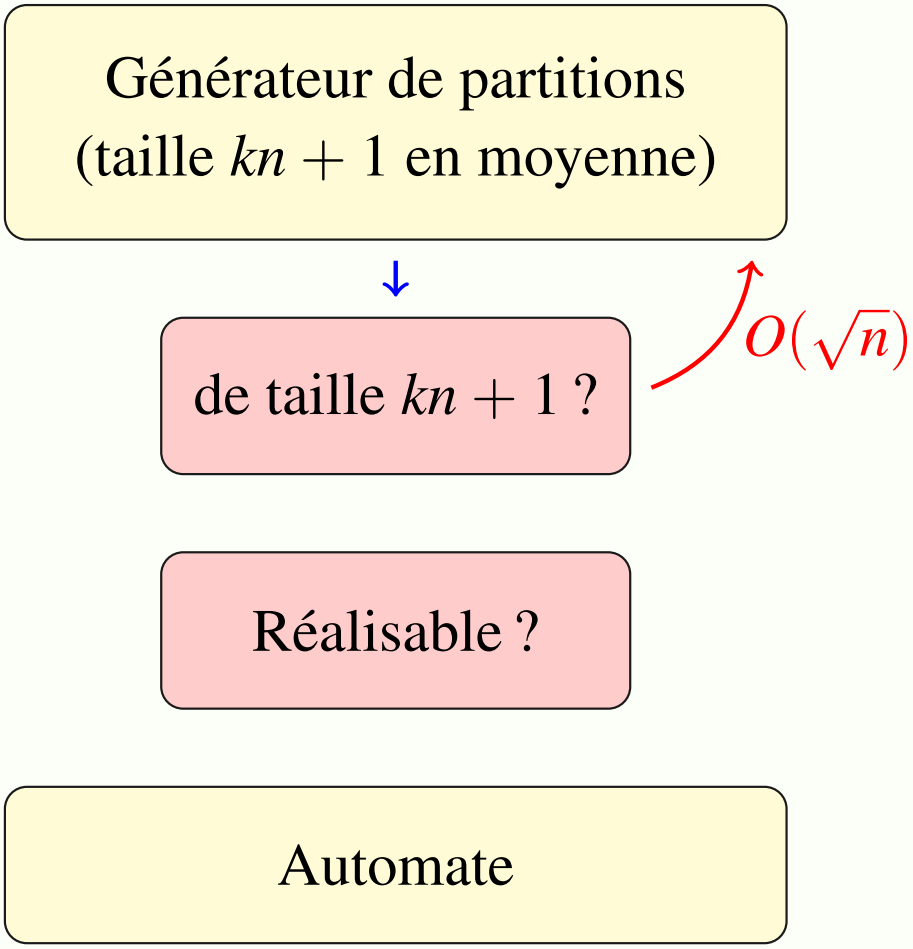
Théorème [Bassino, N. 07 ; Korshunov 78]

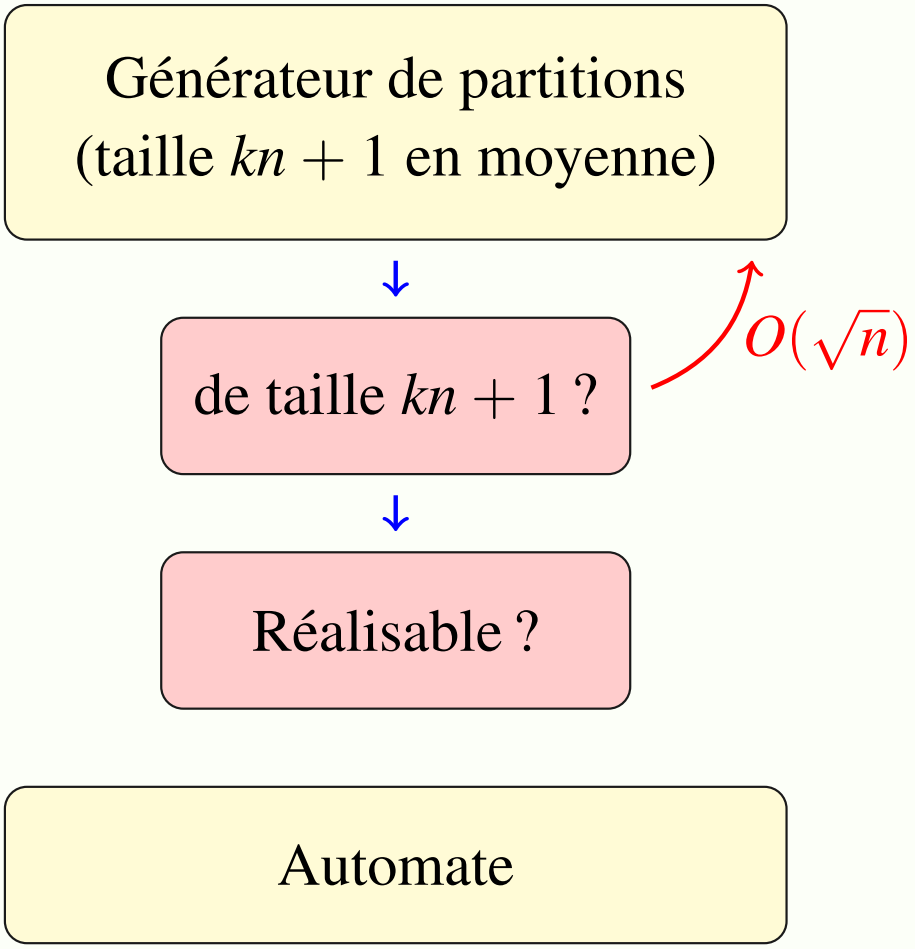
La proportion de partitions réalisables est en $\Theta(1)$.

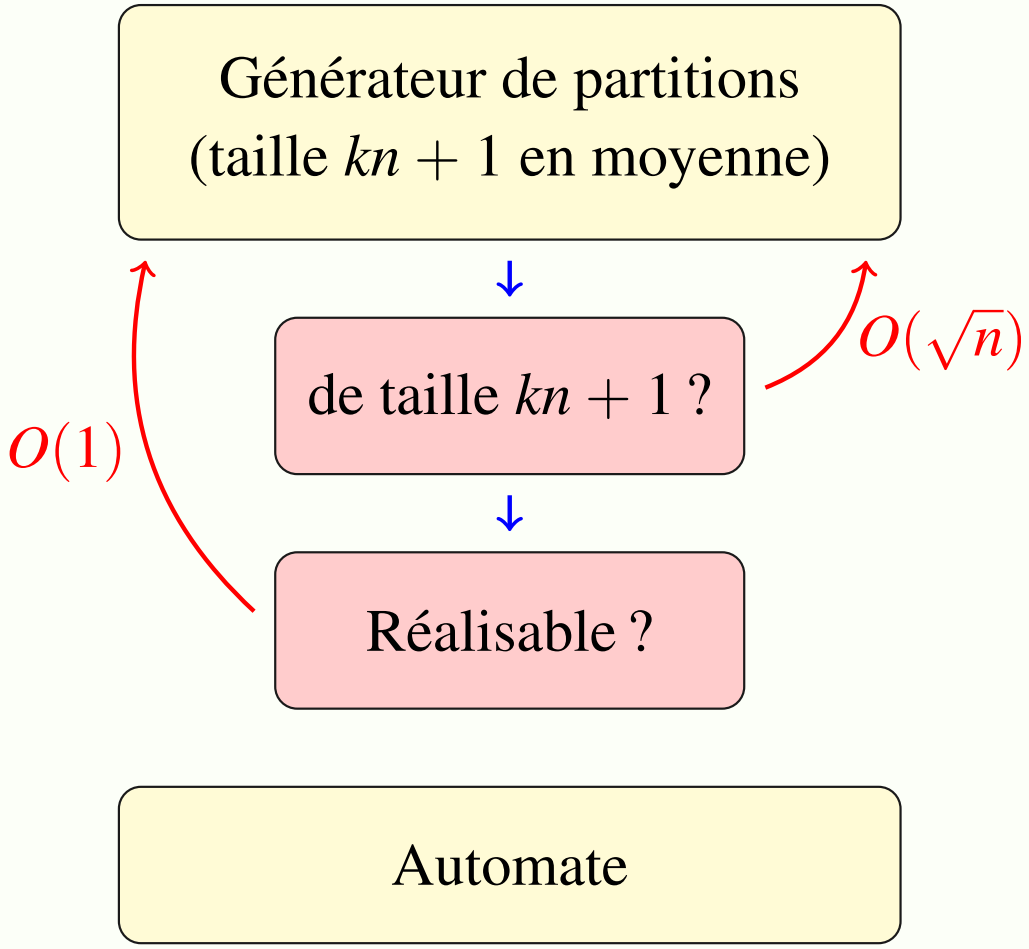
Corollaire

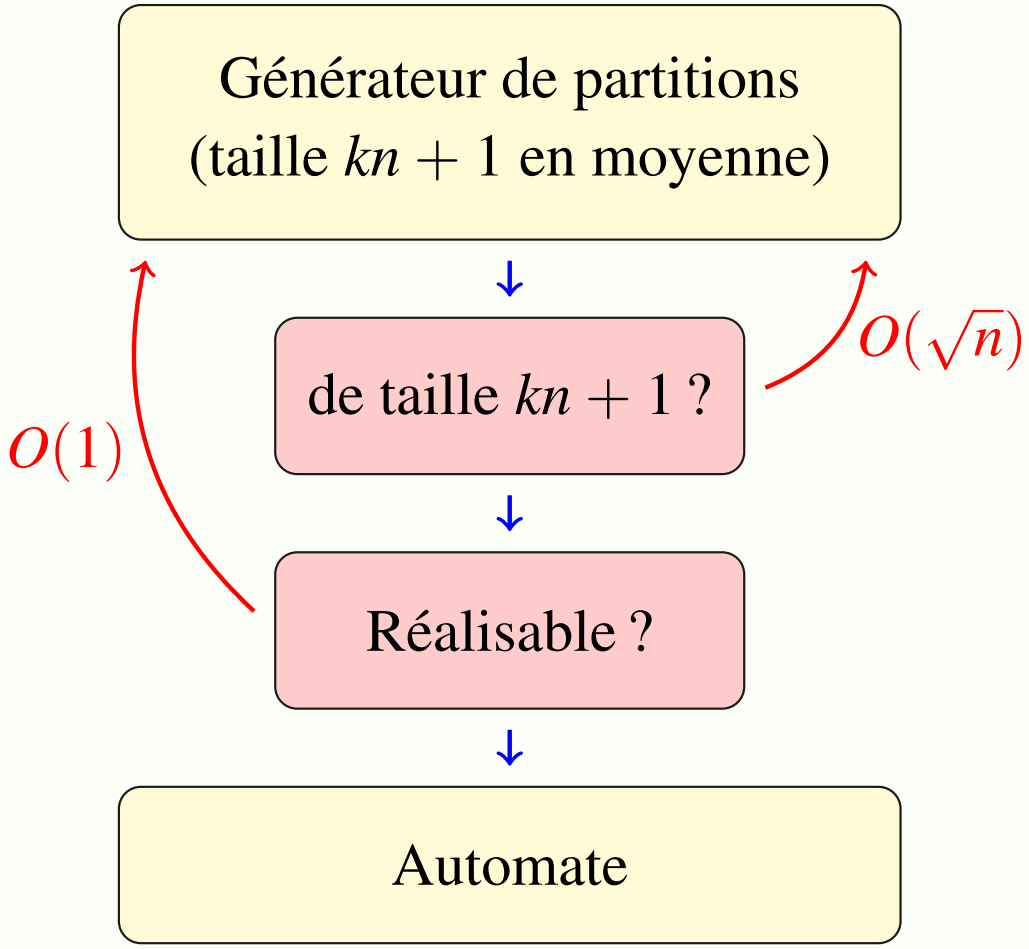
Le nombre d'automates déterministes, accessibles et complets est de l'ordre de $\left\{ {kn+1 \atop n} \right\} 2^n$.


