

Martingales dans l'étude de quelques arbres aléatoires

Brigitte CHAUVIN

Brigitte.C Chauvin@math.uvsq.fr

Monastir, 18-20 Octobre 2010

Contents

| | | |
|----------|--|----------|
| 1 | Introduction | 1 |
| 2 | Arbres binaires de recherche | 2 |
| 2.1 | Définitions | 2 |
| 2.2 | Profil des abr. Martingale discrète | 3 |
| 2.3 | Plongement en temps continu. Arbre de Yule | 4 |
| 2.4 | connexion martingale discrète/martingale continue | 5 |
| 3 | Urnes de Pólya à deux couleurs | 5 |
| 3.1 | Définition d'un processus de Pólya | 5 |
| 3.2 | Martingale discrète vectorielle | 6 |
| 3.3 | Plongement en temps continu. Processus de branchement à deux types | 7 |
| 3.4 | connexion martingale discrète/martingale continue | 8 |
| 4 | Arbres m-aires de recherche | 9 |
| 4.1 | Définition | 9 |
| 4.2 | Martingales discrètes | 9 |
| 4.3 | Plongement en temps continu. Processus de branchement multitype | 11 |

1 Introduction

L'objectif de ce cours est de détailler un même schéma d'étude, pour des applications à trois modèles très utilisés en informatique fondamentale et en analyse d'algorithmes (algorithmes de tri notamment) : (1) les arbres binaires de recherche, (2) les urnes de Pólya et (3) les arbres m -aires de recherche.

Le schéma d'étude est le suivant :

A - Un processus stochastique markovien en temps discret est lié à une structure d'arbre. Il y a de la récursivité. Des martingales discrètes sont mises en évidence ;

B - Le processus est plongé en temps continu, donnant un processus de branchement. Ce processus de branchement est très simple dans l'exemple 1, il est à plusieurs types dans les exemples 2 et 3. Des martingales à temps continu sont associées à ce processus de branchement. Elles sont connectées aux martingales discrètes. Grâce à la propriété de branchement, l'asymptotique de ce processus de branchement est plus accessible.

2 Arbres binaires de recherche

(en abrégé : abr)

2.1 Définitions

On définit les abr en plusieurs étapes :

→ on commence par définir un *arbre binaire* récursivement, par (\emptyset désigne l'arbre vide et \bullet la racine d'un arbre non vide) :

$$\mathcal{B} = \emptyset + (\bullet, \mathcal{B}, \mathcal{B}). \quad (1)$$

Remarquons bien qu'avec cette définition, un arbre binaire n'est pas un arbre... Dans un arbre binaire, on distingue fils gauche et fils droit : les arbres binaires (racine, fils gauche, \emptyset) et (racine, \emptyset , fils droit) sont distincts alors que ce sont le même arbre réduit à la racine et un fils.

La taille de l'arbre est son nombre de sommets, ou de noeuds.

→ dans un *arbre binaire complet*, tout sommet a 0 ou 2 fils. Tout arbre binaire à n noeuds peut être mis en bijection avec un arbre binaire complet. Pour cela, on ajoute $n + 1$ noeuds (dits externes) à l'arbre binaire ; on obtient ainsi un arbre qui a n noeuds internes, $n + 1$ feuilles et $2n + 1$ sommets au total.

→ Autre définition des arbres binaires complets, avec leur étiquetage canonique : soit

$$\mathcal{U} = \{\varepsilon\} \cup \bigcup_{n \geq 1} \{0, 1\}^n$$

qui est l'ensemble des mots sur l'alphabet à deux lettres $\{0, 1\}$, où ε désigne le mot vide. Le fils gauche du mot u est $u0$ et le fils droit est $u1$. Un **arbre binaire complet** T est un sous-ensemble fini de \mathcal{U} tel que

$$\begin{cases} \varepsilon \in T \\ \text{si } uv \in T \text{ alors } u \in T, \\ u1 \in T \Leftrightarrow u0 \in T. \end{cases}$$

Les éléments de T sont les noeuds, la racine est ε , le nombre de lettres d'un noeud u est noté $|u|$, c'est la profondeur de u dans l'arbre ($|\varepsilon| = 0$). On note **BinTree** l'espace des arbres binaires complets.

Notons qu'un arbre binaire complet $T \in \mathbf{BinTree}$ peut être décrit par l'ensemble ∂T de ses feuilles, qui sont par définition les nœuds qui sont dans T mais qui n'ont pas de descendant. Les autres nœuds de $T \setminus \partial T$ sont les nœuds internes.

