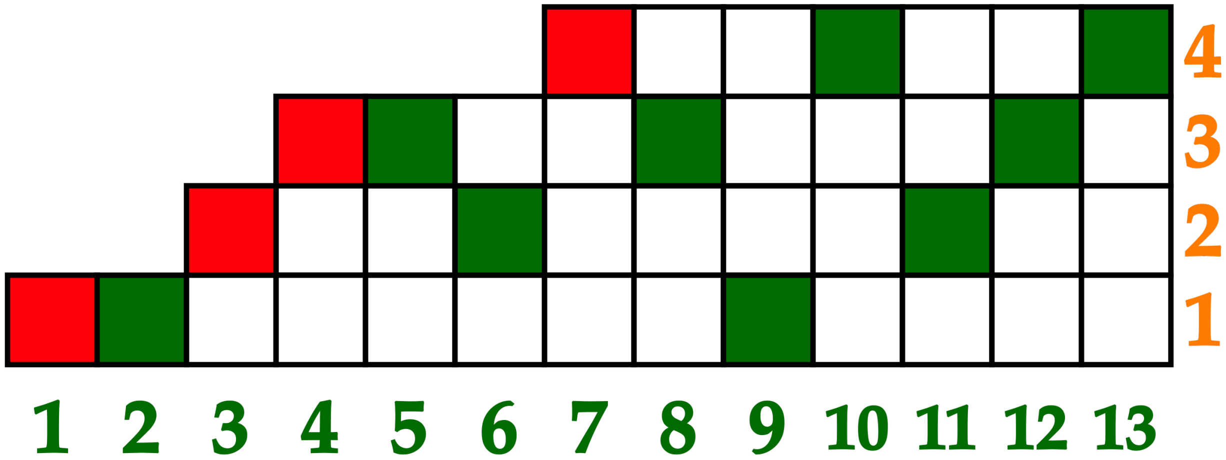
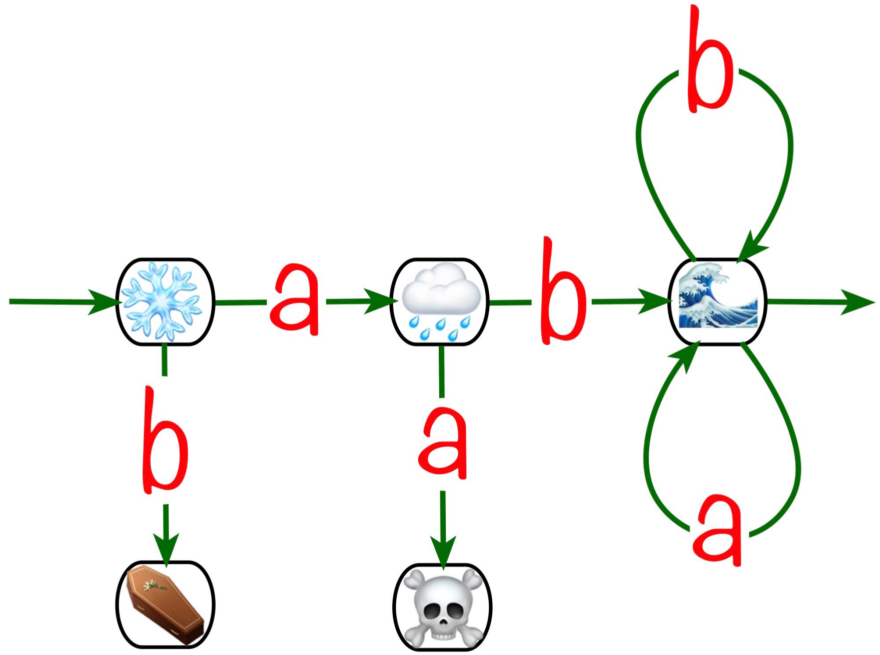
The impatient collector

Anis Amri, Philippe Chassaing

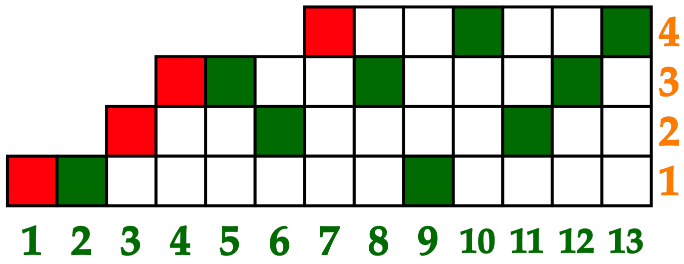


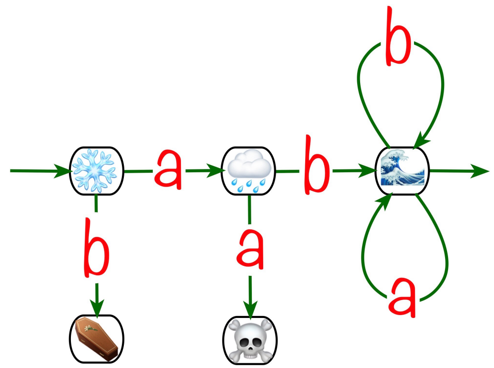


RANDOM GRAPHS AND ITS APPLICATIONS FOR NETWORKS,
2018/10/03-05, Institut Camille Jordan

The impatient collector

Anis Amri, Philippe Chassaing





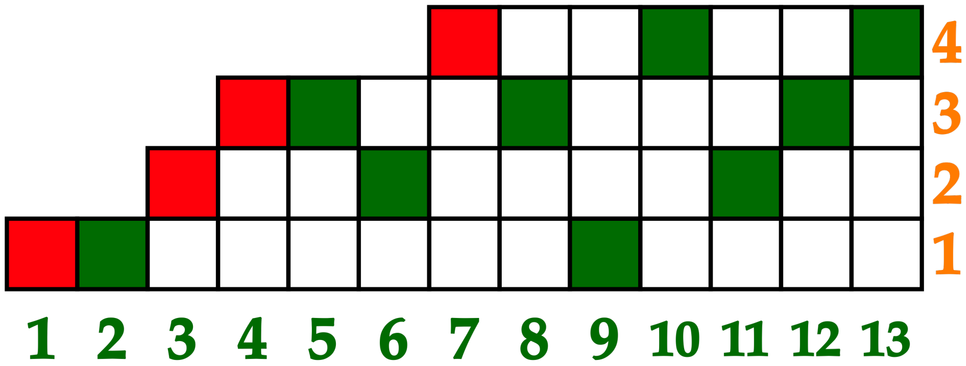
following papers by :

-Cyril Nicaud, Frédérique Bassino, Julien Clément, Andrea Sportiello, et al.

-Devroye and Cai

-Addario-Berry et al.

The "cool & collected" collector



$T_4 = 7$

n coupons (here $n=4$)

$\omega = \omega_1 \omega_2 \omega_3 \cdots \in [\![1,n]\!]^{\mathbb{N}}$ ω_i's are i.i.d. uniform.

$\omega_{[a,b]} = \omega_a \omega_{a+1} \cdots \omega_b$ is a factor of ω.

$T_n(\omega)$: the smallest k s.t.

$\omega_{[1,k]}$ is a surjection onto $[\![1,n]\!]$.

$$\mathbb{E}[T_n] = n H_n \simeq n \ln(n).$$

Erdös Rényi 1961 If $T_n = n \ln n + n Z_n$, then:

$$Z_n \xrightarrow{dist^n} \text{Gumbel (distr}^n\text{ function } e^{-e^{-x}}).$$

ON A CLASSICAL PROBLEM OF PROBABILITY THEORY

by

P. ERDŐS and A. RÉNYI

We consider the following classical "urn-problem". Suppose that there are n urns given, and that balls are placed at random in these urns one after the other. Let us suppose that the urns are labelled with the numbers $1, 2, \ldots, n$ and let ξ_j be equal to k if the j-th ball is placed into the k-th urn. We suppose that the random variables $\xi_1, \xi_2, \ldots, \xi_N, \ldots$ are independent, and $\mathbf{P}(\xi_j = k) = \frac{1}{n}$ for $j = 1, 2, \ldots$ and $k = 1, 2, \ldots, n$. By other words each ball may be placed in any of the urns with the same probability and the choices of the urns for the different balls are independent. We continue this process so long till there are at least m balls in every urn ($m = 1, 2, \ldots$). What can be said about the number of balls which are needed to achieve this goal?

We denote the number in question (which is of course a random variable) by $\nu_m(n)$. The "dixie cup"-problem considered in [1] is clearly equivalent with the above problem. In [1] the mean value $\mathbf{M}(\nu_m(n))$ of $\nu_m(n)$ has been evaluated (here and in what follows $\mathbf{M}(\ )$ denotes the mean value of the random variable in the brackets) and it has been shown that

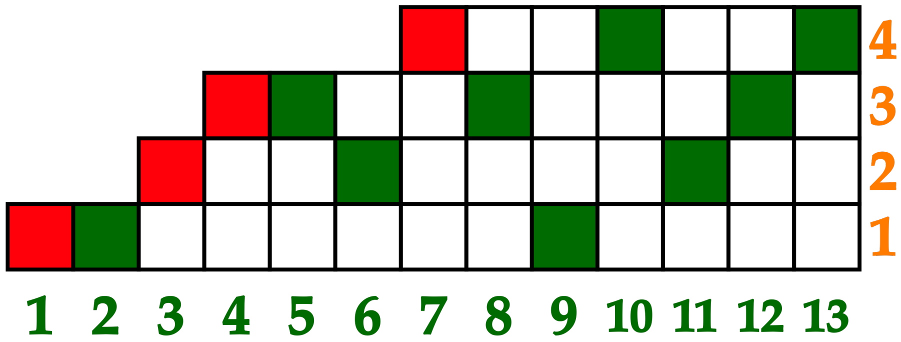
$$\mathbf{M}(\nu_m(n)) = n \log n + (m-1)\, n \log\log n + n \cdot C_m + o(n) \tag{1}$$

where C_m is a constant, depending on m. (The value of C_m is not given in [1]).

In the present note we shall go a step further and determine asymptotically the probability distribution of $\nu_m(n)$; we shall prove that for every real x we have

$$\lim_{n \to +\infty} \mathbf{P}\left(\frac{\nu_m(n)}{n} < \log n + (m-1) \log\log n + x\right) = \exp\left(-\frac{e^{-x}}{(m-1)!}\right). \tag{2}$$

The "cool & collected" collector



$T_4 = 7$

n coupons (here $n=4$)

$\omega = \omega_1 \omega_2 \omega_3 \cdots \in [\![1,n]\!]^{\mathbb{N}}$ ω_i's are i.i.d. uniform.

$\omega_{[a,b]} = \omega_a \omega_{a+1} \cdots \omega_b$ is a factor of ω.

$T_n(\omega)$: the smallest k s.t.

$\omega_{[1,k]}$ is a surjection onto $[\![1,n]\!]$.

$$\mathbb{E}[T_n] = n H_n \simeq n \ln(n).$$

Erdös Rényi 1961 If $T_n = n \ln n + n Z_n$, then:

$Z_n \xrightarrow{\text{dist}^{\text{n}}}$ Gumbel (distr$^{\text{n}}$ function $e^{-e^{-x}}$).

on aura donc à très-peu près pour l'expression du nombre i de tirages, après lesquels la probabilité que tous les numéros seront sortis est $\frac{1}{k}$,

$$i = (\log n - \log\log k).(n - \tfrac{1}{2} + \tfrac{1}{2}\log k) + \tfrac{1}{4}\log k;$$

ON A CLASSICAL PROBLEM OF PROBABILITY THEORY

by

P. ERDŐS and A. RÉNYI

We consider the following classical "urn-problem". Suppose that there are n urns given, and that balls are placed at random in these urns one after the other. Let us suppose that the urns are labelled with the numbers $1, 2, \ldots, n$ and let ξ_j be equal to k if the j-th ball is placed into the k-th urn. We suppose that the random variables $\xi_1, \xi_2, \ldots, \xi_N, \ldots$ are independent, and $\mathbf{P}(\xi_j = k) = \frac{1}{n}$ for $j = 1, 2, \ldots$ and $k = 1, 2, \ldots, n$. By other words each ball may be placed in any of the urns with the same probability and the choices of the urns for the different balls are independent. We continue this process so long till there are at least m balls in every urn ($m = 1, 2, \ldots$). What can be said about the number of balls which are needed to achieve this goal?

