Random trees and Probability.

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

CIMPA Summer School 2014
August 2014[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 school, 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 type or to a multitype *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 all the cases, the limit of the (rescaled) martingale has a non classic distribution. We present some expected properties of these limit distribution (density, support, ...) together with more exciting properties (divergent moment series, fixed point equation, moments, ...).

Sections 2 on binary search trees and Section 3 on m-ary search trees are developped in this course, Pólya urns are developped in Pouyanne's course.

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 [15]).

oral Make precise the relation between the probabilistic definition of a BST and the sorting algorithm Quicksort.

Hereunder 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 $\mathcal{B}$. In a binary complete tree $T \in \mathcal{B}$, a leaf is a node without any children, the set of leaves of T is denoted by ∂T. The other nodes are internal nodes.

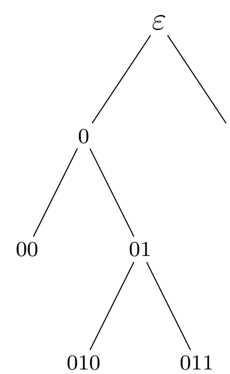


Figure 1: *An example of complete binary tree. At each node is written the word labelling it.*

In the following, we call a random binary search tree the discrete time process $(\mathcal{T}_n)_{n\geq 0}$, with values in $\mathcal{B}$, 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 $\mathcal{T}_n$. See Figure 2.

2.2 Profile of a binary search tree

2.2.1 Level polynomial. BST martingale

A huge literature exists on binary search trees: see Flajolet and Sedgewick [11] for analytic methods, Devroye [9] for more probabilistic ones and Mahmoud [15] for a book on this topics. In this section, let us focus on the *profile* which expresses the shape of the tree. The profile is given par the sequence

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

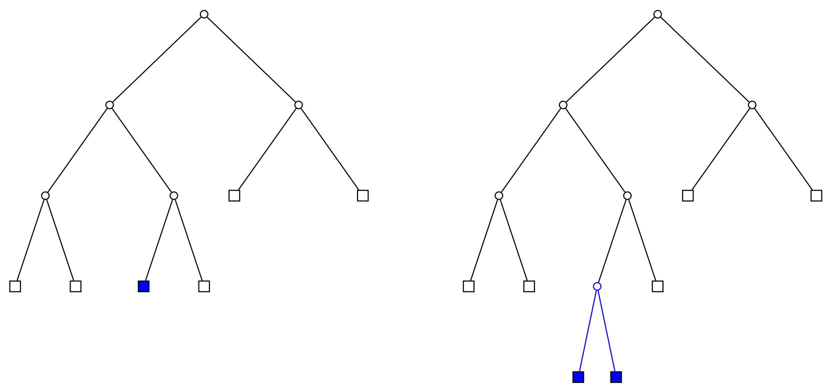


Figure 2: *An example of transition from $\mathcal{T}_5$, a binary search tree of size* 5 *to* $\mathcal{T}_6$*, a binary search tree of size* 6*. The insertion depth equals* 3.

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

$$W_n(z) := \sum_{k=0}^{+\infty} U_k(n)\ z^k = \sum_{u \in \partial\mathcal{T}_n} z^{|u|}. \tag{1}$$

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 *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 with expectation 1*, which can also be written*

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

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*

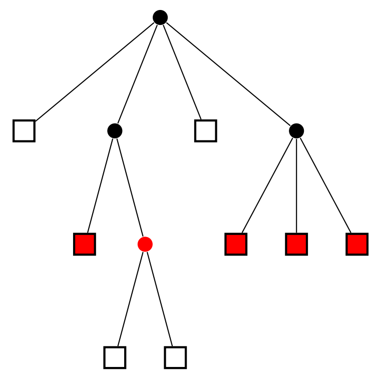


Figure 3: *A non binary tree* τ*, with height* $h(\tau) = 3$*, with profile* $(0, 2, 4, 2)$*. The second generation is in red.*

of equation $z \log(z/2) - z + 2 = 1$*. Numerically,* $z_+ = 4,31107\ldots$ *and* $z_- = 0,3733\ldots$