→ un *arbre binaire de recherche* est un arbre binaire complet dans lequel les nœuds internes contiennent des données (appelées aussi clés) qui sont des nombres. Soit $x_1, x_2, \dots, x_n, \dots$ une suite de nombres distincts. L'abr associé est défini récursivement par

- l'étiquette de la racine est x_1

- le sous-arbre gauche est l'abr associé à $\{x_2, \dots, x_n, \dots\} \cap]-\infty, x_1[$

- le sous-arbre droit est l'abr associé à $\{x_2, \dots, x_n, \dots\} \cap]x_1, +\infty[$.

→ on obtient un *arbre binaire de recherche aléatoire* en rendant aléatoires les données. Soient $(x_i)_{i \geq 1}$ une suite de variables aléatoires indépendantes et de même loi, par exemple de loi uniforme sur l'intervalle $[0, 1]$. On appelle \mathcal{T}_n l'abr associé aux n premières données x_1, x_2, \dots, x_n .

Par conséquent, \mathcal{T}_n est une variable aléatoire à valeurs dans $\mathbf{BinTree}$ et $(\mathcal{T}_n)_n$ est une chaîne de Markov à valeurs dans $\mathbf{BinTree}$. De plus, lorsqu'on insère une donnée x_{n+1} dans \mathcal{T}_n , on a pour toute feuille $u \in \partial \mathcal{T}_n$

$$P(x_{n+1} \in u \mid \mathcal{T}_n) = \frac{1}{n+1}. \quad (2)$$

→ Une vision plus probabiliste de la suite $(\mathcal{T}_n)_n$ est la suivante : $(\mathcal{T}_n)_n$ est un processus à temps discret, à valeurs dans $\mathbf{BinTree}$, qui évolue avec la règle suivante : \mathcal{T}_0 est réduit à une feuille ; pour $n \geq 0$, \mathcal{T}_{n+1} est obtenu à partir de \mathcal{T}_n en insérant uniformément sur une des $(n+1)$ feuilles de \mathcal{T}_n .

2.2 Profil des abr. Martingale discrète

De très nombreux résultats ont été obtenus sur les abr. Voir [12] pour les méthodes analytiques, [10] pour des méthodes plus probabilistes et [15] pour un livre sur le sujet.

Le *profil* donne une idée de la forme de l'abr. Il est donné par la suite

$$U_k(n) := \text{le nombre de feuilles au niveau } k \text{ dans l'arbre } \mathcal{T}_n.$$

L'objectif est de trouver l'asymptotique de ces nombres lorsque $n \rightarrow +\infty$. Pour cela, on introduit le *polynôme de niveau*, défini pour tout paramètre $z \in \mathbb{C}$ par

$$W_n(z) := \sum_{k=0}^{+\infty} U_k(n) z^k = \sum_{u \in \partial \mathcal{T}_n} z^{|u|}.$$

Il s'agit bien d'un polynôme, car (si h_n est la hauteur de l'arbre), pour $k > 1 + h_n$, $U_k(n) = 0$. Bien entendu, c'est une variable aléatoire, puisque les $U_k(n)$ sont aléatoires.

Proposition 1

$$\mathbb{E}(W_{n+1}(z) \mid \mathcal{T}_n) = \frac{n+2z}{n+1} W_n(z).$$

Pour tout nombre complexe $z \in \mathbb{C}$ tel que $z \neq -k, k \in \mathbf{N}$, notons

$$\Gamma_n(z) := \prod_{j=0}^{n-1} \left(1 + \frac{z}{j+1}\right).$$

Théorème 1 (*Jabbour-Hattab, 2001, [13]*)

$$M_n(z) := \frac{W_n(z)}{\mathbb{E}(W_n(z))} = \frac{W_n(z)}{\Gamma_n(2z-1)}$$

est une \mathcal{F}_n -martingale qui s'écrit aussi

$$M_n(z) := \frac{1}{\Gamma_n(2z-1)} \sum_{u \in \partial \mathcal{T}_n} z^{|u|}.$$

Cette martingale converge p.s. pour tout z réel positif.

Elle converge dans L^1 vers une limite notée $M_\infty(z)$ pour tout $z \in]z_-, z_+[$ et elle converge p.s. vers 0 pour tout $z \notin]z_-, z_+[$, où z_- et z_+ sont les solutions de l'équation $z \log(z/2) - z + 2 = 0$.