We denote the number in question (which is of course a random variable) by $\nu_m(n)$. The "dixie cup"-problem considered in [1] is clearly equivalent with the above problem. In [1] the mean value $\mathbf{M}(\nu_m(n))$ of $\nu_m(n)$ has been evaluated (here and in what follows $\mathbf{M}(\ )$ denotes the mean value of the random variable in the brackets) and it has been shown that

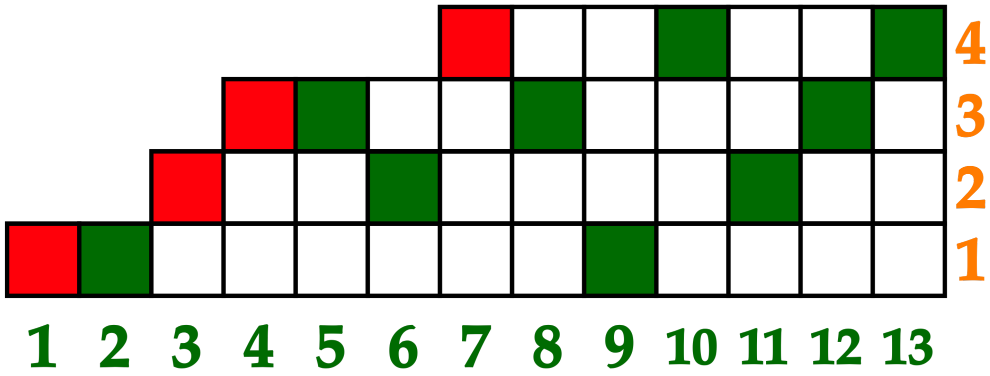
$$\mathbf{M}(\nu_m(n)) = n \log n + (m-1)\, n \log\log n + n \cdot C_m + o(n) \tag{1}$$

where C_m is a constant, depending on m. (The value of C_m is not given in [1]).

In the present note we shall go a step further and determine asymptotically the probability distribution of $\nu_m(n)$; we shall prove that for every real x we have

$$\lim_{n\to+\infty} \mathbf{P}\left(\frac{\nu_m(n)}{n} < \log n + (m-1)\log\log n + x\right) = \exp\left(-\frac{e^{-x}}{(m-1)!}\right). \tag{2}$$

The "cool & collected" collector



$T_4 = 7$

n coupons (here $n = 4$)

$\omega = \omega_1 \omega_2 \omega_3 \cdots \in [\![1, n]\!]^{\mathbb{N}}$ ω_i's are i.i.d. uniform.

$\omega_{[a,b]} = \omega_a \omega_{a+1} \cdots \omega_b$ is a factor of ω.

$T_n(\omega)$: the smallest k s.t.

$\omega_{[1,k]}$ is a surjection onto $[\![1,n]\!]$.

$$\mathbb{E}[T_n] = n H_n \simeq n \ln(n).$$

Erdös Rényi 1961 If $T_n = n \ln n + n Z_n$, then:

$Z_n \xrightarrow{\text{dist}^n}$ Gumbel (distr^n function $e^{-e^{-x}}$).

on aura donc à très-peu près pour l'expression du nombre i de tirages, après lesquels la probabilité que tous les numéros seront sortis est $\frac{1}{k}$,

$$i = (\log n - \log\log k).(n - \tfrac{1}{2} + \tfrac{1}{2}\log k) + \tfrac{1}{4}\log k;$$

THÉORIE

ANALYTIQUE

DES PROBABILITÉS;

PAR M. LE COMTE LAPLACE,

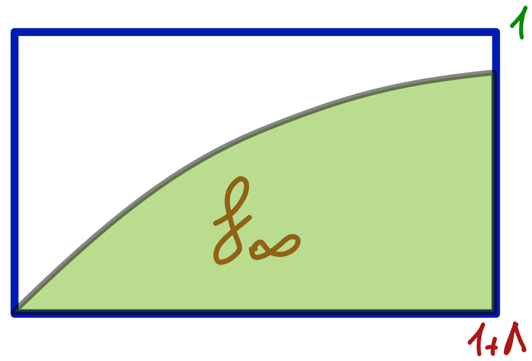
Chancelier du Sénat-Conservateur, Grand-Officier de la Légion d'Honneur; Membre de l'Institut impérial et du Bureau des Longitudes de France; des Sociétés royales de Londres et de Gottingue; des Académies des Sciences de Russie, de Danemarck, de Suède, de Prusse, de Hollande, d'Italie, etc.

The completion curve

$\omega = \omega_1 \omega_2 \omega_3 \cdots$ is "uniform" on $[\![1,n]\!]^{\mathbb{N}}$

$$\xi_n(t,\omega) = \frac{1}{n} \# \{ \omega_k \ , \ 1 \leq k \leq tn \}$$

is the completion curve.



Proposition. (cool case)

$$\lim_n \mathbb{E}\left[\xi_n(t) \right] = 1 - e^{-t}$$

$$\equiv f_\infty(t).$$

The completion curve

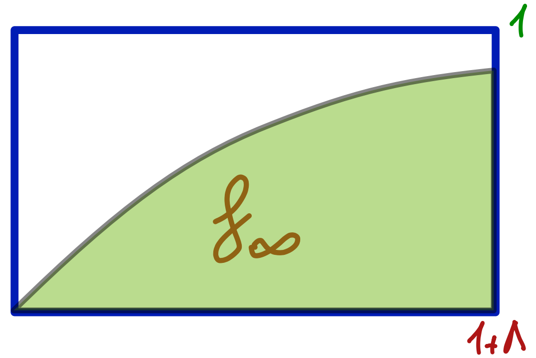
$\omega = \omega_1 \omega_2 \omega_3 \cdots$ is "uniform" on $[\![1,n]\!]^{\mathbb{N}}$

$\xi_n(t,\omega) = \frac{1}{n} \# \{\omega_k \,,\ 1 \leq k \leq tn\}$

is the completion curve.

Proposition. (cool case)

$$\lim_n \mathbb{E}[\xi_n(t)] = 1 - e^{-t} \equiv f_\infty(t). \quad \longleftarrow \quad \mathbb{E}[\ ] = 1 - \left(1 - \frac{1}{n}\right)^{\lfloor tn \rfloor}$$



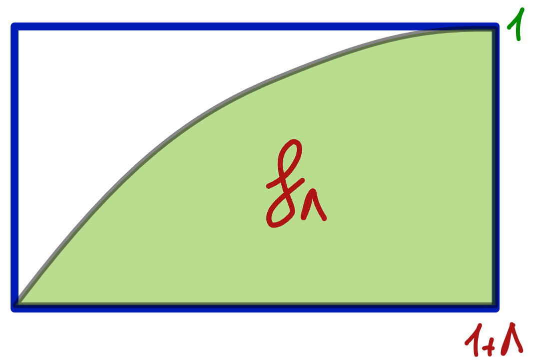
Question (impatient case)

Given that $T_n \leq (1+\Lambda)n$, $(\Lambda > 0)$

asymptotic behaviour of $\xi_n(t)$??

I.e. description of

$$f_\Lambda(t) = \lim_n \mathbb{E}[\xi_n(t) \mid T_n \leq (1+\Lambda)n] \ ??$$



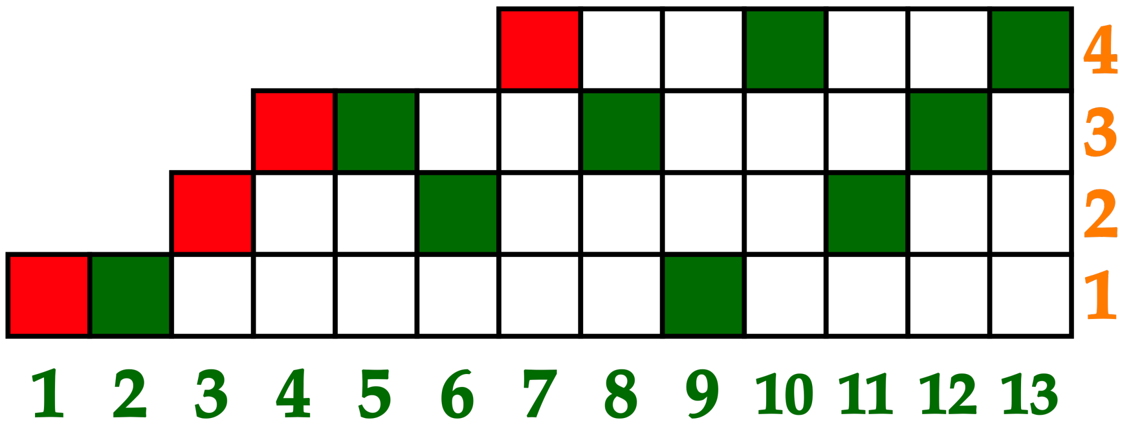
Nota: $\mathbb{P}(T_n \leq (1+\Lambda)n) \approx C(\Lambda)\, e^{-n J(\Lambda)}$ où $J(\Lambda) \underset{+\infty}{\searrow} 0$.

Stirling numbers of the 2nd kind $\left\{ {m \atop \ell} \right\}$

$\left\{ {m \atop \ell} \right\}$: number of partitions of $[\![1,m]\!]$ in ℓ parts, $\ell \leqslant m$.

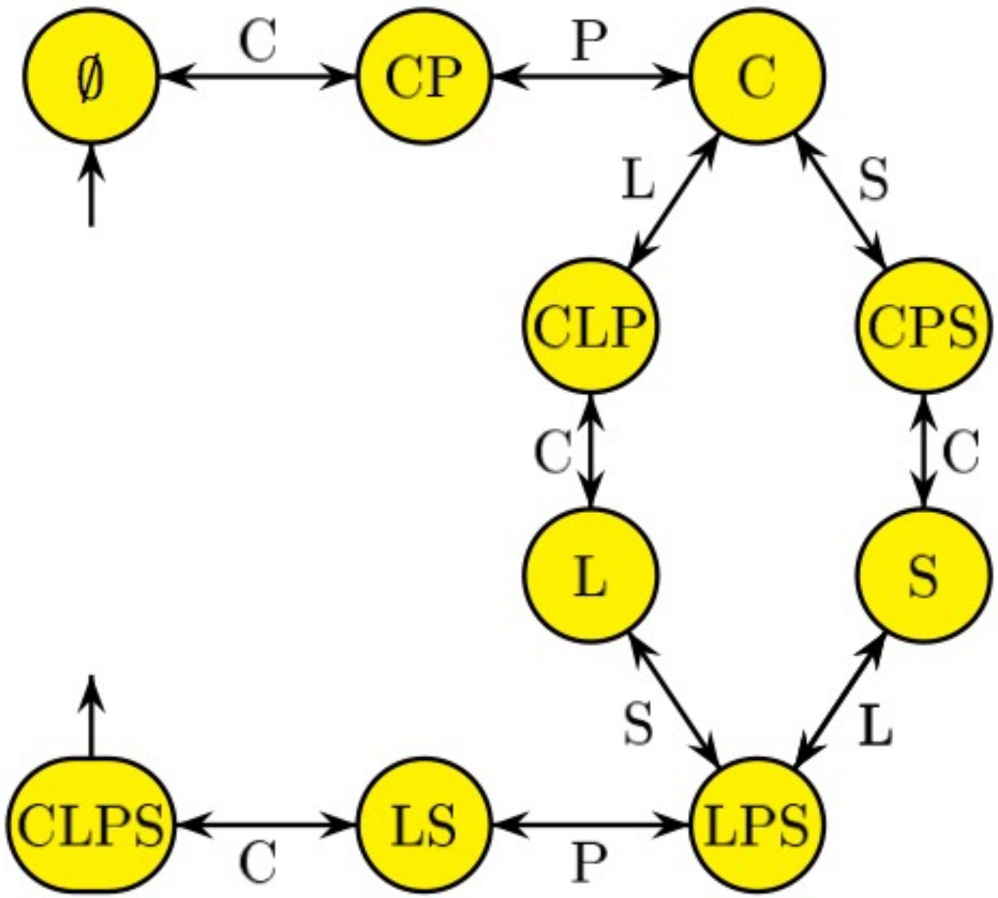
$\left\{ {m \atop \ell} \right\} \times (\ell!)$: number of surjections from $[\![1,m]\!]$ to $[\![1,\ell]\!]$.

$$\left\{ {m \atop \ell} \right\} \times (\ell!) \times \ell^{-m} = \mathbb{P}(T_\ell \leqslant m).$$



