# Martingales, Embedding and Tilting of Binary Trees

B. Chauvin , T. Klein , J-F. Marckert , A. Rouault *

**Abstract**

We are interested in the asymptotic analysis of the binary search tree (BST) under the random permutation model. Two methods are mainly used: the first one is the embedding in continuous time and the second one is the tilting probability method. Combining both gives a commutative scheme between four models:

$$\begin{array}{ccc} \text{BST} & \xrightarrow{\textit{embedding}} & \text{Yule tree / fragmentation} \\ \downarrow \textit{tilting} & & \downarrow \textit{tilting} \\ \text{tilted BST} & \xrightarrow{\textit{embedding}} & \text{tilted Yule tree / tilted fragmentation} \end{array}$$

In this paper we focus on the upper embedding arrow and on the tilting arrows. We thus get new results on the BST and also new proofs of known results. In particular, thanks to the left tilting arrow, we give a conceptual proof (in the sense of Lyons, Pemantle, Peres) of the asymptotic behavior of the profile.

**Key words.** Binary search tree, fragmentation, branching random walk, probability tilting, convergence of martingales, Chinese restaurant.

**A.M.S. Classification.** `60J25, 60J80, 68W40, 60C05, 60G42, 60G44.`

# 1 Introduction

## 1.1 The model of binary search trees

For a convenient definition of trees we are going to work with, let first define

$$\mathbb{U} = \maltese \cup \bigcup_{n \geq 1} \{0,1\}^n$$

as the set of finite words on the alphabet $\{0,1\}$ (with $\maltese$ as an empty word). For $u$ and $v$ in $\mathbb{U}$, denote by $uv$ the concatenation of the word $u$ with the word $v$ (by convention we set, for any $u \in \mathbb{U}$, $\maltese u = u$). If $v \neq \maltese$, we say that $uv$ is a descendant of $u$ and $u$ is an ancestor of $uv$. Moreover $u0$ (resp. $u1$) is called left (resp. right) child of $u$.

A *complete binary tree* $T$ is a finite subset of $\mathbb{U}$ such that

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

*LAMA, Bâtiment Fermat, Université de Versailles F-78035 Versailles. e-mail: Name@math.uvsq.fr

The elements of $T$ are called *nodes* , and ✠ is called the *root* ; $|u|$, the number of letters in $u$, is the *depth* of $u$ (with $|✠| = 0$). Write **BinTree** for the set of complete binary trees.

A tree $T \in$ **BinTree** can be described by giving the set $\partial T$ of its *leaves*, that is, the nodes that are in $T$ but with no descendants in $T$. The nodes of $T \backslash \partial T$ are called *internal* nodes[1].

We study binary search trees (BST), which are widely used to store totally ordered data (the monograph of Mahmoud [28] gives an overview of the state of the art).

Let $A$ be a totally ordered set of elements named keys and for $n \geq 1$, let $(a_1, ..., a_n)$ be picked up without replacement from $A$. The (labeled) binary search tree built from these data is a complete binary tree in which each internal node is associated with a key belonging to $(a_1, ..., a_n)$ in the following way:

The first key $a_1$ is assigned to the root. The next key $a_2$ is assigned to the left child of the root if it is smaller than $a_1$, or it is assigned to the right child of the root if it is larger than $a_1$. We proceed further inserting key by key recursively. We get a labeled complete binary tree with $n$ internal nodes such that the keys of the left subtree of any given node $u$ are smaller than the key of $u$, and the keys of the right subtree are larger than the key of $u$.

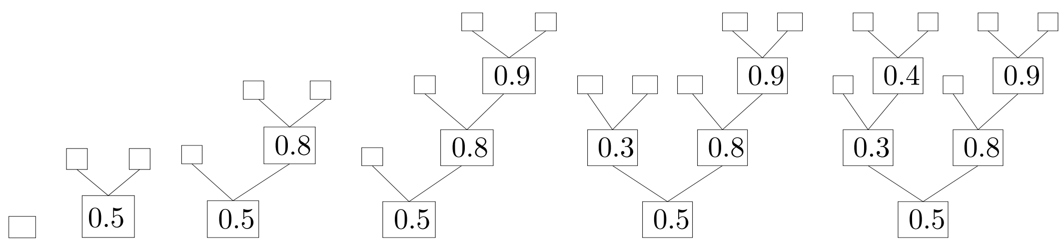


Figure 1: BST built with the sequence of data 0.5, 0.8, 0.9, 0.3, 0.4 (empty squares are leaves).

To study the shape of these trees for large $n$, it is classical to introduce a random model. One usually assumes that the data $(x_i)_{i \geq 1}$ successively inserted are i.i.d. random variables with a continuous distribution $F$. For every $n \geq 1$, the string $x_1, .., x_n$ induces (a.s.) a permutation $\sigma_n$ such that $x_{\sigma_n(1)} < x_{\sigma_n(2)} < \cdots < x_{\sigma_n(n)}$. Since the $x_i$ are exchangeable, $\sigma_n$ is uniformly distributed on the set $\mathcal{S}_n$ of permutations of $\{1, .., n\}$. Since this claim is not sensitive to $F$ we will assume, for the sake of simplicity that $F$ is the uniform distribution on $[0, 1]$. This is the so-called random permutation model.

Again by exchangeability, $\sigma_n$ is independent of the vector $(x_{\sigma_n(1)}, x_{\sigma_n(2)}, \ldots, x_{\sigma_n(n)})$ and we have

$$\begin{aligned} P\big(x_{n+1} \in (x_{\sigma_n}(j), x_{\sigma_n}(j+1)) \,|\, \sigma_n\big) &= P\big(x_{n+1} \in (x_{\sigma_n}(j), x_{\sigma_n}(j+1))\big) \\ &= P(\sigma_{n+1}(j+1) = n+1) = (n+1)^{-1} \end{aligned}$$

for every $j = 0, 1, .., n$, where $x_{\sigma_n(0)} := 0$ and $x_{\sigma_n(n+1)} := 1$. One can also express this property with the help of the sequential ranks of the permutation: the random variables $R_k = \sum_{j=1}^{k} 1\!\!1_{x_j \leq x_k}, k \geq 1$ are independent and $R_k$ is uniform on $\{1, \ldots, k\}$ (see for instance Mahmoud [28], section 2.3) , so that $P(R_{n+1} = j+1 \,|\, R_1, .., R_n) = (n+1)^{-1}$.

[1]Some authors consider non complete binary trees, removing the third condition in the above definition. The boundary is then the set of nodes that are not in $T$ but whose predecessors are in $T$ ([6]). It can be seen as a set of external (or available nodes). Here, we choose to work with complete trees, but this choice has no impact on the results.

In term of binary search tree, this is translated by the fact that the insertion of the $n+1$st key in the tree with $n$ internal nodes is uniform among its $n+1$ leaves.

In this model, the law of the sequence of the underlying (unlabelled) trees is a Markov chain $(\mathcal{T}_n, n \geq 0)$ on **BinTree** defined by $\mathcal{T}_0 = \{\maltese\}$ and

$$\begin{aligned} \mathcal{T}_{n+1} &= \mathcal{T}_n \cup \{D_{n+1}0, D_{n+1}1\}\,, \\ P(D_{n+1} = u \,|\, \mathcal{T}_n) &= (n+1)^{-1}, \qquad u \in \partial\mathcal{T}_n\,, \end{aligned} \tag{1}$$

($D_{n+1}$ is the random node where the $n+1$-st key is inserted). It is a particular case ($\alpha = 1$) of the diffusion-limited aggregation (DLA) on a binary tree, where a constant $\alpha$ is given and the growing of the tree is random with probability of insertion at a leaf $u$ proportional to $\alpha^{-|u|}$ (Aldous-Shields [1], Barlow-Pemantle-Perkins [6]).

To describe the evolution of the BST, two important random variables are the saturation level $h_n$ and the height $H_n$:

$$h_n = \min\{|u| : u \in \partial\mathcal{T}_n\} \quad , \quad H_n = \max\{|u| : u \in \partial\mathcal{T}_n\} \tag{2}$$

which grow logarithmically (see for instance Devroye [14] )

$$\text{a.s.} \quad \lim_{n\to\infty} \frac{h_n}{\log n} = c' = 0.3733... \qquad \lim_{n\to\infty} \frac{H_n}{\log n} = c = 4.31107...\,, \tag{3}$$

where $c'$ and $c$ are the two solutions of the equation $\eta_2(x) = 1$ where

$$\eta_\lambda(x) := x \log \frac{x}{\lambda} - x + \lambda, \qquad x \geq 0\,, \tag{4}$$

is the Cramer transform of the Poisson distribution of parameter $\lambda$. Function $\eta_2$ has its minimum at $x = 2$. It corresponds to the rate of propagation of the insertion depth: $\frac{D_n}{2\log n} \xrightarrow{P} 1$.

A more accurate information on $\mathcal{T}_n$ is provided by the whole profile

$$U_k(n) := \#\{u \in \partial\mathcal{T}_n, |u| = k\} \quad , \quad k \geq 1\,, \tag{5}$$

counting the number of leaves of $\mathcal{T}_n$ at each level. Notice that $U_k(n) = 0$ for $k > H_n$ and for $k < h_n$. To get asymptotic results, it is rather natural to code the profile thanks to the so-called polynomial level $\sum_k U_k(n)z^k$, whose degree is $H_n$.

For $z \notin \frac{1}{2}\mathbb{Z}^- = \{0, -1/2, -1, -3/2, \cdots\}$ let

$$\mathcal{M}_n(z) = \frac{1}{C_n(z)} \sum_{k\geq 0} U_k(n) z^k = \frac{1}{C_n(z)} \sum_{u\in\partial\mathcal{T}_n} z^{|u|} \qquad , \; n \geq 0\,, \tag{6}$$

where

$$C_n(z) = \prod_{k=0}^{n-1} \frac{k+2z}{k+1} = (-1)^n \binom{-2z}{n}, \qquad n \geq 1, \qquad C_0(z) = 1\,, \tag{7}$$

and let $\mathcal{F}_{(n)}$ be the $\sigma$-field generated by all the events $\{u \in \mathcal{T}_j\}_{j\leq n, u\in\mathbb{U}}$. Jabbour [13, 21] proved that $(\mathcal{M}_n(z), \mathcal{F}_{(n)})_n$ is a martingale to which, for the sake of simplicity we refer from now as the BST martingale. If $z > 0$, this positive martingale is a.s. convergent; the limit $\mathcal{M}_\infty(z)$ is positive a.s. if $z \in (z_c^-, z_c^+)$, with

$$z_c^- = c'/2 = 0.186..., \quad z_c^+ = c/2 = 2.155... \tag{8}$$

and $\mathcal{M}_\infty(z) = 0$ for $z \notin [z_c^-, z_c^+]$ (Jabbour [21]). This martingale is also the main tool to prove that the limit profile has a Gaussian shape (see Theorem 1 in [21]).

## 1.2 Embedding of BST in a continuous time model

The aim of the present paper is to revisit the study of this family of martingales, improving results (in the critical case, on the uniformity of convergence), using either the embedding method or the tilting probability method. It allows to get more complete results on the profile of BSTs.

The idea of embedding discrete models (such as urn models) in continuous time branching processes goes back at least to Athreya-Karlin [4]. It is described for instance in Athreya and Ney ([5], section 9) and it has been recently revisited by Janson [22]. For the BST, various embeddings are mentioned in Devroye [14], in particular those due to Pittel [32], and Biggins [10, 11]. Here, we work with a variant of the Yule process, taking into account the tree (or "genealogical") structure.

Let $(u_t)_{t\geq 0}$ be a Poisson point process taking values in $\mathbb{U}$ with intensity measure $\nu_{\mathbb{U}}$, the counting measure on $\mathbb{U}$. Let $(\mathbb{T}_t)_{t\geq 0}$ be a **BinTree** valued process such that $\mathbb{T}_0 = \{\maltese\}$ and $\mathbb{T}_.$ jumps only when $u_.$ jumps. Let $t$ be a jump time for $u_.$; $\mathbb{T}_t$ is obtained from $\mathbb{T}_{t-}$ in the following way:

if $u_t \notin \partial\mathbb{T}_{t-}$ keep $\mathbb{T}_t = \mathbb{T}_{t-}$ and if $u_t \in \partial\mathbb{T}_{t-}$ take $\mathbb{T}_t = \mathbb{T}_{t-} \cup \{u_t 0, u_t 1\}$.
The counting process $(N_t)_{t\geq 0}$ defined by

$$N_t := \#\partial\mathbb{T}_t \tag{9}$$

is the classical Yule (or binary fission) process (Athreya-Ney [5]). In the following, we refer to the continuous-time tree process $(\mathbb{T}_t)_{t\geq 0}$ as the Yule tree process.

We note $0 = \tau_0 < \tau_1 < \tau_2 < ...$ the successive jump times (of $\mathbb{T}_.$),

$$\tau_n = \inf\{t : N_t = n+1\}\,. \tag{10}$$

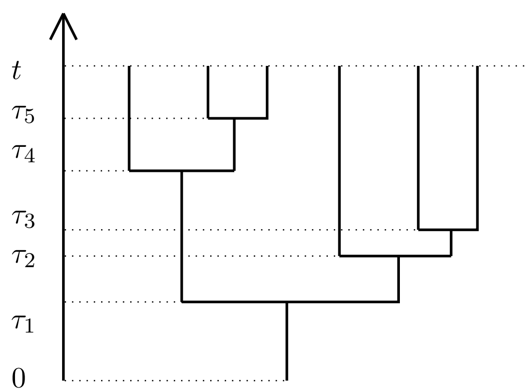


Figure 2: Continuous time binary branching process.

## 1.3 Yule process and fragmentation process

This Yule tree process can also be seen as a fragmentation process. We may encode dyadic open subintervals of $[0,1]$ with elements of $\mathbb{U}$. We set $I_{\maltese} = (0,1)$ and for $u = u_1u_2...u_k \in \mathbb{U}$,

$$I_u = \Big(\sum_{j=1}^{k} u_j 2^{-j}, 2^{-k} + \sum_{j=1}^{k} u_j 2^{-j}\Big).$$

With this coding, the evolution corresponding to the previous process is a very simple example of fragmentation process. This idea goes back to Aldous and Shields ([1] Section 7f and 7g).

In other words, for $t \geq 0$, $F(t)$ is a finite family of intervals. At time 0, we have $F_0 = (0,1)$. Identically independent exponential $\mathcal{E}(1)$ random variables[2] are associated with each intervals of $F(t)$. Each interval in $F(t)$ splits into two parts (with same size) independently of each other after an exponential time $\mathcal{E}(1)$.

