Random trees for analysis of algorithms.

Brigitte CHAUVIN
http://chauvin.perso.math.cnrs.fr/

BMS Summer School 2011

4 oct 2011[1]

Contents

[1]Keywords: branching process, branching property, martingale, analysis of algorithms, Pólya urn, binary search tree, smoothing transformation, fixed point equation, support.

1 Abstract/Introduction

In this course, three examples are developed, involving random trees: binary search trees, Pólya urns and m-ary search trees. For all of them, a same plan runs along the following outline:

(a) A discrete Markovian stochastic process is related to a tree structure. In the three cases, the tree structure is a model coming from computer science and from analysis of algorithms, typically sorting algorithms. The recursive nature of the problem gives rise to *discrete time martingales*.

(b) The process is embedded in continuous time, giving rise to a one (ex. 1) or to a multitype (ex. 2 and ex. 3) *branching process*. The associated continuous time martingales are connected to the previous discrete time martingales. Thanks to the branching property, the asymptotics of this continuous time branching process is more accessible than in discrete time, where the branching property does not hold.

In the three cases, the limit of the (rescaled) martingale has a non classic distribution. We present some expected properties of these limit distribution (density, infinite divisibility, ...) together with more exciting properties (divergent moment series, self decomposability, fixed point equation, ...).

2 Binary search trees

(abridged: bst)

2.1 Definition of a binary search tree

A binary search tree is associated with the sorting algorithm "Quicksort" and several definitions can be given with this algorithm in mind (see Mahmoud [18]). Here we give a more probabilistic definition. Let

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

be the set of finite words on the alphabet $\{0,1\}$, where ε denotes the empty word. Words are written by concatenation, the left children of u is $u0$ and the right children of u is $u1$. A *binary complete tree* T is a finite subset of $\mathcal{U}$ such that

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

The root of the tree is ε. The length of a node u is denoted by $|u|$, it is the depth of u in the tree ($|\varepsilon| = 0$). The set of binary complete trees is denoted by **BinTree**. In a binary complete tree $T \in$ **BinTree**, a leaf is a node without any children, the set of leaves of T is denoted by ∂T. The other nodes are internal nodes.

In the following, we call a random binary search tree the discrete time process $(\mathcal{T}_n)_{n\geq 0}$, with values in **BinTree**, recursively defined by: $\mathcal{T}_0$ is reduced to a single leaf; for $n \geq 0$, $\mathcal{T}_{n+1}$ is obtained from $\mathcal{T}_n$ by a uniform insertion on one of the $(n+1)$ leaves of $\mathcal{T}_n$.

2.2 Profile. Discrete martingale

A huge literature exists on binary search trees: see Flajolet and Sedgewick [12] for analytic methods, Devroye [10] for more probabilistic ones and Mahmoud [18] for a book on this topics. Here, let's focus on the *profile* which expresses the shape of the tree. The profile is given par the sequence

$$U_k(n) := \text{the number of leaves at level } k \text{ in tree } \mathcal{T}_n.$$

What is the asymptotic behavior of these quantities when $n \to +\infty$? To answer, let's introduce the *level polynomial*, defined for any $z \in \mathbb{C}$ by

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

It is indeed a polynomial, since for any level k greater than the height of the tree, $U_k(n) = 0$. It is a random variable, not far from a martingale.

Theorem 2.1 *(Jabbour-Hattab, 2001, [14]) For any complex number $z \in \mathbb{C}$ such that $z \neq -k, k \in \mathbb{N}$, let*

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

and

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

Then, $(M_n^{BST}(z))_n$ is a $\mathcal{F}_n$-martingale which can also be written

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

This martingale is a.s. convergent for any z positive real.

It converges in L^1 to a limit denoted by $M_\infty^{BST}(z)$ for any $z \in]z_-, z_+[$ and it converges a.s. to 0 for any $z \notin]z_-, z_+[$, where z_- and z_+ are the two solutions of equation $z \log(z/2) - z + 2 = 1$.

Proof. The martingale property comes from

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

2.3 Embedding in continuous time. Yule tree

The idea is due to Pittel [19]. Let's consider a continuous time branching process, with an ancestor at time $t = 0$, who lives an exponential time with parameter 1. When he dies, it gives birth to two children who live an exponential time with parameter 1, independently from each other, etc... The tree process thus obtained is called the Yule tree process, it is denoted by $(\mathcal{Y}_t)_t$.

Let's call N_t the number of leaves in $\mathcal{Y}_t$ (at time t) and denote by

$$0 < \tau_1 < \cdots < \tau_n < \ldots$$

the jumping times. It is easy to see that $\tau_n - \tau_{n-1}$ is $\mathcal{E}xp(n)$-distributed and that (it is the embedding principle)

$$\left(\mathcal{Y}_{\tau_n}\right)_n \overset{\mathcal{L}}{=} \left(\mathcal{T}_n\right)_n .$$

On the Yule tree, consider the branching random walk in which the position of a leaf u of $\mathcal{Y}_t$ is given by

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

so that the displacements are (up to a constant) the generations in the tree. Biggins ([4]) and Bertoin and Rouault ([3]) proved

Theorem 2.2 *For any $z \in \mathbb{C}$,*

$$M_t^{YULE}(z) := \sum_{u \in \partial \mathcal{Y}_t} z^{|u|} e^{-t(2z-1)}$$

is a $\mathcal{F}_t$-martingale.

This martingale converges a.s. for all z positive real.