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 $kn+1$?
$O(\sqrt{n})$
$O(1)$
Réalisable ?
Automate


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

Théorème [Bassino, N. 07]

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

Théorème [Bassino, N. 07]

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

Théorème [N. 00 ; Champarnaud, Paranthoën 05]

En adaptant la méthode récursive, on peut générer les mêmes automates en temps linéaire, au prix d'un précalcul en $\Theta(n^2)$ en temps et en espace.

Extensions :

- Automates incomplets [Bassino, David, N. 08]
- Nombre fixé de transitions manquantes [Héam, N., Schmitz 10]
- Transducteurs déterministes en entrée [Héam, N., Schmitz 10]
- Automates cheminants d'arbres [Héam, N., Schmitz 10]
- *etc*.

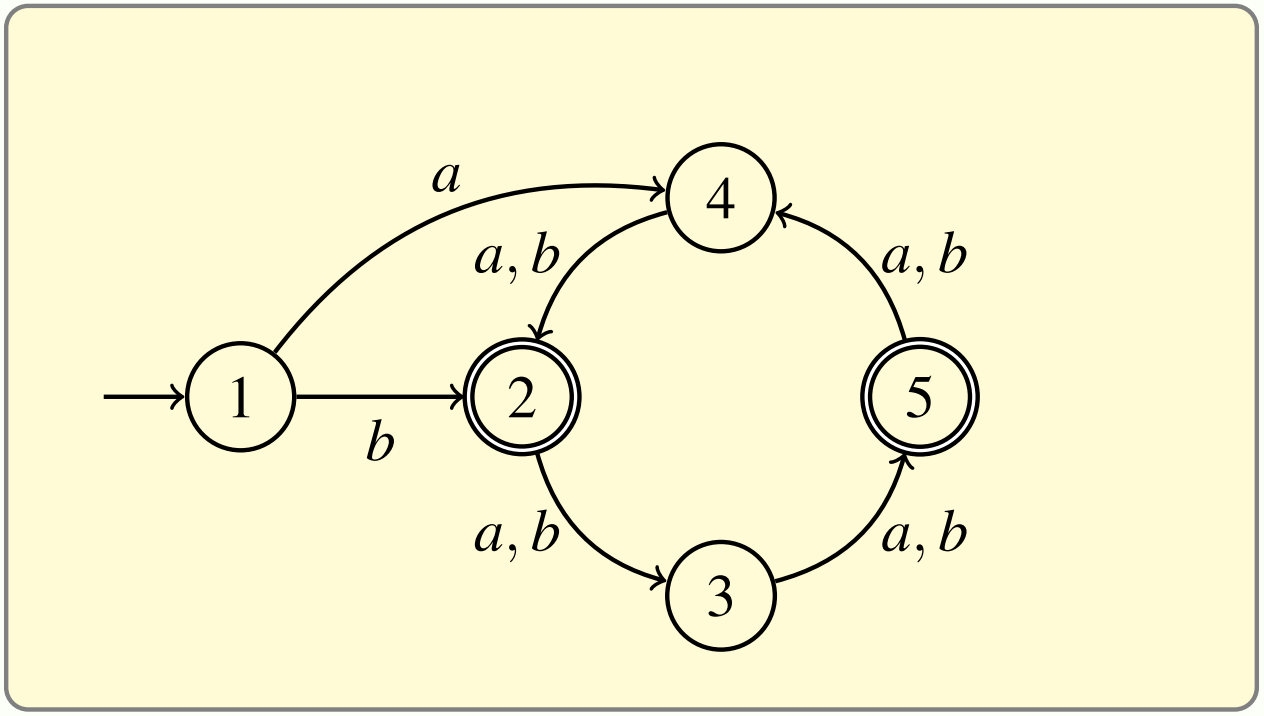
Bibliothèque :

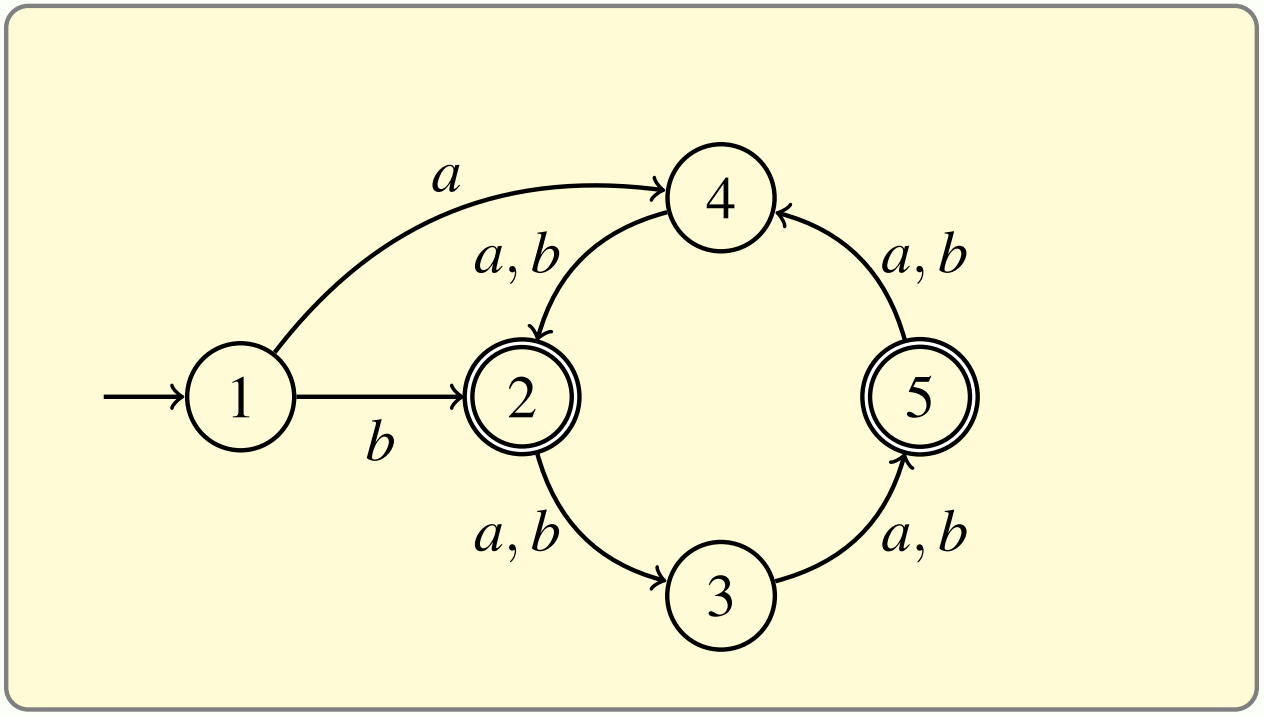
- REGAL, en C++, par J. David.

Question :

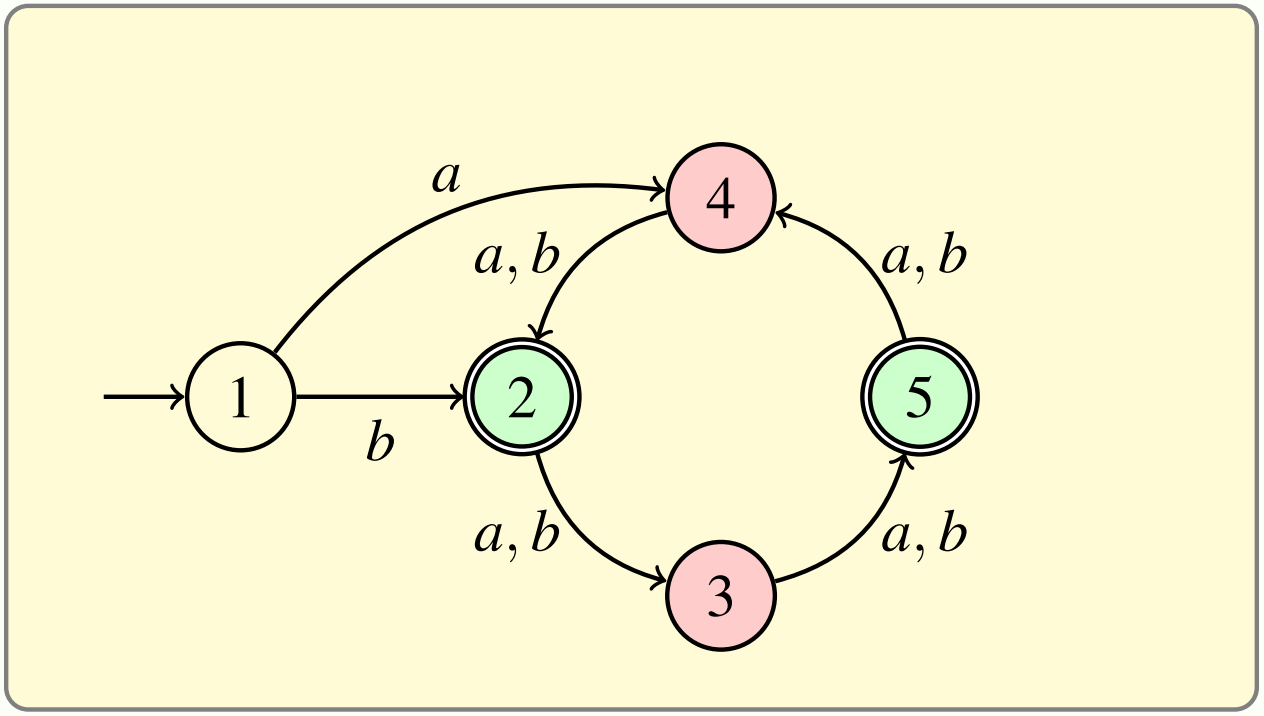
- Peut-on faire mieux que du $\Theta(n\sqrt{n})$ pour les partitions ?