Hence, one has $F(0) = (0,1)$, $F(\tau_1) = ((0,1/2),(1/2,1))$ where $\tau_1 \sim \mathcal{E}(1)$, etc... One can interpret the two fragments $I_{u0}$ and $I_{u1}$ issued from $I_u$ as the two children of $I_u$, one being the left (resp. right) fragment $I_{u0}$ (resp. $I_{u1}$), obtaining thus a tree structure. With this interpretation, one observes that when $n$ fragments are present, each of them will split first equally likely. An interval with length $2^{-k}$ corresponds to a leaf at depth $k$ in the corresponding tree structure.

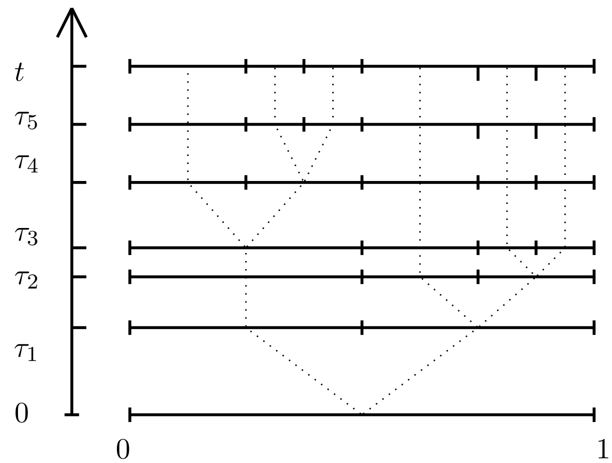


Figure 3: Fragmentation and its tree representation.

The following proposition allows to build on the same probability space, the Yule tree process and the BST. This observation was also made in Aldous-Shields [1] section 1, (see also Kingman [23] p.237 and Tavaré [35] p.164 in other contexts).

**Lemma 1.1** *a) The jump time intervals $(\tau_n - \tau_{n-1})_n$ are independent and satisfy:*

$$\tau_n - \tau_{n-1} \sim \mathcal{E}(n) \text{ for any } n \geq 1. \tag{11}$$

*b) The processes $(\tau_n)_{n\geq 1}$ and $\left(\mathbb{T}_{\tau_n}\right)_{n\geq 1}$ are independent.*

*c)* **(embedding)**

$$\left(\mathbb{T}_{\tau_n}\right)_{n\geq 1} \stackrel{\mathcal{D}}{=} \left(\mathcal{T}_n\right)_{n\geq 1} \tag{12}$$

*where $\stackrel{\mathcal{D}}{=}$ means equality in distribution.*

**Proof:** a) and b) are direct consequences of the properties of Poisson processes: a) comes from the definition of the intensity measure, and b) from the independence of jump chain and jump times. c) is clear since the evolution rules of the two Markov chains are the same in both models.■

A first easy and useful consequence of a) is

$$\mathbb{E}(e^{\tau_n(1-2z)}) = C_n(z)^{-1}\,. \tag{13}$$

[2] $\mathcal{E}(\lambda)$ is the exponential distribution of parameter $\lambda$, $\mathcal{U}([0,1])$ is the uniform distribution on $[0,1]$, and $\mathbf{Be}(p)$ is the Bernoulli distribution of parameter $p$.

If we consider only the size of the fragments, the Yule tree process can be seen as a particular case of branching random walk in continuous time: individuals have an $\mathcal{E}(1)$ distributed lifetime, and at their death, they produce children, whose relative positions are distributed according to a point process $Z$. Individuals do not move during their lives. If we denote the set of individuals alive at time $t$ by $\mathcal{Z}_t$ and for $u \in \mathcal{Z}_t$ the position of individual $u$ by $X_u$, then the classical family of "additive" martingales, parameterized by $\theta$ in $\mathbb{R}$ (sometimes in $\mathbb{C}$) and indexed by $t \geq 0$ is given by

$$m(t,\theta) := \sum_{u \in \mathcal{Z}_t} \exp(\theta X_u - tL(\theta)),$$

where $L(\theta) = E\int e^{\theta x} Z(dx) - 1$ (see [36], [25], and [8] for the fragmentation).

Here, we have $Z = 2\delta_{-\log 2}$, $\mathcal{Z}_t = \partial\mathbb{T}_t$ and $X_u = -|u|\log 2$. For easier use, we set $z = 2^{-\theta}$ and then consider the family of martingales

$$M(t,z) := \sum_{u \in \mathcal{Z}_t} z^{|u|} e^{t(1-2z)}. \tag{14}$$

In particular $M(t,1/2) = 1$ and $M(t,1) = e^{-t}N_t$. A classical result (see Athreya-Ney [5] or Devroye [14] 5.4) says that

$$\xi := \lim_{t\to\infty} e^{-t}N_t \ \sim \mathcal{E}(1)\,. \tag{15}$$

Taking again the (very) particular case $z = 1$, we remark that since $\lim_n \tau_n = \infty$ a.s. (see Lemma 1.1 a) ) we get from (15)

$$\lim_n ne^{-\tau_n} = \xi \quad \text{a.s.}\,. \tag{16}$$

The definitions of the martingales together with the embedding Lemma 1.1 c) give:

**Proposition 1.2 (martingale connection)** *For* $z \notin \frac{1}{2}\mathbb{Z}^-$

$$M(\tau_n, z) = e^{\tau_n(1-2z)} C_n(z) \mathcal{M}_n(z)\,, \tag{17}$$

*where* $\tau_n$ *is independent of* $\mathcal{M}_n(z)$*.*

This connection allows us to transfer known results about the Yule martingales to BST martingales, thus giving a very simple proof of known results (such that in Theorem 2.1 below) about the BST martingale and also getting much more. In particular, in Theorem 2.4 2), we give the answer to the question asked in [21], about critical values of $z$, with a straightforward argument.

### 1.4 Tiltings of the models

We introduce now (and develop in Section 3) the tilting or biasing method which allows us to interpret the martingales as Radon-Nikodym derivatives. In order to do that, we need to "enlarge" the probability space [8, 12, 26]. Roughly speaking it consists in marking at random a special "ray" or branch of the tree, called spine, both in the discrete and in the continuous case. It allows to deduce important properties of the population from the behavior of the spine.

For the fragmentation process $(F(t))_{t\geq 0}$ let us denote by $\mathcal{F}_t$ the $\sigma$-algebra of the interval fragmentation up to time $t$ and $V$ be a $\mathcal{U}([0,1])$ r.v. independent of the filtration $(\mathcal{F}_t)_{t\geq 0}$. Since $\mathbb{P}(V \in \{k2^{-j}, 0 \leq k \leq 2^j, j \in \mathbb{N}, k \in \mathbb{N}\}) = 0$, we may define $\mathbb{P}$-a.s. for every $t$ a unique $S(t) \in \mathbb{U}$

such that $I_{S(t)}$ is an interval of $F(t)$ and $V \in I_{S(t)}$. In other words, $S(t)$ is the element of $\mathbb{U}$ coding the fragment containing $V$, its length is $2^{-|S(t)|}$ and

$$\mathbb{P}(S(t) = u \mid \mathcal{F}_t) = 2^{-|u|} \ , \ \ u \in \partial \mathbb{T}_t \tag{18}$$

(we choose a fragment at random with probability equals to its length, it is the classical size-biasing setting). As a consequence of the general theory of homogeneous fragmentations (see Bertoin [7]) or by a direct computation, we see that $(|S(t)|, t \geq 0)$ is an homogeneous Poisson process with parameter 1. In particular, if

$$\mathcal{E}(t,z) := (2z)^{|S(t)|} e^{t(1-2z)} \tag{19}$$

then $\mathbb{E}\mathcal{E}(t,z) = 1$. Conditionally on $\widehat{\mathcal{F}}_s = \mathcal{F}_s \vee \sigma(S(r), r \leq s)$, the restriction of the fragmentation $F(.+s)$ to the interval $I_{S(s)}$ is distributed as a rescaling of $F(.)$ by a factor $2^{-|S(s)|}$, which entails that $\left(\mathcal{E}(t,z), \widehat{\mathcal{F}}_t\right)_{t\geq 0}$ is a martingale. By the size biasing scheme (18) and the definition (14) we get

$$M(t,z) = \mathbb{E}\left[\mathcal{E}(t,z) \mid \mathcal{F}_t\right]. \tag{20}$$

Coming back to the discrete time, let $\mathcal{F}_{(n)}$ be the $\sigma$-algebra generated by $F(\tau_1), ..., F(\tau_n)$ and let us denote $\mathbf{Spine}_n := S(\tau_n)$ and $s_n := |\mathbf{Spine}_n|$. Applying (18) at the $(\mathcal{F}_t, t \geq 0)$ stopping time $\tau_n$, we get for every leaf $u \in \partial \mathcal{T}_n$ (and $k \geq 1$) :

$$\mathbb{P}(\mathbf{Spine}_n = u \mid \mathcal{F}_{(n)}) = 2^{-|u|} \quad , \quad \mathbb{P}(s_n = k \mid \mathcal{F}_{(n)}) = U_k(n) 2^{-|k|} \ . \tag{21}$$

Let $\widehat{\mathcal{F}}_{(0)}$ be the trivial $\sigma$-algebra, and for $n \geq 1$ let $\widehat{\mathcal{F}}_{(n)}$ be the $\sigma$-algebra obtained from $\mathcal{F}_{(n)}$ by adjunction of $S(\tau_1), ..., S(\tau_n)$. Let us consider $\mathcal{E}_n(z) := \mathbb{E}\Big[\mathcal{E}(\tau_n, z) \mid \widehat{\mathcal{F}}_{(n)}\Big]$ (with $\mathcal{E}_0(z) := 1$). From (13) we have

$$\mathcal{E}_n(z) = (2z)^{s_n}\, C_n(z)^{-1}. \tag{22}$$

From the martingale property of $\mathcal{E}(t,z)$ and the definition of $\mathcal{E}_n(z)$ we see that $\left(\mathcal{E}_n(z), \widehat{\mathcal{F}}_{(n)}\right)$ is a martingale. Like in (20), we get easily

$$\mathcal{M}_n(z) = \mathbb{E}\left[\mathcal{E}_n(z) \mid \mathcal{F}_{(n)}\right], \tag{23}$$

so that the martingales $M(t,z)$ and $\mathcal{M}_n(z)$ are obtained from the "exponential martingales" $\mathcal{E}(z,t)$ and $\mathcal{E}_n(z)$ by projection. All these martingales are precisely the main tool to tilt probabilities. In particular we define $\mathbb{P}^{(2z)}$ on $(\widehat{\mathcal{F}}_t, t \geq 0)$ by

$$\mathbb{P}^{(2z)}_{|_{\widehat{\mathcal{F}}_t}} = \mathcal{E}(t,z)\ \mathbb{P}_{|_{\widehat{\mathcal{F}}_t}}\ , \tag{24}$$

which yields by projection on $(\mathcal{F}_t, t \geq 0)$

$$\mathbb{P}^{(2z)}_{|_{\mathcal{F}_t}} = M(t,z)\ \mathbb{P}_{|_{\mathcal{F}_t}}\ . \tag{25}$$

If ${}^d\mathbb{P}$ (resp. ${}^d\mathbb{P}^{(2z)}$) is the restriction of $\mathbb{P}$ (resp. $\mathbb{P}^{(2z)}$) to $\vee_n \widehat{\mathcal{F}}_{(n)}$, the discrete versions of the above relations are

$${}^d\mathbb{P}^{(2z)}_{|_{\widehat{\mathcal{F}}_{(n)}}} = \mathcal{E}_n(z)\ {}^d\mathbb{P}_{|_{\widehat{\mathcal{F}}_{(n)}}} \quad , \quad {}^d\mathbb{P}^{(2z)}_{|_{\mathcal{F}_{(n)}}} = \mathcal{M}_n(z)\ {}^d\mathbb{P}_{|_{\mathcal{F}_{(n)}}}\ . \tag{26}$$

The probabilities $\mathbb{P}^{(2z)}$ that are given above will have a representation (or an interpretation) further in the paper. In one word, one can say that under $\mathbb{P}^{(2z)}$ the evolution of the fragmentation (or the the one of the Yule tree or BST tree) is biaised. The parameter $z$ serves for a speed-tuner of the spine ($z > 1/2$ corresponds to a speed up and $z < 1/2$ to a slow down).

Let us now explain the content of the paper. In Section 2, we explore the direct consequences of embedding. First, we exhibit (on the same space of probability) a family of uniform r.v. attached to the nodes of the tree. These random variables give, for every node $u$, the limiting proportion of leaves issued from $u$ among those issued from its parent. They are exactly the r.v. called "fictitious" by Devroye in [14] p. 258 in its "backward" construction of a typical realization of $\mathcal{T}_n$ for $n$ fixed. In a second part of Section 2, we study the convergence of the BST martingale $\mathcal{M}_n(z)$. For $z > 0$, the embedding method allows to recover very quickly the behavior of the limit $\mathcal{M}_\infty(z)$: positive when $z \in (z_c^-, z_c^+)$, zero when $z \notin [z_c^-, z_c^+]$. In the critical case $z = z_c^\pm$ the behavior was unknown. We prove that $\mathcal{M}_\infty(z_c^\pm) = 0$ a.s. and get the convergence of the derivative. We also give a strong version of the "quicksort" equation.

In Section 3, we define the biased models (continuous time and discrete time). After enlarging the space, we prove that the growing tree process can be decomposed into a spine evolution together with the evolution of subtrees issued from nodes of the spine. We follow the way initiated by Lyons, Pemantle, Peres ([26],[27]) and followed by many other authors ([3],[12],[6]). In this study, we use several times the Chinese restaurant model of Dubins and Pitman ([31] p.58). In Section 4, we explore the benefits of the tilting method. In a first part, we revisit the behavior of the martingales $\mathcal{M}_n(z)$, giving "conceptual" proofs. In a second part, thanks to this method, we are able to describe the asymptotic behavior of the profile $U_k(n)$ when $k = 2z\log n + o(\sqrt{\log n})$ in the whole range $z \in (z_c^-, z_c^+)$, providing large deviations results around $k = 2\log n$. Previously, the result was known only on a subdomain due to a $L^2$ method ([13]).

# 2 Some benefits of the embedding method

Let us begin with the study of some meaningful random variables arising as a.s limits and playing an important role in the results of Subsection 2.2.