$\left\{ {13 \atop 4} \right\}$

Application : automata



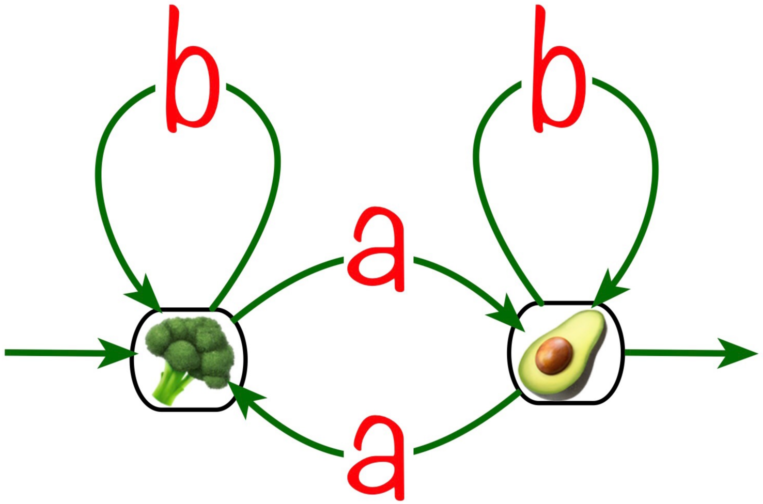
C = chèvre = goat

L = loup = wolf

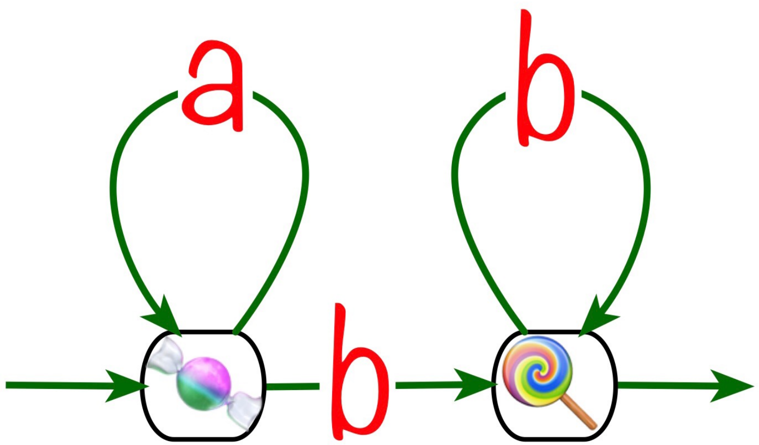
S = salade = cabbage

P = passeur = farmer

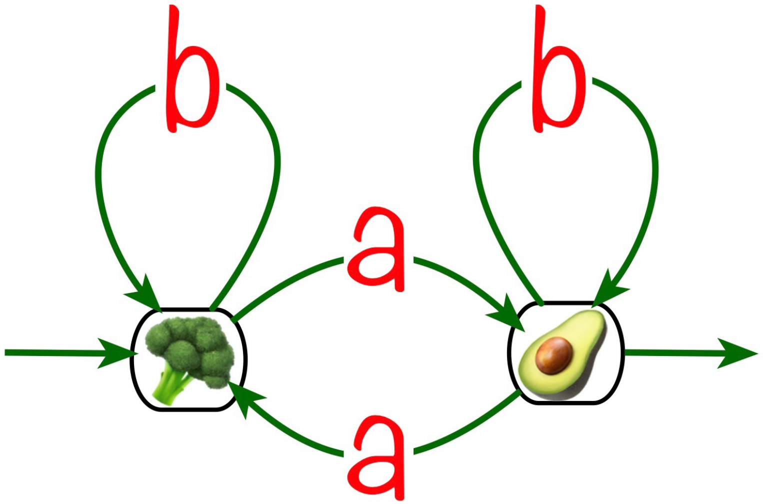
Automata and languages



It checks if a word belongs to a language...

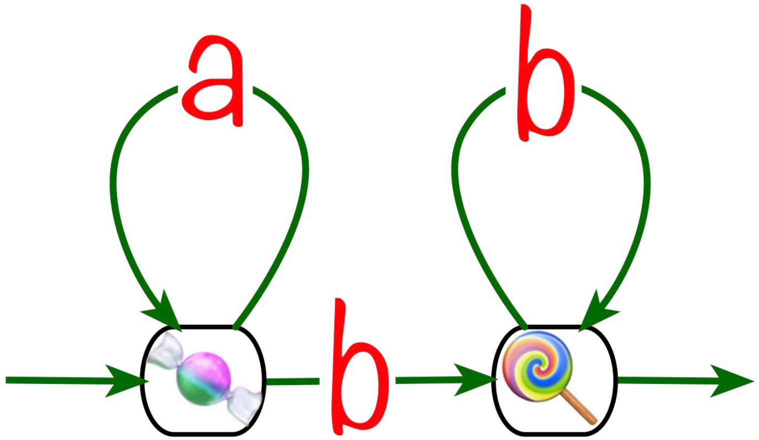


Automata and languages

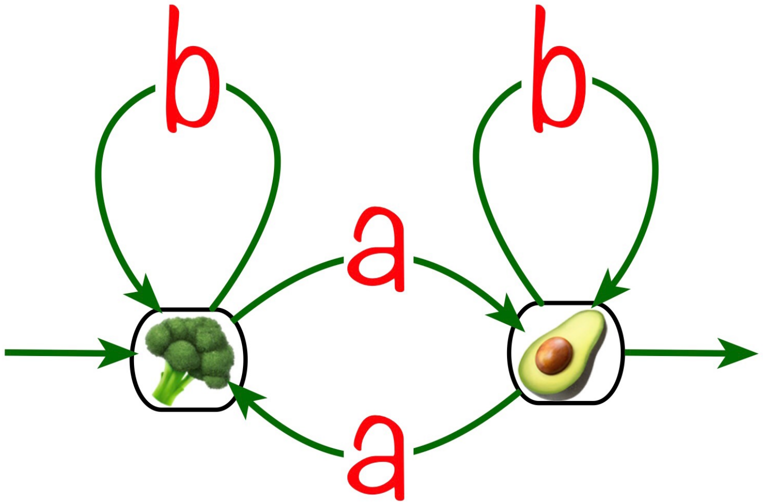


{words with an even number of a's}

It checks if a word belongs to a language...

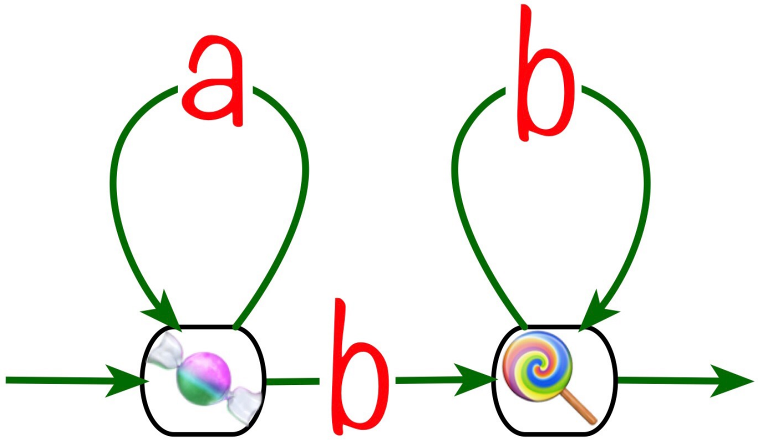


Automata and languages



{words with an even number of a's}

It checks if a word belongs to a language...



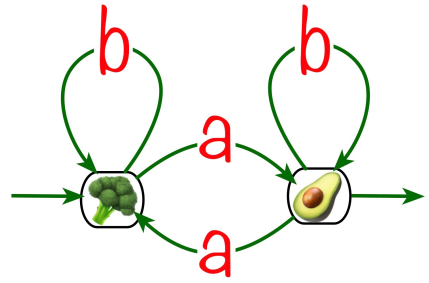
$a^k b^l, \; l \geq 1$

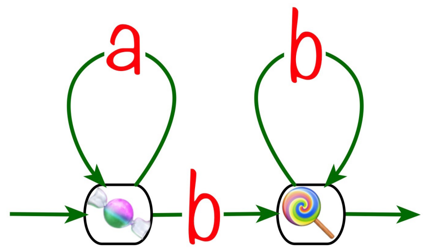
Automata : some terminology

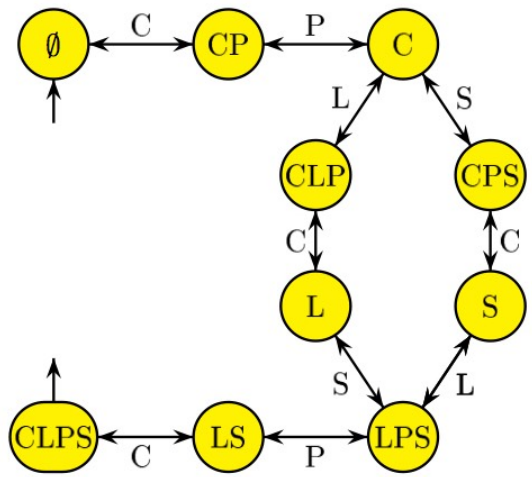
* state : vertex

* alphabet (A) : set of edges' labels.

* starting state, final states



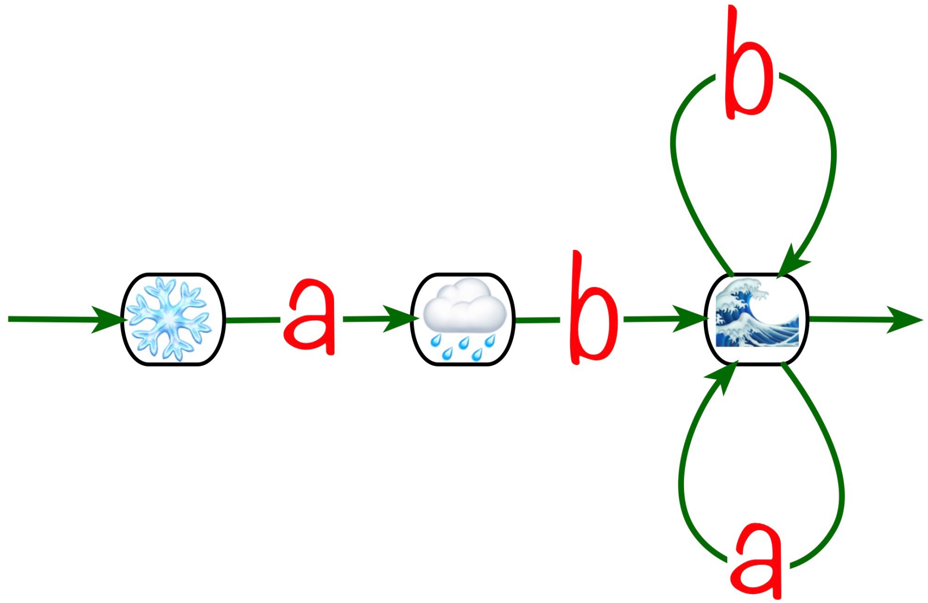




Automata : some terminology

- state : vertex
- alphabet (A) : set of edges' labels.
- starting state, final states
- deterministic automata : $\forall a \in A$, $\forall$ vertex,

  at most one outgoing edge labeled a.



Automata : some terminology

* state : vertex

* alphabet (A) : set of edges' labels.

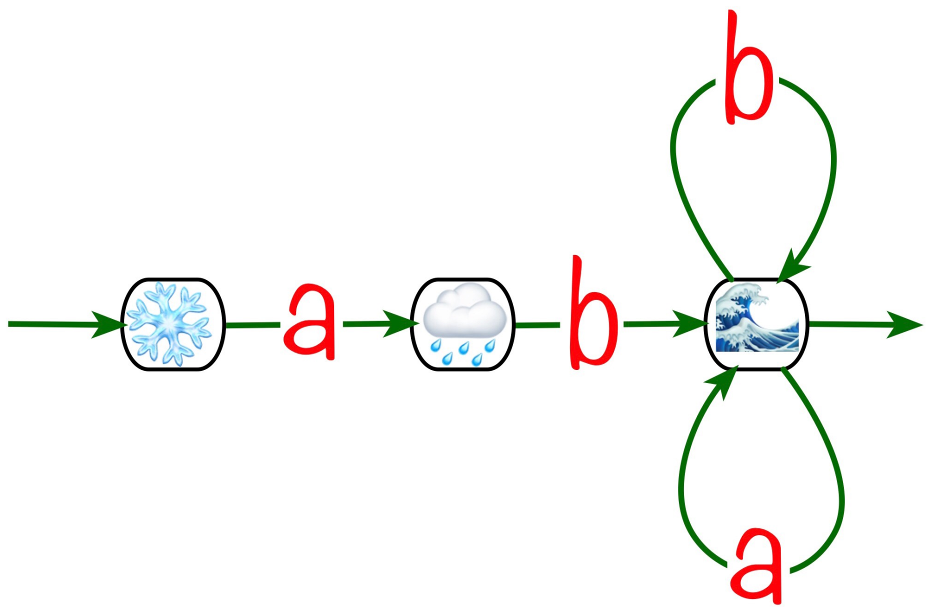
* starting state, final states

* deterministic automata : $\forall a \in A$, $\forall$ vertex,

  at most one outgoing edge labeled a.

* complete automata : " " " exactly

  one outgoing edge ....

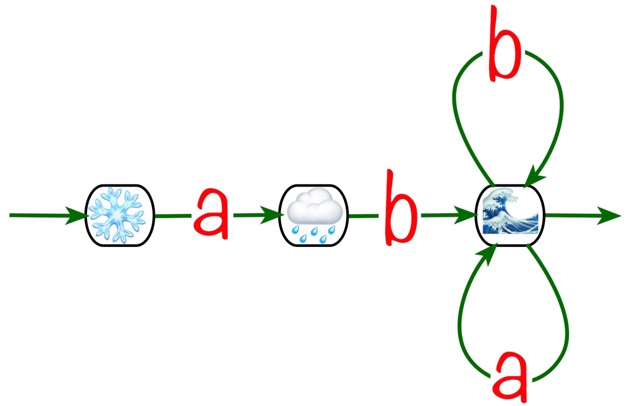


Automata : some terminology

* state : vertex

* alphabet (A) : set of edges' labels.