Partie IV : Algorithme de Moore

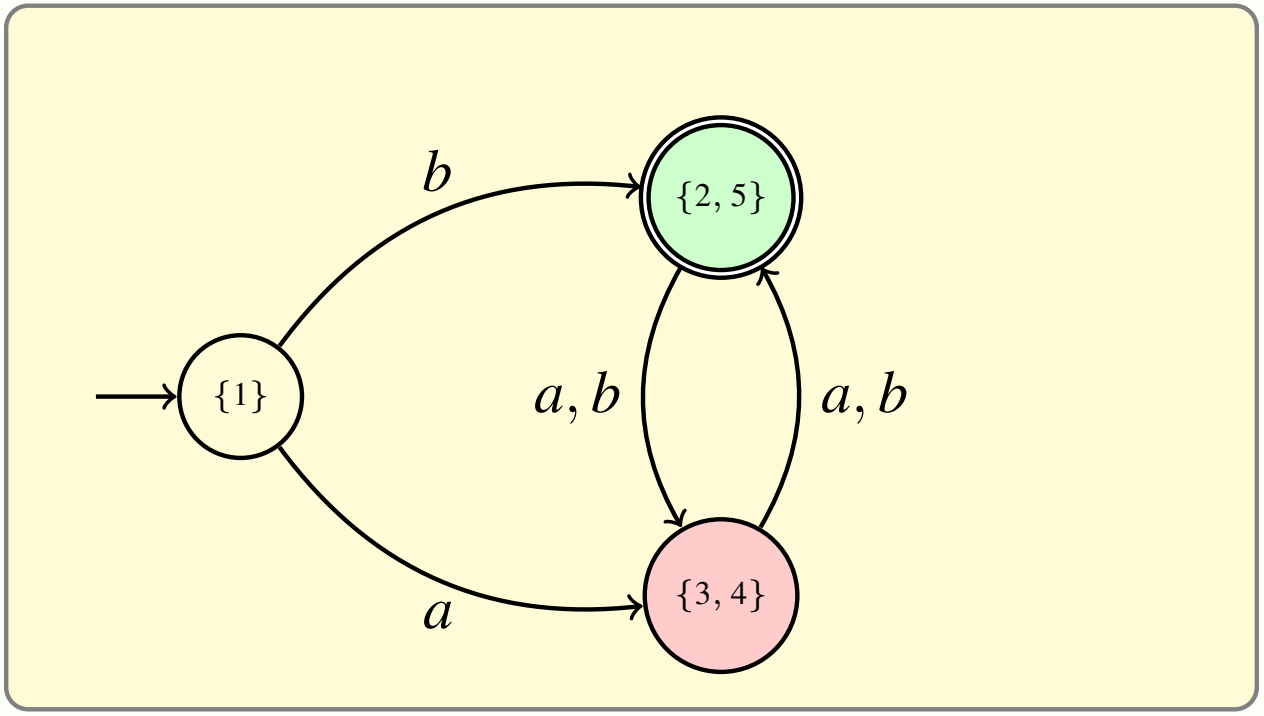

a
4
a, b
a, b
1
2
5
b
a, b
a, b
3



- $L_p = \{u \mid p \cdot u \in F\}$ est le langage reconnu en mettant l'état initial en p.
- $p \sim q$ ssi $L_p = L_q$ (équivalence de Nérode).



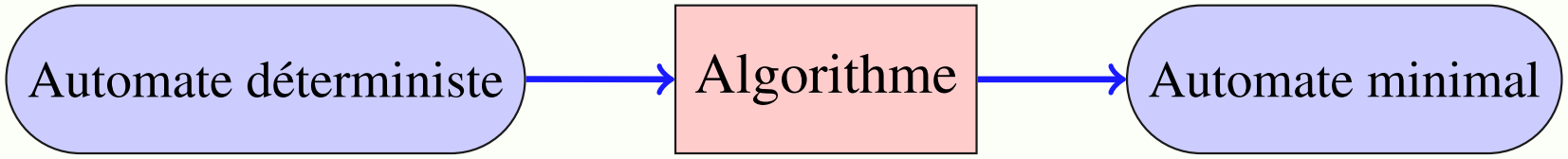
- $L_p = \{u \mid p \cdot u \in F\}$ est le langage reconnu en mettant l'état initial en p.
- $p \sim q$ ssi $L_p = L_q$ (équivalence de Nérode).



- On fait un état par classe d'équivalence.
- On obtient l'automate minimal du langage.

- Tout langage rationnel L admet un unique automate minimal.
- C'est le plus petit automate déterministe qui reconnaît L.

- Tout langage rationnel L admet un unique automate minimal.
- C'est le plus petit automate déterministe qui reconnaît L.



- Tout langage rationnel L admet un unique automate minimal.
- C'est le plus petit automate déterministe qui reconnaît L.

- Algorithme de Moore, en $O(n^2)$ [Moore 56]
- Algorithme de Hopcroft, en $O(n \log n)$ [Hopcroft 71]

Moore($\mathcal{A}$)

1. Calculer $\sim_0$
2. Tant que $\sim_{i+1} \neq \sim_i$
   3. $i := i + 1$
   4. Calculer $\sim_{i+1}$
5. Retourner $\mathcal{A}/\sim_i$

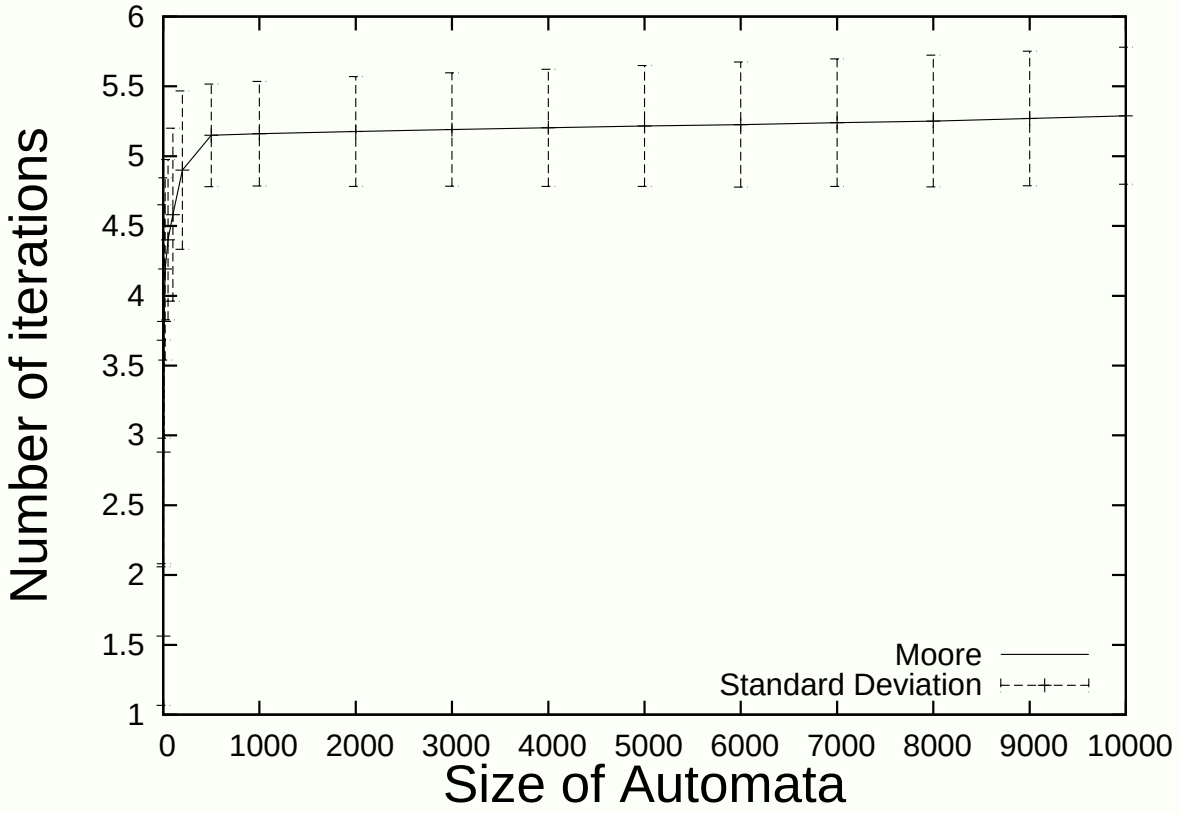
- $p \sim q \Leftrightarrow L_p = L_q$
- $p \sim_i q \Leftrightarrow L_p \cap A^{\leq i} = L_q \cap A^{\leq i}$
- $\sim_{i+1}$ se calcule à partir de $\sim_i$
- Si $\sim_{i+1} = \sim_i$ alors $\sim = \sim_i$

Moore($\mathcal{A}$)

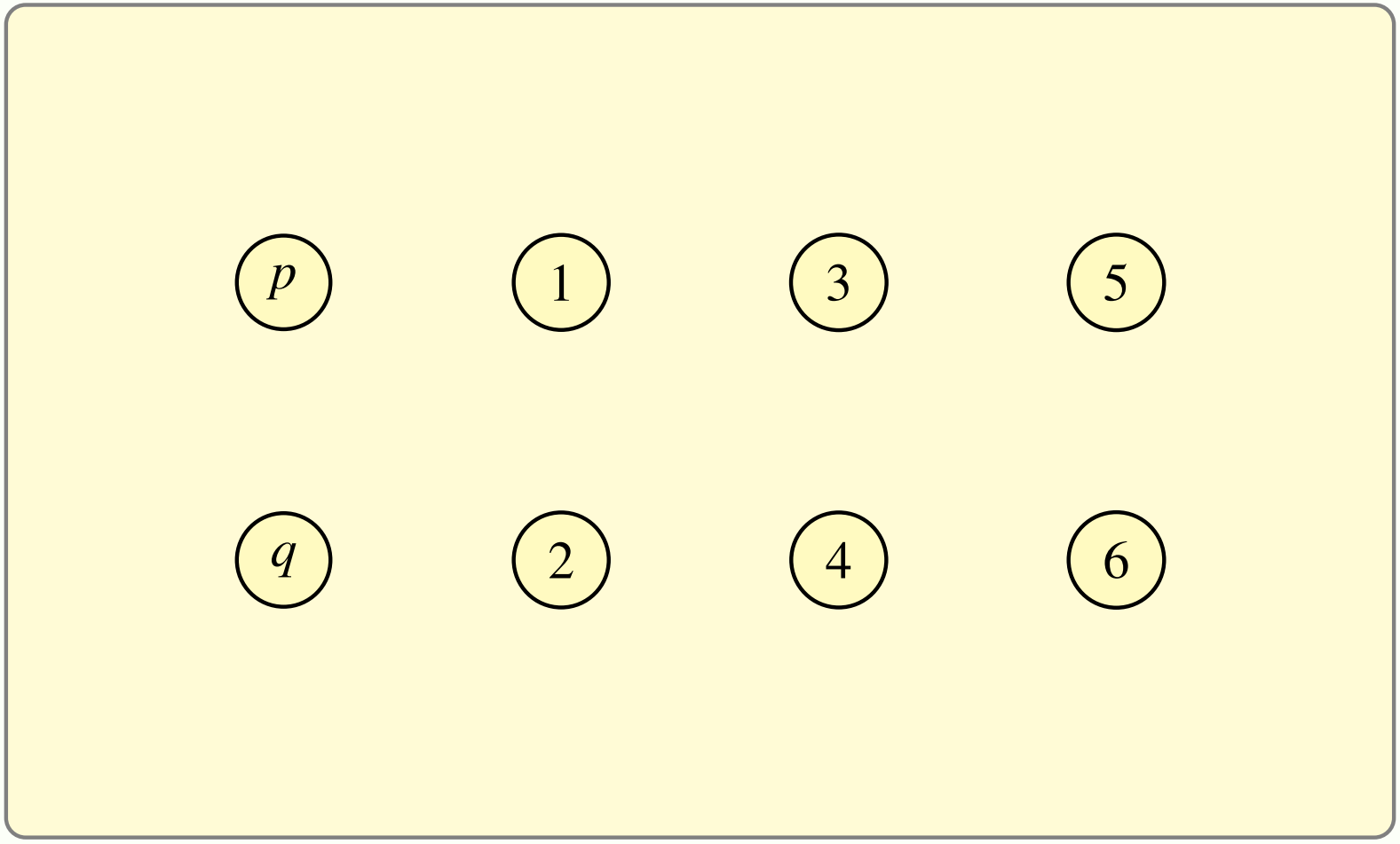
1. Calculer $\sim_0$
2. Tant que $\sim_{i+1} \neq \sim_i$
   3. $i := i + 1$
   4. Calculer $\sim_{i+1}$
5. Retourner $\mathcal{A}/\sim_i$