## 2.1 Uniform r.v. in the BST

For every $u \in \mathbb{U}$, let $\tau^{(u)} = \inf\{t : u \in \mathbb{T}_t\}$ the time (a.s. finite) at which $u$ appears in the tree, and for $t > 0$, let

$$\mathbb{T}_t^{(u)} = \{v \in \mathbb{U} : uv \in \mathbb{T}_{t+\tau^{(u)}}\}$$

the tree process growing from $u$. In particular we denote

$$N_t^{(u)} = \#\partial\mathbb{T}_t^{(u)}\,.$$

For $t > \tau^{(u)}$, the number of leaves at time $t$ in the subtree issued from node $u$ is $n_t^{(u)} := N_{t-\tau^{(u)}}^{(u)}$. The branching property and (15) give that a.s. for every $u \in \mathbb{U}$

$$\lim_{t\to\infty} e^{-t}N_t^{(u)} = \xi_u \quad , \quad \lim_{t\to\infty} e^{-t}n_t^{(u)} = \xi_u\, e^{-\tau^{(u)}}\,, \tag{27}$$

where $\xi_u$ is distributed as $\xi$ i.e. $\mathcal{E}(1)$. Moreover, if $u$ and $v$ are not in the same line of descent, the r.v. $\xi_u$ and $\xi_v$ are independent. Since

$$n_t^{(u)} = n_t^{(u0)} + n_t^{(u1)} \quad \text{and} \quad \tau^{(u0)} = \tau^{(u1)}, \tag{28}$$

a small computation yields

$$\frac{n_t^{(u0)}}{n_t^{(u)}} \xrightarrow{a.s.} U^{(u0)} := \frac{\xi_{u0}}{\xi_{u0}+\xi_{u1}}, \qquad \frac{n_t^{(u1)}}{n_t^{(u)}} \xrightarrow{a.s.} U^{(u1)} := 1 - U^{(u0)} = \frac{\xi_{u1}}{\xi_{u0}+\xi_{u1}}, \tag{29}$$

which allows to attach a $\mathcal{U}([0,1])$ r.v. to each node of $\mathbb{U}$. In particular we set

$$U := U^{(0)} = \frac{\xi_0}{\xi_0+\xi_1} \tag{30}$$

so that

$$\xi := \xi_{\maltese} = e^{-\tau_1}(\xi_0+\xi_1) \ \ , \ \ \xi_0 = U\xi e^{\tau_1} \ \ , \ \ \xi_1 = (1-U)\xi e^{\tau_1}\,. \tag{31}$$

If $u0$ and $u1$ are brother nodes, we have $U^{(u1)} + U^{(u0)} = 1$. We claim that if the finite set of nodes $v_1, \dots, v_k$ does not contain any pair of brothers, the corresponding r.v. $U^{(v_1)}, \dots, U^{(v_k)}$ are independent. When none of the $v_j$ is an ancestor of another ("stopping line" property) it is a consequence of the branching property. In the general case, it is sufficient to prove that $U^{(u)}$ is independent of $(U^{(v)}, v < u)$. To simplify the reading, let us give the details only for $|u| = 2$, for instance $u = 00$. We have, from (28)

$$U^{(00)} = \frac{\xi_{00}}{\xi_{00}+\xi_{01}} \ \ , \ \ U^{(0)} = \frac{(\xi_{00}+\xi_{01})e^{-\tau^{(00)}+\tau^{(0)}}}{(\xi_{00}+\xi_{01})e^{-\tau^{(00)}+\tau^{(0)}} + (\xi_{10}+\xi_{11})e^{-\tau^{(10)}+\tau^{(1)}}}$$

Actually, from the branching property, $\xi_{00}$ and $\xi_{01}$ are independent of $\xi_{10}, \xi_{11}, \tau^{(00)}, \tau^{(0)}, \tau^{(10)}, \tau^{(1)}$. Moreover since $\xi_{00}$ and $\xi_{01}$ are independent and $\mathcal{E}(1)$ distributed, then $\xi_{00}/(\xi_{00}+\xi_{01})$ and $(\xi_{00}+\xi_{01})$ are independent, which allows to conclude that $U^{(00)}$ and $U^{(0)}$ are independent.

Finally, multiplying along the line of the ancestors of a node $u$, we get the representation

$$a.s. \lim_{t\to\infty} \frac{n_t^{(u)}}{N_t} = \prod_{v<u} U^{(v)}\,. \tag{32}$$

Notice that relation (32) gives a strong (which means a.s.) version of the analogy between BST and branching random walks, first given by Devroye [14].

## 2.2 Martingales

For both models a family of martingales plays an essential role: the discrete-time martingale (6) in the BST, and the continuous time "additive martingale" (14) in the Yule tree. They are closely related by the martingale connection of Proposition 1.2. Thus, the embedding method is the key tool for proving and enlarging convergence results on the BST martingale (Theorem 2.4) and its derivative (Theorem 2.5).

### 2.2.1 Known results

The following theorem gives a summary of the main properties of the BST martingale, proved in [21] and [13].

**Theorem 2.1** *1) For $z \in (0,\infty)$, the positive martingale $\mathcal{M}_n(z)$ is a.s. convergent when $n \to \infty$ and the limit denoted $\mathcal{M}_\infty(z)$ satisfies*

$$\mathbb{E}\big(e^{-\theta \mathcal{M}_\infty(z)}\big) = \int_0^1 \mathbb{E}\big(e^{-\theta z x^{2z-1}\mathcal{M}_\infty(z)}\big)\mathbb{E}\big(e^{-\theta z(1-x)^{2z-1}\mathcal{M}_\infty(z)}\big)dx\,; \tag{33}$$

*2) a) for $z \in (z_c^-, z_c^+)$ there exists $p > 1$ such that the $L^p$ convergence holds, and*

$$\mathcal{M}_\infty(z) > 0 \quad a.s.\,,$$

*b) for $z \notin [z_c^-, z_c^+]$*

$$\mathcal{M}_\infty(z) = 0 \ \ a.s.$$

*3) On every compact of $\{z \in \mathbb{C} : |z-1| < \frac{\sqrt{2}}{2}\}$ , $\mathcal{M}_n(z)$ and all its $z$-derivative are a.s. uniformly convergent as $n \to \infty$.*

As a consequence of known results for the branching random walks ([8, 9, 36]), we have for the additive martingale:

**Theorem 2.2** *1) For $z \in (z_c^-, z_c^+)$, the positive martingale $M(t,z)$ is a.s. convergent when $t \to \infty$ and the limit denoted by $M(\infty,z)$ satisfies*

$$M(\infty,z) = ze^{(1-2z)\tau_1}\left(M_0(\infty,z) + M_1(\infty,z)\right) \quad a.s. \tag{34}$$

*where $M_0(\infty,z)$ and $M_1(\infty,z)$ are independent, distributed as $M(\infty,z)$ and independent of $\tau_1$.*
*2) a) For $z \in (z_c^-, z_c^+)$ there exists $p > 1$ such that the $L^p$ convergence holds, and*

$$M(\infty,z) > 0 \quad a.s.\,.$$

*b) For $z \in (0,\infty) \setminus (z_c^-, z_c^+)$, then $M(\infty,z) = 0$ a.s..*

Notice that the zero limit at the critical points $z_c^-$ and $z_c^+$ is known in the continuous-time case and not in the discrete-time case.

The derivative

$$M'(t,z) = \frac{d}{dz}M(t,z)$$

is a martingale which is no more positive. It is called the derivative martingale. Its behavior is ruled by the following theorem.

**Theorem 2.3** *1) For $z \in (z_c^-, z_c^+)$, the derivative martingale is convergent a.s. when $n \to \infty$.*

*2) a) For $z = z_c^-$, the derivative martingale is convergent a.s. to a finite positive limit $M'(\infty, z_c^-)$ and $\mathbb{E}\big(M'(\infty, z_c^-)\big) = +\infty$.*

*b) For $z = z_c^+$, the derivative martingale is convergent a.s. to a finite negative limit $M'(\infty, z_c^+)$ and $\mathbb{E}\big(M'(\infty, z_c^+)\big) = -\infty$.*

### 2.2.2 New results

**Theorem 2.4** *1) For $z \in (0,\infty)$ we have a.s.*

*a)* **(limit martingale connection)**

$$M(\infty,z) = \frac{\xi^{2z-1}}{\Gamma(2z)}\ \mathcal{M}_\infty(z)\,, \tag{35}$$

*where $\xi \sim \mathcal{E}(1)$ is defined in (15), and independent of $\mathcal{M}_\infty(z)$.*

*b)*

$$\mathcal{M}_\infty(z) = z\left(U^{2z-1}\mathcal{M}_{\infty,(0)}(z) + (1-U)^{2z-1}\mathcal{M}_{\infty,(1)}(z)\right) \tag{36}$$

*where $U \sim \mathcal{U}([0,1])$ is defined in (29), $\mathcal{M}_{\infty,(0)}(z), \mathcal{M}_{\infty,(1)}(z)$ are independent (and independent of $U$) and distributed as $\mathcal{M}_\infty(z)$.*

*2) For $z = z_c^\pm$, $\mathcal{M}_\infty(z) = 0$ a.s.*

The results on the derivative martingales

$$\mathcal{M}'_n(z) = \frac{d}{dz}\mathcal{M}_n(z)$$

are given in the following theorem, where $\Psi$ the digamma function is defined by

$$\Psi(x) = \frac{\Gamma'(x)}{\Gamma(x)} = \lim_n \left(\log n - \sum_{j=0}^{n-1}\frac{1}{x+j}\right). \tag{37}$$

**Theorem 2.5** *1) For $z \in [z_c^-, z_c^+]$, $\mathcal{M}'_n(z)$ converges a.s. and its limit $\mathcal{M}'_\infty(z)$ is related to $\mathcal{M}_\infty(z)$ and $M'(\infty,z)$ by*

$$M'(\infty,z) = \frac{\xi^{2z-1}}{\Gamma(2z)}\left(\mathcal{M}'_\infty(z) + 2\left(\log\xi - \Psi(2z)\right)\mathcal{M}_\infty(z)\right) \qquad a.s. \tag{38}$$

*where $\xi \sim \mathcal{E}(1)$ is defined in (15) and is independent of $\mathcal{M}_\infty(z)$ and $\mathcal{M}'_\infty(z)$. Moreover $\mathcal{M}'_\infty(z)$ satisfies a.s.*

$$\begin{aligned}\mathcal{M}'_\infty(z) &= zU^{2z-1}\mathcal{M}'_{\infty,(0)}(z) + z(1-U)^{2z-1}\mathcal{M}'_{\infty,(1)}(z)\\ &+ 2z\left(U^{2z-1}\log U\right)\mathcal{M}_{\infty,(0)}(z) + 2z\left((1-U)^{2z-1}\log(1-U)\right)\mathcal{M}_{\infty,(1)}(z)\\ &+ z^{-1}\mathcal{M}_\infty(z)\end{aligned} \tag{39}$$

*where $U \sim \mathcal{U}([0,1])$ is defined in (29), and the r.v. $\mathcal{M}'_{\infty,(0)}(z)$ and $\mathcal{M}'_{\infty,(1)}(z)$ are independent (and independent of $U$) and distributed as $\mathcal{M}'_\infty(z)$.*

*2) a) $\mathcal{M}'_\infty(z_c^-) > 0$ and $\mathcal{M}'_\infty(z_c^+) < 0$ a.s.*

*b) $\mathbb{E}\big(\mathcal{M}'_\infty(z_c^-)\big) = -\mathbb{E}\big(\mathcal{M}'_\infty(z_c^+)\big) = +\infty$*

*c) For $z = z_c^\pm$, $\mathcal{M}'_\infty(z)$ satisfies the same equation as in Theorem 2.4 b)*

$$\mathcal{M}'_\infty(z) = z\Big(U^{2z-1}\mathcal{M}'_{\infty,(0)}(z) + (1-U)^{2z-1}\mathcal{M}'_{\infty,(1)}(z)\Big) \quad a.s.\,, \tag{40}$$

*where $U$, $\mathcal{M}'_{\infty,(0)}(z), \mathcal{M}'_{\infty,(1)}(z)$ are as above. Moreover*

$$M'(\infty,z) = \frac{\xi^{2z-1}}{\Gamma(2z)}\ \mathcal{M}'_\infty(z)\,. \tag{41}$$

An easy and remarkable consequence of Theorem 2.5 1) is obtained in the following corollary, just taking $z = 1$ in (38) and (39) (remember that $\mathcal{M}_n(1) \equiv 1$). The distribution version (weaker) of (43) below is the subject of a broad literature (see for instance Fill, Janson, Devroye, Neininger, Rösler, Rüschendorf [18, 19, 15, 29, 34, 33]) and some properties of the distribution of $\mathcal{M}'_\infty(1)$ remain unknown.

**Corollary 2.6** *We have*

$$M'(\infty, 1) = \xi\left(\mathcal{M}'_\infty(1) + 2(\log\xi + \gamma - 1)\right) \quad a.s.\,, \tag{42}$$

*where $\gamma$ is the Euler constant, and $\mathcal{M}'_\infty(1)$ satisfies the a.s. version of the* **quicksort** *equation:*

$$\mathcal{M}'_\infty(1) = U\mathcal{M}'_{\infty,(0)}(1) + (1-U)\mathcal{M}'_{\infty,(1)}(1) + 2U\log U + 2(1-U)\log(1-U) + 1\,, \tag{43}$$

*where as above, $\mathcal{M}'_{\infty,(0)}(1)$ and $\mathcal{M}'_{\infty,(1)}(1)$ are independent (and independent of $U$), distributed as $\mathcal{M}'_\infty(1)$ and $U \sim \mathcal{U}([0,1])$.*

The following proposition gives an answer to a natural question asked in [13]: what is the optimal domain *in the complex plane* where the BST martingale is $L^1$−convergent and uniformly convergent? Notice that for $z \in \mathbb{R}$, the notations coincide with those of [21].

**Theorem 2.7** *Let*

$$f(z, q) := 1 + q(2\Re z - 1) - 2|z|^q\ .$$

*Let $\mathcal{V}_q = \{z : f(z,q) > 0\}$ and $\mathcal{V} := \cup_{1<q<2}\mathcal{V}_q$.*

*As $n \to \infty$, $\{\mathcal{M}_n(z)\}$ converges, a.s. and in $L^1$, uniformly on every compact $C$ of $\mathcal{V}$.*

#### 2.2.3 Proofs

In this section we use several times the following lemma.