2.3 Plongement en temps continu. Arbre de Yule

L'idée est due à Pittel [16]. On considère un processus de branchement à temps continu, décrit de la manière suivante : il y a un ancêtre à l'instant $t = 0$, qui vit un temps exponentiel de paramètre 1. Quand il meurt, il donne naissance à deux fils qui vivent chacun un temps exponentiel de paramètre 1, indépendamment l'un de l'autre, etc... Le processus à valeurs arbres ainsi obtenu s'appelle processus d'arbres de Yule, on le note $(T_t)_t$.

On appelle N_t le nombre de feuilles de T_t (à l'instant t) et on note $0 < \tau_1 < \dots < \tau_n < \dots$ les instants de saut. Il est facile de voir que $\tau_n - \tau_{n-1}$ est de loi $\mathcal{Exp}(n)$ et que (principe du plongement)

$$(T_{\tau_n})_n \stackrel{\mathcal{L}}{=} (\mathcal{T}_n)_n.$$

Sur cet arbre de Yule, on construit une marche aléatoire branchante en posant pour tout u feuille de T_t :

$$X_u(t) := -|u| \log 2$$

de sorte que les déplacements de la marche comptent les générations dans l'arbre. Biggins ([3]) et Bertoin et Rouault ([2]) ont démontré

Théorème 2 Pour tout nombre complexe $z \in \mathbb{C}$,

$$M(t, z) := \sum_{u \in \partial T_t} z^{|u|} e^{t(1-2z)}$$

est une \mathcal{F}_t -martingale.

Cette martingale converge p.s. pour tout z réel positif.

Elle converge dans L^1 vers une limite notée $M(\infty, z)$ pour tout $z \in]z_-, z_+[$ et elle converge p.s. vers 0 pour tout $z \notin]z_-, z_+[$, où z_- et z_+ sont les solutions de l'équation $z \log(z/2) - z + 2 = 0$.

2.4 connexion martingale discrète/martingale continue

Proposition 2 Pour tout $z \in]z_-, z_+[$, on a la connexion

$$M(\infty, z) = \frac{\xi^{2z-1}}{\Gamma(2z)} M_\infty(z)$$

où ξ et $M_\infty(z)$ sont indépendantes et ξ suit une loi $\text{Exp}(1)$.

PREUVE. Utiliser le plongement. Cette connexion est un des ingrédients principaux pour obtenir le théorème suivant sur le profil (voir [4]).

Théorème 3 (CKMR 2005) Pour tout compact K contenu dans l'intervalle $]z_-, z_+[$,

$$\frac{U_k(n)}{\mathbb{E}(U_k(n))} - M_\infty\left(\frac{k}{2 \log n}\right) \xrightarrow[n \rightarrow \infty]{} 0 \quad p.s.$$

uniformément sur $\frac{k}{2 \log n} \in K$.

3 Urnes de Pólya à deux couleurs

3.1 Définition d'un processus de Pólya

On définit un processus à temps discret de la manière suivante : à chaque instant, l'urne contient des boules noires et des boules rouges. On tire au hasard (uniformément) une boule, on note sa couleur, on la remet dans l'urne et on ajoute

→ a boules rouges et b boules noires si on a tiré une rouge

→ c boules rouges et d boules noires si on a tiré une noire.

On note $R = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ la matrice de remplacement et on s'intéresse à

$$U^{DT}(n) = \begin{pmatrix} \# \text{ boules rouges au temps } n \\ \# \text{ boules noires au temps } n \end{pmatrix}.$$

qui s'appelle vecteur composition de l'urne. On pose $U^{DT}(0) = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$. On fait désormais l'hypothèse que l'urne est équilibrée, ce qui signifie que le nombre de boules ajoutées à chaque étape est une constante $S = a + b = c + d$, appelée la balance. Par conséquent le nombre total de boules à chaque instant n est déterministe, égal à $\alpha + \beta + nS = u + nS$ (on appelle u le nombre de boules au départ).