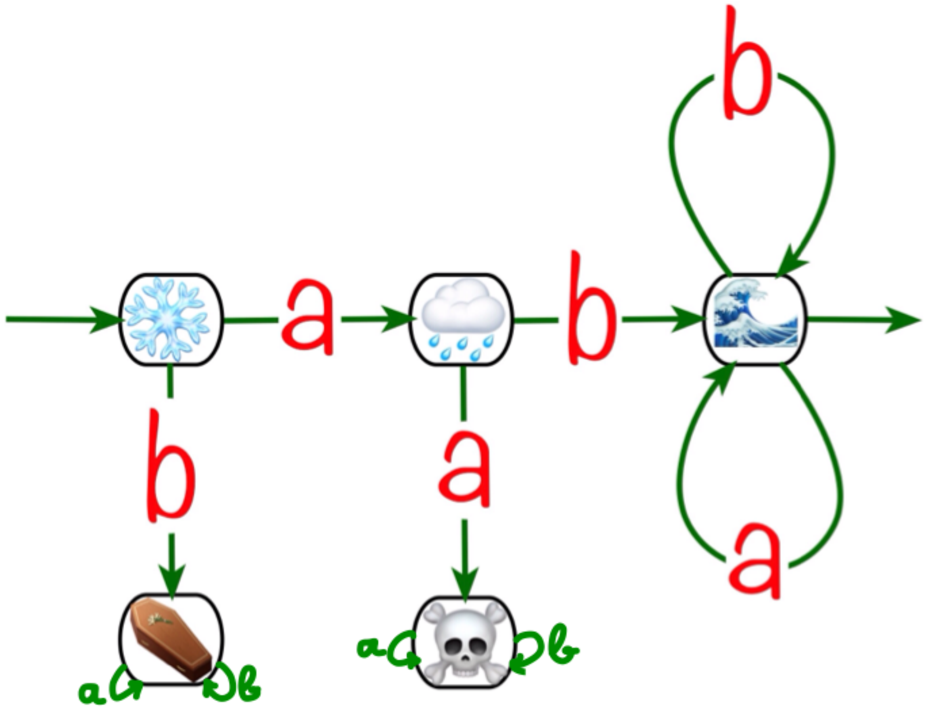
* starting state, final states



* deterministic automata : $\forall a \in A$, $\forall$ vertex,
  at most one outgoing edge labeled a.

* complete automata : " " " exactly
  one outgoing edge ....

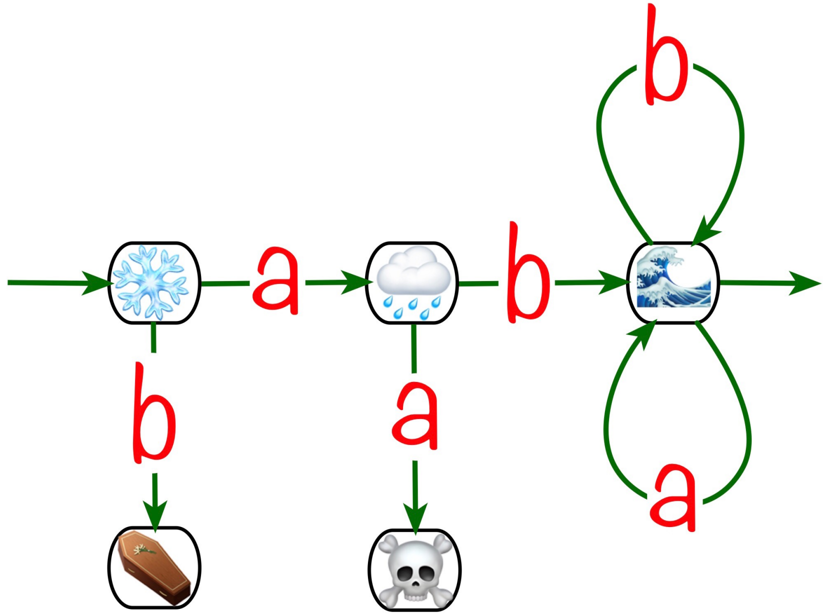
* accessible automata : each state can be
  reached, from the starting state.



Relation automata - coupon collector

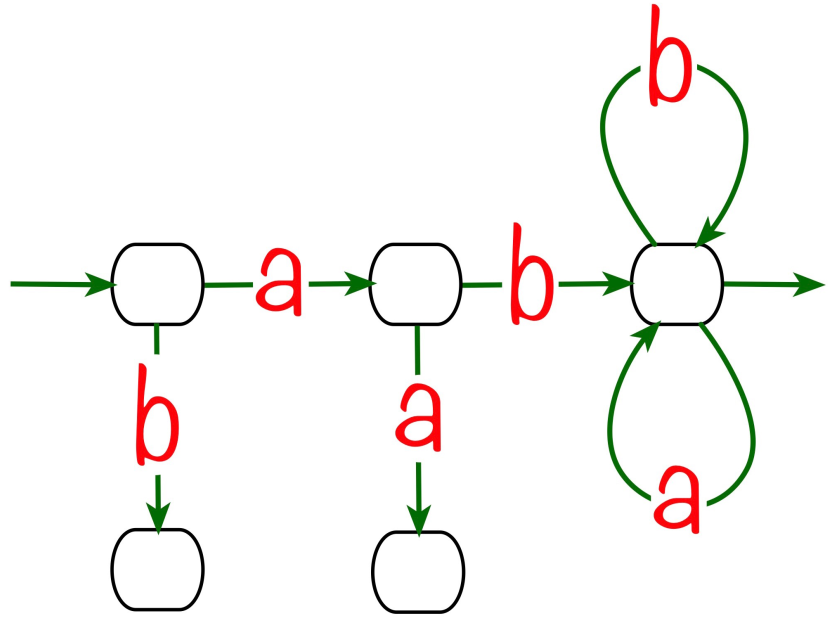
← input for Moore's algorithm

$\mathcal{A}_{n,k}$: set of the deterministic complete accessible automata (DCAA) with n vertices and a k-letters alphabet $\mathcal{A}$.



Koršunov 1978: $\#\mathcal{A}_{n,k} \simeq \left\{ {kn+1 \atop n} \right\} n! \times C(k,n)$

in which $\lim_n C(k,n) = C(k) \in]0,1[$.



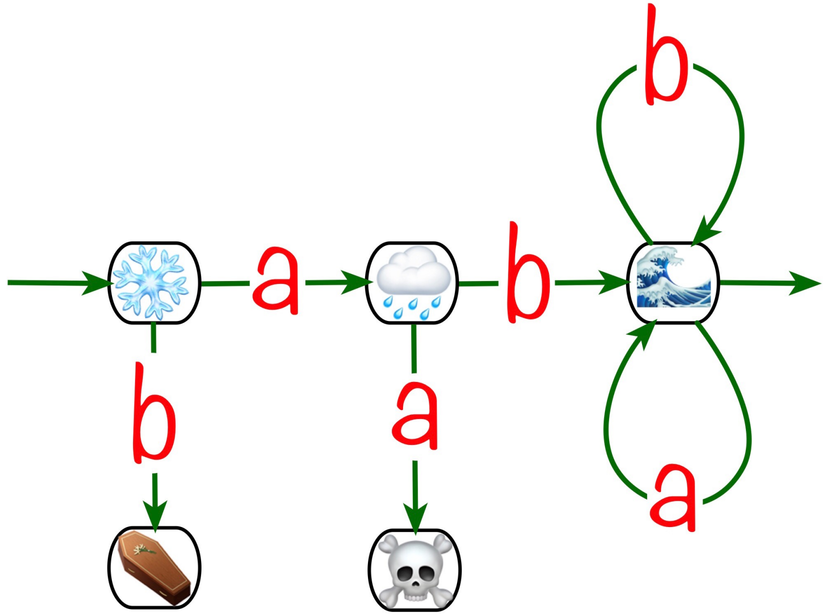
Provided an alphabetic order on $\mathcal{A}$, breadth-first search provides an ordering of the vertices (if accessibility).

Relation automata - coupon collector

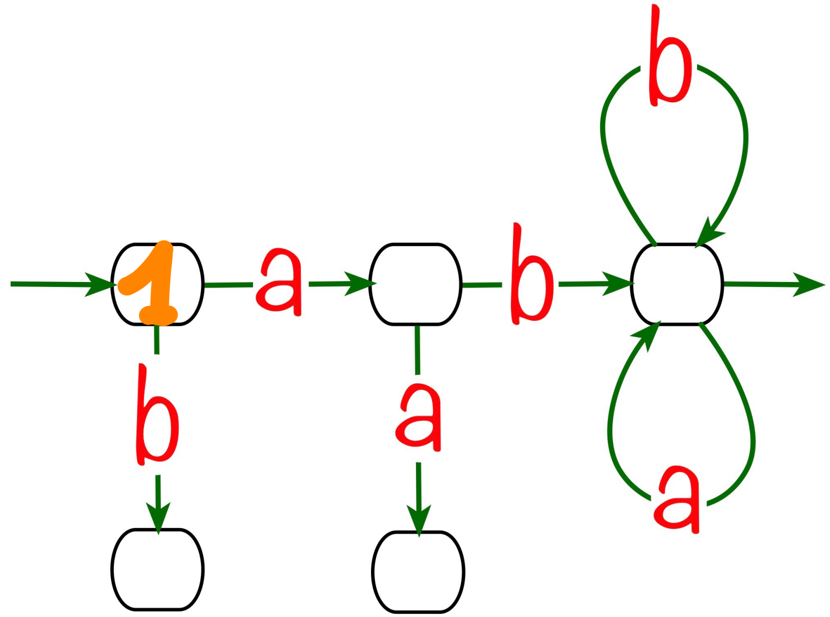
$\mathfrak{a}_{n,k}$: set of the deterministic complete accessible automata (DCAA) with n vertices and a k-letters alphabet $\mathcal{A}$.

Koršunov 1978: $\#\mathfrak{a}_{n,k} \simeq \left\{ {kn+1 \atop n} \right\} n! \times C(k,n)$

in which $\lim_n C(k,n) = C(k) \in]0,1[$.



Provided an alphabetic order on $\mathcal{A}$, breadth-first search provides an ordering of the vertices (if accessibility).

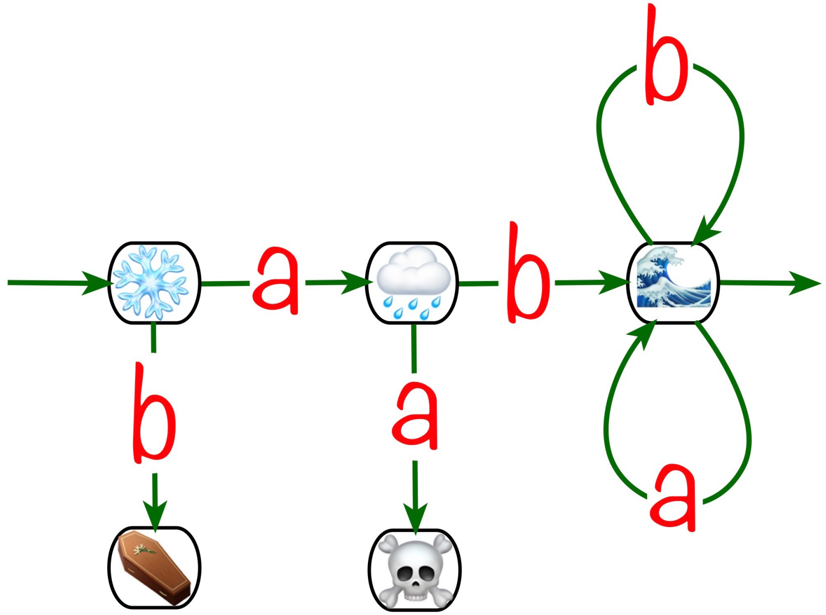


Relation automata - coupon collector

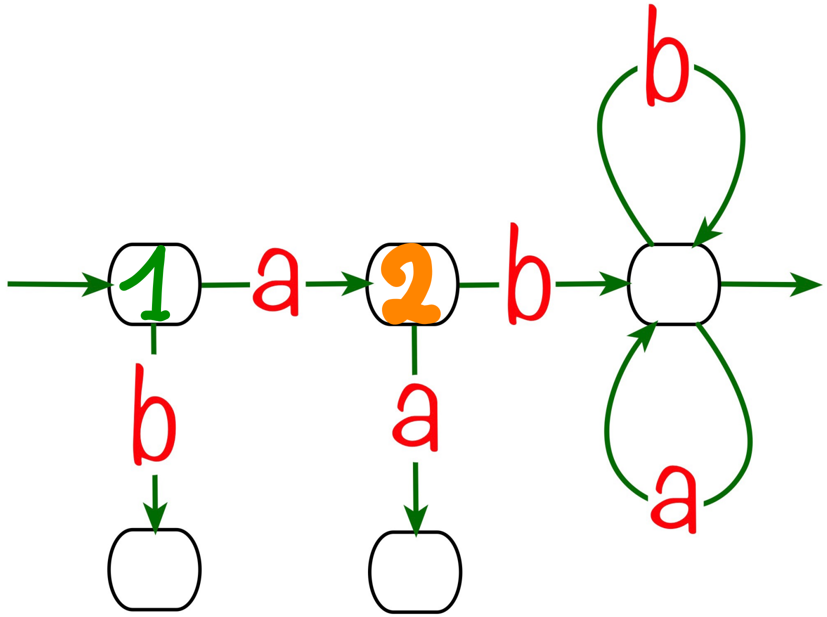
$\mathfrak{A}_{n,k}$: set of the deterministic complete accessible automata (DCAA) with n vertices and a k-letters alphabet $\mathcal{A}$.