**Lemma 2.8** *For $z \notin \frac{1}{2}\mathbb{Z}^-$ we have*

$$a) \qquad C_n(z) \sim \frac{n^{2z-1}}{\Gamma(2z)}\,, \tag{44}$$

$$b) \qquad a.s.\ \lim_n e^{\tau_n(1-2z)}C_n(z) = \frac{\xi^{2z-1}}{\Gamma(2z)}. \tag{45}$$

$$c) \qquad a.s.\ \lim_n \left[\frac{C'_n(z)}{C_n(z)} - 2\tau_n\right] = 2[-\Psi(2z) + \log\xi]\,. \tag{46}$$

**Proof:** a) Use Stirling formula.

b) By (44) and (16) we get $\lim_n e^{\tau_n(1-2z)}n^{2z-1} = \xi^{2z-1}$, a.s..

c) Use

$$\frac{C'_n(z)}{C_n(z)} = \sum_{j=0}^{n-1}\frac{2}{j+2z}\,,$$

(16) and (37). ■

**Proof of Theorem 2.4**:

1) a) Since $M(t,z)$ converges a.s. when $t \to \infty$, and since $\lim_n \tau_n = \infty$ a.s. we have $\lim_n M(\tau_n, z) = M(\infty, z)$. It remains to apply the martingale connection Proposition 1.2 and Lemma 2.8.

b) For $t > \tau_1$ we have the decomposition

$$M(t,z) = ze^{(1-2z)\tau_1}\left[M^{(0)}(t-\tau_1, z) + M^{(1)}(t-\tau_1, z)\right] \tag{47}$$

where for $i = 0, 1$

$$M^{(i)}(s,z) = \sum_{u \in \partial \mathbb{T}_s^{(i)}} z^{|u|} e^{s(1-2z)},$$

and $\mathbb{T}^{(i)}$ is defined in Section 2.1. Take $t = \tau_n$ in (47), condition on the first split time $\tau_1$, apply the branching property, let $n \to \infty$ and apply the limit martingale connection (35) to get

$$\frac{\xi^{2z-1}}{\Gamma(2z)}\mathcal{M}_\infty(z) = ze^{(1-2z)\tau_1}\left(\frac{\xi_0^{2z-1}}{\Gamma(2z)}\mathcal{M}_{\infty,(0)}(z) + \frac{\xi_1^{2z-1}}{\Gamma(2z)}\mathcal{M}_{\infty,(1)}(z)\right) \tag{48}$$

where $\xi_0$ and $\xi_1$ come from section 2.1, which yields b) with the help of (31).

2) The result for critical points comes directly from the limit martingale connection (35) and the analogous known result in continuous time in Theorem 2.2 2) b). ■

**Proof of Theorem 2.5**: 1) Taking derivatives in the martingale connection (Proposition 1.2) gives

$$M'(\tau_n, z) = \left[\frac{C_n'(z)}{C_n(z)} - 2\tau_n\right] e^{\tau_n(1-2z)} C_n(z)\mathcal{M}_n(z) + e^{\tau_n(1-2z)} C_n(z)\mathcal{M}_n'(z). \tag{49}$$

For $z \in [z_c^-, z_c^+]$, let $n \to \infty$ in (49). From Lemma 2.8 and known results in continuous time (Theorem 2.3) , we get that $\mathcal{M}'(\infty)$ satisfies (38).

To prove (39), we differentiate (47) with respect to $z$

$$M'(t,z) = (z^{-1} - 2\tau_1)M(t,z) + ze^{(1-2z)\tau_1}\left[M^{(0)\prime}(t-\tau_1, z) + M^{(1)\prime}(t-\tau_1, z)\right],$$

and we use the same technique as above: take $t = \tau_n$, let $n \to \infty$, apply (38) and its analogs with $(M'^{(i)}, \mathcal{M}^{(i)}, \mathcal{M}'^{(i)}, \xi_i)_{i=0,1}$ instead of $(M', \mathcal{M}, \mathcal{M}', \xi)$, and use (31).

For $z = z_c^\pm$, 2) a) and 2) b) are consequences of Theorem 2.3 2), since $\mathcal{M}_\infty(z_c^\pm) = 0$.

Formula (40) of 2) c) is straightforward from (39) since $\mathcal{M}_\infty(z_c^\pm) = 0$. Formula (41) is (38) for $z = z_c^\pm$. ■

**Proof of Theorem 2.7**: Uniform convergence of martingales in the continuous time BRW has been studied by Biggins [9] Theorem 6. See also Bertoin-Rouault [8].

It is possible to give a proof of the uniform a.s. convergence of $\mathcal{M}_n$ directly from these papers. Actually, for the uniform $L^1$ convergence, we will prove

$$\lim_N \sup_{n \geq N} \mathbb{E} \sup_{z \in C} |\mathcal{M}_n(z) - \mathcal{M}_N(z)| = 0\,. \tag{50}$$

Since $(\sup_{z\in C}|\mathcal{M}_n(z)-\mathcal{M}_N(z)|)_{n\geq N}$ is a submartingale, this will imply also the a.s. uniform convergence. From (13) and the martingale connection (Proposition 1.2), we have

$$\mathcal{M}_n(z)-\mathcal{M}_N(z)=\mathbb{E}[M(\tau_n,z)-M(\tau_N,z)|\mathcal{F}_{(n)}]$$

so that taking supremum and expectation we get

$$\mathbb{E}\sup_{z\in C}|\mathcal{M}_n(z)-\mathcal{M}_N(z)|\leq\mathbb{E}\left(\sup_{z\in C}|M(\tau_n,z)-M(\tau_N,z)|\right).$$

Taking again the supremum in $n$ we get

$$\begin{aligned}\sup_{n\geq N}\mathbb{E}\sup_{z\in C}|\mathcal{M}_n(z)-\mathcal{M}_N(z)| &\leq \mathbb{E}\sup_{n\geq N}\left(\sup_{z\in C}|M(\tau_n,z)-M(\tau_N,z)|\right)\\ &\leq \mathbb{E}\sup_{T\geq\tau_N}\left(\sup_{z\in C}|M(T,z)-M(\tau_N,z)|\right). \qquad (51)\end{aligned}$$

We have for any $t>0$

$$\begin{aligned}\mathbb{E}\sup_{T\geq\tau_N}\left(\sup_{z\in C}|M(T,z)-M(\tau_N,z)|\right) &\leq \mathbb{E}\sup_{T\geq t}\left(\sup_{z\in C}|M(T,z)-M(t,z)|\right)\\ &+ \mathbb{E}\left[\mathbb{1}_{\tau_N<t}\sup_{T\geq\tau_N}\left(\sup_{z\in C}|M(T,z)-M(\tau_N,z)|\right)\right]\\ &:= I(t)+J_N(t) \qquad (52)\end{aligned}$$

On the one hand, $\lim_t I(t)=0$ by [8] Proposition 3. On the other hand, the Hölder inequality yields (for $q\in(1,2)$ such that $\inf_{z\in C}f(z,q)>0$)

$$J_N(t)\leq 2\left(\mathbb{E}\left(\sup_{T\geq 0}\sup_{z\in C}|M(T,z)|^q\right)\right)^{1/q}(\mathbb{P}(\tau_N<t))^{(q-1)/q}. \qquad (53)$$

For any fixed $t$, $\lim_N\mathbb{P}(\tau_N<t)=0$ which allows to end the proof. ∎

# 3 Biased models and tilting probability

In this Section, we construct an enlarged probability space and we describe the tools (spine evolution, Chinese restaurant) which will give the key arguments in the proof of Theorems 4.2 and 4.7 of Section 4.

## 3.1 Construction of biased trees

We call *marked tree* a tree with a distinguished leaf. More precisely let

$$\mathbf{MBinTree}:=\{(T,u);T\in\mathbf{BinTree},u\in\partial T\}.$$

If $\widetilde{T}=(T,u)$ is a marked tree, we say that $u$ is the red leaf of $T$, that $\{v\in\partial T,v\neq u\}$ is the set of *blue* leaves of $T$, and that ancestors of $u$ are red nodes, and other internal nodes are blue. We denote by $\partial\widetilde{T}$ the set of leaves of $T$ with their colors.

Let $z > 0$ be a parameter. We define on $\mathbb{U} \times \{red, blue\} \times \{0,1\}$ a Poisson point process $\tilde{u}_t = (u_t, c_t, \epsilon_t)$ with intensity measure

$$\nu_{\mathbb{U}} \otimes \{2z\delta_{red} + \delta_{blue}\} \otimes \{\frac{1}{2}\delta_0 + \frac{1}{2}\delta_1\}$$

and we denote by $\mathbb{Q}^{(2z)}$ its law. Let us construct continuous time process $(\widetilde{\mathbb{T}}_t)_{t\geq 0}$ with values in **MBinTree** which starts from

$$\widetilde{\mathbb{T}}_0 = (✠, ✠),$$

such that $\widetilde{\mathbb{T}}.$ jumps only when $\tilde{u}.$ jumps. Let $t$ be a jump time for $\tilde{u}.$; $\widetilde{\mathbb{T}}_t$ is obtained from $\widetilde{\mathbb{T}}_{t-}$ in the following way:

– if $(u_t, c_t) \notin \partial\widetilde{\mathbb{T}}_t$, then $\widetilde{\mathbb{T}}_t = \widetilde{\mathbb{T}}_{t-}$

– if $(u_t, c_t) \in \partial\widetilde{\mathbb{T}}_t$, then the new tree is $\mathbb{T}_t = \mathbb{T}_{t-} \cup \{u_t0, u_t1\}$ and its colors are given by

• if $c_t = blue$, the blue leaf $u_t$ becomes a blue node and two (new) blue leaves $u_t0, u_t1$ appear.

• if $c_t = red$, the red leaf $u_t$ becomes a red node, two new leaves appear: $u_t\epsilon_t$ which is red and $u_t(1-\epsilon_t)$ which is blue.

One notes again the successive jump times $(\tilde{\tau}_n)_n$. Once again, $(\tilde{\tau}_n - \tilde{\tau}_{n-1}), n \geq 1$ are independent and exponentially distributed, and

$$\tilde{\tau}_n - \tilde{\tau}_{n-1} \sim \mathcal{E}(n-1+2z)\,.$$

It is clear by construction that the set of red nodes is a branch ; the red branch is called **Spine** and $\mathbf{Spine}_n$ is the red leaf of $\widetilde{\mathbb{T}}_{\tilde{\tau}_n}$. Its length is $s_n = |\mathbf{Spine}_n|$ .

In terms of the second construction of the Yule process, we have now two kinds of nodes (blue and red). With each blue node $u$ is associated a clock $\mathcal{E}(1)$, and at its death it gives two blue nodes $u0$ and $u1$. With each red node $u$ is associated a clock $\mathcal{E}(2z)$ and at its death it gives two nodes a red one $u0$ (resp. $u1$) and a blue one $u1$ (resp. u0) with probability 1/2 (resp. 1/2). Ancestor is assumed red.

We can also see this process as a branching process with immigration, as presented in [35] (see also [31] chap. 10 and [17]). The spine is a Poisson process of rate $2z$ and at each jump time of this process begins a new Poisson process of rate 1 independent of the spine process and independent of all other Poisson processes already running.

One can again define a discrete time process

$$\widetilde{\mathcal{T}}_n = \widetilde{\mathbb{T}}_{\tilde{\tau}_n}$$

containing all the tree structure (and the color) of $\widetilde{\mathbb{T}}_{\tau_n}$. The discrete evolution is as follows:
$\widetilde{\mathcal{T}}_n$ is a complete binary tree with $2n+1$ nodes, $n$ blue leaves and one red one. To construct $\widetilde{\mathcal{T}}_{n+1}$ we choose the red leaf with probability $2z/(n+2z)$ and each blue one with probability $1/(n+2z)$.

– If the chosen leaf $v$ is blue, $\widetilde{\mathcal{T}}_{n+1} = \widetilde{\mathcal{T}}_n \cup \{(v1, blue), (v0, blue)\}$.

– If the leaf chosen $v$ is red, we toss a fair coin. We put $\widetilde{\mathcal{T}}_{n+1} = \widetilde{\mathcal{T}}_n \cup \{(v1, blue), (v0, red)\}$, if the coin is heads and $\widetilde{\mathcal{T}}_{n+1} = \widetilde{\mathcal{T}}_n \cup \{(v0, blue), (v1, red)\}$ if it is tails.

## 3.2 Tilted probability

We use the change of measure defined in (26) that we now recall:

$$d\mathbb{P}^{(2z)}_{|_{\widehat{\mathcal{F}}_{(n)}}} = \frac{(2z)^{s_n}}{C_n(z)}\ d\mathbb{P}_{|_{\widehat{\mathcal{F}}_{(n)}}} \tag{54}$$

so that, in particular ${}^{d}\mathbb{P}^{(1)} = {}^{d}\mathbb{P}$. We often omit the superscript $d$ for simplicity when no confusion is possible. Proposition 3.1 below gives an intuitive interpretation of the change of probability done in (26).