It converges in L^1 to a limit denoted by $M_\infty^{YULE}(z)$ for all $z \in]z_-, z_+[$ and it converges a.s. to 0 for all $z \notin]z_-, z_+[$, where z_- and z_+ are the solutions of equation $z \log(z/2) - z + 2 = 1$; $z_- = 0.186\ldots; z_+ = 2.155\ldots$

2.4 Martingale connection Yule tree - binary search tree

Proposition 2.3 *For any $z \in]z_-, z_+[$, the following connection holds*

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

where ξ and $M_\infty^{BST}(z)$ are independent and ξ is $\mathcal{E}xp(1)$-distributed.

PROOF. Use the embedding principle.

2.5 Asymptotics

The above connection is one of the main tools leading to the following theorem on the profile of binary search trees.

Theorem 2.4 *([5], 2005) For any compact $K \subset]z_-, z_+[$,*

$$\frac{U_k(n)}{\mathbb{E}(U_k(n))} - M_\infty^{BST}(\frac{k}{2\log n}) \underset{n\to\infty}{\longrightarrow} 0 \quad a.s.$$

uniformly on $\frac{k}{2\log n} \in K$.

2.6 Another branching random walk: the bisection

A discrete time branching random walk in $\mathbb{R}$ is recursively defined as follows: an ancestor is at the origin at time 0. At time 1 it has a random number N of children and their positions are $X_1, X_2, \dots$. We denote by Z the point process

$$Z = \sum_{i=1}^{N} \delta_{X_i}$$

(where δ is the Dirac measure). Each individual reproduces independently from each other and the position of each child relatively to its parent is given by an independent copy of Z. Let $Z^{(n)}$ be the point process in $\mathbb{R}$ which describes the positions of the individuals of the n-th generation. $Z^{(0)} = \delta_0; Z^{(1)} = Z$ and

$$Z^{(n)} = \sum_{|u|=n} \delta_{X_u},$$

where X_u is the position of individual u. Individuals are canonically labeled by the words on $\mathcal{U}$ as in Section 2.1. We call $(Z^{(n)})_n$ a branching random walk with point process Z. Additive martingales are simply attached to these processes: let for any parameter $\theta \in \mathbb{R}$

$$\Lambda(\theta) := \log \mathbb{E} \int_{\mathbb{R}} e^{\theta x} Z(dx) = \log \mathbb{E} \sum_{i=1}^{N} e^{\theta X_i},$$

and let

$$M_n(\theta) := \int_{\mathbb{R}} e^{\theta x - n\Lambda(\theta)} Z^{(n)}(dx) = \sum_{|u|=n} e^{\theta X_u - n\Lambda(\theta)}.$$

Then $(M_n(\theta))_n$ is a martingale which converges a.s. to a limit $M_\infty(\theta)$. Under a "$k \log k$" condition, (cf. Biggings [4]),

- if $\theta\Lambda'(\theta) - \Lambda(\theta) < 0$, the convergence is in L^1 and $\mathbb{E}M_\infty(\theta) = 1$
- if $\theta\Lambda'(\theta) - \Lambda(\theta) \geq 0$, then $M_\infty(\theta) = 0$ a.s.

A particular case: the bisection

Consider a segment of length 1 at time 0. At time 1 cut it in two pieces of length U and $1 - U$, where U is uniform on $[0, 1]$. Continue cutting recursively and independently the two pieces. What are the lengths of the 2^n pieces obtained after n steps?

Proposition 2.5 *([9]) Let $(Z^{(n)})_n$ be the branching random walk (called the bisection) with point process $Z = \delta_{-\log U} + \delta_{-\log(1-U)}$ and let $(M_n^{BIS})_n$ be the associated martingale: for any $z \in \mathbb{R}$,*

$$M_n^{BIS}(z) = \sum_{|u|=n} e^{-(2z-1)X_u} z^n.$$

Then, for any $z \in]z_-, z_+[$,

$$M_\infty^{BIS}(z) = M_\infty^{BST}(z) \qquad a.s.$$

3 Pólya urns

3.1 Definition of a Pólya process

Define a discrete time process as follows. At each time, an urn contains black balls and red balls. Pick up at random (which means *uniformly*) one ball, remember its color, put it back in the urn, and add

→ a red balls and b black balls if a red ball has been drawn;

$\to c$ red balls and d black balls if a black ball has been drawn.

Denote by $R = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ the replacement matrix. We are interested in

$$U^{DT}(n) = \begin{pmatrix} \#\text{ red balls at time } n \\ \#\text{ black balls at time } n \end{pmatrix}$$

which is called the composition vector of the urn. Let $U^{DT}(0) = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$ be the initial composition of the urn. In what follows, we assume that the urn is *balanced* which means that the total number of balls added at each step is a constant $S = a + b = c + d$ called the balance. Consequently, the total number of balls at time n is deterministic, equal to $\alpha + \beta + nS = u + nS$ (where u denotes the total number of balls at time 0). Of course the composition of the urn is random.

The asymptotic behavior of the urn depends on the spectrum of matrix R. There are two eigenvalues: the largest eigenvalue is the balance S and the smallest eigenvalue is the integer $m = a - c = d - b$. We assume in the following that

$$S > m$$

consequently, the historical Pólya case where $S = m$ is not covered by these notes. It corresponds to taking $R = S\,\mathrm{Id}$ as replacement matrix. It has been well-known, since Gouet [13], that the composition vector admits an almost sure asymptotics of order one: $U^{DT}(n) = nD + o(n)$ where the random vector D has a Dirichlet density (explicitly given in [13]).