PROOF. Let d_n be the insertion depth of a new node in the tree $\mathcal{T}_n$ of size n. Remember this insertion is uniform on the $n+1$ leaves of $\mathcal{T}_n$. In other words

$$\mathbb{P}(d_n = k \big| \mathcal{F}_n) = \frac{U_k(n)}{n+1}.$$

The number of leaves at level k in the tree $\mathcal{T}_{n+1}$ can be expressed via d_n, see Figure 2:

$$U_k(\mathcal{T}_{n+1}) = U_k(\mathcal{T}_n) - \mathbf{1}_{\{d_n = k\}} + 2\ \mathbf{1}_{\{d_n = k-1\}}.$$

Consequently,

$$\begin{aligned}
\mathbb{E}(W_{n+1}(z) \mid \mathcal{F}_n) &= \mathbb{E}\left(\sum_{k=0}^{+\infty} U_k(\mathcal{T}_{n+1}) z^k \mid \mathcal{F}_n\right) \\
&= \sum_{k=0}^{+\infty} z^k \mathbb{E}\left(U_k(\mathcal{T}_n) - \mathbf{1}_{\{d_n=k\}} + 2\ \mathbf{1}_{\{d_n=k-1\}} \mid \mathcal{F}_n\right) \\
&= \sum_{k=0}^{+\infty} z^k \mathbb{E}\left(U_k(\mathcal{T}_n) - \mathbb{P}(d_n = k \mid \mathcal{F}_n) + 2\mathbb{P}(d_n = k-1 \mid \mathcal{F}_n)\right) \\
&= W_n(z) - \sum_{k=0}^{+\infty} \frac{U_k(\mathcal{T}_n)}{n+1} z^k + 2\sum_{k=1}^{+\infty} \frac{U_{k-1}(\mathcal{T}_n)}{n+1} z^k \\
&= W_n(z) - \frac{1}{n+1} W_n(z) + 2z W_n(z), \\
&= \frac{n+2z}{n+1}\ W_n(z), \qquad (3)
\end{aligned}$$

which gives the martingale property, after scaling: indeed, take the expectation in (3) to obtain par recurrence on n

$$\mathbb{E}(W_n(z)) = \prod_{j=0}^{n-1} \frac{j+2z}{j+1} = \Gamma_n(2z-1).$$

and divide by this expectation in (3) to get

$$\mathbb{E}(M_{n+1}^{BST}(z) \mid \mathcal{F}_n) = \mathbb{E}\left(\frac{W_{n+1}(z)}{\Gamma_{n+1}(2z-1)} \middle| \mathcal{F}_n\right) = \left(1 + \frac{2z-1}{n+1}\right) \frac{W_n(z)}{\Gamma_{n+1}(2z-1)} = M_n^{BST}(z).$$

□

2.2.2 Embedding in continuous time. Yule tree

The idea is due to Pittel [16]. 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

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

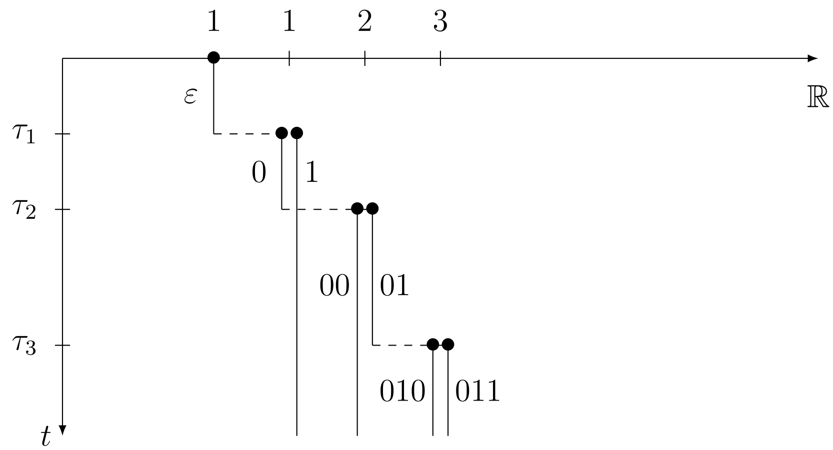


Figure 4: *A representation of a Yule tree. Here $N_t = 4$. The displacements are the generation numbers.*