**Proposition 3.1** *The law of $(\widetilde{\mathcal{T}}_n)_n$ under $\mathbb{Q}^{(2z)}$ is ${}^{d}\mathbb{P}^{(2z)}$. It is called a biased BST and for simplicity we denote by $\widetilde{\mathbb{P}} := {}^{d}\mathbb{P}^{(2z)}$ the biased probability.*

**Proof:** The dynamics we described above yields the following conditional probabilities.
For any blue leaf $v$, $(v, blue) \in \partial\widetilde{\mathcal{T}}_n$ and

$$\mathbb{Q}^{(2z)}(\mathbf{Spine}_{n+1} = \mathbf{Spine}_n, \widetilde{\mathcal{T}}_{n+1} = \widetilde{\mathcal{T}}_n \cup (v0, blue)(v1, blue)|\widetilde{\mathcal{T}}_n) = \frac{1}{n+2z}$$

For the red leaf, $\mathbf{Spine}_n = v$ and

$$\mathbb{Q}^{(2z)}(\mathbf{Spine}_{n+1} = v0, \widetilde{\mathcal{T}}_{n+1} = \widetilde{\mathcal{T}}_n \cup (v0, red)(v1, blue)|\widetilde{\mathcal{T}}_n) = \frac{1}{2}\frac{2z}{n+2z}$$

similarly,

$$\mathbb{Q}^{(2z)}(\mathbf{Spine}_{n+1} = v1, \widetilde{\mathcal{T}}_{n+1} = \widetilde{\mathcal{T}}_n \cup (v0, blue)(v1, red)|\widetilde{\mathcal{T}}_n) = \frac{1}{2}\frac{2z}{n+2z}\,.$$

Summing up, we have for any colored tree $\widetilde{t}_{n+1}$ with $n+1$ nodes that can be obtained from $\widetilde{\mathcal{T}}_n$ by one insertion

$$\mathbb{Q}^{(2z)}(\widetilde{\mathcal{T}}_{n+1} = \widetilde{t}_{n+1}|\widetilde{\mathcal{T}}_n) = \frac{z^{s_{n+1}-s_n}}{n+2z} \tag{55}$$

and

$$\mathbb{Q}^{(1)}(\widetilde{\mathcal{T}}_{n+1} = \widetilde{t}_{n+1}|\widetilde{\mathcal{T}}_n) = \frac{(1/2)^{s_{n+1}-s_n}}{n+1}.$$

It is clear that $\mathbb{Q}^{(2z)}/Q(1) = \prod_{j=0}^{n-1} \frac{(2z)^{s_{j+1}-s_j}(j+1)}{j+2z} = (2z)^{s_n}\, C_n(z)^{-1} = \mathcal{E}_n(z)$. Hence, $d\mathbb{Q}^{(2z)}$ is absolutely continuous with respect to $Q^{(1)}$, with the Radon-Nikodym derivative announced in (54) (see Lemma 1 and 2 in Biggins [12] for a detailed proof in another context). ■

### 3.3 Spine evolution

Thanks to the previous subsections, it appears that

$$s_n = 1 + \sum_{1}^{n-1} \epsilon_k \tag{56}$$

where $(\epsilon_k)_{k\geq 1}$ are independent and for every $k \geq 1$, $\epsilon_k \sim \mathbf{Be}(\frac{2z}{k+2z})$. In particular,

$$\begin{aligned} \mathbb{E}^{(2z)}(s_n) &= 1 + \sum_{1}^{n-1} \frac{2z}{k+2z} \\ \mathrm{Var}^{(2z)}\,(s_n) &= \sum_{1}^{n-1} \frac{2z}{k+2z} - \sum_{1}^{n-1}\left(\frac{2z}{k+2z}\right)^2. \end{aligned} \tag{57}$$

As $n \to \infty$

$$\mathbb{E}^{(2z)}(s_n) = 2z\log n - 2z\Psi(2z) + o(1)$$

$$\mathrm{Var}^{(2z)}\ (s_n) \quad = \quad 2z\log n - 2z\Psi(2z) - 4z^2\Psi'(2z) + o(1)\,. \tag{58}$$

We can now apply known results on sums of independent r.v. or notice that $s_n - \mathbb{E}^{(2z)}(s_n)$ is a martingale, to get the following asymptotic behavior (see [30]).

**Proposition 3.2** *For any parameter $z > 0$,*

*1) (strong law)*

$$\lim \frac{s_n}{\log n} = 2z\,, \qquad \mathbb{P}^{(2z)} - a.s.. \tag{59}$$

*2) (law of the iterated logarithm)* $\mathbb{P}^{(2z)}$*-a.s.*

$$\liminf \frac{s_n - 2z\log n}{2\sqrt{2z\log n\log\log\log n}} = -1\,, \qquad \limsup \frac{s_n - 2z\log n}{2\sqrt{2z\log n\log\log\log n}} = +1. \tag{60}$$

*3) (central limit theorem) The distribution of* $\dfrac{s_n - 2z\log n}{\sqrt{2z\log n}}$ *under* $\mathbb{P}^{(2z)}$ *converges to a standard normal distribution* $\mathcal{N}(0,1)$.

*4) (local limit theorem)*

$$\lim_n \sup_k \left|\sqrt{2\pi V_n}\ \mathbb{P}^{(2z)}(s_n = k) - \exp\Big(-\frac{(k-\mu_n)^2}{2V_n}\Big)\right| = 0 \tag{61}$$

*where* $\mu_n = \mathbb{E}^{(2z)}(s_n)$ *and* $V_n = Var^{(2z)}\ (s_n)$.

*5) (large deviations) The family of distributions of* $(s_n,\ n > 0)$ *under* $\mathbb{P}^{(2z)}$ *satisfies the large deviation principle on* $[0,\infty)$ *with speed* $\log n$ *and rate function* $\eta_{2z}$ *where the function* $\eta_\lambda$ *is defined in (4).*

We give more details in Section 4.2.

To study the growing of the biased BST away from the spine, we need to recall the Chinese restaurant model.

### 3.4 Chinese restaurant model (CRM)

Let $\theta > 0$ be a parameter. We recall here the Pitman $(0,\theta)$ Chinese restaurant model (see Pitman [31] p.58). An initially empty restaurant has an unlimited number of tables numbered 1,2,..., each capable of seating an unlimited number of customers. Customers 1,2,... arrive one by one and are seated according to the following:

Person 1 sits at table 1. For $n \geq 1$ assume that $n$ customers have already entered the restaurant, and are seated in some arrangement, with at least one customer at each of the table $j$, for $j$ from 1 to $k$, where $k$ is the number of tables occupied by the $n$ first customers to arrive. Let $A_j(n)$ the number of customers on table $j$ at time $n$. The $n+1$-st customer sits at table $j$ with probability $A_j(n)/(n+\theta)$ for any $j \leq k$. With probability $\theta/(\theta+n)$ , the $n+1$-st customer sits on the new table $k+1$.

The sequence

$$A(n) = (A_1(n), A_2(n), \dots)$$

is a Markov chain which describes the evolution of the occupation of the Chinese restaurant, we denote by $\mathbf{CR}^{(\theta)}$ its distribution.

Further we will take $\theta = 2z$. So, for $z > 1/2$ the creation of new tables is encouraged. This has to be compared with the speed-tuner of the spine.

### 3.5 Decomposition of the biased BST along the Spine

For every $n$, on $\widetilde{\mathcal{T}}_n$ there is a red branch. Each blue node and each blue leaf has some red ancestors. We class the blue leaves of $\widetilde{\mathcal{T}}_n$ according to their highest red ancestors; in other words, let $u_0, u_1, \dots, u_{s_n}$ be the set of red nodes (with $u_0 = \maltese$ and for $i \geq 1$, $u_i$ is the red child of $u_{i-1}$). We denote by

$$S_i(n) = \{u | u \text{ blue leaf of } \widetilde{\mathcal{T}}_n, u_i \text{ highest red ancestor of } u\}.$$

See Figure 4.

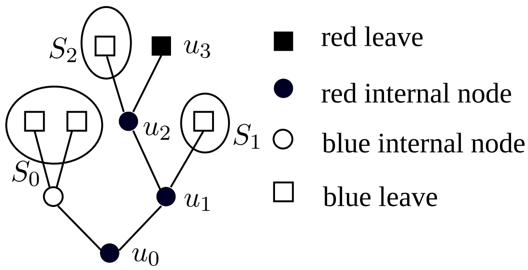


Figure 4: A marked tree and the different classes.

We denote by $|S_i(n)| = \operatorname{card} S_i(n)$ and

$$\mathbf{S}(n) = (|S_0(n)|, |S_1(n)|, \dots)$$

the sequence of classes sizes at time $n$. It satisfies

$$\mathbf{S}(0) = (0, 0, \dots), \quad \mathbf{S}(1) = (1, 0, 0, \dots),$$

and for any $n$,

$$\sum_{i=0}^{+\infty} |S_i(n)| = n.$$

Recall that $s_n$ is the height of the red leaf. At time $n$, the class $S_{s_n}$ is the "first empty class". $(\mathbf{S}(k))_{k\geq 0}$ is a Markov chain whose transition at time $n$ can be described as follows:
$a$) choose class $j$ with probability $|S_j(n)|/(n + 2z)$ and set

$$|S_i(n + 1)| = |S_i(n)| + \delta_{ij}$$

for $i = 0, 1, ...$ where $\delta$ is the Kronecker symbol.
$b$) choose the red leaf with probability $2z/(n + 2z)$ and set

$$\begin{aligned} |S_i(n+1)| &= |S_i(n)| \text{ for } i < s_n\,, \\ |S_{s_n}(n+1)| &= 1\,, \\ s_{n+1} &= s_n + 1\,. \end{aligned} \tag{62}$$

Thus

$$A_i(n) \overset{(d)}{=} |S_{i-1}(n)| \text{ for any } i \geq 1.$$

and we may assert

**Proposition 3.3** *Under $\mathbb{P}^{(2z)}$, the Markov chain $\boldsymbol{S}(n), n \geq 0$ is $\mathbf{CR}^{(2z)}$ distributed.*

In [2] p.52, a Chinese restaurant is also considered for the BST, but associated with the insertion node, which does not allow to keep track of the dynamics of the spine.

We now study some conditional evolutions. Let us denote by $\beta_i = \inf\{j \,|\, u_i \in \widetilde{\mathcal{T}}_j\}$ the birth date of node $u_i$; we have $\beta_0 = 0$, and, for any $l \geq 1$

$$\beta_l = \inf\left\{k \;\middle|\; s_k = l\right\}.$$

It is clear that

$$|S_i(n)| > 0 \iff n \geq \beta_{i+1}.$$

Since at time $\beta_{i+1}$ there are one red leaf, a unique blue leaf on $S_i$, and $\beta_{i+1} - 1$ blue leaves on others sub-trees, the evolution **conditionally on** $\beta_{i+1}$ is given by the following relations:

$$\left\{\begin{array}{rcl} |S_i(\beta_{i+1})| & = & 1, \\ |S_i(\beta_{i+1}+1)| & = & 1 + b_1 \text{ where } b_1 \sim \mathbf{Be}\Big(\dfrac{1}{\beta_{i+1}+2z}\Big) \\ \vdots & & \vdots \\ |S_i(\beta_{i+1}+k)| & = & c_{k-1} + b_k \text{ where } b_k \sim \mathbf{Be}\Big(\dfrac{c_{k-1}}{\beta_{i+1}+k-1+2z}\Big); \end{array}\right. \tag{63}$$

(conditionally on $|S_i(\beta_{i+1}+j)| = c_j, j = 1, \ldots, k-1$ ).

In other words, we have the following proposition.

**Proposition 3.4** *Conditionally on $\beta_{i+1}$, the distribution of $(|S_i(\beta_{i+1}+k)|, k \geq 0)$ under $\mathbb{P}^{(2z)}$ is the same as $(A_1(k), k \geq 0)$ under $\mathbf{CR}^{(2z+\beta_{i+1}-1)}$.*

Another decomposition will be useful in the rest of the paper. We use the notation $\widetilde{\mathbb{P}}$ for ${}^{d}\mathbb{P}^{(2z)}$, $\widetilde{\mathbb{E}}$ for the corresponding expectation, with a superscript if we take a conditional one. We denote by $\Xi = \{\beta_i, i \in \mathbb{N}\}$.

**Proposition 3.5** *Under $\widetilde{\mathbb{P}}^{\Xi}$, for $i$ fixed, the process $(S_i(n), n \geq \beta_{i+1})$ has the same distribution as $\mathcal{T}_{W_i(n)}$, that is a (non-biased) BST with $W_i(n)$ leaves where $(W_i(\beta_{i+1}+k), k \geq 0)$ is an inhomogeneous Markov chain on $\{1, 2, \ldots\}$ with initial state $W_i(\beta_{i+1}) = 1$ and the following transition rule:*

- *If $\beta_{i+1} + k \in \Xi$, then $W_i(\beta_{i+1}+k+1) = W_i(\beta_{i+1}+k)$*
- *If $\beta_{i+1} + k \notin \Xi$, then $W_i(\beta_{i+1}+k+1) - W_i(\beta_{i+1}+k) \sim \mathbf{Be}(\frac{W_i(\beta_{i+1}+k)}{\beta_{i+1}+k})$ .*

Notice that this evolution of $W_i$ does not give a contradiction with (63) since we are conditioning with respect to $\Xi$ which is richer than $\sigma(\beta_{i+1})$.

**Proof:** The fact that $W_i(n)$ is the distribution size of $S_i(n)$ is a consequence that at times $(\beta_k)_{k>i}$ a new class is created and so no new node arrives on $S_i$. It remains to show that knowing $W_i(n)$, $S_i(n)$ is distributed as BST with size $W_i(n)$. This comes from the growing rule of the subtree under $S_i$. Indeed, knowing that "a node arrives" in $S_i$, this node is inserted uniformly among the leaves already present on $S_i$, independently from the past. This growing rule is the same as in the classical BST. ■

# 4 Benefits of tilting

We use the method of the tilted probability to revisit the problem of convergence of the family $\mathcal{M}_n(z)$ (subsection 4.1) and the convergence of profile (subsection 4.2).

The method was initiated by R. Lyons ([26, 27] and developed in several papers involving branching processes or their generalizations ([24, 3, 12, 8]. The main idea consists in changing probability and studying under $\widetilde{\mathbb{P}}$ the spine evolution and the subtrees issued from nodes of the spine. It is a use to call this method "conceptual".

## 4.1 Conceptual proof of convergence of $(\mathcal{M}_n(z), n \geq 1)$

For every $z > 0$, $\mathcal{M}_n(z)$ is a positive $(\mathcal{F}_{(n)}, \mathbb{P})$ martingale and then converges $\mathbb{P}$-a.s. to $\mathcal{M}_\infty(z)$. The $L^1$ convergence is equivalent to $\int \mathcal{M}_\infty(z) d\mathbb{P} = 1$.

The main argument to decide on this convergence lies on the following lemma. It comes from a classical result of measure theory (the most frequently cited reference is [16] Th. 4.3.3, see also [3, 12]).

**Lemma 4.1** *Fix $z > 0$ and let $\bar{\mathcal{M}}(z) = \limsup_n \mathcal{M}_n(z)$ (notice that $\bar{\mathcal{M}}(z) = \mathcal{M}_\infty(z)$ $\mathbb{P}$-a.s.)*

*i) $\mathcal{M}_n(z)^{-1}$ is a $(\mathcal{F}_{(n)}, \widetilde{\mathbb{P}})$ martingale*

*ii) $\int \mathcal{M}_\infty(z) d\mathbb{P} = 1$ if and only if $\widetilde{\mathbb{P}}(\bar{\mathcal{M}}(z) < \infty) = 1$. In that case the two measures $\widetilde{\mathbb{P}}$ and $\mathbb{P}$ are absolutely continuous on $\mathcal{F}_{(\infty)} = \vee_n \mathcal{F}_{(n)}$ with density $\mathcal{M}_\infty(z)$.*

*iii) $\mathcal{M}_\infty(z) = 0$ $\mathbb{P}$-a.s. if and only if $\widetilde{\mathbb{P}}(\bar{\mathcal{M}}(z) = \infty) = 1$. In that case the two measures $\widetilde{\mathbb{P}}$ and $\mathbb{P}$ are are mutually singular on $\mathcal{F}_{(\infty)}$.*

Thanks to this dichotomy, we are now able to give alternative proofs of Theorem 2.1 and Theorem 2.4 2). For an easier reading we summarize these results in the following theorem.