The asymptotic behavior is quite different depending on $\sigma := \dfrac{m}{S} \le \dfrac{1}{2}$ or $\sigma > \dfrac{1}{2}$. Briefly (see Janson [15]):

1. when $\sigma < \frac{1}{2}$, the urn is called *small* and, except when R is triangular, the composition vector is asymptotically Gaussian[2]:

$$\frac{U^{DT}(n) - nv_1}{\sqrt{n}} \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(0, \Sigma^2)$$

where v_1 is a suitable eigenvector of ${}^tR$ relative to S and Σ^2 has a simple closed form;

2. when $\frac{1}{2} < \sigma < 1$, the urn is called *large* and the composition vector has a quite different strong asymptotic form, it is described by a limit of martingales which are introduced in the following section.

[2]The case $\sigma = 1/2$ is similar to this one, the normalisation being $\sqrt{n \log n}$ instead of $\sqrt{n}$.

In the following, we assume that $\sigma > \dfrac{1}{2}$.

3.2 Vectorial discrete martingale

Recall the dynamics of the process $(U^{DT}(n))$: if we call for a while x and y the two coordinates of the vector $U^{DT}(n)$, then

$$\begin{cases} U^{DT}(n+1) = U^{DT}(n) + \begin{pmatrix} a \\ b \end{pmatrix} \text{ with probability } \frac{x}{x+y} \\ U^{DT}(n+1) = U^{DT}(n) + \begin{pmatrix} c \\ d \end{pmatrix} \text{ with probability } \frac{y}{x+y}, \end{cases} \tag{1}$$

and since $x + y = u + nS$,

$$\mathbb{E}^{\mathcal{F}_n}\left(U^{DT}(n+1)\right) = \left(I + \frac{A}{u+nS}\right) U^{DT}(n) \tag{2}$$

where I is the identity matrix of dimension 2 and where

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

The martingale property is immediate from (2) and stated in the proposition hereafter.

Proposition 3.6 *For any positive real x and any positive integer n, denote by $\gamma_{x,n}$ the polynomials*

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

Then, $\gamma_{\frac{u}{S},n}(\frac{A}{S})^{-1}U^{DT}(n)$ is a (vectorial) $\mathcal{F}_n$-martingale.

Let us call (v_1, v_2) a basis of eigenvectors associated with the two eigenvalues S and m respectively and (u_1, u_2) its dual basis of linear forms (or left eigenvectors). It means that $u_1 \circ A = Su_1$ and $u_2 \circ A = mu_2$. They can be explicitly calculated:

$$u_1(x,y) = \frac{1}{S}(x+y), \qquad u_2(x,y) = \frac{1}{S}(bx - cy), \tag{3}$$

$$v_1 = \frac{S}{(b+c)}\begin{pmatrix} c \\ b \end{pmatrix}, \qquad v_2 = \frac{S}{(b+c)}\begin{pmatrix} 1 \\ -1 \end{pmatrix}. \tag{4}$$

By projection of the vectorial martingale onto any of the two eigendirections, we get a one dimensional martingale, for instance

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

is a $\mathcal{F}_n$-martingale which converges (in all the possible ways) when n goes to infinity. With a control of the L^p moments of the martingale $M^{DT}(n)$ (see Pouyanne [20]) one gets the following theorem.

Theorem 3.7 *Assume $\sigma \in]1/2, 1[$. Let*

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

When n goes to infinity,

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

where v_1, v_2, u_1, u_2 are defined by (3) and (4); the convergence happens a.s. and in all the L^p; the moments of W^{DT} can be recursively calculated.

3.3 Embedding in continuous time. Two type branching process

Imagine that red clocks and blacks clocks $\mathcal{E}xp(1)$ distributed are put on red balls and black balls respectively. When a red clock rings, the red ball dies et gives birth to $(a+1)$ red balls and b black balls. When a black clock rings, the black ball dies et gives birth to c red balls and $(d+1)$ black balls. The composition vector of the urn is now

$$U^{CT}(t) = \begin{pmatrix} \#\text{ red balls at time } t \\ \#\text{ black balls at time } t \end{pmatrix}.$$

Start from the same initial composition at time 0: $U^{CT}(0) = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$. The previous process is a two type branching process. Denote by $0 = \tau_0 < \tau_1 < \cdots < \tau_n < \cdots$ the jumping times. It is easy to see that $\tau_{n+1} - \tau_n$ is $\mathcal{E}xp(u+nS)$ distributed and that (embedding principle)

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

The benefit of this embedding consists in providing a branching process in continuous time where the subtrees of the root are independent, which was not the case in the discrete time process. For such a multitype branching process, it is classical (see Athreya-Ney [1]) to see that

Proposition 3.8 $$\left(e^{-tA}U^{CT}(t)\right)_{t\geq 0}$$

is a $\mathcal{F}_t$-martingale (a vectorial one). By projection on the eigenlines (the same ones as previously) we get real-valued convergent martingales and

$$\xi = \lim_t e^{-St}u_1\left(U^{CT}(t)\right) \qquad \text{and} \qquad W^{CT} := \lim_t e^{-mt}u_2\left(U^{CT}(t)\right)$$

With a control on the L^p moments of these martingales, we get the asymptotic behavior of the continuous time process in the following theorem ([8]).