Le spectre de la matrice R détermine le comportement asymptotique de l'urne. Il y a deux valeurs propres : S et $m = a - c = d - b$ et on supposera que $S > m$ (on ne considère pas le cas historique de Pólya où $S = m$). Le comportement est différent selon que $\sigma := \frac{m}{S} \leq \frac{1}{2}$ ou que $\sigma > \frac{1}{2}$. Dans le premier cas il y a un théorème central limite et $U^{DT}(n)$ est asymptotiquement gaussien (voir Janson [14]). Dans le second cas, les martingales discrètes introduites ci-après sont plus utiles. Dans la suite, on supposera que $\sigma > \frac{1}{2}$.

3.2 Martingale discrète vectorielle

Le vecteur composition de l'urne évolue avec la dynamique suivante :

si on pose $U^{DT}(n) = \begin{pmatrix} x \\ y \end{pmatrix}$

$$\begin{cases} U^{DT}(n+1) = U^{DT}(n) + \begin{pmatrix} a \\ b \end{pmatrix} & \text{avec proba } \frac{x}{x+y} \\ U^{DT}(n+1) = U^{DT}(n) + \begin{pmatrix} c \\ d \end{pmatrix} & \text{avec proba } \frac{y}{x+y}, \end{cases} \quad (3)$$

et comme $x + y = u + nS$,

$$\mathbb{E}^{\mathcal{F}_n}(U^{DT}(n+1)) = \left(I + \frac{A}{u + nS} \right) U^{DT}(n) \quad (4)$$

où I est la matrice identité de dimension 2 et où

$$A := {}^tR = \begin{pmatrix} a & c \\ b & d \end{pmatrix}.$$

La propriété de martingale est alors immédiate.

Proposition 3 *Pour tout réel positif x et tout entier n , notons $\gamma_{x,n}$ le polynôme*

$$\gamma_{x,n}(t) := \prod_{k=0}^{n-1} \left(1 + \frac{t}{x+k} \right).$$

Alors $\gamma_{\frac{u}{S},n} \left(\frac{A}{S} \right)^{-1} U^{DT}(n)$ est une \mathcal{F}_n -martingale (vectorielle).

En projetant successivement sur les deux droites propres, le long de v_1 et le long de v_2 , on obtient en particulier que

$$M^{DT}(n) := \gamma_{\frac{u}{S}, n} \left(\frac{m}{S}\right)^{-1} u_2(U^{DT}(n))$$

est une \mathcal{F}_n -martingale (u_2 dénote la forme linéaire propre ou vecteur propre à gauche associée à la valeur propre m) qui converge lorsque n tend vers l'infini et on pose

$$W^{DT} := \lim_{n \rightarrow +\infty} \frac{1}{n^\sigma} u_2(U^{DT}(n)).$$

En bornant les moments d'ordre p de cette martingale, on obtient (voir Pouyanne [17]) le théorème suivant

Théorème 4 *Supposons que $\sigma \in]1/2, 1[$. Alors quand n tend vers l'infini,*

$$U^{DT}(n) = nv_1 + n^\sigma W^{DT} v_2 + o(n^\sigma),$$

où v_1 et v_2 sont deux vecteurs propres associés aux valeurs propres S et m ; W^{DT} est défini plus haut ; $o(\cdot)$ signifie une convergence p.s. et dans tous les $L^p, p \geq 1$; les moments de W^{DT} peuvent se calculer récursivement.

3.3 Plongement en temps continu. Processus de branchement à deux types

Mettons des horloges rouges et noires de loi $\mathcal{Exp}(1)$ sur les boules respectivement rouges et noires. Quand une horloge rouge sonne, la boule rouge meurt et donne naissance à $(a + 1)$ boules rouges et b boules noires. Quand une horloge noire sonne, la boule noire meurt et donne naissance à c boules rouges et $(d + 1)$ boules noires. Le vecteur composition de l'urne est maintenant

$$U^{CT}(t) = \begin{pmatrix} \# \text{ boules rouges au temps } t \\ \# \text{ boules noires au temps } t \end{pmatrix}.$$

et l'on part de la même situation à l'instant 0 : $U^{CT}(0) = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$. C'est un processus de branchement à deux types. Appelons $0 = \tau_0 < \tau_1 < \dots < \tau_n < \dots$ les instants de saut. Il est facile de voir que $\tau_n - \tau_{n-1}$ est de loi $\mathcal{Exp}(u + nS)$ et que (principe du plongement)

$$(U^{CT}(\tau_n))_n \stackrel{\mathcal{L}}{=} (U^{DT}(n))_n.$$

Le bénéfice de cette opération est de récupérer en temps continu un processus de branchement, pour lequel il y a indépendance entre les sous-arbres, ce qui n'était pas le cas en temps discret. Pour un processus de branchement multitype, il est classique (voir Athreya-Ney [1]) de voir que

