A.S. CONVERGENCE FOR INFINITE COLOUR PÓLYA URNS ASSOCIATED WITH RANDOM WALKS

SVANTE JANSON

ABSTRACT. We consider Pólya urns with infinitely many colours that are of a random walk type, in two related version. We show that the colour distribution a.s., after rescaling, converges to a normal distribution, assuming only second moments on the offset distribution. This improves results by Bandyopadhyay and Thacker (2014–2017; convergence in probability), and Mailler and Marckert (2017; a.s. convergence assuming exponential moment).

1. INTRODUCTION

Pólya urns with an infinite set $\mathcal{S}$ of possible colurs of the balls have been studied by Bandyopadhyay and Thacker [4, 5, 6] and Mailler and Marckert [25]. We consider here two special types of Pólya urns with the colour space $\mathcal{S} = \mathbb{R}^d$ ($d \geqslant 1$), which both are associated with random walks on $\mathbb{R}^d$. To distinguish them, we call them *single ball addition random walk (SBARW) Pólya urns* and *deterministic addition random walk (DARW) Pólya urns*. The DARW type is the Pólya urn considered in [4; 5], and it is included among the more general urns in [6; 25] and studied further there. The SBARW Pólya urn differs from the urns considered in [4; 5; 6; 25] by having random replacements, but it is closely related to the DARW type. We begin by giving a brief definition of an SBARW Pólya urn, and refer to Section 2 for the DARW type and for further details, including the connection between the two models, as well as a definition of more general Pólya urns.

For simplicity, we first consider an important special case of an SBARW Pólya urn. In this case, the urn contains a (finite) number of balls, each labelled with a vector $X_i \in \mathbb{R}^d$, and starts at time 0 with a single ball labelled with 0. Furthermore, the urn evolves by drawing a ball uniformly at random from the urn, noting its label, $\widehat{X}_n$ say, and replacing it together with a new ball which is labelled with $X_{n+1} := \widehat{X}_n + \eta_n$, where η_n are i.i.d. random variables with some given distribution in $\mathbb{R}^d$. At time $n \geqslant 0$, this urn contains $n + 1$ balls. We describe the composition (= state) of the urn by the measure (on $\mathbb{R}^d$)

$$\mu_n := \sum_{i=0}^{n} \delta_{X_i}, \tag{1.1}$$

where δ_x (the Dirac delta) denotes a point mass at x, and $X_0, \ldots, X_n$ are the balls in the urn.

The general SBARW Pólya urn is an extension of this model; the evolution proceeds in the same way, but the initial number of balls may be different from 1; in fact, it may be any real number $\rho > 0$, and the initial labels may be described by an arbitrary measure μ_0 on $\mathbb{R}^d$ with $\mu_0(\mathbb{R}^d) = \rho \in (0, \infty)$, see Section 2 for details. To draw a ball from an urn with composition μ_n means that we pick a colour with the normalized distribution $\widetilde{\mu}_n$, where we for any non-zero finite measure μ on a space $\mathcal{S}$ define its normalization by

$$\widetilde{\mu} := \mu(\mathcal{S})^{-1}\mu. \tag{1.2}$$

In the case above, where (1.1) holds, $\widetilde{\mu}_n$ is the empirical distribution of the sequence of colours $X_0, \ldots, X_n$.

We assume that the offsets η_n have a finite second moment, i.e., $\mathbb{E}\,|\eta|^2 < \infty$. (We use η to denote a generic offset η_n.) Then Bandyopadhyay and Thacker [4, 5, 6] and Mailler and Marckert [25] proved, for the DARW model, that the normalized compositions $\widetilde{\mu}_n$ are asymptotically normal. To state this formally, we rescale the distributions, using the following notation from [25]. Let $\mathcal{P}(\mathbb{R}^d)$ be the space of Borel probability measures on $\mathbb{R}^d$. If $a > 0$ and $b \in \mathbb{R}^d$, let $\Theta_{a,b} : \mathcal{P}(\mathbb{R}^d) \to \mathcal{P}(\mathbb{R}^d)$ be the rescaling mapping defined by:

$$\text{if} \quad X \sim \mu \in \mathcal{P}(\mathbb{R}^d), \qquad \text{then} \qquad \frac{X - b}{a} \sim \Theta_{a,b}(\mu). \tag{1.3}$$

Note that if μ_n is given by (1.1), and thus $\widetilde{\mu}_n = \frac{1}{n+1}\sum_0^n \delta_{X_i}$, then rescaling $\widetilde{\mu}_n$ by $\Theta_{a,b}$ is the same as rescaling all X_i in the natural way:

$$\Theta_{a,b}(\widetilde{\mu}_n) = \frac{1}{n+1}\sum_{i=0}^n \delta_{(X_i - b)/a}. \tag{1.4}$$

We consider in this paper random probability measures, which formally are described by probability measures on a space of probability measures. This can be confusing. In order to minimize confusion, we use, in statements about limits in $\mathcal{P}(\mathbb{R}^d)$, the notation $\mathcal{L}(N(0, \Sigma))$ to denote the normal distribution $N(0, \Sigma)$ regarded as an element of $\mathcal{P}(\mathbb{R}^d)$; we emphasize that this is a deterministic object.