the successive jumping times. For any time t there exists a unique integer n such that $\tau_{n-1} \leq t < \tau_n$ and

$$\{N_t = n\} = \{\tau_{n-1} \leq t < \tau_n\}.$$

Due to the lack of memory of the exponential distribution, $\tau_n - \tau_{n-1}$ is the first time when one of the n living particles splits. Consequently, $\tau_n - \tau_{n-1}$ is the minimum of n independent random variables $\mathcal{E}xp(1)$-distributed, so it is $\mathcal{E}xp(n)$-distributed. Moreover, the splitting particle is uniformly chosen among the n living particles. Finally, the continuous-time process stopped at time τ_n and the binary search tree have the same growing dynamics, so that (it is the embedding principle)

$$\left(\mathcal{Y}_{\tau_n}\right)_n \stackrel{\mathcal{L}}{=} \left(\mathcal{T}_n\right)_n. \tag{4}$$

From now, we consider that both processes (the binary search tree and the Yule tree) are built on the same probability space, so that equality in distribution becomes almost sure equality.

2.2.3 Connection Yule tree - binary search tree

On the Yule tree, let us define the "position" of an individual u living at time y by

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

so that the displacements are (up to the constant $\log 2$) like the generation numbers in the tree. See Figure 4. It can be proved (coming from the theory of branching random walks, see Biggins [4] and Bertoin and Rouault [3]) that

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, with expectation 1. 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 $z \log(z/2) - z + 2 = 1$; $z_- = 0.186\ldots$; $z_+ = 2.155\ldots$

Moreover, this martingale is connected to the BST martingale $M_n^{BST}(z)$. Indeed, writing $M_n^{BST}(z)$ like in (2), and taking the Yule martingale at time $t = \tau_n$ gives, thanks to the embedding principle (4)

$$\begin{aligned} M_{\tau_n}^{YULE}(z) &= \sum_{u \in \partial \mathcal{Y}_{\tau_n}} z^{|u|} e^{-\tau_n(2z-1)} \\ &= e^{-\tau_n(2z-1)} \sum_{u \in \mathcal{T}_n} z^{|u|} \\ &= e^{-\tau_n(2z-1)} \Gamma_n(2z-1) M_n^{BST}(z). \end{aligned}$$

It is not difficult to pass to the limit in the preceding equality, when n tends to infinity, when the parameter z belongs to the L^1-convergence domain of the martingales. In a Yule process, it is known (see for instance Athreya and Ney [1]) that $e^{-t}N_t$ tends to a random limit ξ which is $\mathcal{E}xp(1)$-distributed, when t tends to infinity. Since the stopping times τ_n go to infinity when n goes to infinity, we deduce that $ne^{-\tau_n}$ converges to ξ when n goes tio infinity. Finally let us use the Stirling formula to get the estimate

$$\Gamma_n(2z-1) \sim \frac{n^{2z-1}}{\Gamma(2z)},$$

so that we have proved the following proposition.

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.

2.2.4 Asymptotics of the profile

The above connection is one of the main tools leading to the following theorem on the profile of binary search trees. This theorem expresses that, after scaling, the profile tends to the random limit M_∞^{BST}. The asymptotics of the profile is concentrated on the levels k proportional to $\log n$.

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

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

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

2.3 Path length of a binary search tree

Definition 2.5 (path length of a BST) *The (external) path length L_n of a binary search tree $\mathcal{T}_n$ is the sum of the levels of the leaves of the tree.*

$$L_n := \sum_{u \in \partial\mathcal{T}_n} |u|.$$

This parameter of the tree is interesting in analysis of algorithms, since it represents a cost: $\frac{L_n}{n+1}$ is the mean cost of an insertion in the tree of size n.