Koršunov: $\#\mathfrak{A}_{n,k} \simeq \left\{ {kn+1 \atop n} \right\} n! \times C(k,n)$
1978

in which $\lim_n C(k,n) = C(k) \in]0,1[$.

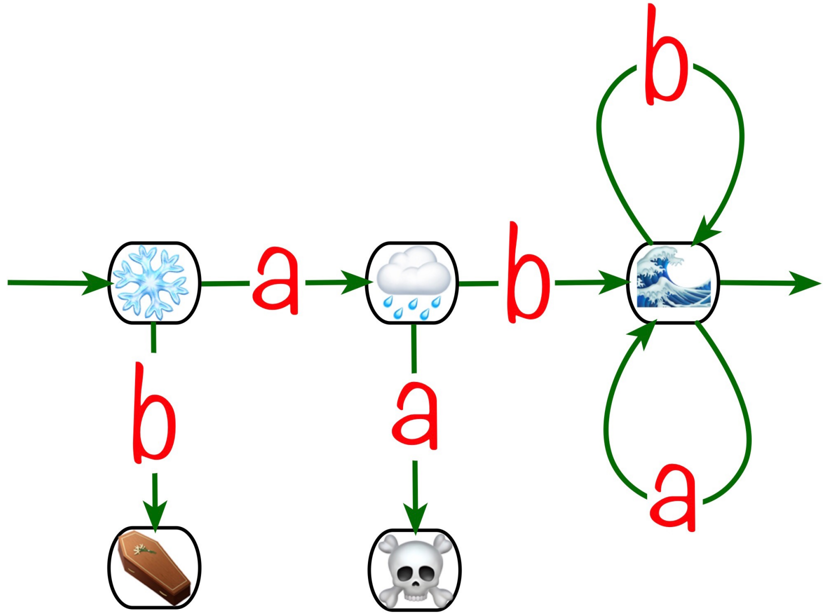


Provided an alphabetic order on $\mathcal{A}$, breadth-first search provides an ordering of the vertices (if accessibility).



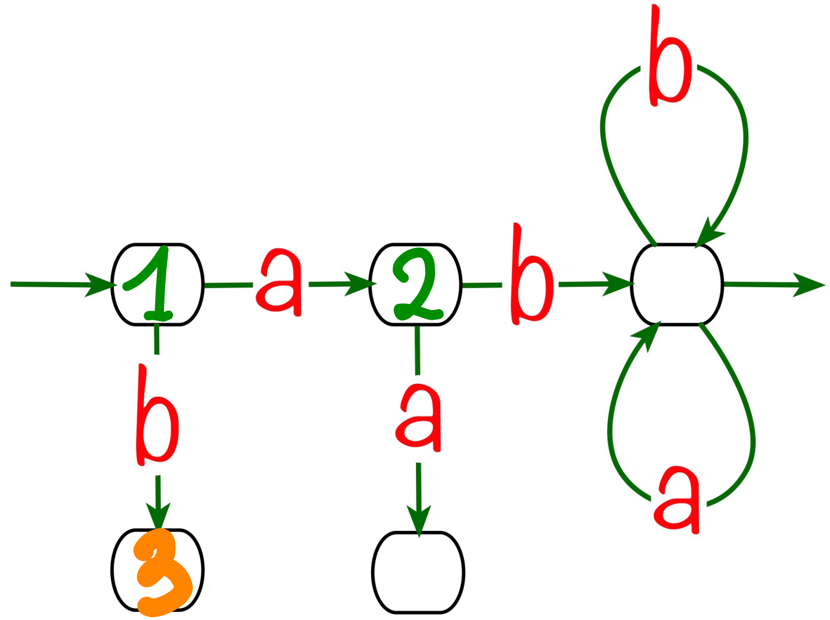
Relation automata - coupon collector

$\mathfrak{A}_{n,k}$: set of the deterministic complete accessible automata (DCAA) with n vertices and a k-letters alphabet $\mathcal{A}$.



Koršunov 1978: $\#\mathfrak{A}_{n,k} \simeq \left\{ {kn+1 \atop n} \right\} n! \times C(k,n)$

in which $\lim_{n} C(k,n) = C(k) \in]0,1[$.



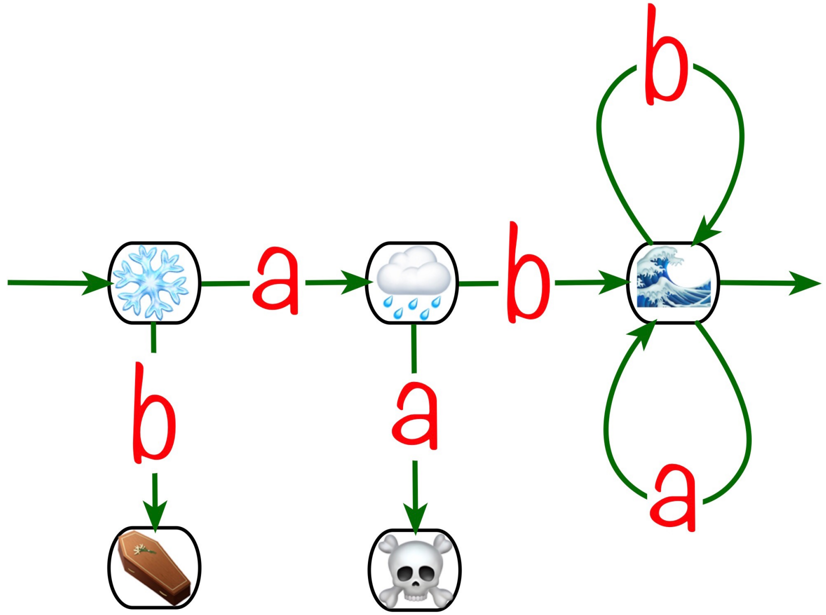
Provided an alphabetic order on $\mathcal{A}$, breadth-first search provides an ordering of the vertices (if accessibility).

Relation automata - coupon collector

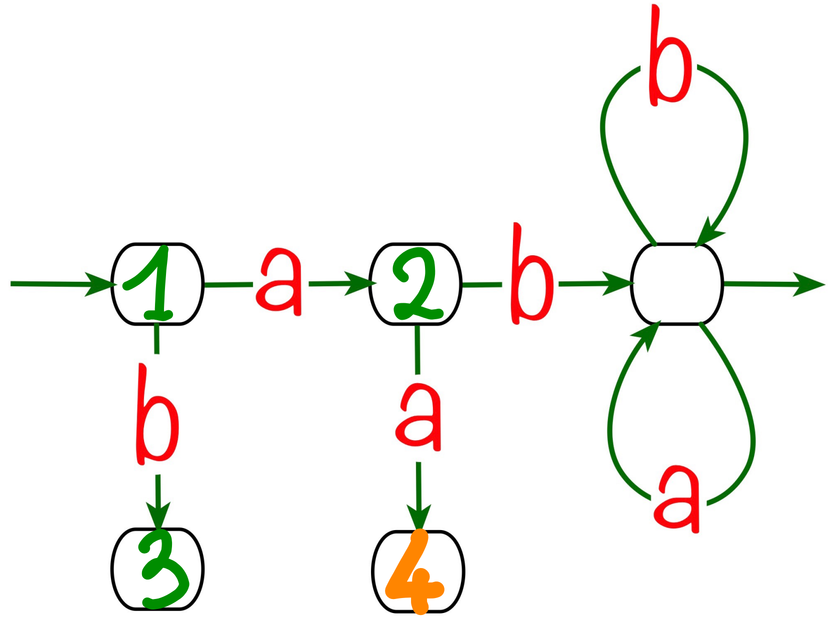
$\mathcal{A}_{n,k}$: set of the deterministic complete accessible automata (DCAA) with n vertices and a k-letters alphabet $\mathcal{A}$.

Koršunov 1978: $\# \mathcal{A}_{n,k} \simeq \left\{ {kn+1 \atop n} \right\} n! \times C(k,n)$

in which $\lim_n C(k,n) = C(k) \in]0,1[$.

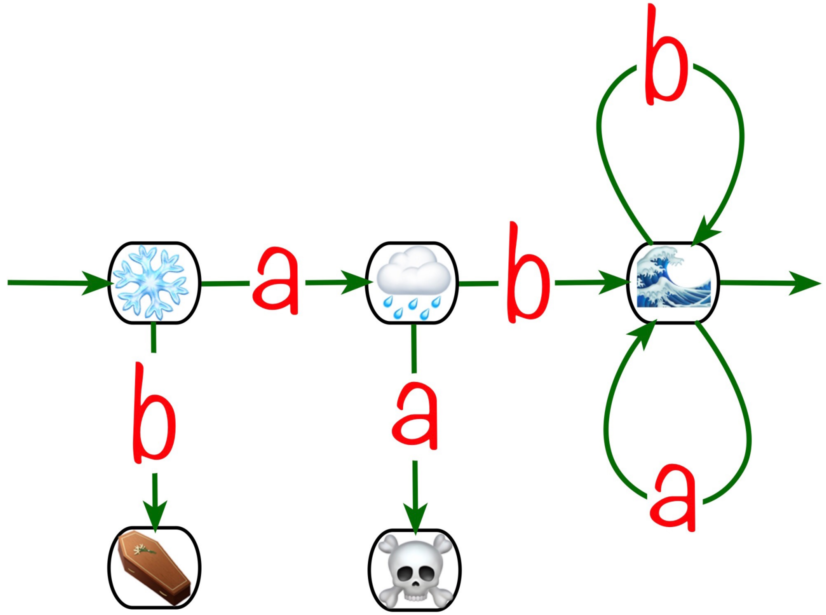


Provided an alphabetic order on $\mathcal{A}$, breadth-first search provides an ordering of the vertices (if accessibility).



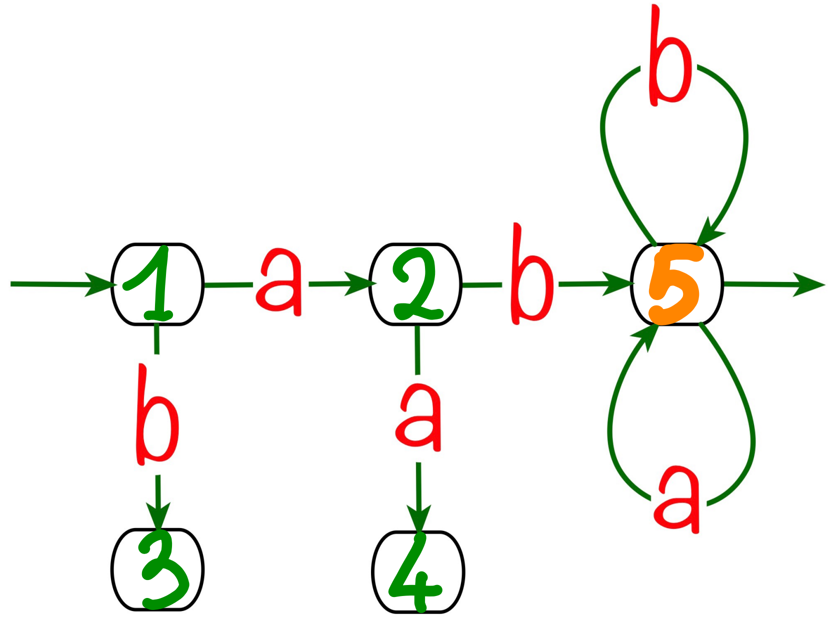
Relation automata - coupon collector

$\mathcal{A}_{n,k}$: set of the deterministic complete accessible automata (DCAA) with n vertices and a k-letters alphabet $\mathcal{A}$.



Koršunov: 1978 $\#\mathcal{A}_{n,k} \simeq \left\{ {kn+1 \atop n} \right\} n! \times C(k,n)$

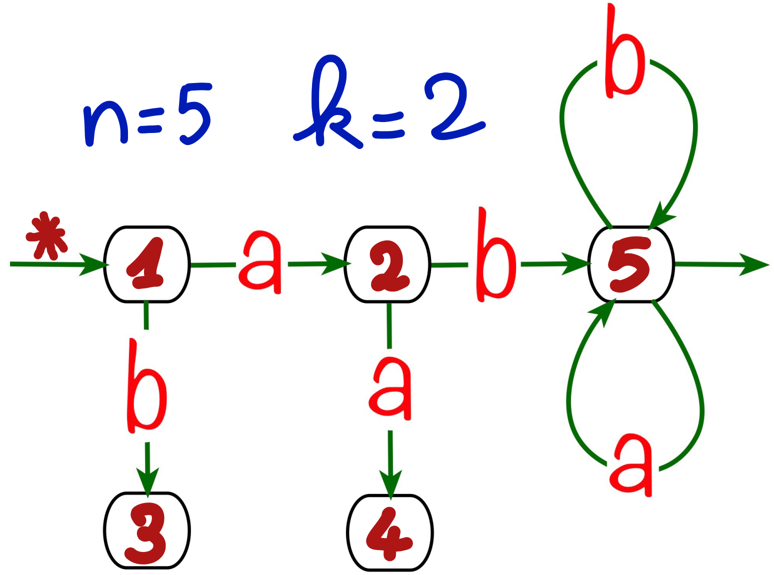
in which $\lim_n C(k,n) = C(k) \in]0,1[$.