Proposition 4

$$(e^{-tA}U^{CT}(t))_{t \geq 0}$$

est une \mathcal{F}_t -martingale vectorielle. En projetant sur les droites propres, on a des martingales à valeurs réelles convergentes

$$\xi = \lim_t e^{-St} u_1 (U^{CT}(t)) \quad \text{et} \quad W^{CT} := \lim_t e^{-mt} u_2 (U^{CT}(t))$$

En travaillant sur les moments d'ordre p de ces martingales, on obtient l'asymptotique en temps continu du processus ([7]).

Théorème 5 *Quand t tend vers l'infini,*

$$U^{CT}(t) = e^{St} \xi v_1 (1 + o(1)) + e^{mt} W^{CT} v_2 (1 + o(1)), \quad (5)$$

où ξ et W^{CT} sont des variables aléatoires réelles, le petit o signifie une convergence p.s. et dans tous les L^p , $p \geq 1$. De plus, ξ est de loi Gamma(u/S), où $u = \alpha + \beta$ est le nombre total de boules à l'instant 0.

3.4 connexion martingale discrète/martingale continue

Proposition 5 *On a la connexion*

$$W^{CT} = \xi^\sigma W^{DT} \quad p.s.$$

où ξ et W^{DT} sont indépendantes et ξ suit une loi Gamma(u/S).

PREUVE. Utiliser le plongement.

Ensuite, comme W^{CT} est la limite d'un processus de branchement projeté et renormalisé (c'est la Proposition 7), on exploite la propriété de branchement au premier instant de saut τ_1 (qui devient τ pour alléger): appelons

$$\begin{cases} X_t := U^{CT}(t) \text{ partant de } \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ Y_t := U^{CT}(t) \text{ partant de } \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \end{cases} \quad (6)$$

de sorte que

$$\forall t > \tau, \begin{cases} X_t \stackrel{\mathcal{L}}{=} [a+1]X_{t-\tau} + [b]Y_{t-\tau} \\ Y_t \stackrel{\mathcal{L}}{=} [c]X_{t-\tau} + [d+1]Y_{t-\tau}, \end{cases} \quad (7)$$

où la notation $[n]X$ signifie somme de n copies de la variable aléatoire X . Après projection et renormalisation, on obtient pour les limites notées X et Y

$$\forall t > \tau, \begin{cases} X \stackrel{\mathcal{L}}{=} e^{-m\tau} ([a+1]X + [b]Y) \\ Y \stackrel{\mathcal{L}}{=} e^{-m\tau} ([c]X + [d+1]Y). \end{cases} \quad (8)$$

Il s'agit d'un système d'équations de point fixe, en loi, que l'on peut traduire sur les transformées de Fourier de ces lois, et que l'on peut résoudre. On obtient ainsi des renseignements sur les lois de X et de Y , cad sur W^{CT} . Pour les détails, voir [7].

La question se pose de voir si le même raisonnement permet de traiter le cas des urnes à plus de deux couleurs. Ce raisonnement produit également un système d'équations de point fixe, mais qui, en général n'est pas résoluble. Néanmoins, dans un cas particulier, correspondant à un problème d'analyse d'algorithme classique, celui des arbres m -aires de recherche, nous allons obtenir des résultats sur les lois limites.

4 Arbres m -aires de recherche

4.1 Définition

Les arbres m -aires de recherche, $m \geq 3$, sont une généralisation des arbres binaires de recherche : on place au maximum $m - 1$ données dans un noeud de l'arbre et chaque noeud a m descendants. L'alea est produit par une suite de données x_1, \dots, x_n, \dots indépendantes et de même loi, uniforme sur l'intervalle $[0, 1]$.