- $p \sim q \Leftrightarrow L_p = L_q$
- $p \sim_i q \Leftrightarrow L_p \cap A^{\leq i} = L_q \cap A^{\leq i}$
- $\sim_{i+1}$ se calcule à partir de $\sim_i$
- Si $\sim_{i+1} = \sim_i$ alors $\sim = \sim_i$

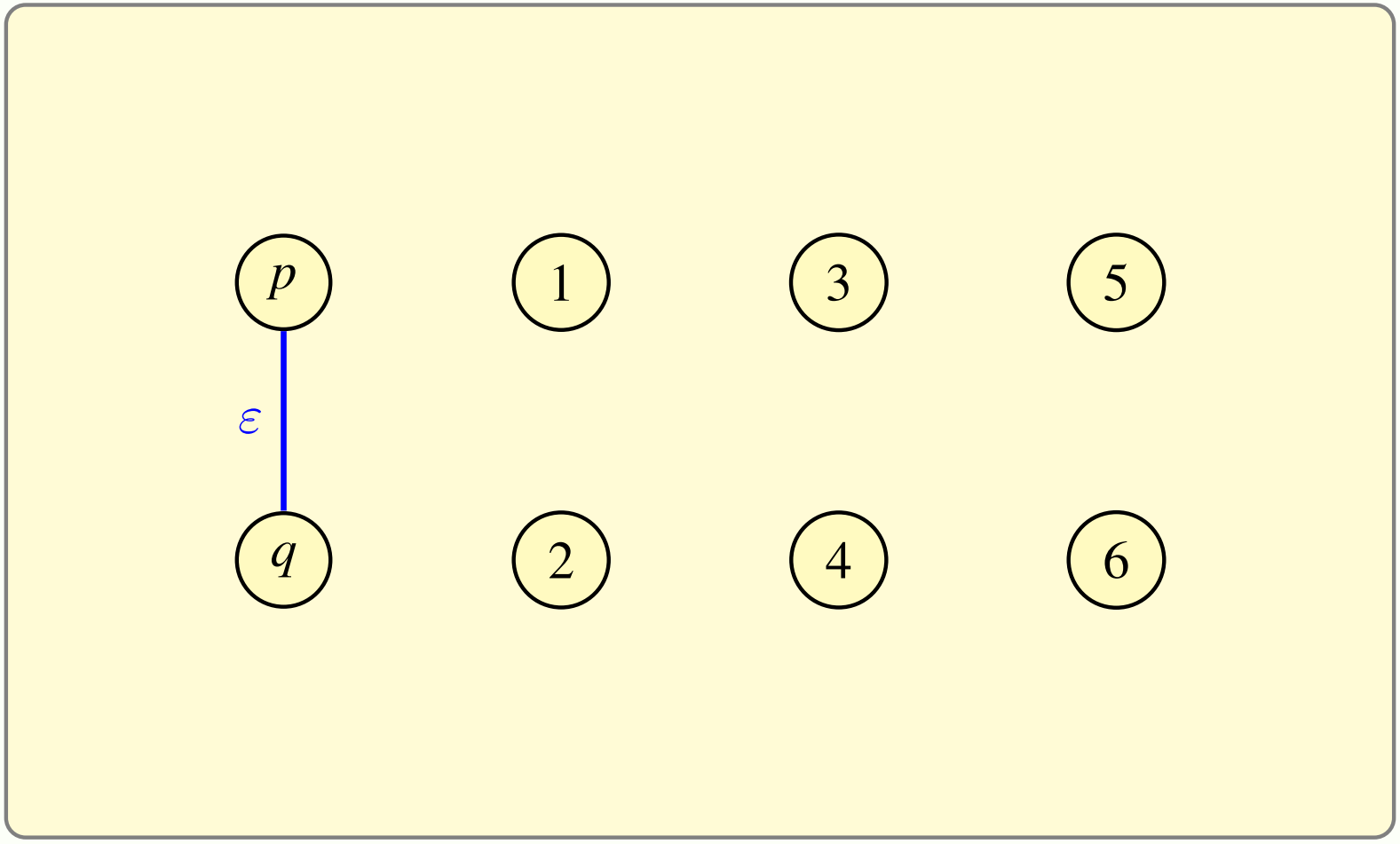
- Calculer $\sim_{i+1}$ se fait en temps $\Theta(kn)$
- Il y a au plus $n - 1$ itérations


Number of iterations
Size of Automata
Moore
Standard Deviation
6
5.5
5
4.5
4
3.5
3
2.5
2
1.5
1
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000

On fixe le graphe de l'automate.
$u = abaab$ le plus court mot qui sépare L_p et L_q : $p \sim_4 q$ et $p \nsim_5 q$

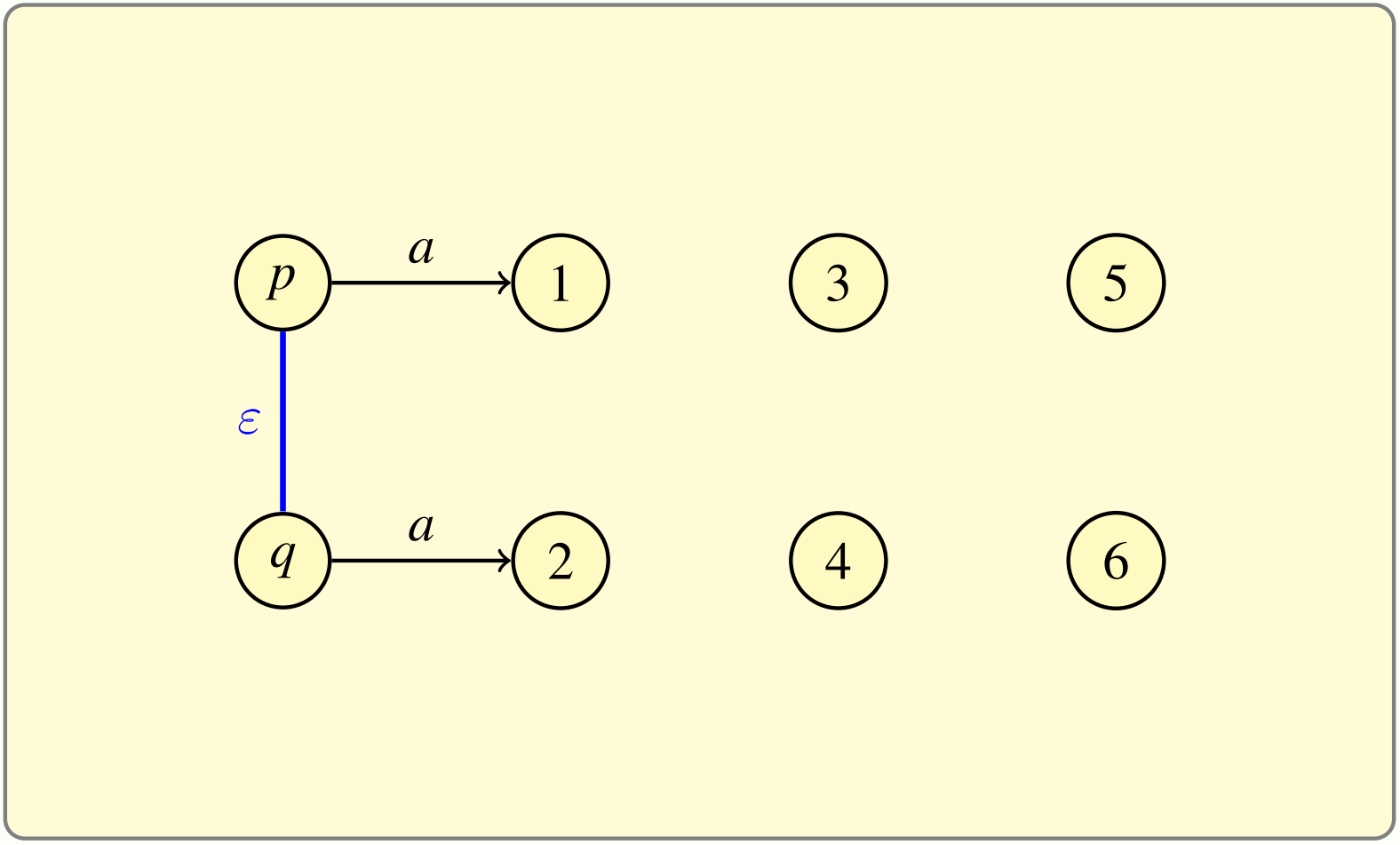


On fixe le graphe de l'automate.
$u = abaab$ le plus court mot qui sépare L_p et L_q : $p \sim_4 q$ et $p \nsim_5 q$



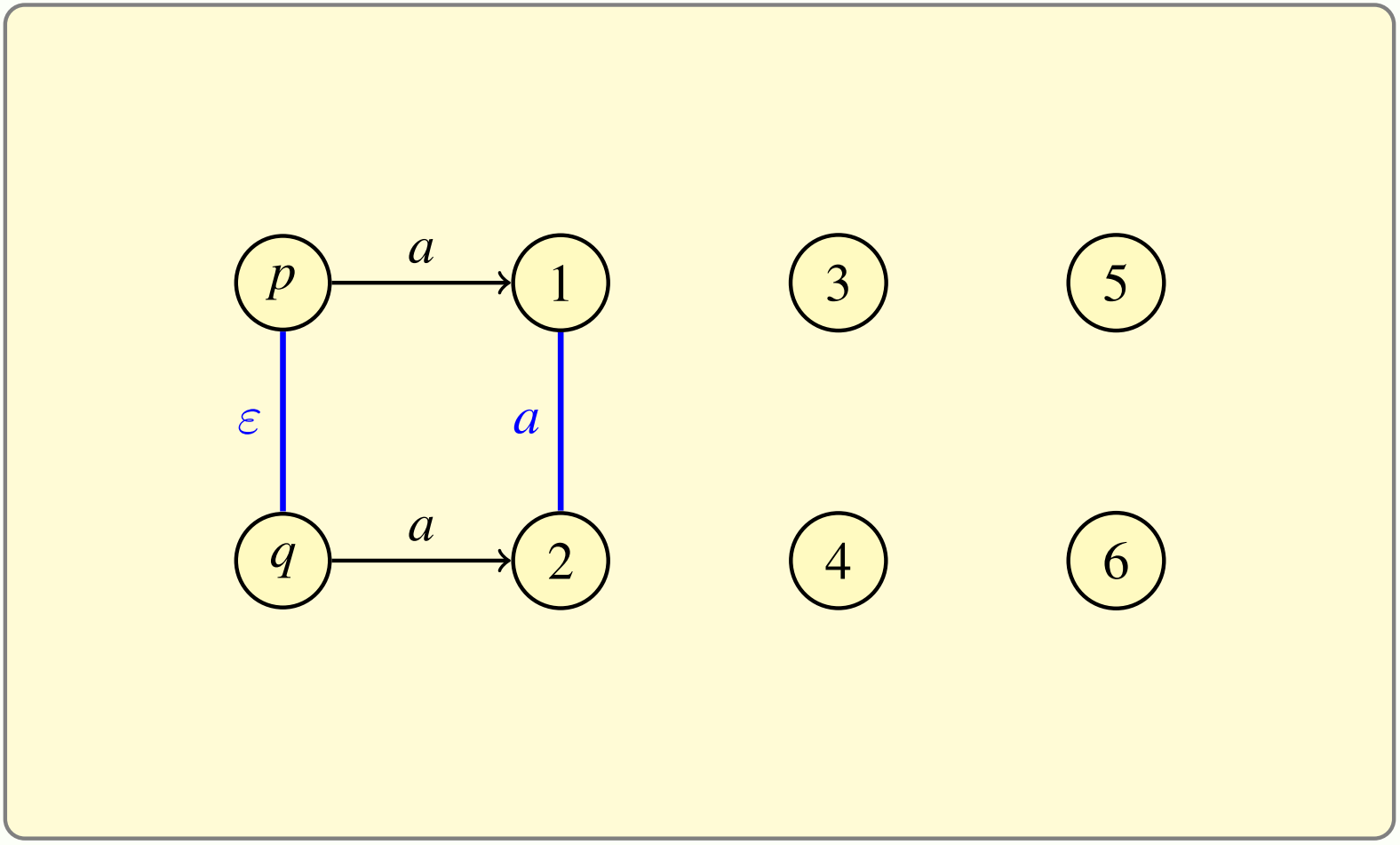
On fixe le graphe de l'automate.