Provided an alphabetic order on $\mathcal{A}$, breadth-first search provides an ordering of the vertices (if accessibility).

Relation automata - coupon collector

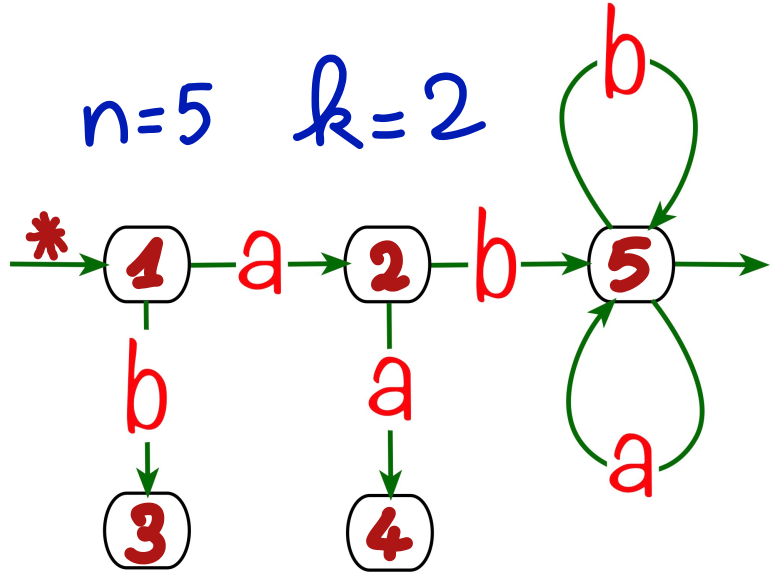
$n=5 \quad k=2$



* →	1a	1b	2a	2b	3a	3b	4a	4b	5a	5b
1	2	3	4	5	3	3	4	4	5	5

Relation automata - coupon collector

$n=5$ $k=2$



* →	1a	1b	2a	2b	3a	3b	4a	4b	5a	5b
1	2	3	4	5	3	3	4	4	5	5

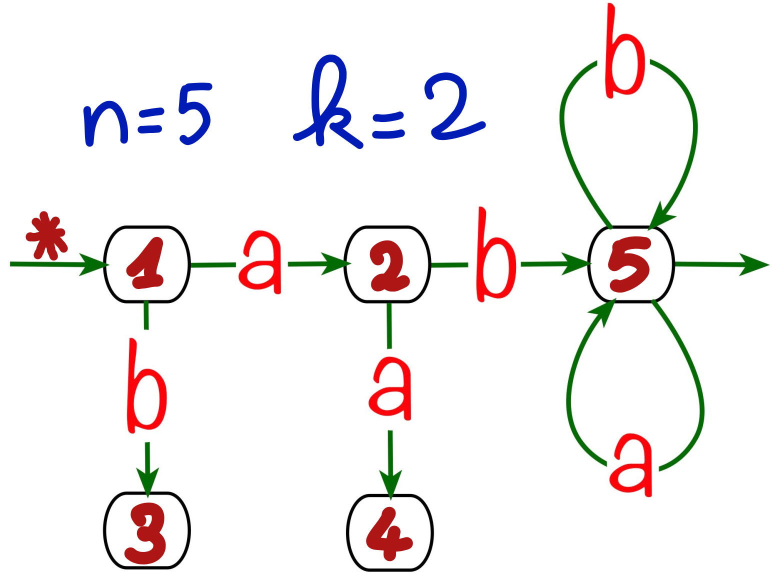
automata in $\mathcal{A}_{k,n}$

$\updownarrow$

map from $[\![1, 1+kn]\!]$ to $[\![1,n]\!]$.

Relation automata - coupon collector

$n=5 \quad k=2$



$*\rightarrow$	1a	1b	2a	2b	3a	3b	4a	4b	5a	5b
1	2	3	4	5	3	3	4	4	5	5

automata in $\mathcal{A}_{k,n}$

$\updownarrow$

map from $[\![1, 1+kn]\!]$ to $[\![1, n]\!]$.

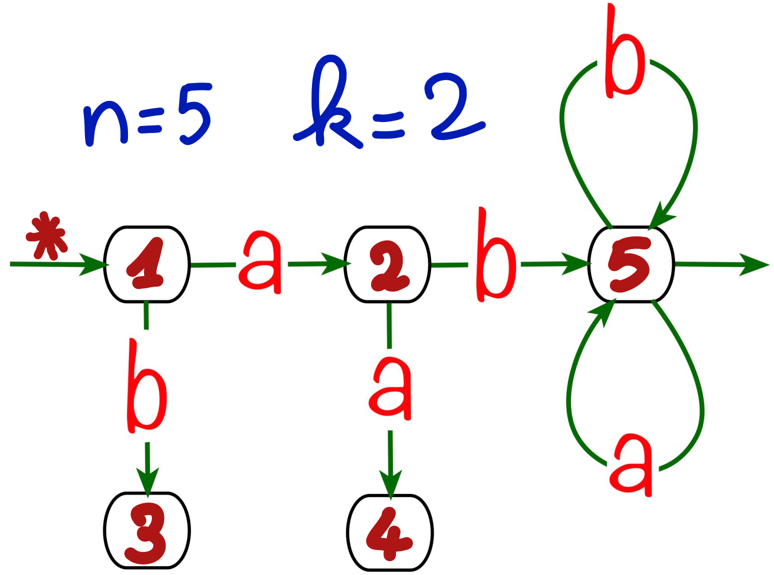
accessible

$\Downarrow$

surjection !

Relation automata - coupon collector

$n=5 \quad k=2$



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

automata

$\overset{?}{=}$

impatient collector

$\lambda = k$

$$\# \mathcal{A}_{k,n} \overset{?}{=} \left\{ {1+kn \atop n} \right\} \times n!$$

$\Uparrow$

accessible

$\Downarrow$

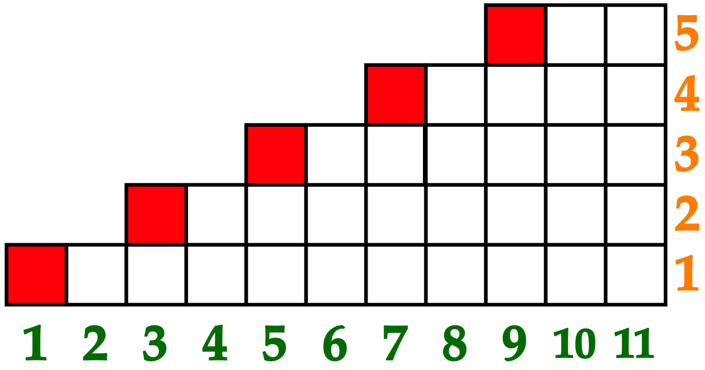
surjection !

automata in $\mathcal{A}_{k,n}$

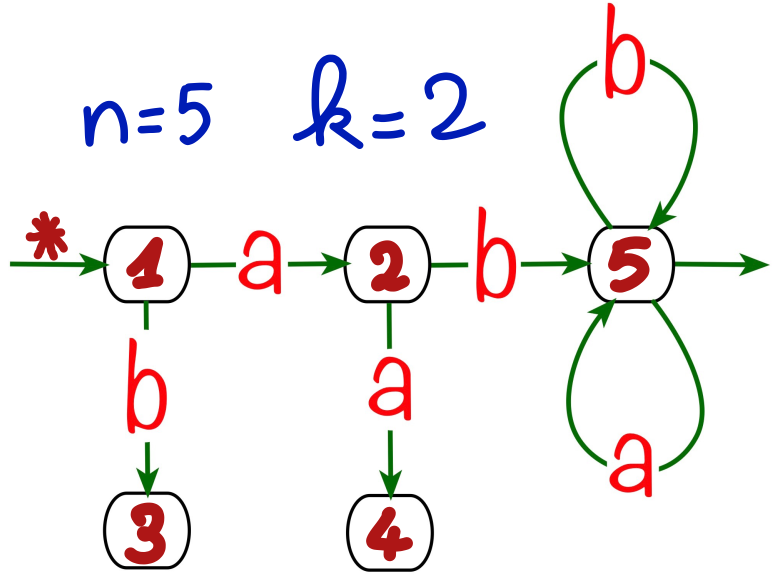
$\updownarrow$

map from $[\![1, 1+kn]\!]$ to $[\![1,n]\!]$.

Relation automata - coupon collector



$n=5$ $k=2$



automata $\overset{?}{=}$ impatient collector

$\lambda = k$

$\# \mathcal{A}_{k,n} \overset{?}{=} \left\{ {1+kn \atop n} \right\} \times n!$

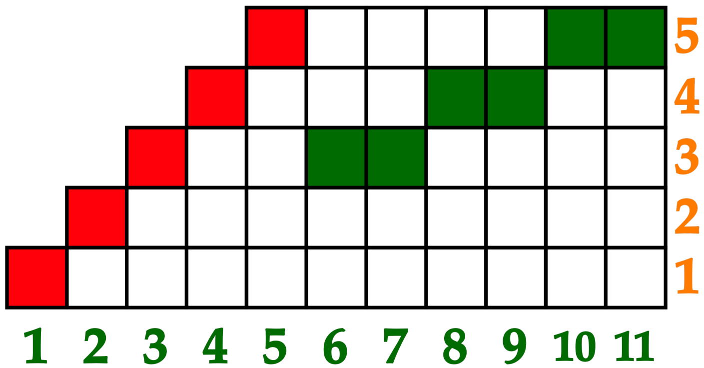
$\Uparrow$

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

automata in $\mathcal{A}_{k,n}$

$\updownarrow$

map from $[\![1, 1+kn]\!]$ to $[\![1,n]\!]$.

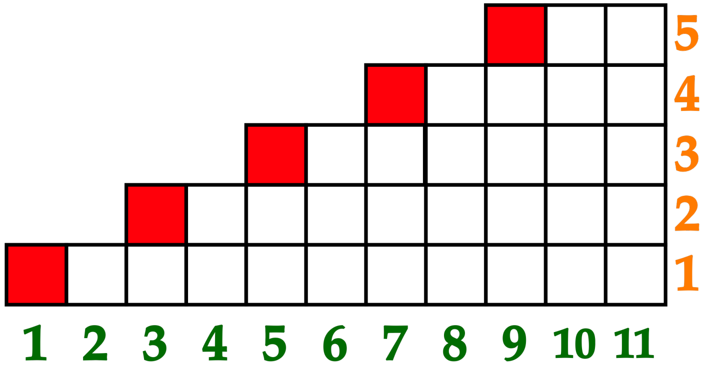


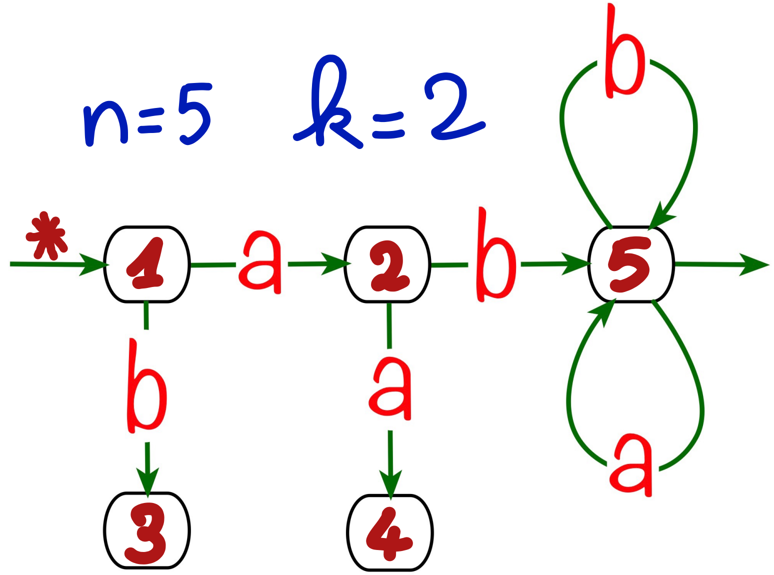
accessible

$\Downarrow$

surjection !