**Theorem 4.2** *a) If $z \notin [z_c^-, z_c^+]$ (subcritical case), then $\mathbb{P}$-a.s. $\lim_n \mathcal{M}_n(z) = 0$.*

*b) If $z = z_c^\pm$ (critical case), then $\mathbb{P}$-a.s. $\lim_n \mathcal{M}_n(z) = 0$.*

*c) If $z \in (z_c^-, z_c^+)$ (supercritical), then $\mathbb{P}$-a.s. $\lim_n \mathcal{M}_n(z) > 0$.*

### 4.1.1 Subcritical and critical cases (proof of Theorem 4.2 a)b))

For any $z \geq 0$, we start from

$$\mathcal{M}_n(z) \geq \frac{z^{s_n}}{C_n(z)} \tag{64}$$

and we consider the right hand side under $\widetilde{\mathbb{P}}$. From (59) and (44) we get

$$\lim_n \frac{s_n \log z - \log C_n(z)}{\log n} = \eta_2(2z) - 1\,, \qquad \widetilde{\mathbb{P}} - a.s.$$

In the "subcritical case", i.e. $z \notin [z_c^-, z_c^+]$, we have $\eta_2(2z) > 1$ (see (4)) and

$$\lim_n \frac{z^{s_n}}{C_n(z)} = \infty \quad \widetilde{\mathbb{P}} - a.s.. \tag{65}$$

In the critical case, $\eta_2(2z) = 1$, we use directly in (64) the law of iterated logarithm (60) instead of the strong law (59). This yields in both cases $\bar{\mathcal{M}}(z) = \limsup_n \mathcal{M}_n(z) = +\infty$, $\widetilde{\mathbb{P}}$- a.s. and by Lemma (4.1) iii), $\mathcal{M}_\infty(z) = 0$, $\mathbb{P}$- a.s.

### 4.1.2 Supercritical case (proof of Theorem 4.2 c) for $z \in (1/2, z_c^+)$)

We will show that for $z \in (1/2, z_c^+)$

$$\liminf_n \mathcal{M}_n(z) < \infty\,,\ \widetilde{\mathbb{P}}\text{- a.s.}\,. \tag{66}$$

This is sufficient since by Lemma 4.1 i) $\mathcal{M}_n(z)^{-1}$ is $\widetilde{\mathbb{P}}$-a.s. convergent. Its limit is nonzero according to (66). It will imply that $\mathcal{M}_n(z)$ converges $\widetilde{\mathbb{P}}$- a.s. and allows to conclude with Lemma (4.1)ii). We stress that for technical reasons, we were able to reach this aim only for $z \in (1/2, z_c^+)$.

Consider $\sigma_n = \sigma(\beta_i \mathbb{1}_{\beta_i \le n}, i \ge 1)$ the $\sigma$-algebra containing the birth date of the red nodes (before time $n$). By Fatou's lemma, (66) is a consequence of the following result.

**Proposition 4.3** *For $z \in (1/2, z_c^+)$,*

$$\limsup_n \widetilde{\mathbb{E}}^{\sigma_n}(\mathcal{M}_n(z)) < +\infty \ \ \widetilde{\mathbb{P}}\text{- } a.s..$$

*Proof:* With the previous decomposition along the spine and by definition of $S_i(n)$, we may write

$$\sum_{u\in\partial\mathcal{T}_n} z^{|u|} = z^{s_n} + \sum_{i=0}^{s_n-1} \sum_{u\in S_i(n)} z^{|u|}$$

hence

$$\widetilde{\mathbb{E}}^{\sigma_n} \sum_{u\in\partial\mathcal{T}_n} z^{|u|} = z^{s_n} + \sum_{i=0}^{s_n-1} \widetilde{\mathbb{E}}^{\sigma_n} \sum_{u\in S_i(n)} z^{|u|}\,.$$

For every $i \le s_n - 1$, we have

$$\widetilde{\mathbb{E}}^{\sigma_n} \sum_{u\in S_i(n)} z^{|u|} = \widetilde{\mathbb{E}}^{\sigma_n} \widetilde{\mathbb{E}}^{\Xi,|S_i(n)|} \sum_{u\in S_i(n)} z^{|u|}\,.$$

From Proposition 3.5

$$\widetilde{\mathbb{E}}^{\Xi,|S_i(n)|} \sum_{u\in S_i(n)} z^{|u|} = z^i C_{|S_i(n)|}(z)\,,$$

hence

$$\widetilde{\mathbb{E}}^{\sigma_n} \sum_{u\in\partial\mathcal{T}_n} z^{|u|} = z^{s_n} + \sum_{i=0}^{s_n-1} z^i\, \widetilde{\mathbb{E}}^{\sigma_n} C_{|S_i(n)|}(z)\,. \tag{67}$$

The main problem is that, knowing $\sigma_n$, $|S_i(n)|$ is difficult to handle. Since

$$k \mapsto C_k(z) \text{ is decreasing for } z < 1/2 \text{ and } k \mapsto C_k(z) \text{ is non decreasing for } z \ge 1/2, \tag{68}$$

we introduce a new sequence of random variables that will bound $|S_i(n)|$ for the stochastic order.

Recall that if $X$ and $Y$ are two random variables, we say $X$ dominates $Y$ for the stochastic order if for any $x \in \mathbb{R}$, $\mathbb{P}(X \geq x) \geq \mathbb{P}(Y \geq x)$. It implies that for any increasing function $g$ we have $Eg(X) \geq Eg(Y)$.

From Proposition 3.5, the law of $|S_i(n)|$ conditionally on $\widetilde{\mathbb{P}}^{\Xi}$ is stochastically dominated by the law of $A_1(n-\beta_{i+1})$ under $\mathbf{CR}^{(\beta_{i+1})}$, and a fortiori by the law of $A_1(n)$.

Since $z \geq 1/2$, thanks to (68),

$$\widetilde{\mathbb{E}}^{\sigma_n}\left(C_{|S_i(n)|}(z)\right) \leq \mathbf{CR}^{\beta_{i+1}}(C_{A_1(n)}(z)) \,. \tag{69}$$

Following Barbour & al. [2] page 93 (equations (4.73), (4.74)),

$$\mathbf{CR}^{(\lambda)}(A_1(k) = a) = \frac{\lambda \lambda^{(k-a)}}{\lambda^{(k)}} \frac{(k-1)!}{(k-a)!}$$

where $x^{(n)} = x(x+1)\dots(x+n-1)$ (this is sometimes called the Polya distribution). In the sequel of this proof $2z = \theta$ so that $\theta^{(n)} = C_n(z)n!$. Finally, one obtains (denoting $\beta = \beta_{i+1}$)

$$\begin{aligned}\mathbf{CR}^{(\beta)}(C_{A_1(n)}(z)) &= \frac{\beta(n-1)!}{\beta^{(n)}} \sum_{a=1}^{n} \frac{\theta^{(a)}}{a!} \frac{\beta^{(n-a)}}{(n-a)!} \\ &= \frac{\beta}{n}\left(\frac{(\theta+\beta)^{(n)}}{\beta^{(n)}} - 1\right) ,\end{aligned} \tag{70}$$

where for the last display we applied Chu-Vandermonde's formula :

$$\sum_{j=0}^{n} \binom{n}{j} x^{(j)} y^{(n-j)} = (x+y)^{(n)} \,. \tag{71}$$

From (67), (69) and (70) we get

$$\widetilde{\mathbb{E}}^{\sigma_n} \sum_{u \in \partial \mathcal{T}_n} z^{|u|} \leq z^{s_n} + \sum_{i=1}^{s_n - 1} z^i \frac{\beta_{i+1}}{n} \frac{(\theta + \beta_{i+1})^{(n)}}{\beta_{i+1}^{(n)}}. \tag{72}$$

Dividing by $C_n(z)$ and setting

$$a_n(\beta, \theta) := \beta \frac{(\theta+\beta)^{(n)}}{\beta^{(n)}} \frac{(n-1)!}{\theta^{(n)}} = \frac{\Gamma(\theta)\Gamma(\theta+\beta+n)\Gamma(\beta+1)\Gamma(n)}{\Gamma(\theta+\beta)\Gamma(\beta+n)\Gamma(\theta+n)},$$

equation (72) can be rewritten

$$\widetilde{\mathbb{E}}^{\sigma_n} \mathcal{M}_n(z) \leq \frac{z^{s_n}}{C_n(z)} + \sum_{i=1}^{s_n-1} a_n(\beta_{i+1}, \theta) z^i \,. \tag{73}$$

Since $\lim_{x\to\infty} \dfrac{x^\theta \Gamma(x)}{\Gamma(x+\theta)} = 1$, one can find $C(\theta) > 0$ such that for every $x \geq 1$

$$\frac{1}{C(\theta) x^\theta} \leq \frac{\Gamma(x)}{\Gamma(x+\theta)} \leq \frac{C(\theta)}{x^\theta} \,,$$

which yields

$$a_n(\beta,\theta) \le C(\theta)^2\Gamma(\theta)\frac{(\theta+\beta+n)^\theta}{\beta^\theta n^\theta}.$$

For $\beta \le n$ and $n > \theta$, this gives $a_n(\beta,\theta) \le C'\beta^{-\theta}$, where $C'$ is a constant depending only on $\theta$. Since $\beta_{i+1} \le n$ for $i \le s_n - 1$, this yields, coming back to the notation $\theta = 2z$

$$\widetilde{\mathbb{E}}^{\sigma_n}\mathcal{M}_n(z) \le \frac{z^{s_n}}{C_n(z)} + C'\sum_{i=1}^{+\infty}(\beta_{i+1})^{1-2z}z^i. \tag{74}$$

Since $s_{\beta_{l-1}} \le l \le s_{\beta_{l+1}}$ the strong law (59) gives

$$\lim_l \frac{\log \beta_l}{l} = \frac{1}{2z} \quad \widetilde{\mathbb{P}}\text{- a.s.}$$

(recall that $\widetilde{\mathbb{P}} = \mathbb{P}^{(2z)}$), hence

$$\lim_n \left((\beta_{i+1})^{1-2z}z^i\right)^{1/i} = e^{(\eta_2(2z)-1)/2z} < 1 \quad \widetilde{\mathbb{P}}\text{- a.s.}$$

(see (4)) and the series in the right hand side of (74) converges $\widetilde{\mathbb{P}}$-a.s..

For the same reasons, $\lim_n z^{s_n}C_n(z)^{-1} = 0$, $\widetilde{\mathbb{P}}$-a.s. This ends the proof of Proposition 4.3 and then the proof of Theorem 4.2. ■

## 4.2 Convergence of profiles

### 4.2.1 Random measures and profiles

The profile of the tree $\mathcal{T}_n$ is the sequence

$$\left(U_k(n), k \ge 1\right).$$

Jabbour in [21] introduced the random measure counting the heights of leaves in $\mathcal{T}_n$

$$r_n := \sum_k U_k(n)\delta_k = \sum_{u\in\partial\mathcal{T}_n}\delta_{|u|}.$$

Extreme points of the support of $r_n$ are $h_n$ and $H_n$. We are interested in the asymptotic behavior of $r_n$ and of its "local" contributions $U_k(n), k \ge 1$. It is related to the behavior of its intensity $\mathbb{E}r_n$ (which is a non-random measure). We may also look at the random measure counting the heights of leaves in the Yule tree:

$$\rho_t = \sum_{u\in\partial\mathbb{T}_t}\delta_{|u|}.$$

As it is clear from (3) and as it appears below, the convenient scalings are $(\log n)^{-1}$ for the BST and $t^{-1}$ for the Yule tree process.

Our purpose is, from the one hand to explore some direct links between $\rho_t$ as $t \to \infty$ and $r_n$ as $n \to \infty$, and from the other hand, to give a conceptual proof of the convergence of profiles. A first result concerns the intensity of these measures.

**Proposition 4.4** *a) For $x > 2$*

$$\begin{aligned}\lim_{t\to\infty} \frac{1}{t} \log \mathbb{E}\big(\rho_t(]xt, \infty[)\big) &= 1 - \eta_2(x) &(75)\\ \lim_{n\to\infty} \frac{1}{\log n} \log \mathbb{E}\big(r_n(]x\log n, \infty[)\big) &= 1 - \eta_2(x)\,. &(76)\end{aligned}$$

*b) For $x < 2$, replacing $]xt, \infty[$ (resp. $]x\log n, \infty[$) by $]0, xt[$ (resp. $]0, x\log n[$) the same results hold.*

**Remark** 4.5
– for $x \in ]c', c[$ , $\eta_2(x) < 1$, so there are (in mean) about $n^{1-\eta_2(x)}$ leaves of height $\simeq x\log n$. We call this interval $]c', c[$ "supercritical area".
– for $x \in [0, c'[\cup]c, \infty[$, $\eta_2(x) > 1$, so there are (in mean) a very small number of leaves of height $\simeq x\log n$. We call this set "subcritical area".

We call the set $\{c', c\}$ "critical area" .

*Proof:* Relation (75) is easy to obtain, noticing first that by size biasing, for any nonnegative bounded function $f$,

$$\int f(x)\rho_t(dx) = \mathbb{E}\big[2^{|S(t)|} f(|S(t)|) \mid \mathcal{F}_t\big]$$

so that

$$\mathbb{E}\int f(x)\rho_t(dx) = \mathbb{E}\big[2^{|S(t)|} f(|S(t)|)\big]$$

and then using large deviations for the Poisson process $(|S(t)|, t \geq 0)$.

For the BST, we have similarly

$$\int f(x) r_n(dx) = \mathbb{E}\big[2^{s_n} f(s_n) \mid \mathcal{F}_{(n)}\big]$$

so that

$$\mathbb{E}\int f(x) r_n(dx) = \mathbb{E}\big[2^{s_n} f(s_n)\big]\,,$$

and then (76) follows using large deviations for $(s_n)$ of Proposition 3.2.

Notice that the limit in (76) is related to (3) of [13]. ■

In the supercritical area we have a.s. convergences:

**Theorem 4.6** *a) For $x \in (2, c)$ a.s.*

$$\begin{aligned}\lim_{t\to\infty} \frac{1}{t} \log \rho_t(]xt, \infty[) &= 1 - \eta_2(x) &(77)\\ \lim_{n\to\infty} \frac{1}{\log n} \log r_n(]x\log n, \infty[) &= 1 - \eta_2(x) &(78)\end{aligned}$$

*b) For $x \in (c', 2)$ replacing $]xt, \infty[$ (resp. $]x\log n, \infty[$) by $]0, xt[$ (resp. $]0, x\log n[$) the same results hold.*

A sharp (non logarithmic) version of relation (77) is proved in section 5 of [9] (see also Theorem 5 in [8]). Relation (78) is proved by Jabbour [21] in his Theorem 1 (in an equivalent variant for $\nu_n = (n+1)^{-1} r_n$ ), using Gartner-Ellis theorem. Let us explain shortly how (78) can be deduced from (77) by embedding.