$u = abaab$ le plus court mot qui sépare L_p et L_q : $p \sim_4 q$ et $p \nsim_5 q$



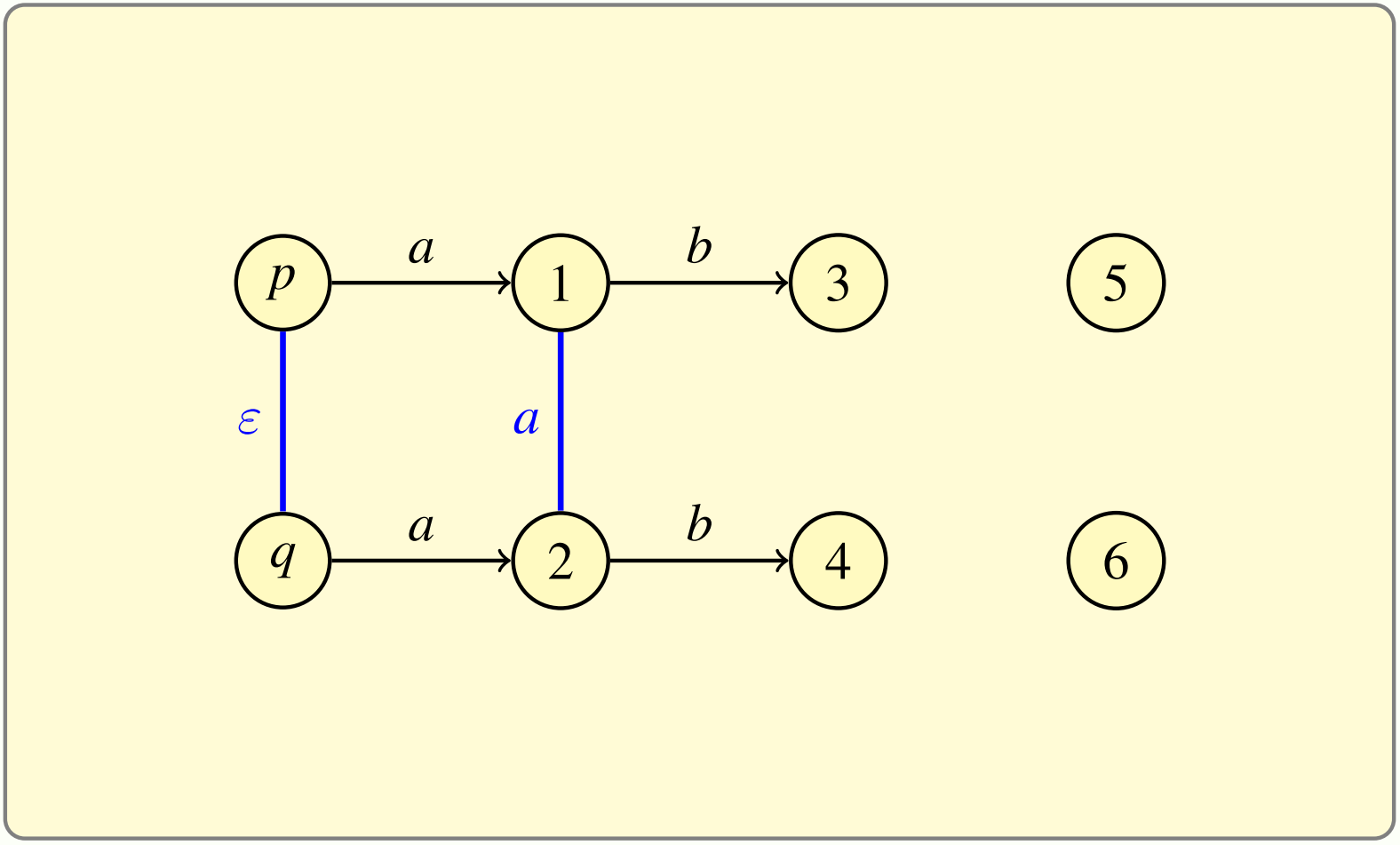
On fixe le graphe de l'automate.

$u = abaab$ le plus court mot qui sépare L_p et L_q : $p \sim_4 q$ et $p \nsim_5 q$

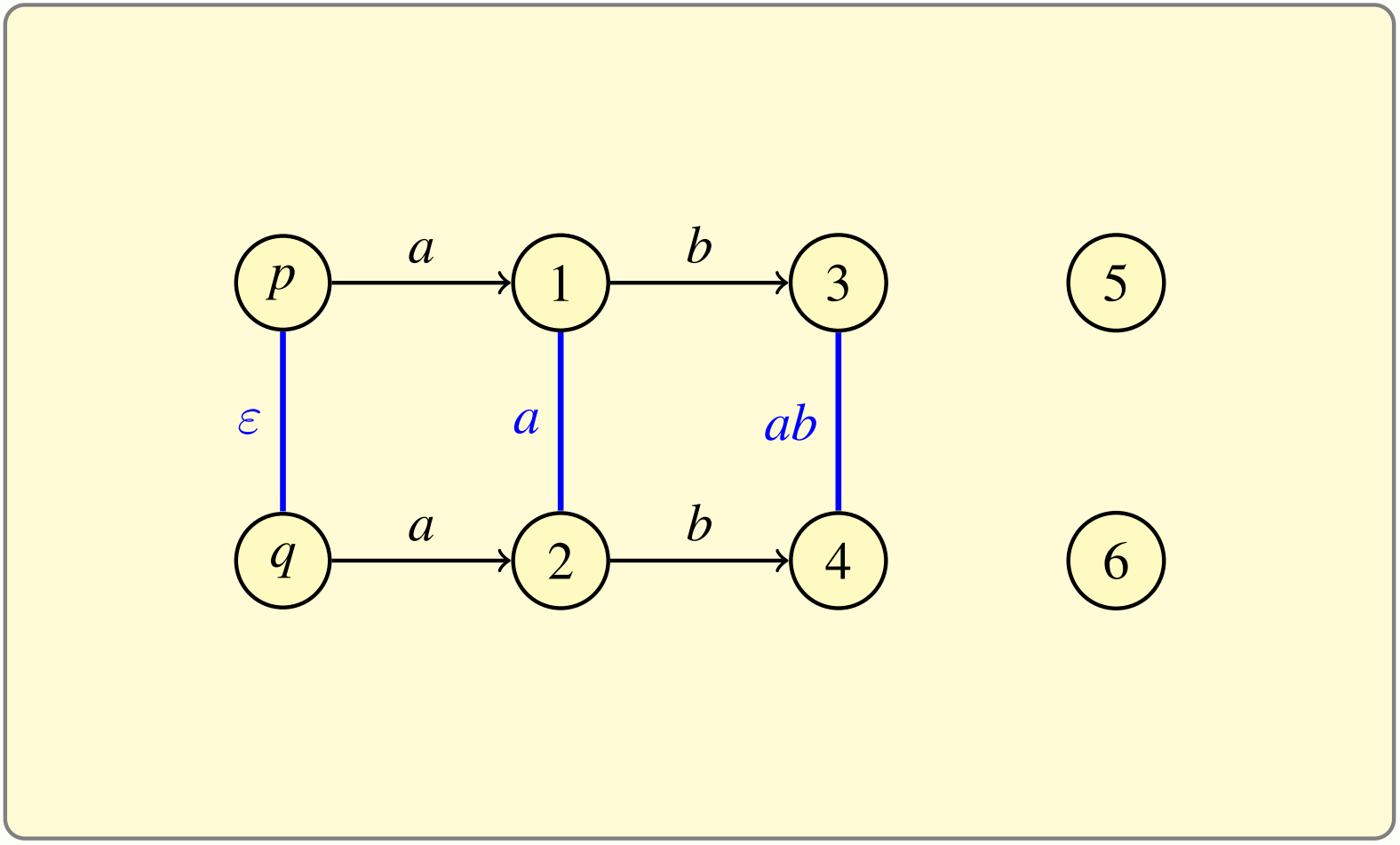


On fixe le graphe de l'automate.

$u = abaab$ le plus court mot qui sépare L_p et L_q : $p \sim_4 q$ et $p \nsim_5 q$

