

B-urnes

Brigitte CHAUVIN

Université de Versailles St-Quentin, LMV

Travail en cours avec

Danièle GARDY

Nicolas POUYANNE

Dai-Hai TON-THAT

Processus de
branchement
 $(X(t))_t$

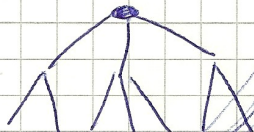
$(X_m)_m$

Urne de Pólya
 $(G_m)_m$

$t \rightarrow +\infty$

W

feuilles d'un
arbre - B



Arbres-B

Un **arbre-B** de paramètre $m \geq 2$ est un processus d'arbres $(\mathcal{T}_n)_n$ planaires où chaque arbre \mathcal{T}_n est un

- **arbre de recherche** contenant n clés ;
- **toutes les feuilles au même niveau**

Dans l'algorithme "optimiste" :

$$m \leq \# \text{ enfants par nœud} \leq 2m - 1$$

$$m - 1 \leq \# \text{ clés par nœud} \leq 2m - 2$$

sauf pour la racine (entre 2 et $2m - 1$ enfants ; entre 1 et $2m - 2$ clés).

- **l'insertion** d'une nouvelle clé se fait par les feuilles

Aléa :

les clés sont une suite de variables aléatoires i.i.d. par exemple $\text{Unif}([0, 1])$
(les $n + 1$ feuilles de \mathcal{T}_n sont équiprobables)

$$m = 3$$

$$3 \leq \# \text{ enfants} \leq 5$$

$$2 \leq \# \text{ clés} \leq 4$$

Insertion des clés :

0.3 ; 0.1 ; 0.4 ; 0.15 ; 0.9 ; ...