Taking $t = \tau_n$ in (77) we have

$$\text{a.s.} \quad \lim_{n\to\infty} \frac{1}{\tau_n} \log \rho_{\tau_n}(]x\tau_n, \infty[) = 1 - \eta_2(x) \tag{79}$$

Now since $\lim ne^{-\tau_n} = \xi$ (cf. (16)) we get

$$\text{a.s.} \quad \lim_{n\to\infty} \frac{1}{\log n} \log r_n(]x\tau_n, \infty[) = 1 - \eta_2(x)\,. \tag{80}$$

Taking into account that for every $\gamma' > 1 > \gamma$ a.s. there exists $n_0$ such that for $n > n_0$, $\gamma' x \log n < x\tau_n < \gamma x \log n$, and using the monotonicity of $a \mapsto r_n(]a, \infty[)$ we get the result.

Let us now consider sharp results for the profile. It is well known that

$$\mathbb{E}\big(U_k(n)\big) = \frac{2^k}{n!} S_n^{(k)}$$

where $S_n^{(k)}$ is the Stirling number of the first kind, so from Hwang ([20]), we get, for any $\ell > 0$ as $n \to \infty$ and $k \to \infty$ such that $r = k/\log n \le \ell$

$$\mathbb{E}\, U_k(n) = \frac{(2\log n)^k}{k!\, n\, \Gamma(r)}(1 + o(1)), \tag{81}$$

which yields easily

$$\mathbb{E}\, U_k(n) \sim \frac{n^{1-\eta_2(\frac{k}{\log n})}}{\Gamma(\frac{k}{\log n})\sqrt{2\pi k}} \tag{82}$$

(see also [13]).

At the level of random variables Jabbour et al. proved in [13] that $\mathbb{P}-$ a.s.

$$\lim_n \frac{U_k(n)}{\mathbb{E} U_k(n)} = \mathcal{M}_\infty(z)\,, \tag{83}$$

for $k = 2z\log n + o(\log n)$ and $z \in [0.6,\ 1.4]$. Since their approach laid on $L^2$ estimations, they guessed that the range $[0.6,\ 1.4]$ may be extended to $I := (1-2^{-1/2}, 1+2^{-1/2}) = (0.293..., 1.707...)$ which is the maximal interval corresponding to a $L^2$ convergence of $\mathcal{M}_n(z)$. In the following subsection we extend the validity of the above result to the entire supercritical interval $(z_c^-, z_c^+)$. Our method consists in adapting the proof of the analogous result in the fragmentation model ([8]). Its main interest is that the random limit $\mathcal{M}_\infty(z)$ appears naturally as the usual Radon-Nikodym derivative $d\widetilde{\mathbb{P}}/d\mathbb{P}$, so it is a "conceptual" proof in the Lyons sense ([26]). Although one can obtain a sharper result adapting for instance the Biggins' method ([9]), we prefere to use the conceptual method to illustrate once more its strength.

### 4.2.2 Main result

**Theorem 4.7** *For $k = 2z\log n + o(\sqrt{\log n})$ and $z \in (z_c^-, z_c^+)$ then*

$$\lim_n \frac{U_k(n)}{\mathbb{E}U_k(n)} = \mathcal{M}_\infty(z)\,,$$

*holds in $\mathbb{P}$-probability.*

**Proof:** Let $\widetilde{\mathbb{P}} := \mathbb{P}^{(2z)}$ as defined in Section 3 especially by formulas (26).

We actually will prove that

$$\lim_n \frac{U_k(n)}{\mathcal{M}_n(z)\mathbb{E}\big(U_k(n)\big)} = 1 \tag{84}$$

in $L^2(\widetilde{\mathbb{P}})$, which will entail (83). Indeed, the variables $U_k(n)$ and $\mathcal{M}_n(z)$ are $\mathcal{F}_{(\infty)}$-measurable and $\mathcal{M}_n(z)$ converges $\mathbb{P}$-a.s. to $\mathcal{M}_\infty(z)$. Since $z$ is supercritical, the probabilities $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ are equivalent on $\mathcal{F}_{(\infty)}$ , and then $\dfrac{U_k(n)}{\mathbb{E}U_k(n)}$ will converge in $\mathbb{P}$-probability to $\mathcal{M}_\infty(z)$.

To prove (84) we first remark that

$$\widetilde{\mathbb{E}}\left(\frac{U_k(n)}{\mathcal{M}_n(z)\mathbb{E}(U_k(n))}\right) = \mathbb{E}\left(\frac{U_k(n)}{\mathbb{E}(U_k(n))}\right) = 1 \tag{85}$$

so that it is enough to prove that

$$\limsup_n \widetilde{\mathbb{E}}\left(\frac{U_k(n)^2}{\mathcal{M}_n(z)^2(\mathbb{E}(U_k(n)))^2}\right) \le 1\,. \tag{86}$$

Using again the change of probability and the size-biasing, especially formula (21), we get

$$\begin{aligned}
\widetilde{\mathbb{E}}\left(\frac{U_k(n)^2}{\mathcal{M}_n(z)^2(\mathbb{E}(U_k(n)))^2}\right) &= \mathbb{E}\Big(\frac{U_k(n)^2}{\mathcal{M}_n(z)(\mathbb{E}(U_k(n)))^2}\Big)\\
&= \mathbb{E}\Big[\frac{U_k(n)}{\mathcal{M}_n(z)(\mathbb{E}(U_k(n)))^2}\mathbb{E}\left(\mathbb{1}_{|s_n|=k}2^{|s_n|}|\mathcal{F}_{(n)}\right)\Big]\\
&= \mathbb{E}\Big[\frac{U_k(n)}{\mathcal{M}_n(z)(\mathbb{E}(U_k(n)))^2}\mathbb{1}_{|s_n|=k}2^{|s_n|}\Big]\\
&= \widetilde{\mathbb{E}}\Big[\frac{U_k(n)}{\mathcal{M}_n(z)(\mathbb{E}(U_k(n)))^2}\mathbb{1}_{|s_n|=k}C_n(z)z^{-|s_n|}\Big]\,.
\end{aligned} \tag{87}$$

Setting

$$B(k,n) = \frac{C_n(z)z^{-k}}{\mathbb{E}(U_k(n))}\ , \quad A(k,n) = \frac{U_k(n)}{\mathbb{E}(U_k(n))}B(k,n)\,, \tag{88}$$

the last display of (87) becomes

$$\widetilde{\mathbb{E}}\left(\frac{U_k(n)^2}{\mathcal{M}_n(z)^2(\mathbb{E}(U_k(n)))^2}\right) = \widetilde{\mathbb{E}}\left(\frac{A(k,n)}{\mathcal{M}_n(z)}\mathbb{1}_{|s_n|=k}\right)\,. \tag{89}$$

The idea is now to replace $A(k,n)$ (resp. $\mathcal{M}_n(z)$) by a similar quantity $\widehat{A}(k,n)$ (resp. $\widehat{\mathcal{M}}_n(z)$), computed with elements above some "dotted line", then apply the Markov property to $\dfrac{\widehat{A}(k,n)}{\widehat{\mathcal{M}}_n(z)}\mathbb{1}_{|s_n|=k}$

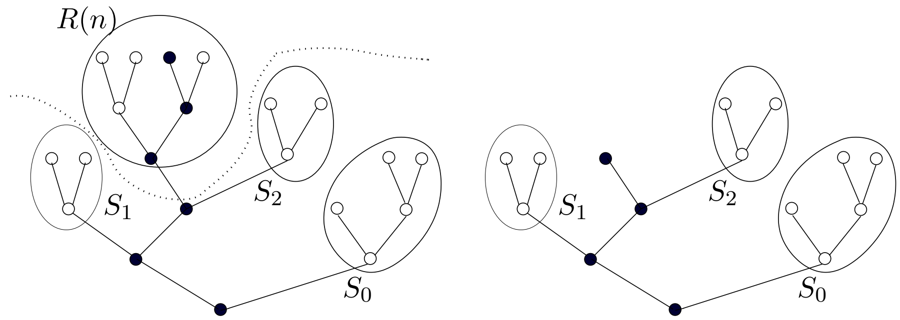


Figure 5: A marked tree and the different classes. On the right, what is known in $\mathcal{A}_n$ (and the apparition date of the red points).

and the local central limit theorem to the remaining part of the trajectory of the spine. Consider again a marked BST . Let $(\alpha_n)_{n\in\mathbb{N}}$ any increasing sequence of integers and let $\mathcal{A}_n$ be the $\sigma$-field

$$\mathcal{A}_n := \sigma\{\beta_0, \beta_1, \dots, \beta_{\alpha_n}, |S_0(n)|, \dots, |S_{\alpha_n}(n)|\};$$

$\mathcal{A}_n$ contains the birth date of the $\alpha_n$ first classes and the number of nodes of each of these classes. Consider $R(n)$, the subtree rooted at $u_{\alpha_n}$ (the red node at level $\alpha_n$); see Figure 5. It contains a red branch and its number of blue leaves is

$$J_n = n - \sum_{i=1}^{\alpha_n} |S_i(n)|.$$

**Lemma 4.8** *a) Let $\theta \in (0, \infty)$. Under $\mathbb{P}^{(\theta)}$ and conditionally on $\mathcal{A}_n$, $R(n)$ is distributed as $\widetilde{\mathcal{T}}_{J_n}$ (under $\mathbb{P}^{(\theta)}$).*

*b) Under $\mathbb{P}$ and conditionally on $\{\beta_{\alpha_n} = \beta\}$, the distribution of $J_n$ is that of $A_1(n-\beta)$ under $\mathbf{CR}^{(\beta)}$ for any $\beta \leq n-1$.*

**Proof:** a) It is clear that $R(n)$ has $J_n$ nodes. The problem is to show that $R(n)$ has the good "tree structure" distribution. Insertions in the subtree rooted in $u_{\alpha_n}$ occur at times which are not $\mathcal{A}_n$ measurable. But, as a matter of fact, these insertion times are not important. Suppose that at time $j$ an insertion occurs in the subtree rooted in $u_{\alpha_n}$. At time $j-1$, there were (say) $k$ blue leaves and one red one in this subtree and (say) $m$ blue leaves in the whole tree ($m = j-2$). A simple computation shows that knowing that the insertion occurs in the subtree rooted in $u_{\alpha_n}$, the insertion occurs on the red node with probability $\theta/(\theta + k)$ and on each blue leave of the subtree with probability $1/(k+\theta)$. These probabilities do not depend on $j$. The evolution of the tree structure of the subtree is the same as the one of the usual marked tree.

b) It is the result of Proposition 3.4 with $z = 1/2$. ■

Let us choose $\alpha_n = \lfloor\sqrt{\log n}\rfloor$ and denote by $\widehat{\zeta}_n = \{v \in \partial\mathcal{T}_n : v \in S_i(n) \text{ for some } i \leq \alpha_n\}$ the set of leaves below the "dotted line". Let

$$\widehat{\mathcal{M}}_n(z) := \sum_{u\in\widehat{\zeta}_n} \frac{z^{|u|}}{C_n(z)}\ , \quad \widehat{A}(k,n) := \sum_{u\in\widehat{\zeta}_n} \mathbb{I}_{|u|=k} \frac{C_n(z) z^{-k}}{(\mathbb{E}U_k(n))^2}\,.$$

The cost of taking $\widehat{A}$ and $\widehat{M}$ instead of $A$ and $M$ is given by the following lemma.

**Lemma 4.9** *For every $q > 0$,*

$$\mathbb{E}\left(|\widehat{A}(k,n) - A(k,n)|\right) = o((\log n)^{-q}) \ , \ \mathbb{E}\left(|\widehat{\mathcal{M}}_n(z) - \mathcal{M}_n(z)|\right) = o((\log n)^{-q}) \, , \tag{90}$$

*which implies*

$$\widetilde{\mathbb{E}}\Big(\Big|\frac{A(k,n)}{\mathcal{M}_n(z)} - \frac{\widehat{A}(k,n)}{\widehat{\mathcal{M}}_n(z)}\Big|\Big) = o((\log n)^{-q}) \, . \tag{91}$$

**Proof of Lemma 4.9** We have

$$\mathcal{M}_n(z) - \widehat{\mathcal{M}}_n(z) = \sum_{u \in \zeta_n - \widehat{\zeta}_n} \frac{z^{|u|}}{C_n(z)}$$

Using Lemma 4.8 a) with $\theta = 1$ we get

$$\mathbb{E}^{\mathcal{A}_n}\Big(\sum_{u \in \zeta_n - \widehat{\zeta}_n} z^{|u|}\Big) = z^{\alpha_n} C_{J_n}(z) \, .$$

From Lemma 4.8 b) and formula (70) we have for $\beta \le n-1$

$$\mathbb{E}(C_{J_n}(z) | \beta_{\alpha_n} = \beta) = \mathbf{CR}^{(\beta)}(C_{A_1(n-\beta)}) = \frac{\beta(2z+\beta)^{(n-\beta)}}{(n-\beta)\beta^{(n-\beta)}} - \frac{\beta}{(n-\beta)} \tag{92}$$

which from (70) gives

$$\mathbb{E}\left(\mathcal{M}_n(z) - \widehat{\mathcal{M}}_n(z)\right) = \frac{z^{\alpha_n}}{C_n(z)} \mathbb{E}\left(\frac{\beta}{(n-\beta)} \frac{(2z+\beta)^{(n-\beta)}}{\beta^{(n-\beta)}} - \frac{\beta}{(n-\beta)}\right) , \tag{93}$$

and since

$$\frac{\beta(2z+\beta)^{(n-\beta)}}{(n-\beta)\beta^{(n-\beta)}} = \frac{n}{n-\beta} \frac{C_n(z)}{C_\beta(z)}$$

we get

$$\mathbb{E}\left(\mathcal{M}_n(z) - \widehat{\mathcal{M}}_n(z)\right) = z^{\alpha_n} \mathbb{E}\left(\frac{1}{(n-\beta)}\left(\frac{n}{C_\beta(z)} - \frac{\beta}{C_n(z)}\right)\right)$$

We use here the exponential martingale (introduced in (22)) $(\mathcal{E}_k(z), k \ge 1)$ and its stopping time $\beta_{\alpha_n}$, so that the above formula becomes

$$\mathbb{E}\Big(\mathcal{M}_n(z) - \widehat{\mathcal{M}}_n(z)\Big) = 2^{-\alpha_n} \mathbb{E}\left(\mathcal{E}_{\beta_{\alpha_n}}(z) \Pi_n(z)\right) \tag{94}$$

where (writing $\beta$ for $\beta_n$)

$$\Pi_n(z) := \frac{1}{n-\beta}\left(n - \beta \frac{C_\beta(z)}{C_n(z)}\right) \, .$$

We write $\Pi_n$ for $\Pi_n(z)$ and we will prove that $\Pi_n$ is bounded by a (deterministic) constant which will give

$$\mathbb{E}\Big(\mathcal{M}_n(z) - \widehat{\mathcal{M}}_n(z)\Big) = \mathcal{O}(2^{-\alpha_n}) \, . \tag{95}$$

For $2z < 1$ and $\beta < n$ we have $C_\beta(z) > C_n(z)$ (see (68) so that $\Pi_n \leq 1$.