Obviously, the path length is related to the level polynomial $W_n(z)$, since

$$L_n = \sum_{k\geq 1} kU_k(n) = W_n'(1).$$

Consequently, elementary computations lead to

$$\mathbb{E}(L_n) = 2(n+1)(H_{n+1} - 1) \qquad ; \qquad M_n'(1) = \frac{1}{n+1}\left(L_n - \mathbb{E}(L_n)\right).$$

oral Computations to be done at the blackboard. Comments on the alternative method, by analytic combinatorics.

Now, the derivative of a martingale is still a martingale, and $z = 1$ is in the L^1-convergence domain of the BST martingale $M_n(z)$, so that it is straightforward to obtain the following theorem.

Theorem 2.6 *After scaling, the path length of a binary search tree, defined by*

$$Y_n := \frac{1}{n+1}\left(L_n - \mathbb{E}(L_n)\right)$$

is a $\mathcal{F}_n$-martingale with mean 0. It converges almost surely and in L^1 to a random limit denoted by Y.

The law of Y is sometimes called the "law of Quicksort". It can be viewed as a solution of a distributional equation, in the spirit of Section 4.

3 m-ary search trees

3.1 Definition

For $m \geq 3$, m-ary search trees are a generalization of binary search trees (see for instance Mahmoud [15]). 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. The process $(T_n)_{n\geq 0}$ is recursively built, where T_n is the m-ary tree of size n, i.e. containing n keys. See Figure 5.

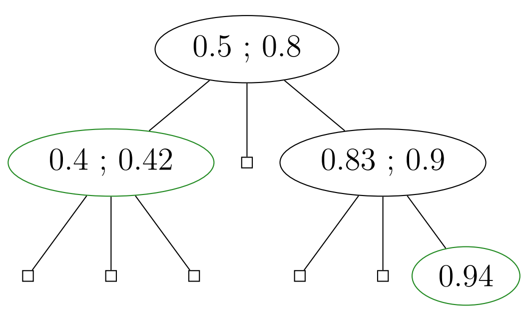


Figure 5: *A m-ary search tree ($m = 3$) of size* 7*, with* 8 *gaps,* 4 *nodes; among them, fringe nodes are in green. The tree has been built with the successive keys:* $0.8; 0.5; 0.9; 0.4; 0.42; 0.83; 0.94$.

To describe such a tree, let us introduce the so-called composition vector of the tree, X_n, which counts the nodes of differents types in the tree. This composition vector of the m-ary search tree provides a model for the space requirement of the sorting algorithm. More precisely, for each $i = \{1, \dots, m\}$ 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. Counting the number of keys in T_n with the $X_n^{(i)}$, we get the relation:

$$n = \sum_{i=1}^{m} (i-1) X_n^{(i)},$$

which allows to only study $m-1$ variables $X_n^{(i)}$ instead of m. We choose to forget the saturated nodes, which are internal nodes and to only count the non saturated nodes, which are at the fringe of the tree.

When the data are i.i.d. random variables, one gets a *random* m-ary search tree. 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.

oral Remark here the urn model, when considering the *gaps*. Call the gap process $(G_n)_n$. Write the replacement matrix at the blackboard. Notice that $G_n^{(i)} = i X_n^{(i)}$.

3.2 Vectorial discrete martingale

The dynamics of the nodes is illustrated by Figure 6 and it gives the expression

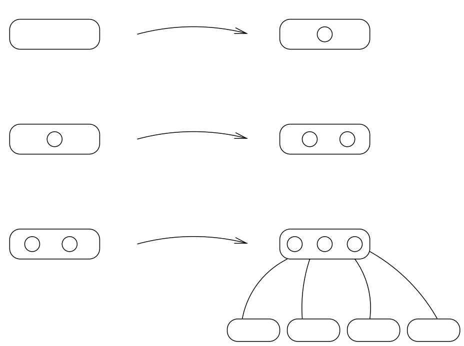


Figure 6: Dynamics of insertion of data, in the case $m = 4$.