Theorem 3.9 *When t goes to infinity,*

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

where ξ and W^{CT} are real valued random variables defined in Proposition 3.8; the convergence happens a.s. and in all the L^p; the law of ξ is Gamma(u/S)*, where $u = \alpha + \beta$ is the total number of balls at time* 0*.*

3.4 Martingale connection

Proposition 3.10 *The following connection holds*

$$W^{CT} = \xi^\sigma\ W^{DT} \qquad a.s.$$

where ξ and W^{DT} are independent and the law of ξ is Gamma(u/S)*.*

Proof. Use the embedding principle.

This connection indicates that informations on W^{CT} provide informations on W^{DT}. To get more on W^{CT}, let's take advantage of the independence properties in continuous time, already mentioned; they are exploited in the following section.

3.5 Asymptotics

Remember (with Proposition 3.8) that W^{CT} is the limit of a branching process after projection and scaling. The branching property can be applied at the first jumping time τ_1 (denoted by τ_1 for simplicity). Adopt the notation

$$\begin{cases} X_t := U^{CT}(t) \text{ starting from } \begin{pmatrix}1\\0\end{pmatrix} \\ Y_t := U^{CT}(t) \text{ starting from } \begin{pmatrix}0\\1\end{pmatrix}, \end{cases} \tag{6}$$

so that

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

where the notation $[n]X$ stands for the sum of n independent copies of the random variable X. After projection and scaling, the limits denoted by X and Y respectively satisfy

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

It can be seen as a system of fixed point distributional equations. Translating it on the Fourier transforms, it gives a system of differential equations. This system can be solved, giving informations on the distribution of X and Y : these distributions have a density on the whole real axis, they are infinitely divisible, their Laplace series (exponential series of the moments) have a radius of convergence equal to 0, ... Details can be found in [8].

A natural extension of the method should provide results for Pólya urns with more than two colors. Indeed, we still get a system of fixed point distributional equations, which in general is impossible to be solved. Nevertheless, for a particular case, corresponding to the classical sorting algorithm of so-called m-ary search trees, some results on the limit distribution can be obtained.

4 m-ary search trees

4.1 Definition

For $m \geq 3$, m-ary search trees are a generalization of binary search trees (see for instance Mahmoud [18]). A sequence $(T_n, n \geq 0)$ of m-ary search trees grow by successive insertions of keys in their leaves. Each node of these trees contains at most $(m-1)$ keys. Keys are i.i.d. random variables $x_i, i \geq 1$ with any diffusive distribution on the interval $[0,1]$. The tree $T_n, n \geq 0$, is recursively defined as follows:

T_0 is reduced to an empty node-root; T_1 is reduced to a node-root which contains x_1, T_2 is reduced to a node-root which contains x_1 and x_2, ... , T_{m-1} has a node-root containing $x_1, \dots x_{m-1}$. As soon as the $m-1$-st key is inserted in the root, m empty subtrees of the root are created, corresponding from left to right to the m ordered intervals $I_1 =]0, x_{(1)}[, \dots, I_m =]x_{(m-1)}, 1[$ where $0 < x_{(1)} <$

$\dots < x_{(m-1)} < 1$ are the ordered $(m-1)$ first keys. Each following key $x_m, \dots$ is recursively inserted in the subtree corresponding to the unique interval I_j to which it belongs. As soon as a node is saturated, m empty subtrees of this node are created.

Ex : for $m = 4$, draw the tree containing the following data:
0.3, 0.1, 0.4, 0.15, 0.9, 0.2, 0.6, 0.5, 0.35

For each $i = \{1, \dots, m-1\}$ and $n \geq 1$, let

$$X_n^{(i)} := \text{ number of nodes in } T_n \text{ which contain } (i-1) \text{ keys (and } i \text{ gaps).}$$

Such nodes are named nodes of type i. We don't worry about the number of saturated nodes. The vector X_n^{DT} is called the composition vector of the m-ary search tree. It provides a model for the space requirement of the sorting algorithm.

When the data are i.i.d. random variables, one gets a *random* m-ary search tree, and with this dynamics, the insertion of a new key is *uniform* on the gaps. We want to describe the asymptotic behavior of the vector X_n^{DT} as n tends to infinity.

4.2 Vectorial discrete martingale

The dynamics of the nodes is illustrated by Figure 1 and it gives the expression of X_{n+1} as a function of X_n. The $(n+1)$-st data is inserted in a node of type i, $i = 1, \dots, m-1$ with probability $\dfrac{iX_n^{(i)}}{n+1}$ and in this case, the node becomes a node of type $i+1$ for $i = 1, 2, \dots, m-2$, and gives m nodes of type 1, if $i = m-1$.

In other words, for $i = 1, \dots, m-1$, let

$$\begin{cases} \Delta_1 = (-1, 1, 0, 0, \dots) \\ \Delta_2 = (0, -1, 1, 0, \dots) \\ \quad \vdots \\ \Delta_{m-2} = (0, \dots, 0, -1, 1) \\ \Delta_{m-1} = (m, 0, \dots, 0, -1). \end{cases},$$

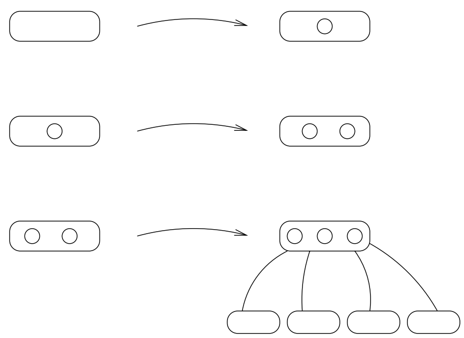

Figure 1: Dynamics of insertion of data, in the case $m = 4$.