Exemple : construire pour $m = 4$ l'arbre contenant dans cet ordre les données suivantes : 0.3, 0.1, 0.4, 0.15, 0.9, 0.2, 0.6, 0.5, 0.35

On s'intéresse à

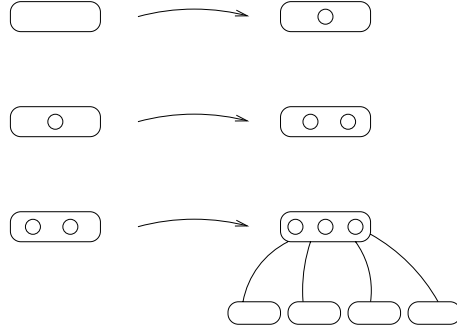
$X_n^{(i)} :=$ le nombre de feuilles de type i , cad contenant $i - 1$ clés à l'instant n

et au comportement asymptotique du vecteur

$$X_n = \begin{pmatrix} X_n^{(1)} \\ \vdots \\ X_n^{(m-1)} \end{pmatrix}.$$

4.2 Martingales discrètes

La dynamique de croissance de l'arbre est la suivante :



de sorte que X_{n+1} s'écrit en fonction de X_n et si l'on pose

$$A = \begin{pmatrix} -1 & & & & & & m(m-1) \\ 1 & -2 & & & & & \\ & 2 & -3 & & & & \\ & & \ddots & \ddots & & & \\ & & & \ddots & -(m-2) & & \\ & & & & m-2 & -(m-1) & \end{pmatrix}$$

on a

$$\mathbb{E}^{\mathcal{F}_n}(X_{n+1}) = \left(I + \frac{A}{n+1} \right) X_n$$

et donc en posant

$$\Gamma_n(z) := \prod_{j=0}^{n-1} \left(1 + \frac{z}{j+1} \right),$$

on a immédiatement

Proposition 6 (CP, 2004, [6])

$(\Gamma_n(A)^{-1}X_n)_n$ est une \mathcal{F}_n -martingale vectorielle.

Le spectre de la matrice A détermine le comportement asymptotique de X_n . Les valeurs propres de A sont solutions de l'équation caractéristique

$$\prod_{k=1}^{m-1} (\lambda + k) = m!$$

Toutes les valeurs propres sont simples, conjuguées, 1 est valeur propre. On appelle λ_2 la valeur propre de partie imaginaire positive et de partie réelle σ_2 qui est la plus grande des parties réelles des valeurs propres différentes de 1. Le comportement asymptotique de X_n est différent selon que $\sigma_2 \leq \frac{1}{2}$ ou que $\sigma_2 > \frac{1}{2}$. Les résultats énoncés dans le théorème suivant se trouvent dans [15, 14, 6, 17].

Théorème 6

- Quand $\sigma_2 < \frac{1}{2}$, $m \leq 26$ et

$$\frac{X_n - nv_1}{\sqrt{n}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \Sigma^2)$$

où v_1 est un vecteur propre pour la valeur propre 1, et où Σ^2 se calcule.

- Quand $1 > \sigma_2 > \frac{1}{2}$, $m \geq 27$ et

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

où v_1, v_2 sont des vecteurs propres déterministes, non réels ; W^{DT} est une variable aléatoire à valeurs dans \mathbb{C} , limite de martingale ; $o(\cdot)$ signifie une convergence p.s. et dans tous les L^p , $p \geq 1$; les moments de W^{DT} peuvent se calculer récursivement.

4.3 Plongement en temps continu. Processus de branchement multitype

On définit un processus de branchement $(X(t))_t$ à $m - 1$ types, pour $m \geq 3$,

$$X(t) = \begin{pmatrix} X(t)^{(1)} \\ \vdots \\ X(t)^{(m-1)} \end{pmatrix}$$

avec $X(t)^{(j)} = \#$ particules de type j en vie au temps t . Chaque particule de type j porte une horloge de loi $\mathcal{Exp}(j)$. Quand elle sonne, la particule de type j meurt et donne naissance à

→ une particule de type $j + 1$, quand $j \leq m - 2$

→ m particules de type 1 quand $j = m - 1$.

Appelons $0 = \tau_0 < \tau_1 < \dots < \tau_n < \dots$ les instants de saut. Il est facile de voir que $\tau_n - \tau_{n-1}$ est de loi $\mathcal{Exp}(u + n - 1)$ où $u = \sum_{k=1}^{m-1} kX(0)^{(k)}$ et que (principe du plongement)

$$(X(\tau_n))_n \stackrel{\mathcal{L}}{=} (X_n)_n.$$

Pour ce processus de branchement multitype, il est classique (voir Athreya-Ney [1] ou Janson [14]) de voir que

Proposition 7

$$(e^{-tA} U^{CT}(t))_{t \geq 0}$$

est une \mathcal{F}_t -martingale vectorielle.