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

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

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. Notice also that $\sum_{i=1}^{m-1} \frac{iX_n^{(i)}}{n+1} = 1$. When we note

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

then

$$\mathbb{E}\Big(X_{n+1}|X_n\Big) = \sum_{i=1}^{m-1} (X_n + \Delta_i) \frac{iX_n^{(i)}}{n+1} = \Big(I + \frac{A}{n+1}\Big) X_n.$$

We call Γ_n the polynomial

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

we deduce first, taking the expectation, and by induction that: $\mathbb{E}(X_n) = \Gamma_n(A)X_0$. Dividing

Proposition 3.7 *Let $(X_n)_n$ be the composition vector of a m-ary search tree. Then, $\big(\Gamma_n(A)^{-1}X_n\big)_n$ is a $\mathcal{F}_n$ vectorial martingale.*

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

$$\chi_A(\lambda) = \prod_{k=1}^{m-1}(\lambda + k) - m! = \frac{\Gamma(\lambda+m)}{\Gamma(\lambda+1)} - m! \tag{5}$$

where Γ denotes Euler's Gamma function. In other words, each eigenvalue λ is a solution of the so-called characteristic equation

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

oral Figure at the blackboard.

All eigenvalues are simple, 1 being the one having the largest real part. Let λ_2 be the eigenvalue with a positive imaginary part τ_2 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 [15, 12, 7, 17].

Theorem 3.8

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

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

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

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

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

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.*

3.3 Embedding in continuous time. Multitype branching process

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

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

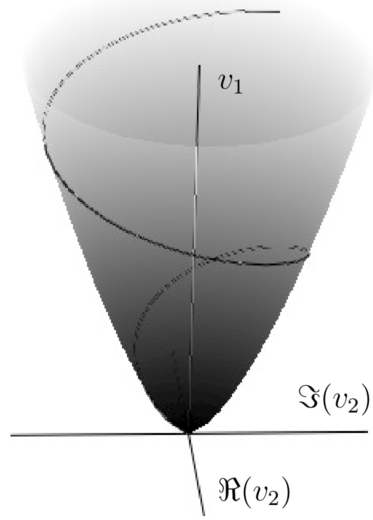
Geometrically speaking: let us denote by φ any argument of the complex number W^{DT}. The trajectory of the random vector X_n, projected in the 3-dimensional real vector space spanned by the vectors $(\Re(v_2), \Im(v_2), v_1)$ is almost surely asymptotic to the (random) spiral

$$\begin{cases} x_n = |W| n^{\sigma_2} \cos(\tau_2 \log n + \varphi), \\ y_n = -|W| n^{\sigma_2} \sin(\tau_2 \log n + \varphi), \\ z_n = n, \end{cases}$$

drawn on the (random) revolution surface

$$|W|^2 z^{2\sigma_2} = x^2 + y^2,$$

when n tends to infinity.



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

$\to$ a particle of type $j+1$ when $j \leq m-2$

$\to$ m particles of type 1 when $j = m-1$.

Call $0 = \tau_0 < \tau_1 < \cdots < \tau_n < \cdots$ the successive jumping times. The arguments on the exponential distribution are the same ones as for binary search trees embedding. Considering the process of *gaps* instead of nodes, it is easy to see that $\tau_n - \tau_{n-1}$ is $\mathcal{E}xp(u+n-1)$-distributed, where $u = \sum_{k=1}^{m-1} k X^{CT}(0)^{(k)}$ is the numbers of gaps at time 0.

The embedding principle can be expressed

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

and as for BST, we consider that both processes are built on the same probability space, so that this equality holds almost surely. For this multitype branching process, it is classical to see that

Proposition 3.9

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

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 3.10 *([5], Janson [12])*

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

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.

The geometric interpretation with a random curve on a spiral can be done like in discrete time. Nonetheless, notice the random first term in the expansion of $X^{CT}(t)$.