Then

$$\mathbb{P}(X_{n+1} = X_n + \Delta_i | X_n) = \frac{iX_n^{(i)}}{n+1}.$$

The remarkable fact is that the transition from X_n to X_{n+1} is *linear* in X_n. Consequently, denote

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

so that

$$\mathbb{E}\Big(X_{n+1}|X_n\Big) = \Big(I + \frac{A}{n+1}\Big) X_n.$$

Denoting

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

we immediately deduce

Proposition 4.11 *(2004, [7])*
$(\Gamma_n(A)^{-1}X_n)_n$ *is a* $\mathcal{F}_n$ *vectorial martingale.*

The spectrum of matrix A gives the asymptotic behavior of X_n. The eigenvalue are the roots of the characteristic polynomial

$$\chi_A(\lambda) = \prod_{k=1}^{m-1}(\lambda + k) - m! = \frac{\Gamma(\lambda + m)}{\Gamma(\lambda + 1)} - m! \tag{9}$$

where Γ denotes Euler's Gamma function. In other words, each eigenvalue λ is a solution of

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

All eigenvalues are simple, 1 being the one having the largest real part. Let λ_2 be the eigenvalue with a positive imaginary part and with the greatest real part σ_2 among all the eigenvalues different from 1. The asymptotic behaviour of X_n is different depending on $\sigma_2 \leq \frac{1}{2}$ or $\sigma_2 > \frac{1}{2}$. The results in the following theorem can be found in [18, 15, 7, 20].

Theorem 4.12

- *When* $\sigma_2 < \frac{1}{2}$, $m \leq 26$ *and*

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

where v_1 *is an eigenvector for the eigenvalue* 1, *and where* Σ^2 *can be calculated.*

- *When* $1 > \sigma_2 > \frac{1}{2}$, $m \geq 27$ *and*

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

where v_1, v_2 *are deterministic, nonreal eigenvectors;* W^{DT} *is a* $\mathbb{C}$*-valued random variable which a martingale limit;* $o(\ )$ *means a convergence a.s. and in all the* L^p, $p \geq 1$*; the moments of* W^{DT} *can be recursively calculated.*

4.3 Embedding in continuous time. Multitype branching process

For $m \geq 3$, define a continuous time multitype branching process, with $m - 1$ types

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

with $X^{CT}(t)^{(j)} = \#$ particles of type j alive at time t.

Each particle of type j is equipped with a clock $\mathcal{E}xp(j)$-distributed. When this clock rings, the particle of type j dies and gives birth to

→ a particle of type $j+1$ when $j \leq m-2$
→ m particles of type 1 when $j = m-1$.

Call $0 = \tau_0 < \tau_1 < \cdots < \tau_n < \cdots$ the successive jumping times. It is easy to see that $\tau_n - \tau_{n-1}$ is $\mathcal{E}xp(u+n-1)$-distributed, where $u = \sum_{k=1}^{m-1} kX^{CT}(0)^{(k)}$ is the numbers of free places at time 0.

The embedding principle can be expressed

$$\left(X^{CT}(\tau_n)\right)_n \overset{\mathcal{L}}{=} \left(X_n\right)_n .$$

For this multitype branching process, it is classical to see that

Proposition 4.13

$$\left(e^{-tA}X^{CT}(t)\right)_{t\geq 0}$$

is a $\mathcal{F}_t$ vectorial martingale.

By projection on the eigenlines (v_1, v_2 are eigenvectors and u_1, u_2 are eigen linear forms), we get

Theorem 4.14 *([6], Janson [15])*

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

where ξ is a real-valued random variable Gamma(u)-distributed;

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

is a complex valued random variable, which admits moments of any order $p \geq 1$; $o(\ )$ means a convergence a.s. and in all the L^p, $p \geq 1$. Moreover, the following martingale connection holds

$$W^{CT} = \xi^{\lambda_2}\ W^{DT} \qquad a.s.$$

with ξ and W^{DT} independent.

4.4 Asymptotics

Notations

In the following, we denote

$$T = \tau_{(1)} + \cdots + \tau_{(m-1)}. \tag{10}$$

where the $\tau_{(j)}$ are independent of each other and each $\tau_{(j)}$ is $\mathcal{E}xp(j)$ distributed. Let us make precise some elementary properties of T. It is easy to see that T has

$$f_T(u) = (m-1)e^{-u}(1-e^{-u})^{m-2}\mathbf{1}_{\mathbb{R}_+}(u), \quad u \in \mathbb{R}, \tag{11}$$

as a density, so that e^{-T} has a Beta distribution with parameters 1 and $m-1$. A straightforward change of variable under the integral shows that for any complex number λ such that $\Re(\lambda) > 0$,

$$\mathbb{E}e^{-\lambda T} = \int_0^{+\infty} e^{-\lambda u} f_T(u)du = (m-1)B(1+\lambda, m-1) \tag{12}$$

$$= \frac{(m-1)!}{\prod_{k=1}^{m-1}(\lambda+k)}, \tag{13}$$

where B denotes Euler's Beta function:

$$B(x,y) = \int_0^1 u^{x-1}(1-u)^{y-1}du = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}, \quad \Re x > 0, \Re y > 0. \tag{14}$$

In particular,

$$m\mathbb{E}|e^{-\lambda T}| \begin{cases} < 1 \text{ if } \Re(\lambda) > 1, \\ = 1 \text{ if } \Re(\lambda) = 1, \\ > 1 \text{ if } \Re(\lambda) < 1. \end{cases} \tag{15}$$

A complete description of the $\mathbb{C}$-valued random variable W^{CT} is wanted. Since it is a limit of a branching process after projection and scaling, the branching property applied at the first splitting time provides fixed point equations on the limit distributions. Denoting by W_1 the distribution of W^{CT} starting from a particle of type 1 (which is indeed the case for the m-ary search tree), then W_1 is a solution of the following equation in distribution

$$W \overset{\mathcal{L}}{=} e^{-\lambda_2 T}(W^{(1)} + \cdots + W^{(m)}),$$

where T is defined in (10), $W^{(i)}$ are $\mathbb{C}$-valued independent copies of W, which are also independent of T.