En projetant sur les droites propres, on obtient

Théorème 7 (*Janson [14], CLP [5]*)

$$X(t) = e^t(1 + o(1))\xi v_1 + \Re(e^{\lambda_2 t} W^{CT} v_2) + o(e^{\sigma_2 t})$$

où ξ est une variable aléatoire réelle de loi $\text{Gamma}(u)$;

$$W^{CT} := \lim_t e^{-\lambda_2 t} u_2(X(t))$$

est une variable aléatoire à valeurs complexes, qui admet des moments de tous ordres $p \geq 1$; $o(\cdot)$ signifie une convergence p.s. et dans tous les L^p , $p \geq 1$. De plus, on a la connexion

$$W^{CT} = \xi^{\lambda_2} W^{DT} \quad p.s.$$

avec ξ et W^{DT} indépendantes.

Ensuite, comme W^{CT} est la limite d'un processus de branchement projeté et renormalisé, on utilise la propriété de branchement au premier instant de saut pour obtenir des équations de point fixe sur les lois limites. En appelant W la loi de W^{CT} lorsqu'on part d'une particule de type 1, il est facile de voir que W est solution de l'équation en loi (sur \mathbb{C})

$$W \stackrel{\mathcal{L}}{=} e^{-\lambda_2 T} [m]W$$

où e^{-T} est de loi $\text{Beta}(1, m - 1)$. Différents résultats peuvent se déduire de cette équation : existence et unicité de solutions, support de la loi (détails dans [5]).

References

- [1] K.B. ATHREYA AND P. NEY, *Branching processes*, Springer, 1972.
- [2] J. BERTOIN AND A. ROUAULT Discretization methods for homogeneous fragmentations, *J. London Math. Soc.* Vol 72, n1 (2005)
- [3] J.D.BIGGINS. Martingale convergence in the branching random walk. *Adv in Appl. Probab.* 10, 62-84, 1978
- [4] B. CHAUVIN, T. KLEIN, J.F. MARCKERT, AND A. ROUAULT. Martingales and Profile of Binary Search Trees. *Electronic J. Probab.* Vol. 10 Paper n 12, 420-435 (2005)
- [5] B. CHAUVIN, Q. LIU, N. POUYANNE, Limit distributions for multitype branching processes associated with m -ary search trees when $m \geq 27$. (2010) Preprint.
- [6] B. CHAUVIN, N. POUYANNE, m -ary search trees when $m > 26$: a strong asymptotics for the space requirements *Random Structures and Algorithms* **24**, 2, 133-154 (2004)

- [7] B. CHAUVIN, N. POUYANNE, R. SAHNOUN, Limit distributions for large Pólya urns. To appear, *Annals of Applied Probability* (2010)
- [8] B. CHAUVIN AND A. ROUAULT. Connecting Yule process, Bisection and Binary Search Tree via Martingales. *J. of the Iranian Stat. Society* Vol. 3 , n 2 (2004)
- [9] H.-H. CHERN, H.-K. HWANG. Phase changes in random m -ary search trees and generalized quicksort. *Random Structures and Algorithms* **19** (2001), 316–358.
- [10] L. DEVROYE. Branching Processes and Their Applications in the Analysis of Tree Structures and Tree Algorithms. In M. Habib et al., editor, *Probabilistic Methods for Algorithmic Discrete Mathematics*. Springer, 1998.
- [11] J. A. FILL, N. KAPUR, The space requirement of m -ary search trees: distributional asymptotics for $m \geq 27$. (2004) *Proceedings of the 7th Iranian Conference. arXiv:math.PR/0405144*
- [12] P. FLAJOLET AND R. SEDGEWICK, *Analytic combinatorics*. Cambridge University Press, 2009.
- [13] J. JABBOUR-HATTAB, Martingales and large deviations for binary search trees. *Random Structure and Algorithms*, 19, 112-127, (2001)
- [14] S. JANSON, Functional limit theorem for multitype branching processes and generalized Pólya urns. *Stochastic Processes and Applications*, **110** (2004), 177–245.
- [15] H. M. MAHMOUD, *Evolution of Random Search Trees*, John Wiley & Sons, New York, 1992.
- [16] B. PITTEL, On growing random binary trees. *J. Math. Anal. Appl.*, 103(2), 461-480, 1984.
- [17] N. POUYANNE, An algebraic approach to Pólya processes (2008). *Ann. Inst. Henri Poincaré*, Vol. 44, No. 2, 293–323.