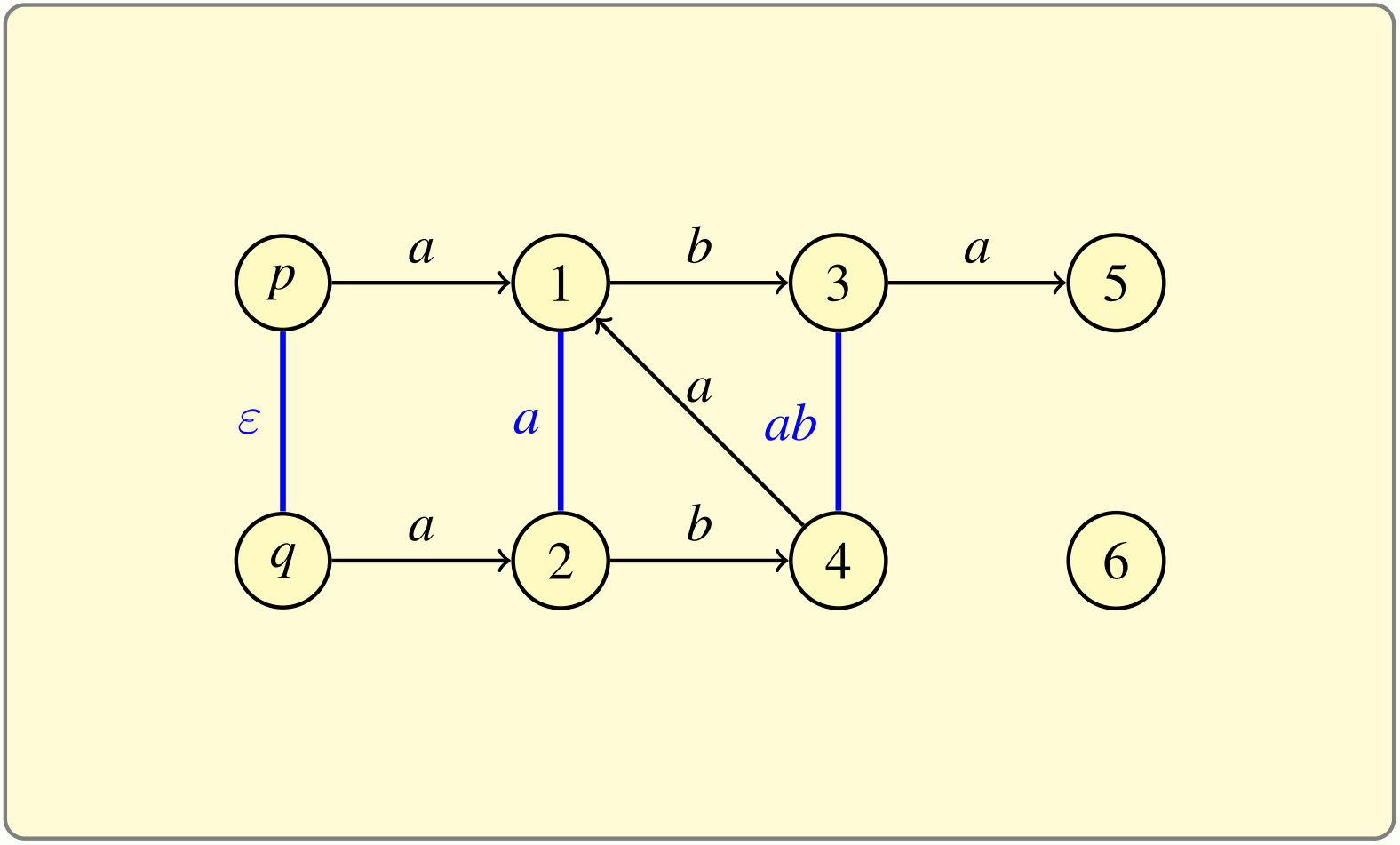


On fixe le graphe de l'automate.
$u = abaab$ le plus court mot qui sépare L_p et L_q : $p \sim_4 q$ et $p \nsim_5 q$



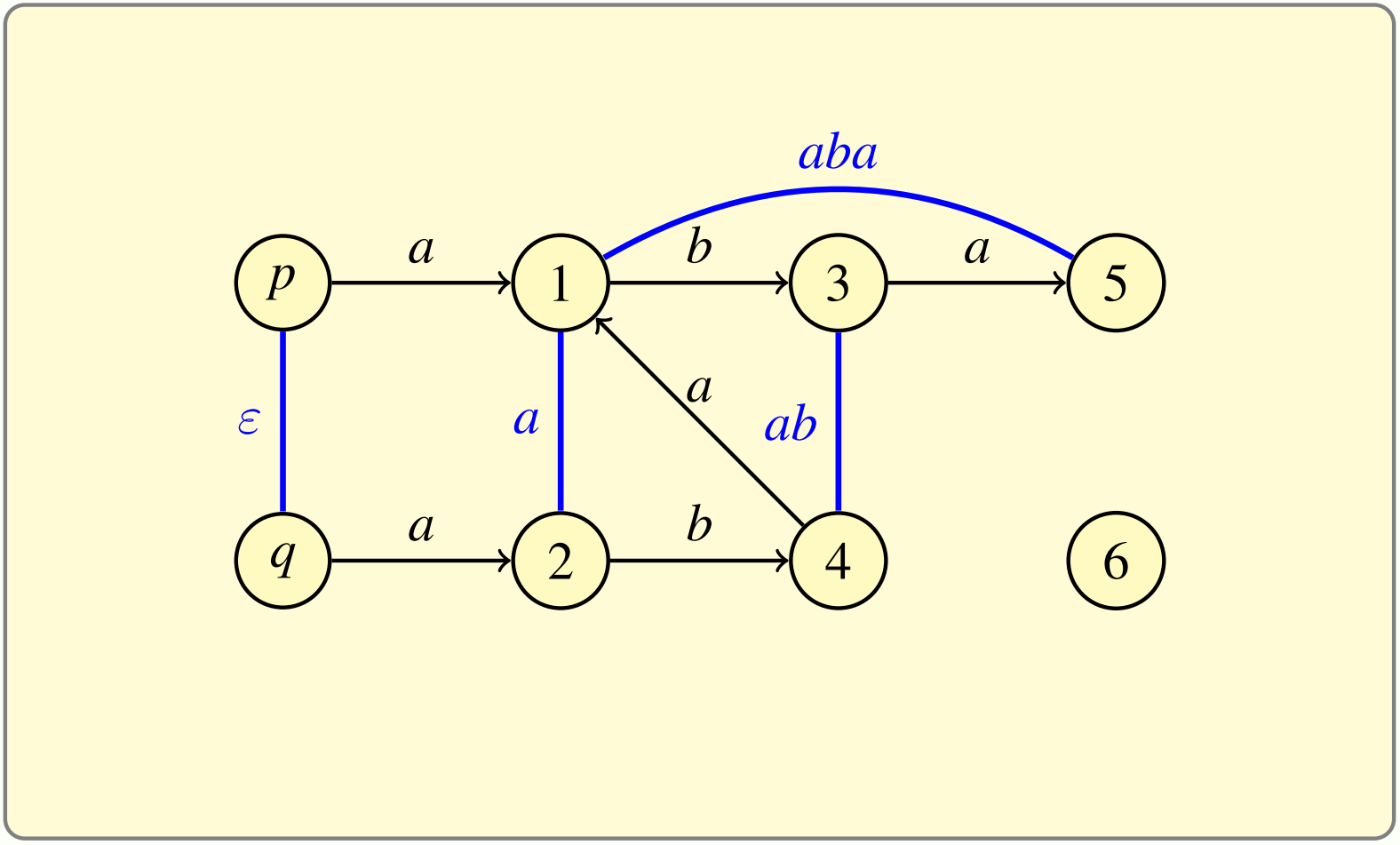
On fixe le graphe de l'automate.

$u = abaab$ le plus court mot qui sépare L_p et L_q : $p \sim_4 q$ et $p \nsim_5 q$



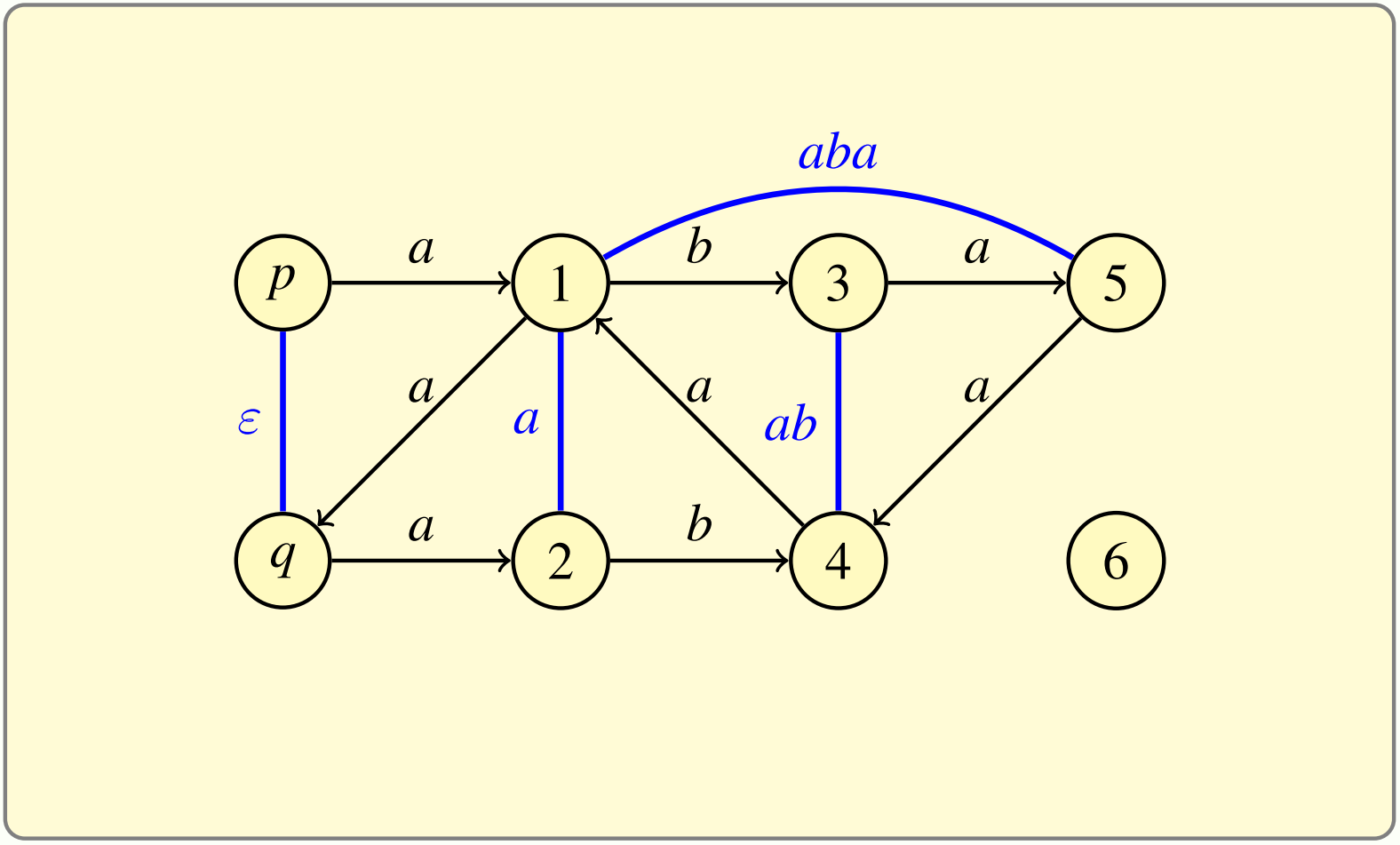
On fixe le graphe de l'automate.