Several results can be deduced from this equation, namely the existence and the unicity of solutions, properties of the support. Some are described in the following section.

5 Smoothing transformation

In this section, inspired from [6], the following fixed point equation coming from the previous multitype branching process is studied, thanks to several methods. The different methods are general ones, they are used for other distributional equations, the following is just a motivation in relation with Section 4.

$$W \overset{\mathcal{L}}{=} e^{-\lambda T}(W^{(1)} + \cdots + W^{(m)}), \tag{16}$$

where $\lambda \in \mathbb{C}$, T is defined in (10), $W^{(i)}$ are $\mathbb{C}$-valued independent copies of W, which are also independent of T. We successively see the contraction method (to prove existence and unicity of a solution, in a suitable space of probability measure), some analysis on the Fourier transforms in order to prove that W has a density, and a cascade type martingale which is a key tool to obtain the existence of exponential moments for W.

5.1 Contraction method

This method has been developed in [21] and [22] for many examples in analysis of algorithms. The idea is to get existence and unicity of a solution of equation (16) thanks to the Banach fixed point Theorem. The key point is to chose a suitable metric space of probability measures on $\mathbb{C}$ where the hereunder transformation $K : \mu \mapsto K\mu$ is a contraction.

$$K\mu := \mathcal{L}\Big(e^{-\lambda T}(X^{(1)} + \cdots + X^{(m)})\Big), \tag{17}$$

where T is given by (10), $X^{(i)}$ are independent random variables of law μ, which are also independent of T.

First step: the metric space.

For any complex number C, let $\mathcal{M}_2(C)$ be the space of probability distributions on $\mathbb{C}$ admitting a second absolute moment and having C as expectation. The first work is to see that K maps $\mathcal{M}_2(C)$ into itself.

Then, define d_2 as the Wasserstein distance on $\mathcal{M}_2(C)$ (see for instance Dudley [11]): for $\mu, \nu \in \mathcal{M}_2(C)$,

$$d_2(\mu,\nu) = \left(\min_{(X,Y)} \mathbb{E}\left(|X-Y|^2\right)\right)^{\frac{1}{2}}, \tag{18}$$

where the minimum is taken over couples of random variables (X, Y) having respective marginal distributions μ and ν; the minimum is attained by the Kantorovich-Rubinstein Theorem – see for instance Dudley [11], p. 421. With this distance d_2, $\mathcal{M}_2(C)$ is a complete metric space.

Second step: K is a contraction on $(\mathcal{M}_2(C), d_2)$.

It is a small calculation, taking some care when choosing the random variables: let (X, Y) be a couple of complex-valued random variables such that $\mathcal{L}(X) = \mu$, $\mathcal{L}(Y) = \nu$ and $d_2(\mu, \nu) = \sqrt{\mathbb{E}|X-Y|^2}$. Let $(X_i, Y_i), i = 1, \ldots, m$ be m independent copies of the d_2-optimal couple (X, Y), and T be a real random variable with density f_T defined by (11), independent from any (X_i, Y_i). Then,

$$\mathcal{L}(e^{-\lambda T} \sum_{i=1}^{m} X_i) = K\mu \quad \text{and} \quad \mathcal{L}(e^{-\lambda T} \sum_{i=1}^{m} Y_i) = K\nu,$$

so that

$$\begin{aligned}
d_2(K\mu, K\nu)^2 &\leq \mathbb{E}\left|\left(e^{-\lambda T}\sum_{i=1}^{m} X_i\right) - \left(e^{-\lambda T}\sum_{i=1}^{m} Y_i\right)\right|^2 \\
&= \mathbb{E}\left|e^{-\lambda T}\sum_{i=1}^{m}(X_i - Y_i)\right|^2 \\
&= \mathbb{E}\left|e^{-\lambda T}\right|^2 \mathbb{E}\left|\sum_{i=1}^{m}(X_i - Y_i)\right|^2 \\
&= \mathbb{E}\left|e^{-\lambda T}\right|^2 \left(\sum_{i=1}^{m}\mathbb{E}\left|X_i - Y_i\right|^2 + \sum_{i \neq j}\mathbb{E}\left(X_i - Y_i\right)\left(\overline{X_j} - \overline{Y_j}\right)\right) \\
&= m\mathbb{E}\left|e^{-2\lambda T}\right| \; d_2(\mu, \nu)^2.
\end{aligned}$$

Since $2\Re(\lambda) > 1$, we have $m\mathbb{E}\left|e^{-2\lambda T}\right| < 1$ (see (15)). Therefore K is a contraction on $\mathcal{M}_2(C)$. We have proved the following theorem.