3.4 Asymptotics

3.4.1 Notations

In the following, we denote

$$T = \tau_{(1)} + \cdots + \tau_{(m-1)}. \tag{6}$$

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. By induction on m, let us prove that T has

$$f_T(u) = (m-1)e^{-u}(1 - e^{-u})^{m-2} \mathbf{1}_{\mathbb{R}_+}(u), \quad u \in \mathbb{R}, \tag{7}$$

as a density. Indeed, it is true for $m = 2$; when X and Y have f_X and f_Y as densities respectively, then the convolution formula gives that $Z = X + Y$ has f_Z as a density, where

$$f_Z(z) = \int_0^z f_X(z-y) f_Y(y) dy.$$

Consequently, taking $X = T$ having f_T given by (7), and $Y = \tau_{(m)}$ having $f_Y(y) = me^{-my}$ as a density, we get

$$\begin{aligned} f_Z(z) &= \int_0^z (m-1)e^{-(z-y)}\left(1-e^{-(z-y)}\right)^{m-2} me^{-my}dy && (8)\\ &= m(m-1)e^{-z}\int_0^z e^{-y}\left(e^{-y}-e^{-z}\right)^{m-2}dy && (9)\\ &= m(m-1)e^{-z}\left[-\frac{e^{-y}-e^{-z}}{m-1}\right]_0^z && (10)\\ &= me^{-z}(1-e^{-z})^{m-1}. && (11)\end{aligned}$$

We deduce from (7) that e^{-T} has a Beta distribution with parameters 1 and $m-1$. A straightforward change of variable ($x = e^{-u}$) under the integral shows that for any complex number λ such that $\Re(\lambda) > -1$,

$$\begin{aligned} \mathbb{E}e^{-\lambda T} &= \int_0^{+\infty} e^{-\lambda u} f_T(u)du = (m-1)B(1+\lambda, m-1) && (12)\\ &= \frac{(m-1)!}{\prod_{k=1}^{m-1}(\lambda+k)}, && (13)\end{aligned}$$

where B denotes Euler's Beta function:

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

In particular,

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

3.4.2 Dislocation equations

We would like a complete description of the $\mathbb{C}$-valued random variable W^{CT}. It is a limit of a branching process after projection and scaling, remember that

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

Let us see now how the branching property applied at the first splitting time provides fixed point equations on the limit distributions.

Let us write dislocation equations for the continuous time branching process at finite time t. We write $X_j(t)$ for $X^{CT}(t)$ when the process starts from $X^{CT}(0) = e_j$, where e_j denotes the j-th vector of the canonical basis of $\mathbb{R}^{m-1}$ (whose j-th component is 1 and all the others are 0). This means that the process starts from an ancestor of type j.

Notice that the distribution of the first splitting time τ_1 depends on the ancestor's type; denote by $\tau_{(j)}, j = 1, \dots, m-1$, the first splitting time when the process starts from $X(0) = e_j$. Thus $\tau_{(j)}$ is $\mathcal{E}xp(j)$ distributed.

The branching property applied at the first splitting time gives:

$$\forall t > \tau_1, \quad \begin{cases} X_1(t) \overset{\mathcal{L}}{=} X_2(t - \tau_{(1)}), \\ X_2(t) \overset{\mathcal{L}}{=} X_3(t - \tau_{(2)}), \\ \dots \\ X_{m-2}(t) \overset{\mathcal{L}}{=} X_{m-1}(t - \tau_{(m-2)}), \\ X_{m-1}(t) \overset{\mathcal{L}}{=} [m]X_1(t - \tau_{(m-1)}), \end{cases} \tag{16}$$

where the notation $[m]X$ denotes the sum of m independent copies of the random variable X.

After projections of variables $X_j(t)$ with u_2, scaling with $e^{-\lambda_2 t}$ and taking the limit when t goes to infinity, we get the variables

$$W_j := \lim_{t \to +\infty} e^{-\lambda_2 t} u_2(X_j(t)),$$

so that the system (16) on $X_j(t)$ leads to the following system of distributional equations on W_j:

$$\begin{cases} W_1 \overset{\mathcal{L}}{=} e^{-\lambda_2 \tau_{(1)}} W_2, \\ W_2 \overset{\mathcal{L}}{=} e^{-\lambda_2 \tau_{(2)}} W_3, \\ \dots \\ W_{m-2} \overset{\mathcal{L}}{=} e^{-\lambda_2 \tau_{(m-2)}} W_{m-1}, \\ W_{m-1} \overset{\mathcal{L}}{=} e^{-\lambda_2 \tau_{(m-1)}} [m] W_1. \end{cases} \tag{17}$$