$u = abaab$ le plus court mot qui sépare L_p et L_q : $p \sim_4 q$ et $p \nsim_5 q$

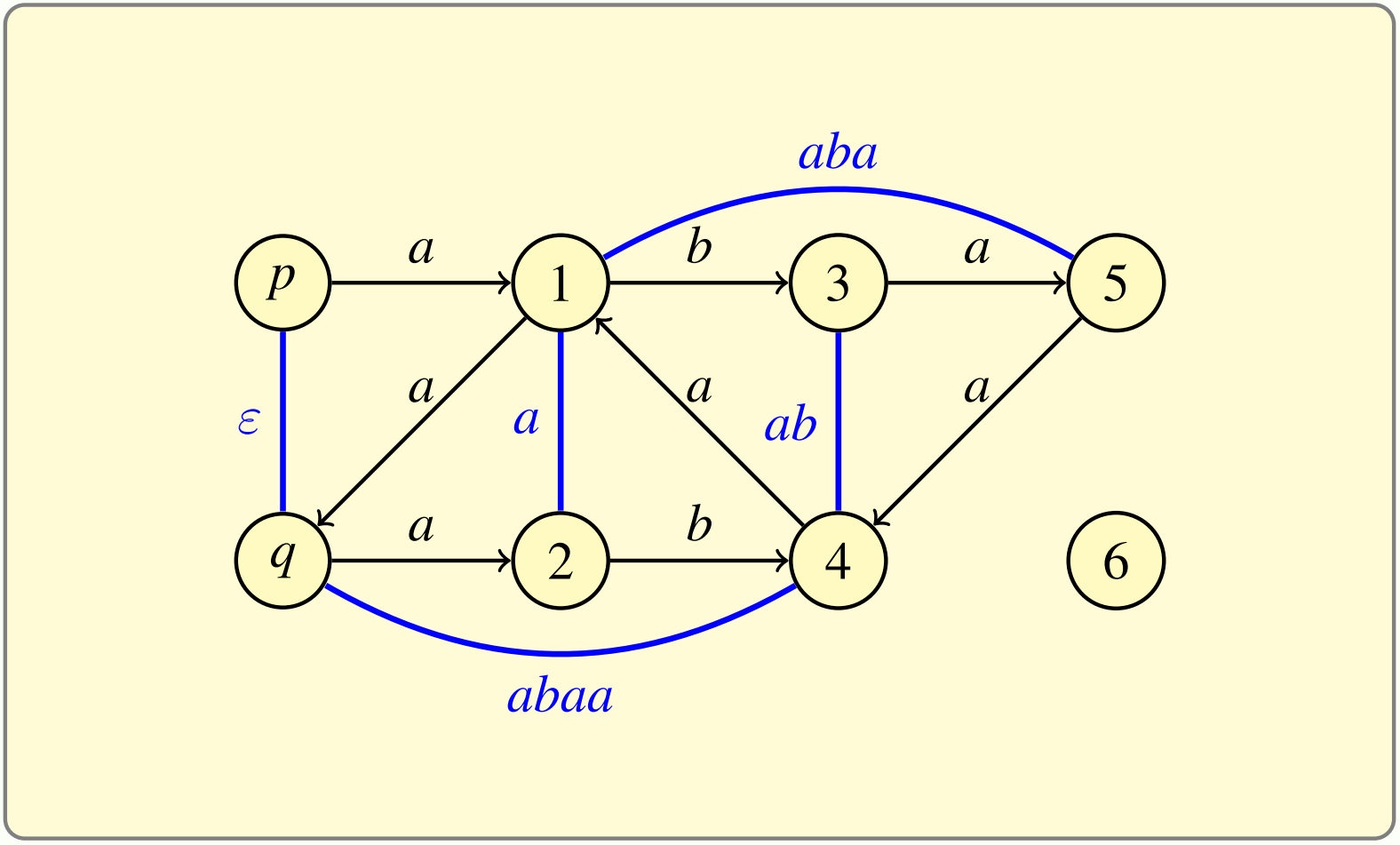


On fixe le graphe de l'automate.

$u = abaab$ le plus court mot qui sépare L_p et L_q : $p \sim_4 q$ et $p \not\sim_5 q$

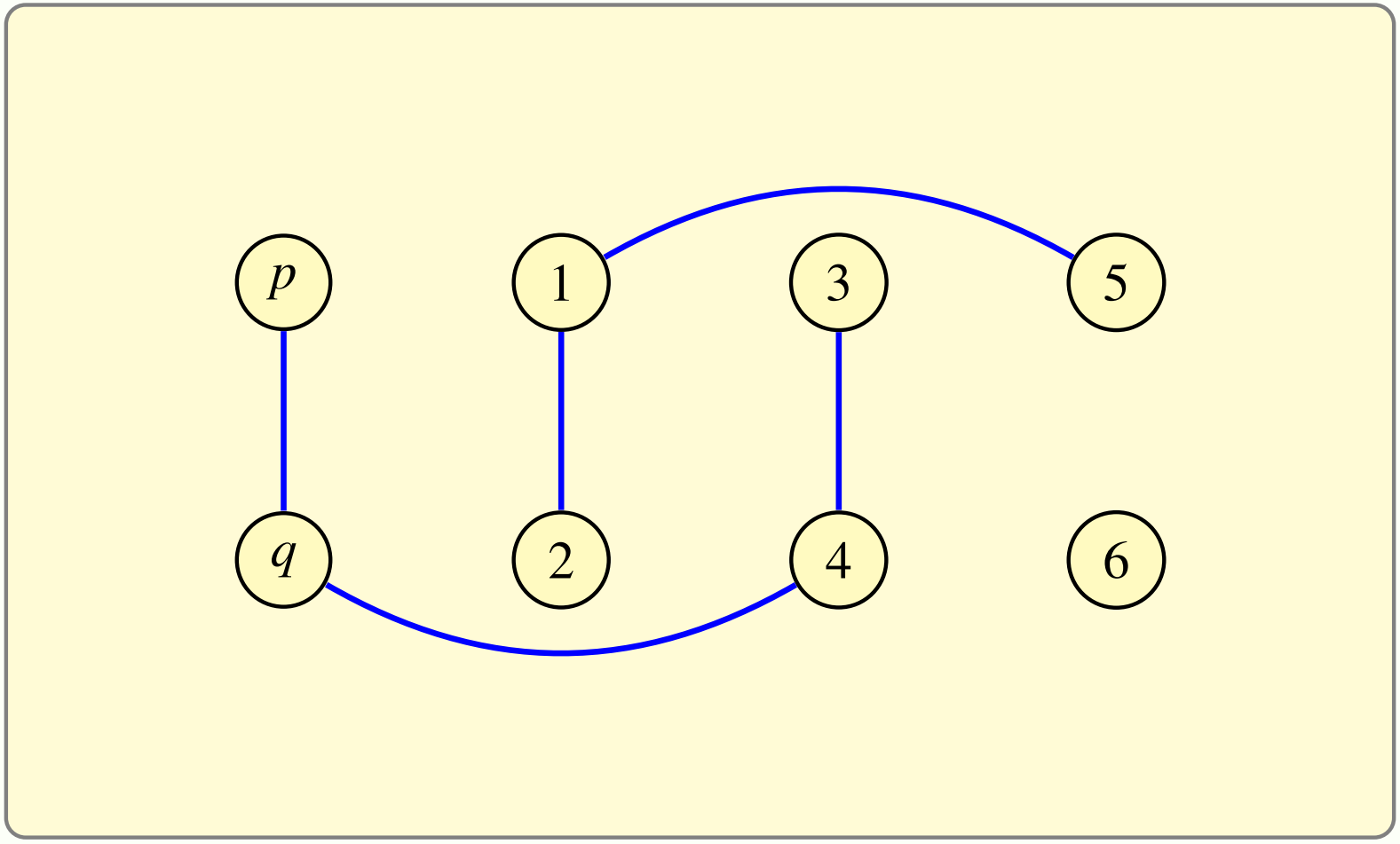


On fixe le graphe de l'automate.
$u = abaab$ le plus court mot qui sépare L_p et L_q : $p \sim_4 q$ et $p \nsim_5 q$

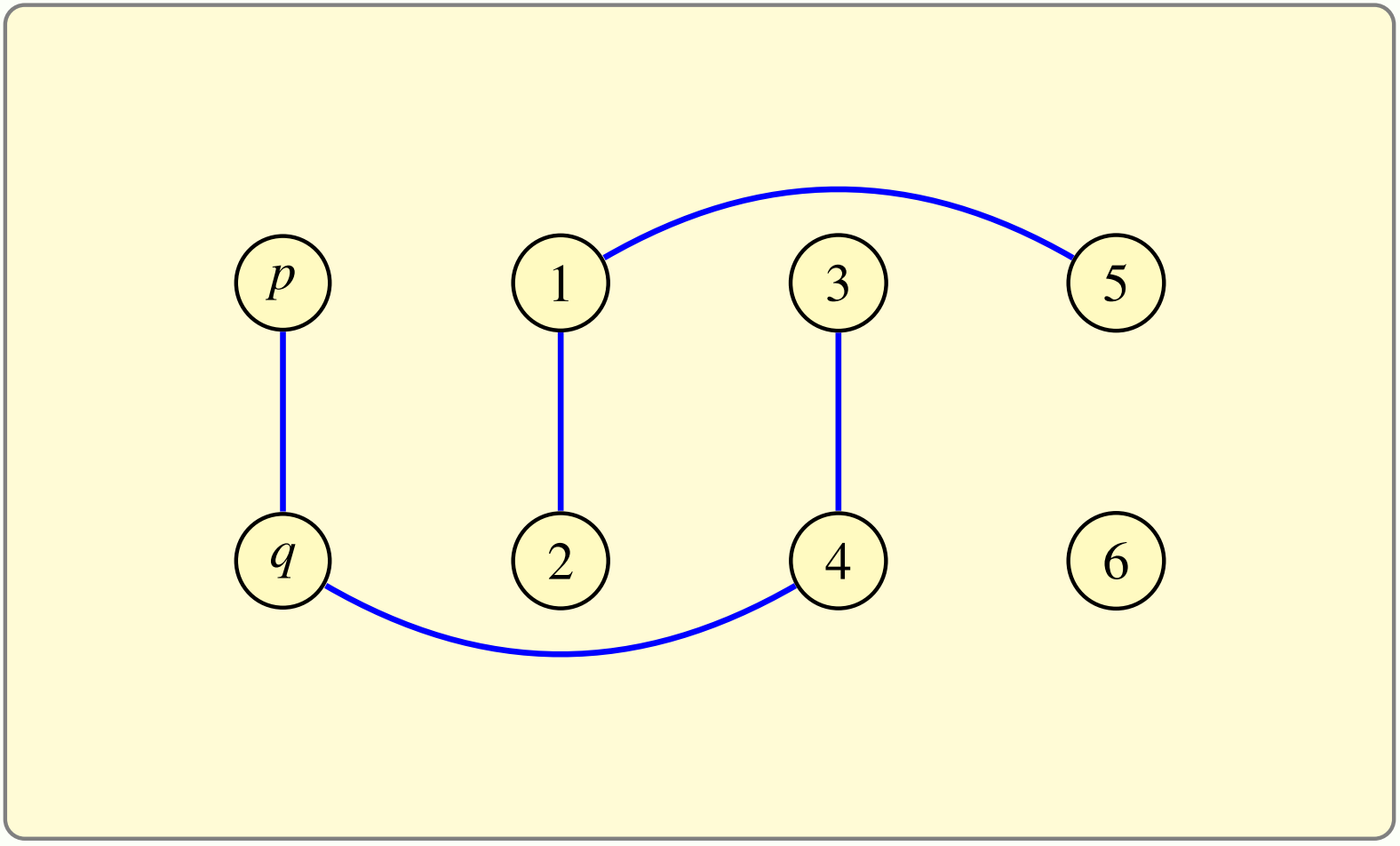


On fixe le graphe de l'automate.

$u = abaab$ le plus court mot qui sépare L_p et L_q : $p \sim_4 q$ et $p \not\sim_5 q$



Probabilité : $2^{3-8} = 2^{-5}$ pour un mot de longueur 5



Proposition

(un peu faux) Si on fixe le graphe de l'automate et deux états p et q, $p \sim_{\ell-1} q$ et $p \nsim_{\ell} q$ avec probabilité au plus $2^{-\ell}$, en tirant uniformément l'ensemble d'états terminaux.

Proposition

(un peu faux) Si on fixe le graphe de l'automate et deux états p et q, $p \sim_{\ell-1} q$ et $p \not\sim_{\ell} q$ avec probabilité au plus $2^{-\ell}$, en tirant uniformément l'ensemble d'états terminaux.

Théorème [Bassino, David, N. 09]

Pour la distribution uniforme sur les automates déterministes complets et accessibles, le nombre moyen d'itérations de l'algorithme de Moore est en $O(\log n)$. Sa complexité moyenne est donc en $O(n \log n)$.