Theorem 5.15 *Let $\lambda \in \mathbb{C}$ be a root of the characteristic polynomial (9) such that $\Re(\lambda) > \frac{1}{2}$, and let $C \in \mathbb{C}$. Then K is a contraction on the complete metric space $(\mathcal{M}_2(C), d_2)$, and the fixed point equation (16) has a unique solution W in $\mathcal{M}_2(C)$.*

Another distance on $\mathcal{M}_2(C)$

Take $\mu, \nu \in \mathcal{M}_2(C)$ and denote respectively φ and ψ their characteristic functions. By definition of $\mathcal{M}_2(C)$, both φ and ψ admit the expansion $\varphi(t) =$

$1 + i\langle t, C\rangle + O\left(|t|^2\right)$ when t tends to 0. Therefore, one can define $d_2^*(\mu, \nu)$ by

$$d_2^*(\mu, \nu) = \sup_{t \in \mathbb{C} \setminus \{0\}} \frac{|\varphi(t) - \psi(t)|}{|t|^2}.$$

Clearly, $d_2^*(\mu, \nu) < \infty$, and d_2^* is a distance on $\mathcal{M}_2(C)$. It can be easily checked that $(\mathcal{M}_2(C), d_2^*)$ is a complete metric space.

The following result is a counterpart of Theorem 5.15. It gives an alternative proof for the existence and uniqueness in the distributional equation (16) in the class of probability measures on $\mathbb{C}$ with a given mean and finite second moments.

Theorem 5.16 *Let $\lambda \in \mathbb{C}$ be a root of the characteristic polynomial (9) such that $\Re(\lambda) > \frac{1}{2}$, and let $C \in \mathbb{C}$. Then K is a contraction on the complete metric space $(\mathcal{M}_2(C), d_2^*)$, and the fixed point equation (16) has a unique solution W in $\mathcal{M}_2(C)$.*

5.2 Analysis on Fourier transforms

The aim is to prove that W solution of equation (16) has the whole complex plane $\mathbb{C}$ as its support and that W has a density with respect to the Lebesgue measure on $\mathbb{C}$. The method relies on [16] and [17] adapted in [6] for $\mathbb{C}$-valued variables. It runs along the following lines.

Let φ be the Fourier transform of any solution W of (16). It is a solution of the functional equation

$$\varphi(t) = \int_0^{+\infty} \varphi^m(te^{-\bar{\lambda}u}) f_T(u) du, \quad t \in \mathbb{C}, \tag{19}$$

where f_T is defined by (11).

It is sufficient to prove that φ is in $L^2(\mathbb{C})$ because it is dominated by $|t|^{-a}$ for some $a > 1$ so that the inverse Fourier transform provides a density for W. For a distributional equation in $\mathbb{R}$, it is proved that φ is in $L^1(\mathbb{R})$. ref ?

To prove that $\varphi(t) = O(|t|^{-a})$ when $|t| \to \infty$, for some $a > 1$, we use a Gronwall-type technical Lemma which holds as soon as $A := e^{-\lambda T}$ has good moments and once we prove that $\lim_{|t| \to +\infty} \varphi(t) = 0$. It is the same to prove that $\lim_{r \to +\infty} \psi(r) = 0$ where $\psi(r) := \max_{|t|=r} |\varphi(t)|$. This comes from iterating the distributional equation (19) so that

$$\psi(r) \leq \mathbb{E}(\psi^m(r|A|)).$$

By Fatou lemma, we deduce that $\limsup_r \psi(r)$ equals 0 or 1. And it cannot be 1 because of technical considerations and because the only point where $\psi(r) = 1$ is $r = 0$. This key fact comes from a property of the support of W strongly related to the distributional equation with a non lattice type assumption: as soon as a point z is in the support of W, then the whole disc $D(0, |z|)$ is contained in the support of W. Finally, the result is

Theorem 5.17 *Let W be a complex-valued random variable solution of the distributional equation*

$$W \overset{\mathcal{L}}{=} e^{-\lambda T}(W^{(1)} + \cdots + W^{(m)}),$$

where λ is a complex number, $W^{(i)}$ are independent copies of W, which are also independent of T. Assume that $\lambda \neq 1$, $\Re(\lambda) > 0$, $\mathbb{E}W < \infty$ and $\mathbb{E}W \neq 0$. Then

(i) The support of W is the whole complex plane $\mathbb{C}$;

(ii) the distribution of W has a density with respect to the Lebesgue measure on $\mathbb{C}$.

5.3 Cascade type martingales

The distributional equation (16) suggests to use Mandelbrot's cascades in the complex setting (see [2] for independent interest about complex Mandelbrot's cascades).

As in Section 4, take $\lambda \in \mathbb{C}$ be a root of the characteristic polynomial (9) with $\Re(\lambda) > 1/2$. Still denote $A = e^{-\lambda T}$. Then $m\mathbb{E}A = 1$ because λ is a root of the characteristic polynomial (9) and $m\mathbb{E}|A|^2 < 1$ because $\Re(\lambda) > 1/2$ (see (15)). Let $A_u, u \in U$ be independent copies of A, indexed by all finite sequences of integers

$$u = u_1 ... u_n \in U := \bigcup_{n \geq 1} \{1, 2, \ldots, m\}^n$$

and set $Y_0 = 1$, $Y_1 = mA$ and for $n \geq 2$,

$$Y_n = \sum_{u_1 \ldots u_{n-1} \in \{1,\ldots,m\}^{n-1}} mAA_{u_1}A_{u_1u_2}\ldots A_{u_1\ldots u_{n-1}}. \tag{20}$$