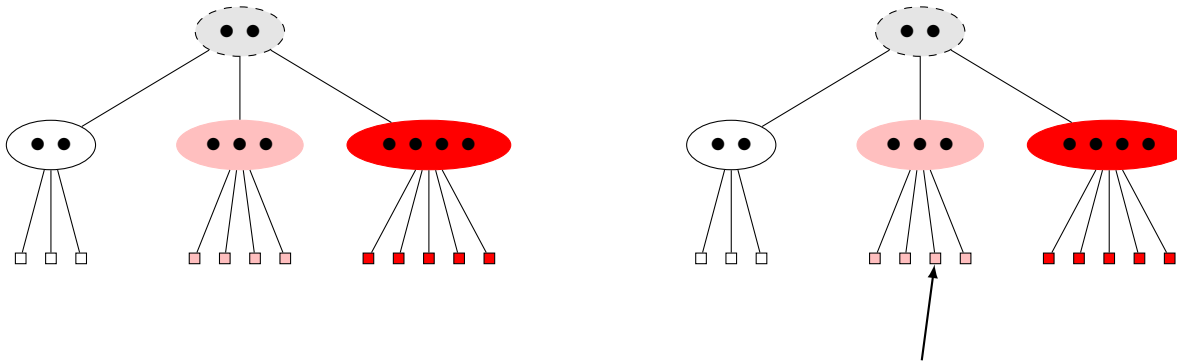
$m = 3$

$\mathcal{T}_{11} \longrightarrow \mathcal{T}_{12}$

$3 \leq \# \text{ enfants} \leq 5$

$2 \leq \# \text{ clés} \leq 4$

3 couleurs: blanc, rose, rouge



$$m = 3$$

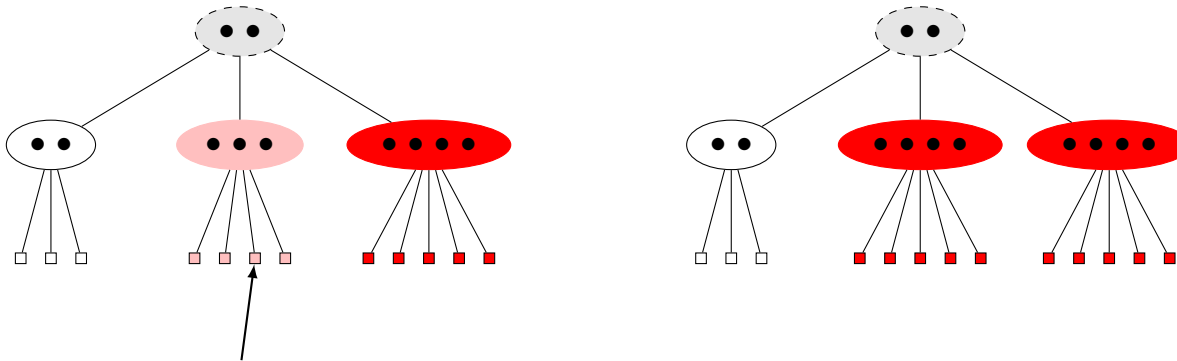
$$\mathcal{T}_{11} \longrightarrow \mathcal{T}_{12}$$

$$3 \leq \# \text{ enfants} \leq 5$$

$$2 \leq \# \text{ clés} \leq 4$$

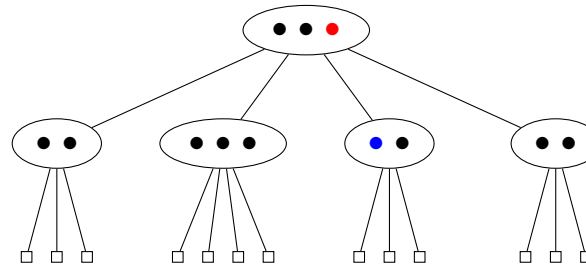
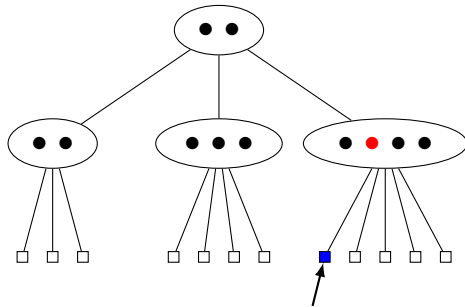
3 couleurs: blanc, rose, rouge

Insertion facile



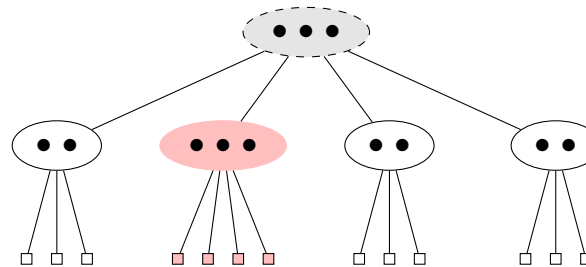
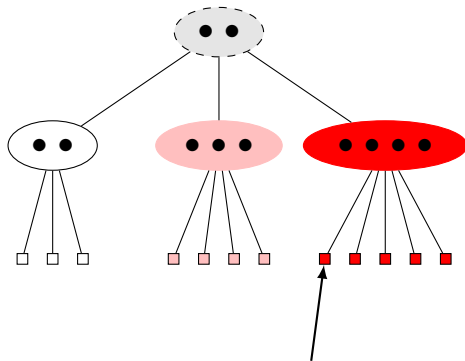
Bilan sur les gaps : -4 roses et $+5$ rouges, pour une insertion sur un gap rose ;
(-3 blancs et $+4$ roses pour une insertion sur un gap blanc)