Relation automata - coupon collector





$\#\mathcal{A}_{k,n} = \left\{ {1+kn \atop n} \right\} \times n!$ $\times C(k,n) \in]0,1[$

~~automata~~

~~$\overset{?}{=}$~~

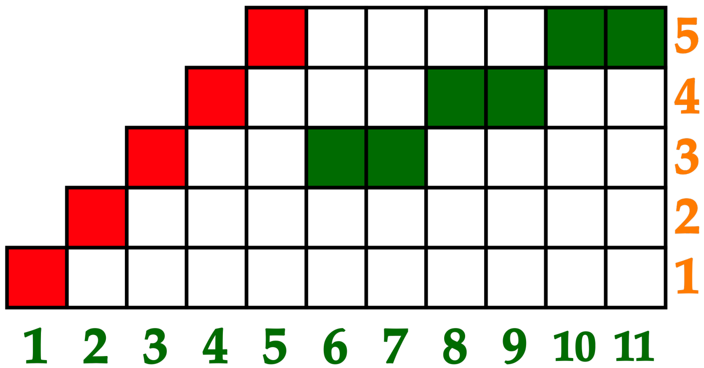
~~impatient collector~~

~~$1+k$~~

~~$\#\mathcal{A}_{k,n} \overset{?}{=} \left\{ {1+kn \atop n} \right\} \times n!$~~

~~$\Uparrow$~~

* →	1a	1b	2a	2b	3a	3b	4a	4b	5a	5b
1	2	3	4	5	3	3	4	4	5	5



accessible

$\Downarrow$ $\quad \cancel{\Uparrow}$

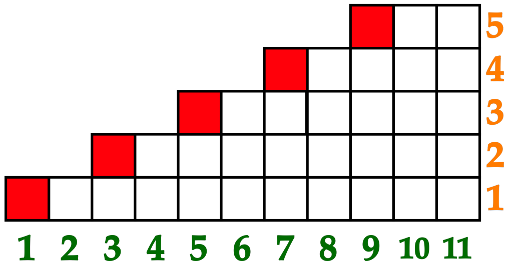
surjection !

automata in $\mathcal{A}_{k,n}$

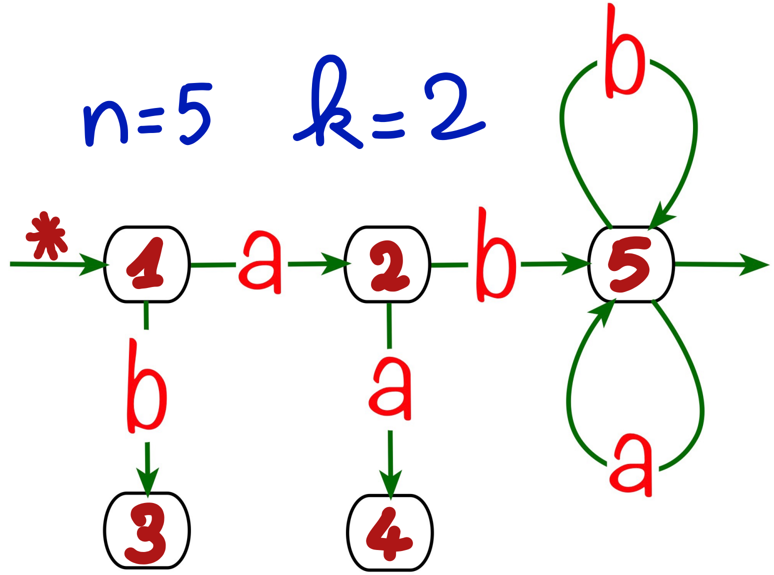
$\Updownarrow$

map from $[\![1, 1+kn]\!]$ to $[\![1,n]\!]$.

Relation automata - coupon collector



$n=5 \quad k=2$



$$\#\mathcal{A}_{k,n} = \left\{ {1+kn \atop n} \right\} \times n! \times \underbrace{C(k,n)}_{\downarrow} \in \,]0,1[$$

some probability in the impatient collector model

$* \rightarrow$	1a	1b	2a	2b	3a	3b	4a	4b	5a	5b
1	2	3	4	5	3	3	4	4	5	5

~~automata $\overset{?}{=}$ impatient collector~~

~~$n \leq k$~~

~~$\#\mathcal{A}_{k,n} \overset{?}{=} \left\{ {1+kn \atop n} \right\} \times n!$~~

~~$\Uparrow$~~

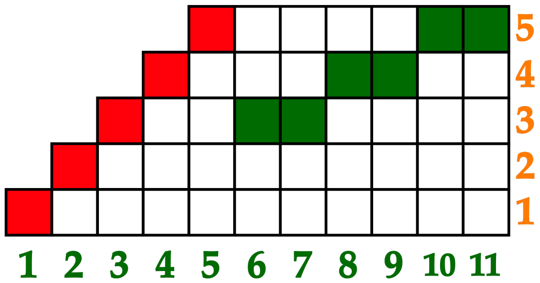
automata in $\mathcal{A}_{k,n}$

$\updownarrow$

accessible

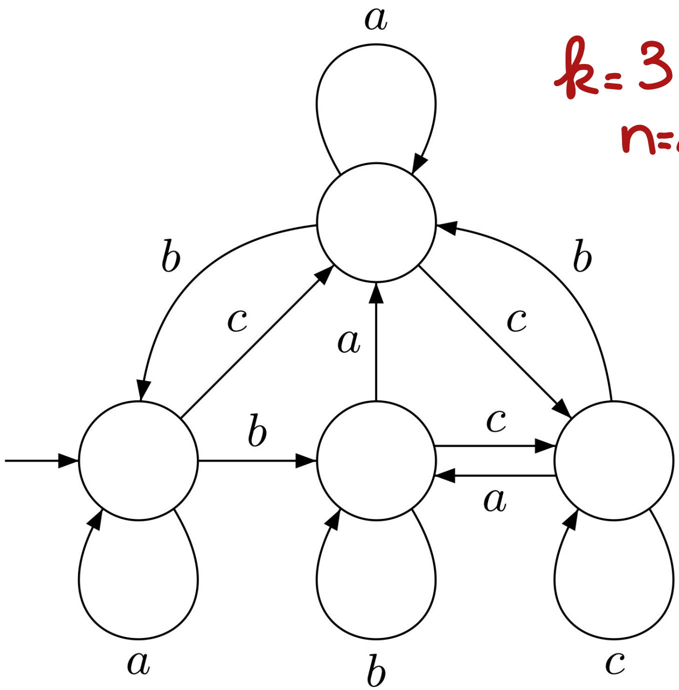
$\Downarrow \quad \cancel{\Uparrow}$

surjection ! ~~map~~ map from $[\![1, 1+kn]\!]$ to $[\![1,n]\!]$.

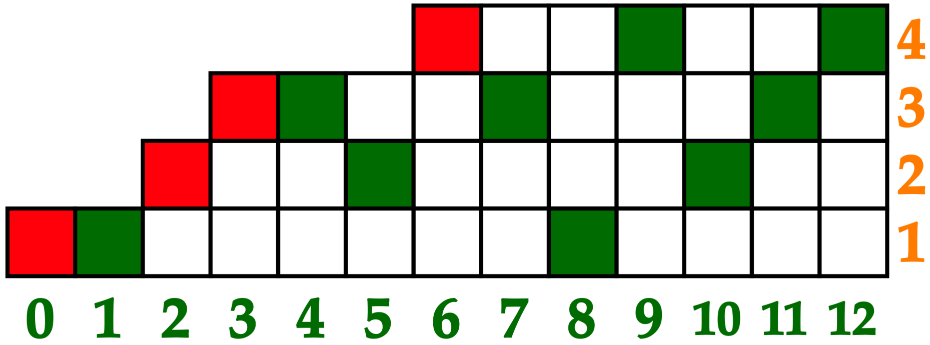


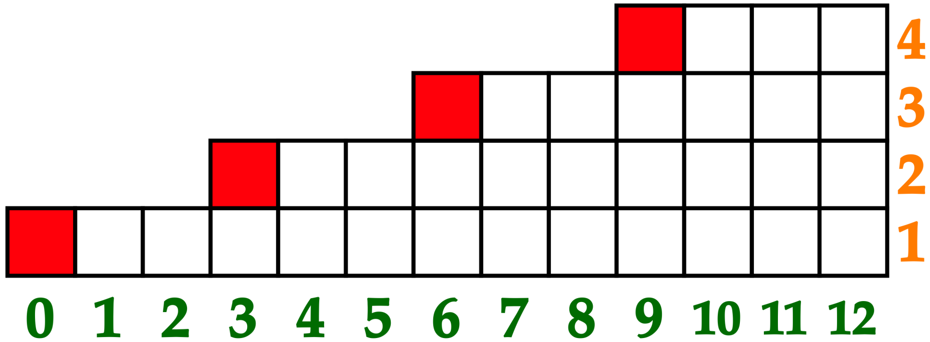
Koršunov 1978

$$\lim_{n} \frac{\#\mathcal{A}_{k,n}}{\left\{ {1+kn \atop n} \right\} \times n!} = C(k) \in \,]0,1[$$