As $m\mathbb{E}A = 1$, $(Y_n)_n$ is a martingale. This martingale has been studied by many authors in the real random variable case, especially in the context of Mandelbrot's cascades, see for example [17] and the references therein. It can be easily seen that

$$Y_{n+1} = A \sum_{i=1}^{m} Y_{n,i} \tag{21}$$

where $Y_{n,i}$ for $1 \le i \le m$ are independent of each other and independent of A and each has the same distribution as Y_n. Therefore for $n \ge 1$, Y_n is square-integrable and

$$\operatorname{Var} Y_{n+1} = (\mathbb{E}|A|^2 m^2 - 1) + m\mathbb{E}|A|^2 \operatorname{Var} Y_n,$$

where $\operatorname{Var} X = \mathbb{E}\left(|X - \mathbb{E}X|^2\right)$ denotes the variance of X. Since $m\mathbb{E}|A|^2 < 1$, the martingale $(Y_n)_n$ is bounded in L^2, so that the following result holds.

$$Y_n \to Y_\infty \text{ a.s. and in } L^2$$

where Y_∞ is a (complex-valued) random variable with

$$\operatorname{Var}(Y_\infty) = \frac{\mathbb{E}|A|^2 m^2 - 1}{1 - m\mathbb{E}|A|^2}.$$

Notice that, passing to the limit in (21) gives a new proof of the existence of a solution W of Eq. (16) with $\mathbb{E}W = 1$ and finite second moment whenever $\Re(\lambda) > 1/2$.

The previous convergence allows to think on Y_∞ instead of W and a technical lemma then leads to the following theorem, showing that the exponential moments of W exist in a neighborhood of 0, so that the characteristic function of W is analytic at 0.

Theorem 5.18 *Let $\lambda \in \mathbb{C}$ be a root of the characteristic polynomial (9) with $\Re(\lambda) > 1/2$ and let W be a solution of Equation (16). There exist some constants $C > 0$ and $\varepsilon > 0$ such that for all $t \in \mathbb{C}$ with $|t| \le \varepsilon$,*

$$\mathbb{E}e^{\langle t, W\rangle} \le e^{\Re(t) + C|t|^2} \quad \textit{and} \quad \mathbb{E}e^{|tW|} \le 4e^{|t| + 2C|t|^2}.$$

References

[1] K.B. Athreya and P. Ney. *Branching processes.* Springer, 1972.

[2] J. Barral, X. Jin, and B. Mandelbrot. Convergence of complex multiplicative cascades. *Ann. Appl. Probab.*, 20(4):1219–1252, 2010.

[3] J Bertoin and A. Rouault. Discretization methods for homogeneous fragmentations. *J. London Math. Soc.*, 72(1):91–109, 2005.

[4] J.D. Biggins. Martingale convergence in the branching random walk. *Adv in Appl. Prob.*, 10:62–84, 1978.

[5] B. Chauvin, T. Klein, J-F. Marckert, and A. Rouault. Martingales and profile of binary search trees. *EJP*, 10:420–435, 2005.

[6] B. Chauvin, Q. Liu, and N. Pouyanne. Limit distributions for multitype branching processes of m-ary search trees. *Preprint*, 2011.

[7] B. Chauvin and 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.

[8] B. Chauvin, N. Pouyanne, and R. Sahnoun. Limit distributions for large pólya urns. *Annals Applied Prob.*, 21(1):1–32, 2011.

[9] B. Chauvin and A. Rouault. Connecting yule process, bisection and binary search tree via martingales. *J. of the Iranian Stat. Society .*, 3(2):arXiv:math/0410318v1 [math.PR], 2004.

[10] L. Devroye. *Branching Processes and Their Applications in the Analysis of Tree Structures and Tree Algorithms*. Probabilistic Methods for Algorithmic Discrete Mathematics. Springer, m. habib et al. edition, 1998.

[11] R.M. Dudley. *Real Analysis and Probability.* Cambridge University Press, 2002 edition.

[12] P. Flajolet and R. Sedgewick. *Analytic combinatorics*. Cambridge University Press, 2009.

[13] R. Gouet. Strong convergence of proportions in a multicolor pólya urn. *J. Appl. Prob.*, 34:426–435, 1997.

[14] J. Jabbour-Hattab. Martingales and large deviations for binary search trees. *Random Structure and Algorithms*, 19:112–127, 2001.

[15] S. Janson. Functional limit theorem for multitype branching processes and generalized pólya urns. *Stochastic Processes and their Applications*, 110:177–245, 2004.

[16] Q. Liu. Asymptotic properties of supercritical age-dependent branching processes and homogeneous branching random walks. *Stochastic Processes and their Applications*, 82(1):61–87, 1999.

[17] Q. Liu. Asymptotic properties and absolute continuity of laws stable by random weighted mean. *Stochastic Processes and their Applications*, 95:83–107, 2001.

[18] H.M. Mahmoud. *Evolution of Random Search Trees.* John Wiley & Sons Inc., New York, 1992.

[19] B. Pittel. On growing random binary trees. *J. Math. Anal. Appl.*, 103(2):461–480, 1984.

[20] N. Pouyanne. An algebraic approach to pólya processes. *Ann. Inst. Henri Poincaré*, 44(2):293–323, 2008.

[21] U. Rösler. A fixed point theorem for distributions. *Stochastic Processes and their Applications*, 42:195–214, 1992.

[22] U. Rösler and Rüschendorf. The contraction method for recursive algorithms. *Algorithmica*, 29(1-2):3–33, 2001.