Théorème [David 10]

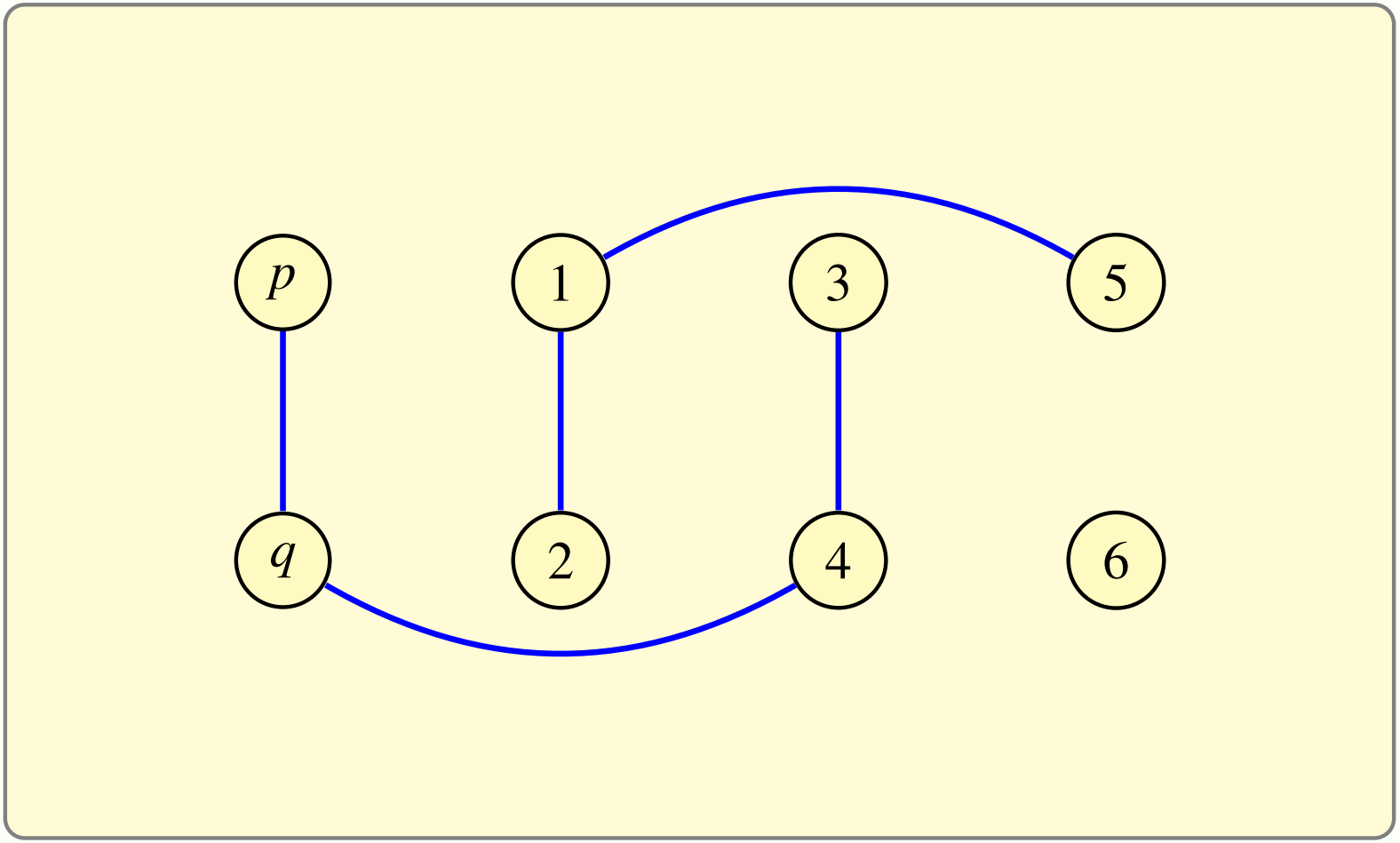
Pour la distribution uniforme sur les automates déterministes complets (pas nécessairement accessibles), le nombre moyen d'itérations de l'algorithme de Moore est en $O(\log \log n)$. Sa complexité moyenne est donc en $O(n \log \log n)$.

Conjecture

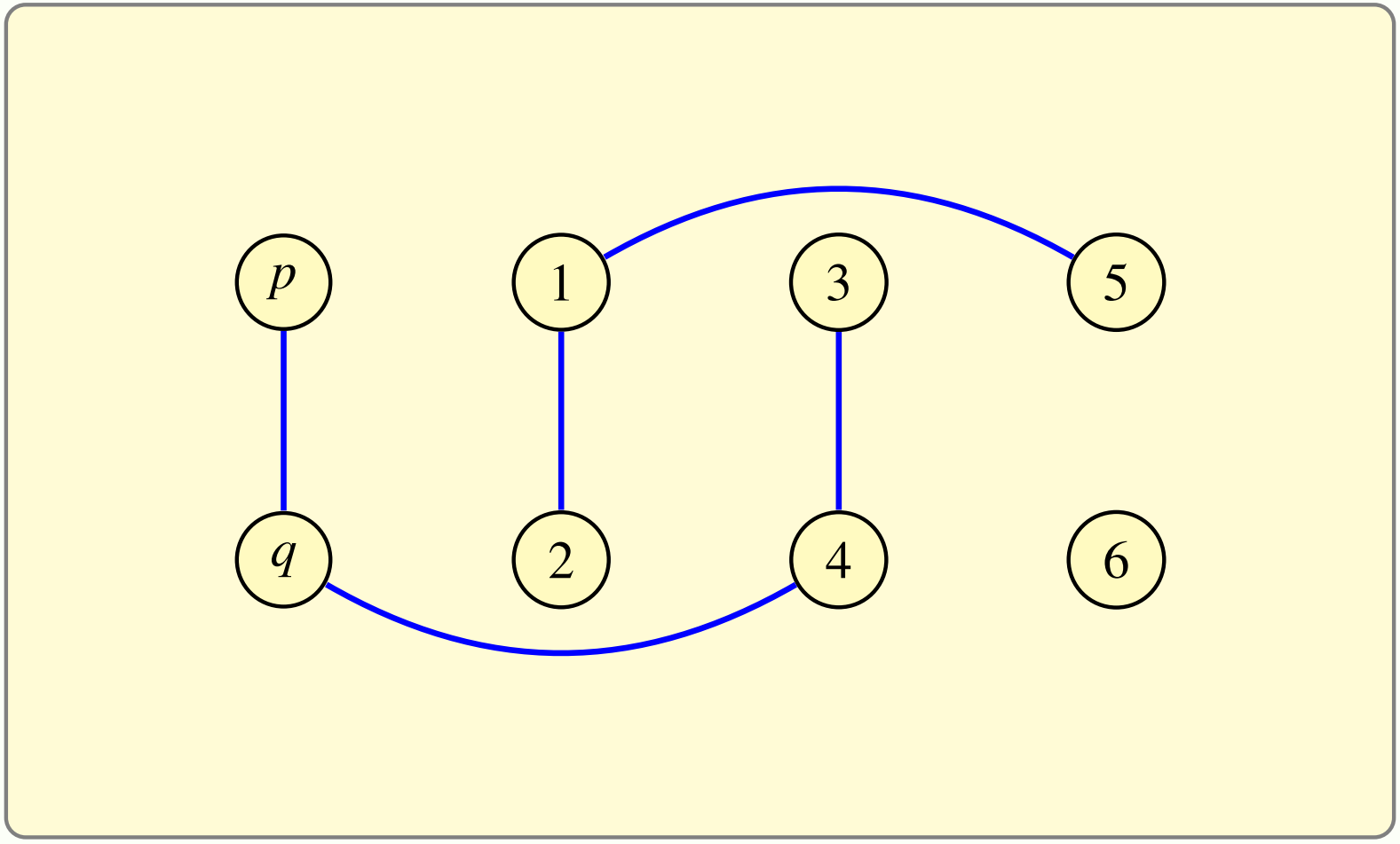
Le résultat est vrai pour la distribution uniforme sur les automates déterministes complets et accessibles.

Julien David utilise la structure typique d'un automate aléatoire pour la distribution uniforme.

Conditions nécessaires : ℓ contraintes.



Conditions nécessaires : ℓ contraintes.



Condition sur la distribution d'état finaux : la probabilité de respecter ℓ contraintes décroît exponentiellement avec ℓ.

Théorème [Bassino, David, N. 10]

Soit une distribution sur les automates déterministes telle que :

- Le graphe et les états terminaux sont choisis indépendamment ;
- La probabilité sur les ensembles d'états terminaux vérifie la condition de décroissance exponentielle.

Alors, le nombre moyen d'itérations de l'algorithme de Moore est en $O(\log n)$. Sa complexité moyenne est donc en $O(n \log n)$.

Corollaire [Bassino, David, N. 10]

Moore est de complexité moyenne $O(n \log n)$ pour les distributions suivantes :

- Uniforme sur les automates déterministes, complets et accessibles.
- Uniforme sur les automates déterministes et accessibles.
- Uniforme sur les automates déterministes, acycliques et accessibles.
- Uniforme sur les automates inversibles, déterministes et accessibles.
- N'importe lequel au-dessus, avec probabilité $p \in]0, 1[$ pour chaque état d'être terminal.
- N'importe lequel au-dessus, en interdisant d'avoir tous les états terminaux ou aucun état terminal.
- *etc*

Théorème [Bassino, David, N. 10]

Soit une distribution sur les automates déterministes telle que :

- Le graphe et les états terminaux sont choisis indépendamment ;
- La probabilité sur les ensembles d'états terminaux vérifie la condition de décroissance exponentielle.

Alors, le nombre moyen d'itérations de l'algorithme de Moore est en $O(\log n)$. Sa complexité moyenne est donc en $O(n \log n)$.

Partie V : Conclusion

Autres résultats :

- Factorisation standard de mots de Lyndon [Bassino, Clément, N. 05]
- Sous-groupes finiment engendrés d'un groupe libre [Bassino, N., Weil 08] [Bassino, Martino, N., Ventura, Weil 10]
- Génération aléatoire par approximation binomiale [Gouyou-Beauchamps, N. 10]

- On a pu obtenir des résultats en moyenne en théorie des langages.

- On a pu obtenir des résultats en moyenne en théorie des langages.
- On a utilisé les générateurs aléatoires pour guider l'analyse.

- On a pu obtenir des résultats en moyenne en théorie des langages.
- On a utilisé les générateurs aléatoires pour guider l'analyse.
- On a utilisé des techniques de combinatoire analytique.

- On a pu obtenir des résultats en moyenne en théorie des langages.
- On a utilisé les générateurs aléatoires pour guider l'analyse.
- On a utilisé des techniques de combinatoire analytique.

- Le choix de la distribution est crucial :
  - Essayer de couvrir en largeur (énoncés qui captent le plus de modèles probabilistes possibles).
  - Discuter des distributions qui arrivent dans les domaines qui utilisent les objets étudiés.

- On a pu obtenir des résultats en moyenne en théorie des langages.
- On a utilisé les générateurs aléatoires pour guider l'analyse.
- On a utilisé des techniques de combinatoire analytique.

- Le choix de la distribution est crucial :
  - Essayer de couvrir en largeur (énoncés qui captent le plus de modèles probabilistes possibles).
  - Discuter des distributions qui arrivent dans les domaines qui utilisent les objets étudiés.

- Si la taille d'un langage rationnel est le nombre d'états de son automate minimal, quelle est la taille moyenne de l'intersection de deux langages rationnels ?