Since W_1 is the distribution of W^{CT} starting from a particle of type 1 (which is indeed the case for the m-ary search tree), this shows that W_1 is a solution of the following fixed point equation:

$$Z \overset{\mathcal{L}}{=} e^{-\lambda_2 T}(Z^{(1)} + \cdots + Z^{(m)}), \tag{18}$$

where T is defined in (6) and where $Z^{(i)}$ are independent copies of Z, 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.

In terms of the Fourier transform

$$\varphi(t) := \mathbb{E}\exp\{i\langle t, Z\rangle\} = \mathbb{E}\exp\{i\Re(\bar{t}Z)\}, \quad t \in \mathbb{C},$$

where $\langle x, y\rangle = \Re(\overline{x}y) = \Re(x)\Re(y) + \Im(x)\Im(y)$, equation (18) reads

$$\varphi(t) = \int_0^{+\infty} \varphi^m(te^{-\overline{\lambda_2}u}) f_T(u)du, \quad t \in \mathbb{C}, \tag{19}$$

where f_T is defined by (7). Notice that this functional equation can also be written in a convolution form: if $\Phi(t) := \varphi(e^{\overline{\lambda_2}t})$ for any $t \in \mathbb{C}$, then Φ satisfies the following functional equation:

$$\Phi(t) = \int_0^{+\infty} \Phi^m(t-u) f_T(u)du, \quad t \in \mathbb{C}. \tag{20}$$

4 Smoothing transformation

In this section, inspired from the case of m-ary search trees (see [5]), the following fixed point equation coming from the previous multitype branching process is studied, thanks to several methods. These methods are general ones, they are used for other distributional equations. Let us just mention analogous results for

- binary search trees, where the quicksort distribution is studied in Rösler [18];
- Pólya urns where the limit distribution occuring for large urns is studied in [8, 6].

The following smoothing equation comes from m-ary search trees, studied in Section 3.

$$W \overset{\mathcal{L}}{=} e^{-\lambda T}(W^{(1)} + \cdots + W^{(m)}), \tag{21}$$

where $\lambda \in \mathbb{C}$, T is defined in (6), $W^{(i)}$ are $\mathbb{C}$-valued independent copies of W, which are also independent of T. We successively see

• the contraction method, in order to prove existence and unicity of a solution, in a suitable space of probability measure, in Section 4.1;

• some analysis on the Fourier transforms in order to prove that W has a density, in Section 4.2;

• a cascade type martingale which is a key tool to obtain the existence of exponential moments for W, in Section 4.3.

oral Mixing in the fixed point equation, spiral given by $e^{-\lambda T}$.

4.1 Contraction method

This method has been developed in Rösler [18] and Rösler and Rüschendorf [19] for many examples in analysis of algorithms. The idea is to get existence and unicity of a solution of equation (21) thanks to the Banach fixed point Theorem. Notice that we already have the existence, thanks to Section 3. 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.

$$K\mu := \mathcal{L}\big(e^{-\lambda T}(X^{(1)} + \cdots + X^{(m)})\big), \tag{22}$$

where T is given by (6), $X^{(i)}$ 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 point is to be sure that K maps $\mathcal{M}_2(C)$ into itself, this is given by the following lemma.

Lemma 4.11 *If λ is a root of the characteristic polynomial (5) such that $\Re(\lambda) > -\frac{1}{2}$ and if C is any complex number, then K maps $\mathcal{M}_2(C)$ into itself.*

Proof. Since $\Re(\lambda) > -1$, the random variable $e^{-\lambda T}$ has an expectation. See Section 3.4.1. Furthermore, by (12), $m\mathbb{E}e^{-\lambda T} = 1$ as λ is a root of (5). This ensures the conservation of the expectation by K. Since $\Re(\lambda) > -\frac{1}{2}$, then $\mathbb{E}|e^{-\lambda T}|^2 < \infty$ and $K\mu$ admits a second absolute moment whenever μ does. Therefore $K\mu \in \mathcal{M}_2(C)$ whenever $\mu \in \mathcal{M}_2(C)$. □