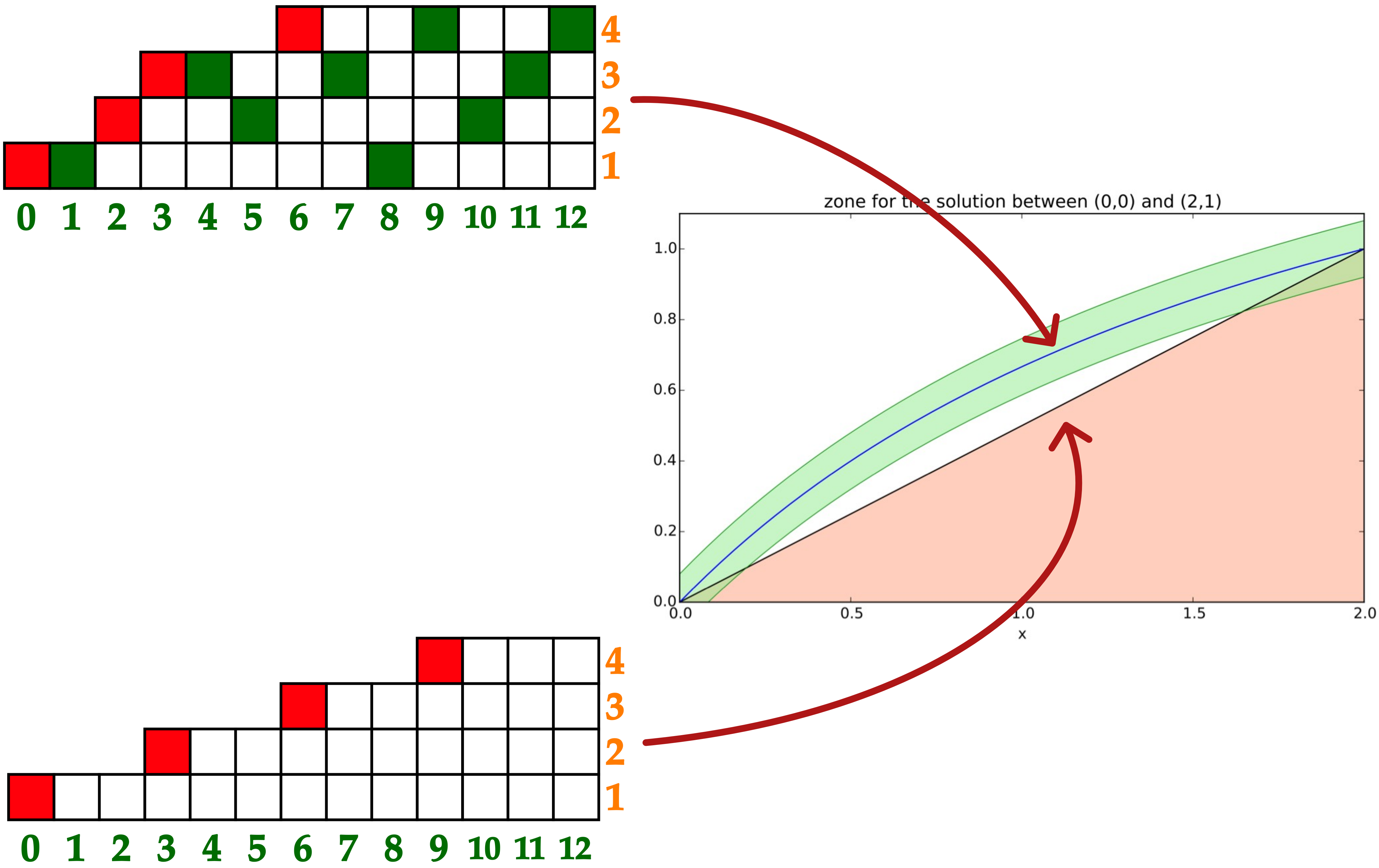


$k=3$
$n=4$

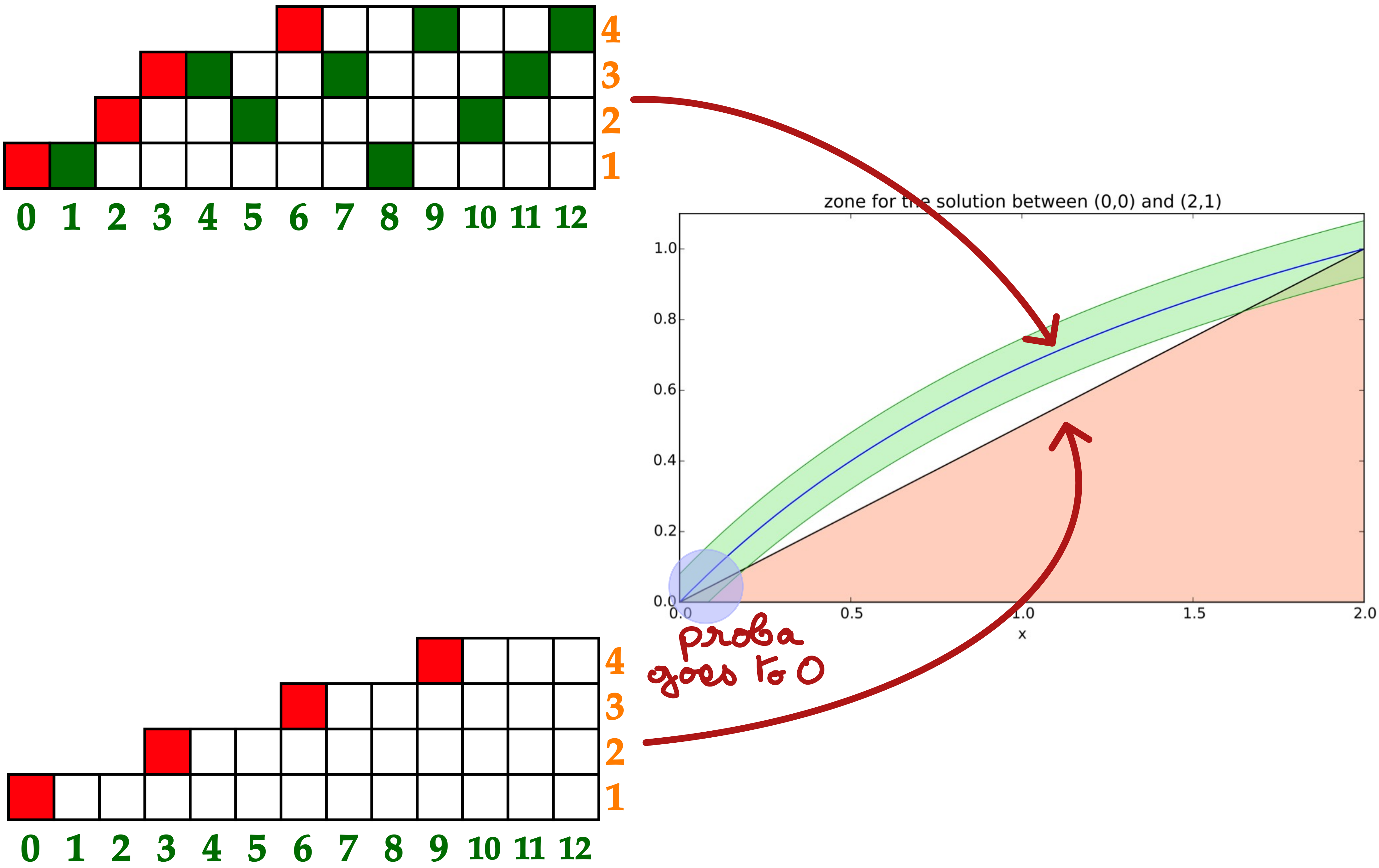




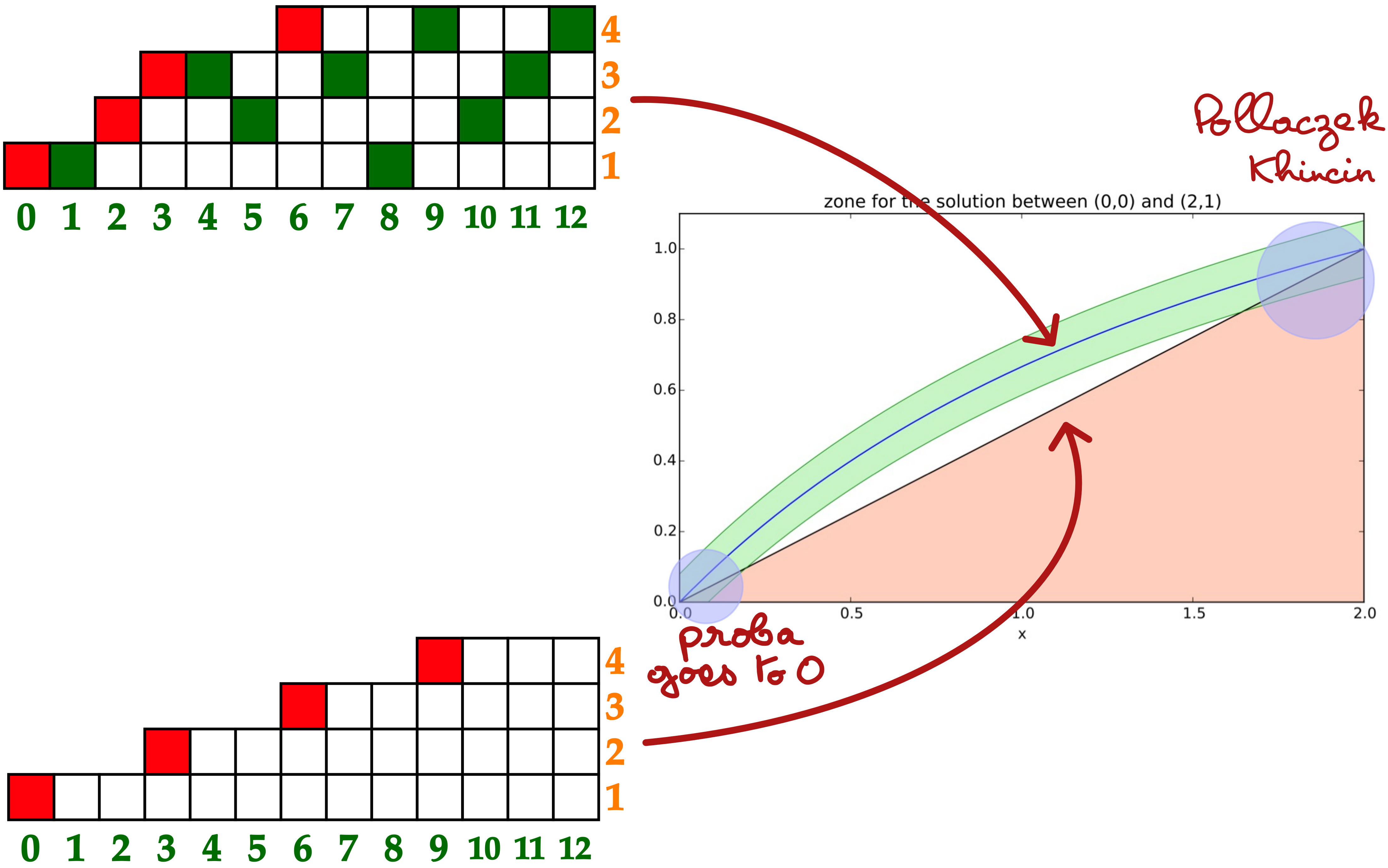
Limit profile & forbidden zone



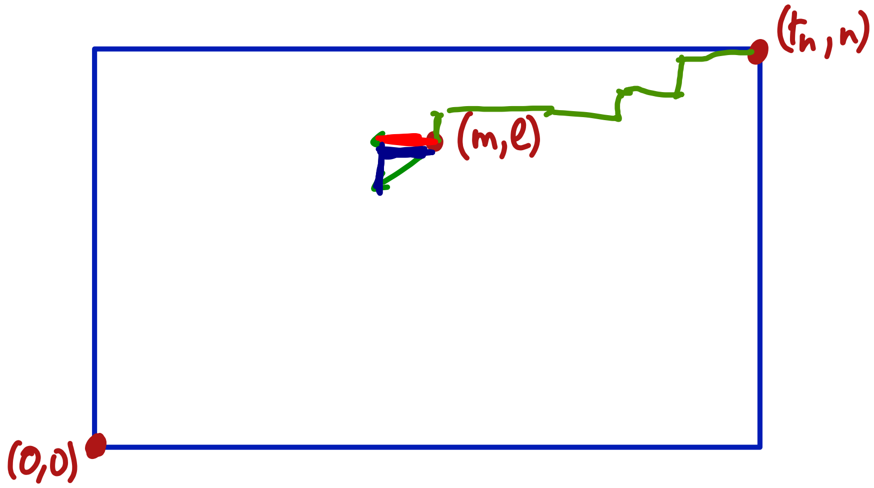
Limit profile & forbidden zone



Limit profile & forbidden zone



3) Loi de la marche aléatoire.



total : ${t_n \brace n}\, n!$

(m,ℓ) : ${m \brace \ell}\, n^{\downarrow \ell} \times C_{m,\ell} = a$

$(m,\ell) \to (m-1,\ell)$: ${m-1 \brace \ell}\, n^{\downarrow \ell}\, \ell \times C_{m,\ell} = b$

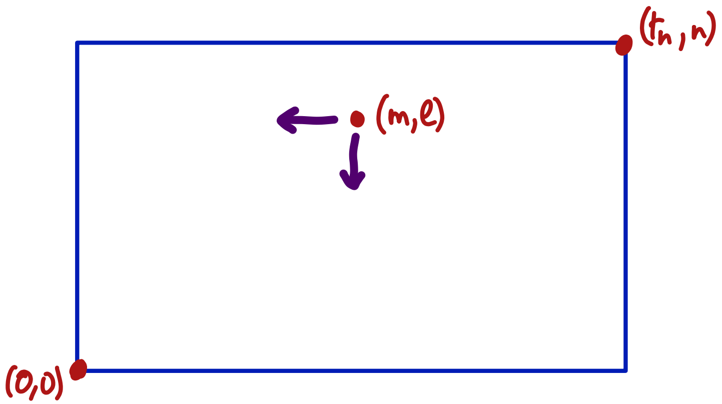
$(m,\ell) \to (m-1,\ell-1)$: ${m-1 \brace \ell-1}\, n^{\downarrow \ell-1}\, (n-\ell+1) \times C_{m,\ell} = c$

$a = b + c \to$ probabilités $\xleftarrow{b/a}$ et $\swarrow^{c/a}$

$$\boxed{b/a = \frac{{m-1 \brace \ell}\,\ell}{{m \brace \ell}}} \qquad \boxed{c/a = \frac{{m-1 \brace \ell-1}}{{m \brace \ell}}}$$

Nota : ${m \brace \ell} = {m-1 \brace \ell}\,\ell + {m-1 \brace \ell-1}$ est un classique du genre $\binom{m+\ell}{\ell} = \binom{m+\ell-1}{\ell-1} + \binom{m+\ell-1}{m-1}$

qui correspond a une autre marche classique dans le rectangle :



$$\frac{c}{a} = \binom{m+\ell-1}{\ell-1} \Big/ \binom{m+\ell}{\ell} = \frac{\ell}{m+\ell}$$

et

$$b/a = \frac{m}{m+\ell}$$

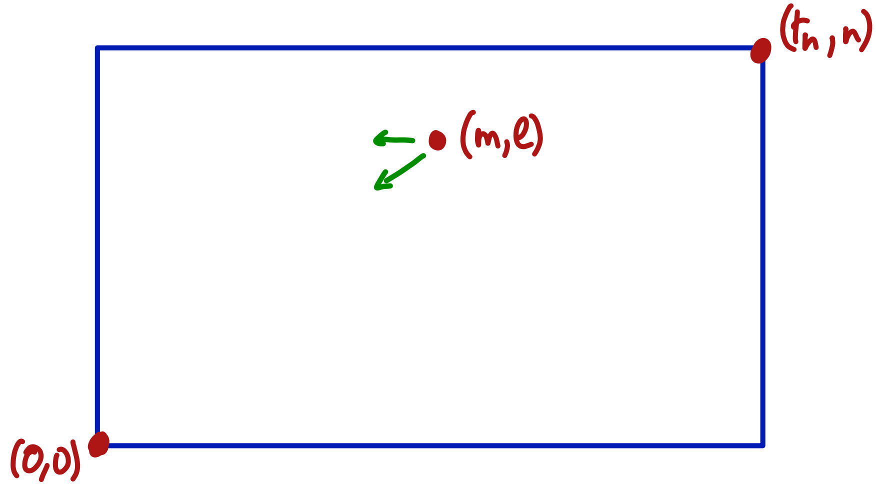
Comportement asymptotique de $g(m,\ell) = \dfrac{\left\{ {m-1 \atop \ell-1} \right\}}{\left\{ {m \atop \ell} \right\}}$??

Good 1961 : $\left\{ {m \atop \ell} \right\} \simeq \psi(m,\ell) = \dfrac{m!}{\ell!} \left(\dfrac{e^\zeta - 1}{\zeta^{1+\lambda}}\right)^\ell C(\lambda) \dfrac{1}{\sqrt{\ell}}$

où $1+\lambda = \dfrac{m}{\ell}$, $\zeta = 1+\lambda + W_0(-(1+\lambda)e^{-1-\lambda})$ i.e. $\zeta = (1-e^{-\zeta})(1+\lambda)$.

et où $C(\lambda) = [2\pi(1+\lambda)(\zeta-\lambda)]^{-1/2}$.

$$\boxed{\text{Th} \quad |g(m,\ell) - e^{-\zeta}| = O\left(\frac{1}{\ell}\right)}$$



En 0_+, $\zeta = 2\lambda + o(\lambda)$, et en $+\infty$, $\zeta = 1+\lambda - O(\lambda e^{-\lambda})$.

$$\boxed{\mathbb{P}_{\alpha,n}\left(\sup_{[\varepsilon,R]} |\xi_n - \xi| \geq C n^{-1/3}\right) \leq n^{1/3 - 1/2 \ln(n)}}$$

où ξ est la solution de $\ln(y'(x)) = -\zeta\left(\dfrac{x-y}{y}\right)$ et $\xi(1) = 1+\alpha$.