Insertion moins facile



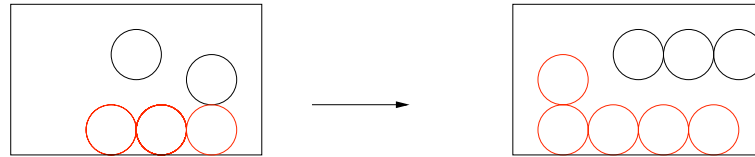
$2 \leq \# \text{ clés} \leq 4$

3 couleurs: blanc, rose, rouge



Bilan sur les gaps : -5 rouges et $2 \times 3 = +6$ blancs pour une insertion sur un gap rouge ;
 (-4 roses et $+5$ rouges, pour une insertion sur un gap rose)
 (-3 blancs et $+4$ roses pour une insertion sur un gap blanc)

Urne de Pólya à 2 couleurs



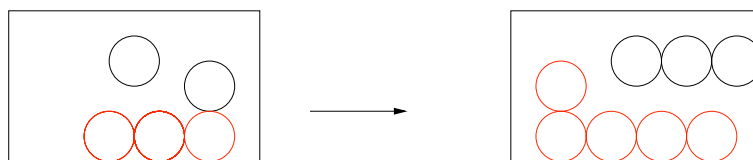
Matrice de remplacement $R = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, a, b, c, d entiers.

Vecteur composition $U(n) = \begin{pmatrix} \# \text{ rouges à l'instant } n \\ \# \text{ noires à l'instant } n \end{pmatrix}$;

$$U(0) = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

d couleurs : $R = (a_{i,j})_{i,j=1\dots d}$

Quel comportement asymptotique de $U(n)$, quand $n \rightarrow +\infty$?



Matrice de remplacement $R = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, a, b, c, d entiers.

$$U(n) = \begin{pmatrix} \# \text{ rouges à l'instant } n \\ \# \text{ noires à l'instant } n \end{pmatrix}; \quad U(0) = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

d couleurs : $R = (a_{i,j})_{i,j=1\dots d}$

Hypothèse de balance

$$\forall i = 1 \dots d, \quad \sum_{j=1}^d a_{i,j} := S$$

$$\# \text{ à l'instant } n = \alpha + \beta + Sn.$$

Les grandes lignes : la transition de phase

Matrice de remplacement $R = (a_{i,j})_{i,j=1\dots d}$

Hypothèse de balance S

Comportement asymptotique du vecteur composition $U(n)$:

(pour $S = 1$)

Soit $\sigma_2 :=$ plus grande partie réelle des valeurs propres $\neq 1$

- si $\sigma_2 < \frac{1}{2}$, c'est une "petite" urne

$$\frac{U(n) - nv_1}{\sqrt{n}} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, \Sigma)$$

- si $\sigma_2 > \frac{1}{2}$, c'est une "grande" urne

$$U(n) = nv_1 + \text{termes en } n^{\sigma_2} + o(n^{\sigma_2}) \quad p.s.$$

Un exemple de grande urne de Pólya : la frange des arbres-B

Etude de la frange des arbres-B

$G_n = \begin{pmatrix} G_n^{(1)} \\ \vdots \\ G_n^{(m)} \end{pmatrix}$ compte le # de feuilles (**gaps**) dans un **arbre-B** de taille n .

$(G_n)_n$ est une urne de Pólya

- à m couleurs ;
- de balance $S = 1$;
- de matrice de remplacement

$$R_m = \begin{pmatrix} -m & m+1 & & & & & \\ & -(m+1) & m+2 & & & & \\ & & \ddots & & & & \\ & & & \ddots & & & \\ & & & & -(2m-2) & 2m-1 & \\ 2m & & & & & -(2m-1) & \end{pmatrix}.$$

- viable (tenable)
- Spectre

$$(\lambda + m) \dots (\lambda + 2m - 1) = \frac{(2m)!}{m!}$$

$\lambda_1 = 1$; $\lambda_2, \bar{\lambda}_2 = \text{vp conjuguées de partie réelle max et } < 1$; $\sigma_2 := \Re(\lambda_2)$.