Now, define d_2 as the Wasserstein distance on $\mathcal{M}_2(C)$ (see for instance Dudley [10]): for $\mu, \nu \in \mathcal{M}_2(C)$,

$$d_2(\mu, \nu) = \left(\min_{(X,Y)} \mathbb{E}\left(|X - Y|^2\right)\right)^{\frac{1}{2}}, \tag{23}$$

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 [10], 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 $(X_i,Y_i), i = 1,\dots,m$ 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 (7), 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

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

With Equation (15), we know that $m\mathbb{E}\left|e^{-2\lambda T}\right| < 1 \iff \Re(\lambda) > \dfrac{1}{2}$, which happens for a large urn. Therefore K is a contraction on $\mathcal{M}_2(C)$. We have proved the following theorem.

Theorem 4.12 *Let $\lambda \in \mathbb{C}$ be a root of the characteristic polynomial (5) 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 (21) has a unique solution W in $\mathcal{M}_2(C)$.*

4.2 Analysis on Fourier transforms

The aim is to prove that W solution of equation (21) 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 Liu [13, 14] adapted in [5] for $\mathbb{C}$-valued variables. It runs along the following lines.

Let φ be the Fourier transform of any solution W of (21). It is a solution of the functional equation

$$\varphi(t) = \int_0^{+\infty} \varphi^m(te^{-\bar{\lambda}u}) f_T(u)du, \quad t \in \mathbb{C}, \tag{24}$$

where f_T is defined by (7). **oral** Calculation on the blackboard

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 $L^1(\mathbb{R})$.

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 (24) 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 4.13 *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 $\mathbb{C}$.

4.3 Cascade type martingales

The distributional equation (21) suggests to use Mandelbrot's cascades in the complex setting (see Barral [2] for independent interest about complex Mandelbrot's cascades).

As in Section 3, take $\lambda \in \mathbb{C}$ be a root of the characteristic polynomial (5) 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 (5) 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

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

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

$$Y_n = \sum_{u_1 ... u_{n-1} \in \{1,\dots,m\}^{n-1}} mAA_{u_1}A_{u_1u_2} \dots A_{u_1 ... u_{n-1}}. \tag{25}$$

As $m\mathbb{E}A = 1$, $(Y_n)_n$ is a martingale, with expectation 1.

oral Detail the proof, which is immediate.

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 [14] and the references therein. It can be easily seen that

$$Y_{n+1} = A \sum_{i=1}^{m} Y_{n,i} \tag{26}$$

where $Y_{n,i}$ for $1 \leq i \leq m$ are independent of each other and independent of A and each has the same distribution as Y_n.

oral Detail the proof, which is immediate. Notice both arguments, forward and backward. In the variance calculation, be careful with complex numbers.

Therefore for $n \geq 1$, 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 (see Theorem 2.14 in Mailler's course) 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

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

Notice that, passing to the limit in (26) gives a new proof of the existence of a solution W of Eq. (21) 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 4.14 *Let $\lambda \in \mathbb{C}$ be a root of the characteristic polynomial (5) with $\Re(\lambda) > 1/2$ and let W be a solution of Equation (21). There exist some constants $C > 0$ and $\varepsilon > 0$ such that for all $t \in \mathbb{C}$ with $|t| \leq \varepsilon$,*

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

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, Q. Liu, and N. Pouyanne. Limit distributions for multitype branching processes of m-ary search trees. *Ann. Inst. Henri Poincaré*, to appear, 2013.

[6] B. Chauvin, C. Mailler, and N. Pouyanne. Smoothing equations for large Pólya urns. *Submitted JTP*, 2013.

[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] 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.

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

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

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

[13] 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.

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

[15] H.M. Mahmoud. *Evolution of Random Search Trees*. John Wiley & Sons Inc., 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. *Ann. Inst. Henri Poincaré*, 44(2):293–323, 2008.

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

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