When $2z > 1$, we have

$$(n-\beta)\Pi_n = n\Big(1 - \prod_{k=\beta}^{n-1}\Big(1+\frac{2z}{k}\Big)^{-1}\Big)$$

and using the inequality $1 + \frac{2z}{k} \leq \left(1-\frac{1}{k}\right)^{-2z}$ we get

$$\Pi_n \leq \frac{n}{(n-\beta)(n-1)^{2z}}\left((n-1)^{2z} - (\beta-1)^{2z}\right) \leq \frac{2zn}{n-1},$$

where the last inequality comes from Taylor formula. So, estimation (95) holds in any case.

For $A$ we may use the same method. We have, for ever $u \in \mathbb{U}$,

$$\mathbb{1}_{|u|=k} \leq z^{\alpha_n - k} z^{|u|-\alpha_n}. \tag{96}$$

Adding and taking conditional expectations, we get

$$\mathbb{E}^{\mathcal{A}_n}\left(\sum_{u\in\zeta_n - \widehat{\zeta}_n} \mathbb{1}_{|u|=k}\right) \leq z^{\alpha_n - k} C_{J_n}(z) \tag{97}$$

so that

$$\mathbb{E}\left(|\widehat{A}_n(z) - A_n(z)|\right) = \mathcal{O}(2^{-\alpha_n})B(k,n)^2 \tag{98}$$

where $B(k,n)$ was defined in (88). From (81)

$$B(k,n) \sim \sqrt{2\pi k}\ e^{(2z\log n - k)}\left(\frac{2z\log n}{k}\right)^{-k}. \tag{99}$$

In particular if $k = 2z\log n + o(\sqrt{\log n})$ then

$$B(k,n) \sim \sqrt{2\pi k} \tag{100}$$

so that

$$\mathbb{E}\left(|\widehat{A}_n(z) - A_n(z)|\right) = \mathcal{O}(2^{-\alpha_n})k \tag{101}$$

which, joined with (95) proves the first part of the lemma.

To get (91), it is enough to tilt again and use the triangular inequality. ■

*Proof of Theorem 4.7 (end):*

From (89) and (91) we have

$$\widetilde{\mathbb{E}}\left(\frac{U_k(n)^2}{\mathcal{M}_n(z)^2(EU_k(n))^2}\right) = \widetilde{\mathbb{E}}\left(\frac{\widehat{A}(k,n)}{\widehat{\mathcal{M}}_n(z)}\mathbb{1}_{s_n=k}\right) + o((\log n)^{-q}). \tag{102}$$

By conditioning on the dotted line and applying Lemma 4.8 a), we may replace the above indicator function by $\widetilde{\mathbb{P}}(s_{n'} = k')$ where $n' = n - \beta_{\alpha_n}$ and $k' = k - \alpha_n$. However, to control $n'$ which is random, we first split the main term of (102) into

$$\widetilde{\mathbb{E}}\left(\frac{\widehat{A}(k,n)}{\widehat{\mathcal{M}}_n(z)}\mathbb{I}_{s_n=k}\right) \leq \widetilde{\mathbb{E}}\left(\frac{\widehat{A}(k,n)}{\widehat{\mathcal{M}}_n(z)}\mathbb{I}_{\beta_{\alpha_n}\geq\gamma_n}\right) + \widetilde{\mathbb{E}}\left(\frac{\widehat{A}(k,n)}{\widehat{\mathcal{M}}_n(z)}\mathbb{I}_{s_n=k,\ \beta_{\alpha_n}<\gamma_n}\right). \tag{103}$$

On the one hand, we will prove further that, for $\gamma_n = \exp(z^{-1}\alpha_n)$,

$$\widetilde{\mathbb{E}}\left(\frac{\widehat{A}(k,n)}{\widehat{\mathcal{M}}_n(z)}\mathbb{I}_{\beta_{\alpha_n}\geq\gamma_n}\right) = o(1). \tag{104}$$

On the other hand, the second term of (103) becomes

$$\widetilde{\mathbb{E}}\left(\frac{\widehat{A}(k,n)}{\widehat{\mathcal{M}}_n(z)}\mathbb{I}_{s_n=k,\ \beta_{\alpha_n}<\gamma_n}\right) = \widetilde{\mathbb{E}}\left(\frac{\widehat{A}(k,n)}{\widehat{\mathcal{M}}_n(z)}\mathbb{I}_{\beta_{\alpha_n}<\gamma_n}\widetilde{\mathbb{P}}(s_{n'}=k')\right). \tag{105}$$

Taking into account the local central limit theorem (61), we may, for any $\varepsilon > 0$, find $r_0 > 0$ such that for $r \geq r_0$

$$\widetilde{\mathbb{P}}(s_r = k') \leq \frac{1+\varepsilon}{\sqrt{2\pi V_r}} \tag{106}$$

and since $V_r = 2z\log r + o(\log r)$ we may assume $r_0$ large enough to ensure $V_r > 2z(1-\varepsilon)\log r$. Choose $n_0$ such that $n - \gamma_n \geq r_0$ for $n \geq n_0$. It entails

$$\widetilde{\mathbb{E}}\left(\frac{\widehat{A}(k,n)}{\widehat{\mathcal{M}}_n(z)}\mathbb{I}_{\beta_{\alpha_n}<\gamma_n}\widetilde{\mathbb{P}}(s_{n'}=k')\right) \leq \frac{1+\eta}{\sqrt{4\pi z(1-\eta)(\log(n-\gamma_n))}}\,\widetilde{\mathbb{E}}\left(\frac{\widehat{A}(k,n)}{\widehat{\mathcal{M}}_n(z)}\right). \tag{107}$$

Now, again by (90) and (88)

$$\widetilde{\mathbb{E}}\left(\frac{\widehat{A}(k,n)}{\widehat{\mathcal{M}}_n(z)}\right) \leq \widetilde{\mathbb{E}}\left(\frac{A(k,n)}{\mathcal{M}_n(z)}\right) + o((\log n)^{-q} = B(k,n) + o((\log n)^{-q}.$$

From (100) this gives

$$\limsup_n \widetilde{\mathbb{E}}\left(\frac{\widehat{A}(k,n)}{\widehat{\mathcal{M}}_n(z)}\mathbb{I}_{s_n=k,\ \beta_{\alpha_n}<\gamma_n}\right) \leq \frac{1+\varepsilon}{\sqrt{1-\varepsilon}},$$

for any $\varepsilon$. If we admit (104) for a while, equations (103) and (102) lead to (86) which ends the proof of the theorem.

It remains to prove (104). By (91)

$$\widetilde{\mathbb{E}}\left(\frac{\widehat{A}(k,n)}{\widehat{\mathcal{M}}_n(z)}\mathbb{I}_{\beta_{\alpha_n}\geq\gamma_n}\right) = \widetilde{\mathbb{E}}\left(\frac{A(k,n)}{\mathcal{M}_n(z)}\mathbb{I}_{\beta_{\alpha_n}\geq\gamma_n}\right) + o((\log n)^{-q}). \tag{108}$$

Now, since $U(k,n) \leq z^{-k}C_n(z)\mathcal{M}_n(z)$ we have $A(k,n) \leq B(k,n)^2\mathcal{M}_n(z)$ and then

$$\widetilde{\mathbb{E}}\left(\frac{A(k,n)}{\mathcal{M}_n(z)}\mathbb{I}_{\beta_{\alpha_n}\geq\gamma_n}\right) \leq B(k,n)^2\ \widetilde{\mathbb{P}}(\beta_{\alpha_n} > \gamma_n) \leq B(k,n)^2\ \widetilde{\mathbb{P}}(s_{\gamma_n} < \alpha_n) \tag{109}$$

(for the last inequality see the definition of $\beta$).

As said in Proposition 3.2 5), the family $(s_\ell,\ \ell > 0)$ satisfies under $\widetilde{\mathbb{P}}$ the large deviation principle on $[0,\infty)$ with speed $\log \ell$ and rate function $\eta_{2z}$ . Therefore, taking $\gamma_n = \exp \frac{\alpha_n}{z}$ we get

$$\limsup \frac{1}{\alpha_n} \log \widetilde{\mathbb{P}}(s_{\gamma_n} < \alpha_n) \leq -\eta_{2z}(z) = -z(1 - \log 2) < 0$$

which, joined with (100) gives (104). This ends the proof of the theorem. ■

## References


[1] D. Aldous and P. Shields. A diffusion limit for a class of randomly-growing binary trees. *Probab. Theory Related Fields*, 79:509–542, 1988.

[2] R. Arratia, A. Barbour, and S. Tavaré. Logarithmic combinatorial structures: a probabilistic approach. Preprint (book) available at http://www-hto.usc.edu/books/tavare/index.html.

[3] K. Athreya. Change of measures for Markov chains and the L log L theorem for branching processes. *Bernoulli*, 6:323–338, 1999.

[4] K. B. Athreya and S. Karlin. Embedding of urn schemes into continuous time Markov branching processes and related limit theorems. *Ann. Math. Statist.*, 39:1801–1817, 1968.

[5] K. B. Athreya and P. E. Ney. *Branching processes.* Springer-Verlag, New York, 1972.

[6] M.T. Barlow, R. Pemantle, and E.A. Perkins. Diffusion-limited aggregation on a tree. *Probab. Theory Relat. Fields*, 107:1–60, 1997.

[7] J. Bertoin. Homogeneous fragmentation processes. *Probab. Theory Related Fields*, 121(3):301–318, 2001.

[8] J. Bertoin and A. Rouault. Additive martingales and probability tilting for homogeneous fragmentations. Preprint PMA-808 available at http://www.proba.jussieu.fr/mathdoc/preprints/index.html#2003.

[9] J. D. Biggins. Uniform convergence of martingales in the branching random walk. *Ann. Probab.*, 20(1):137–151, 1992.

[10] J. D. Biggins. How fast does a general branching random walk spread? In *Classical and modern branching processes (Minneapolis, MN, 1994)*, volume 84 of *IMA Vol. Math. Appl.*, pages 19–39. Springer, New York, 1997.

[11] J. D. Biggins and D. R. Grey. A note on the growth of random trees. *Statist. Probab. Lett.*, 32(4):339–342, 1997.

[12] J.D. Biggins and A.E. Kyprianou. Measure change in multitype branching. Preprint available at http://www.shef.ac.uk/ st1jdb/mcimb.html, 2001.

[13] B. Chauvin, M. Drmota, and J. Jabbour-Hattab. The profile of binary search trees. *Ann. Appl. Prob.*, 11:1042–1062, 2001.

[14] 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.

[15] L. Devroye, J.A. Fill, and R. Neininger. Perfect simulation from the quicksort limit distribution. *Electronic Communications in Probability*, 5:95–99, 2000.

[16] R Durrett. *Probability: Theory and Exxamples*. Duxbury, Belmont (CA), 1996.

[17] S. Feng and F.M. Hoppe. Large deviation principles for some random combinatorial structures in population genetics and Brownian motion. *The Annals of Prob.*, 8:975–994, 1998.

[18] J.A. Fill and S. Janson. Approximating the limiting quicksort distribution. In *Special Issue of Analysis on Algorithms*, volume 19, pages 376–406, 2001.

[19] J.A. Fill and S. Janson. Quicksort asymptotics. In *Special Issue of Analysis on Algorithms*, volume 44, pages 4–28, 2002.

[20] H.K. Hwang. Asymptotic expansions for the Stirling numbers of the first kind. *J. Combin. Theory Ser. A*, 71(2):343–351, 1995.

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

[22] S. Janson. Functional limit theorems for multitype branching processes and generalized Pólya urns. Preprint available at http://www.math.uu.se/∼svante/papers/index.html.

[23] J.F.C. Kingman. The coalescent process. *Stochastic Process. Appl.*, 13:235–248, 1982.

[24] T.G. Kurtz, R. Lyons, R. Pemantle, and Y. Peres. A conceptual proof of the kesten-stigum theorem for mutitype branching processes. In P. Jagers K.B. Athreya, editor, *Classical and Modern Branching Processes*, volume 84, pages 181–186. IMA Volumes in Mathematics and its Applications, Springer, 1997.

[25] A. E. Kyprianou. A note on branching Lévy processes. *Stochastic Process. Appl.*, 82(1):1–14, 1999.

[26] R. Lyons. A simple path to Biggins' martingale convergence for the branching random walk. In P. Jagers K.B. Athreya, editor, *Classical and Modern Branching Processes*, volume 84, pages 217–222. IMA Volumes in Mathematics and its Applications, Springer, 1997.

[27] R. Lyons, R. Pemantle, and Y. Peres. Conceptual proofs of L log L criteria for mean behavior of branching processes. *Ann. Probab.*, 23:1125–1138, 1995.

[28] H. Mahmoud. *Evolution of Random Search Trees*. John Wiley, New York, 1992.

[29] R. Neininger and L. Rüschendorf. A general limit theorem for recursive algorithms and combinatorial structures. *Annals of App. Probab., to appear*, 2003.

[30] V.V. Petrov. *Sums of independent random variables*. Springer Verlag, 1975.

[31] J. Pitman. Cours de Saint-Flour 2002: Combinatorial Stochastic Processes. available at http://stat-www.berkeley.edu/users/pitman/bibliog.html.

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

[33] U. Rösler. A limit theorem for "quicksort". *RAIRO, Inform. Théor. Appl.*, 25(1):85–100, 1991.

[34] U. Rösler. On the analysis of stochastic divide and conquer algorithms. *Algorithmica*, 29(1-2):238–261, 2001. Average-case analysis of algorithms (Princeton, NJ, 1998).

[35] S. Tavaré. The birth process with immigration, and the genealogical structure of large populations. *J. Math. Biol.*, 25(2):161–168, 1987.

[36] K. Uchiyama. Spatial growth of a branching process of particles living in R. *Ann. Probab.*, 10(4):896–918, 1982.