σ_2 dépend de m

σ_2 croît avec m et vaut

$$0,4953 \dots \text{ si } m = 59$$

et

$$0,5037 \dots \text{ si } m = 60.$$

Théorème (\sim Janson):

(cas Gaussien) si $m \leq 59$,

$$\frac{G_n - nv_1}{\sqrt{n}} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} G$$

(cas non Gaussien) si $m \geq 60$

$$G_n = nv_1 + 2\Re(n^{\lambda_2} W v_2) + o(n^{\sigma_2})$$

où

- v_1, v_2 sont des vecteurs déterministes ;
- W est une limite de martingale, à valeurs dans \mathbb{C} , $\mathbb{E}(W) = \frac{m!}{\Gamma(m + \lambda_2)}$

$$W = \lim_{n \rightarrow \infty} \frac{1}{n^{\lambda_2}} u_2(G_n).$$

- $o(\)$ signifie p.s. et dans tout $L^p, p \geq 1$.

Questions : loi de W ? propriétés ?

Théorème : W est solution de l'équation de point fixe en loi :

$$W \stackrel{\mathcal{L}}{=} V_1^{\lambda_2} W^{(1)} + V_2^{\lambda_2} W^{(2)} \quad \text{où } (V_1, V_2) \stackrel{\mathcal{L}}{=} \text{Dirichlet}(m, m)$$

Preuve : structure d'arbre et loi limite d'une urne de Pólya originelle
(demander à Cécile Mailler)

Question

quelles propriétés de W solution de

$$W \stackrel{\mathcal{L}}{=} V_1^{\lambda_2} W^{(1)} + V_2^{\lambda_2} W^{(2)}$$

où λ_2 est solution de l'équation

$$(\lambda + m) \dots (\lambda + 2m - 1) = \frac{(2m)!}{m!}$$

- Th 1 : existence et unicité (méthode de contraction)
- Th 2 : existence et moments exponentiels (cascade)
- Th 3 : support et densité % mesure de Lebesgue sur \mathbb{C} (méthode de Liu)

Théorème 1 *L'équation en loi*

$$W \stackrel{\mathcal{L}}{=} V_1^{\lambda_2} W^{(1)} + V_2^{\lambda_2} W^{(2)}$$

admet une unique solution W dans l'espace $\mathcal{M}_2(C)$ des mesures de probabilités admettant un second moment et une espérance C .

Éléments de preuve :

→ théorème de point fixe de Banach sur $\mathcal{M}_2(C)$ muni de la distance de Wasserstein.

→ petits calculs faciles :

$$2\mathbb{E}(V^s) = \frac{(2m) \dots (m+1)}{(2m-1+s) \dots (m+s)}$$

• λ_2 solution de $(\lambda + m) \dots (\lambda + 2m - 1) = \frac{(2m)!}{m!} \implies \mathbb{E}(V^{\lambda_2}) = \frac{1}{2}$

• La “smoothing transform” est $\frac{1}{\frac{1}{2} + \Re(\lambda_2)}$ -Lipschitz et

$\Re(\lambda_2) > \frac{1}{2} \implies$ contraction

Davantage d'informations sur W ?

- Sa loi s'exprime-t-elle avec des lois classiques ?
- Taille exacte des moments ?
- Allure de la densité ?

Illustrations (en cours)

- pour $m = 10$,

l'histogramme des quantités $\frac{\langle G_n - nv_1, e_1 \rangle}{\sqrt{n}}$ doit ressembler à une courbe Gaussienne ;
et la trajectoire $k \mapsto \frac{1}{\sqrt{k}} (\langle G_k - kv_1, e_1 \rangle)$, $0 \leq k \leq n$, doit ressembler à un mouvement brownien.

- pour $m = 100$

la trajectoire $k \mapsto \Re \left(\frac{u_2(G_k)}{k^{\lambda_2}} \right)$, $0 \leq k \leq n$ converge vers une limite différente pour chaque expérience. L'histogramme doit donner une idée de la densité de $\Re(W)$ et $\Im(W)$.