We regard all vectors as column vectors. It is proved in [4; 5; 6; 25] that, for DARW Pólya urns, if $m := \mathbb{E}\,\eta$, then

$$\Theta_{\sqrt{\log n},\, m\log n}(\widetilde{\mu}_n) \xrightarrow{\text{p}} \mathcal{L}\left(N\left(0, \mathbb{E}[\eta\eta^{\mathsf{t}}]\right)\right), \tag{1.5}$$

with convergence in $\mathcal{P}(\mathbb{R}^d)$ with the usual weak topology; furthermore Mailler and Marckert [25] showed that the convergence in (1.5) hold a.s. if $|\eta|$ has a finite exponential moment. One of the main purposes of the present paper is to show that a.s. convergence always holds, assuming only a second moment, for both types of Pólya urns associated with random walks considered here.

Theorem 1.1. *Consider an SBARW Pólya urn, with i.i.d. offsets $\eta_n \in \mathbb{R}^d$ and an initial composition μ_0 that is an arbitrary non-zero finite measure on $\mathbb{R}^d$. Assume that $\mathbb{E}\,|\eta|^2 < \infty$ and let $m := \mathbb{E}\,\eta$. Then, as $n \to \infty$,*

$$\Theta_{\sqrt{\log n},\, m\log n}(\widetilde{\mu}_n) \xrightarrow{\text{a.s.}} \mathcal{L}\left(N\left(0, \mathbb{E}[\eta\eta^{\mathsf{t}}]\right)\right), \tag{1.6}$$

in $\mathcal{P}(\mathbb{R}^d)$ with the usual weak topology.

The same result holds also for DARW Pólya urns.

As remarked above, the right-hand side in (1.6) is non-random; it is a fixed distribution in $\mathcal{P}(\mathbb{R}^d)$. Note also that the variance in the limit in (1.6) is $\mathbb{E}[\eta\eta^{\mathsf{t}}]$ and not $\mathrm{Var}[\eta] = \mathbb{E}[\eta\eta^{\mathsf{t}}] - mm^{\mathsf{t}}$, cf. Example 6.4.

Remark 1.2. If we inspect the urn by drawing a ball at random from the urn at time n (without interfering with the urn process), and let X_n^* be is its colour, then, conditionally on the urn composition μ_n, the distribution of X_n^* is $\widetilde{\mu}_n$. Hence, recalling (1.3), (1.6) can also be written as

$$\mathcal{L}\Big(\frac{X_n^* - m\log n}{\sqrt{\log n}} \,\Big|\, \mu_n\Big) \xrightarrow{\text{a.s.}} \mathcal{L}\big(N\big(0, \mathbb{E}[\eta\eta^{\mathsf{t}}]\big)\big) \tag{1.7}$$

in $\mathcal{P}(\mathbb{R}^d)$. We can also rewrite (1.7) as a conditional convergence in distribution:

$$\text{Conditioned on } (\mu_n)_1^\infty, \text{ a.s.,} \quad \frac{X_n^* - m\log n}{\sqrt{\log n}} \xrightarrow{\text{d}} N\big(0, \mathbb{E}[\eta\eta^{\mathsf{t}}]\big). \tag{1.8}$$

Remark 1.3. By unconditioning in (1.8), it follows that

$$\frac{X_n^* - m\log n}{\sqrt{\log n}} \xrightarrow{\text{d}} N\big(0, \mathbb{E}[\eta\eta^{\mathsf{t}}]\big). \tag{1.9}$$

This is a much simpler result, which e.g. easily follows from the correspondence with random recursive trees used below and the asymptotic normal distribution of the depth of a random node in a random recursive tree, see [14, Theorem 6.17], together with the usual central limit theorem for i.i.d. variables.

In the language of statistical physics, we study in (1.7)–(1.8) the *quenched* version of the problem, where we fix a realization of the urn process (μ_n), and then consider the random variable X_n^*, obtaining results for a.e. realization of the urn process. The corresponding *annealed* version, where we just consider X_n^* as a random variable obtained by randomly constructing the urn μ_n and choosing a ball X_n^* in it, is the much simpler (1.9). Note that the distribution of X_n^* in the annealed version is $\mathbb{E}\,\widetilde{\mu}_n$, the expectation of the random measure $\widetilde{\mu}_n$ defined above. Hence, (1.9) can be written as

$$\Theta_{\sqrt{\log n},\, m\log n}(\mathbb{E}\,\widetilde{\mu}_n) \to \mathcal{L}\big(N\big(0, \mathbb{E}[\eta\eta^{\mathsf{t}}]\big)\big). \tag{1.10}$$

We can regard (1.9)–(1.10) as the annealed version of Theorem 1.1. Similar unconditioning to annealed versions can be done in the theorems for random trees in Sections 6 and 10.

The proofs by [4; 5; 6] and [25] are based on a connection between Pólya urns and the random recursive tree, see Section 5. We do the same in the present paper; we also introduce a weighted modification of the random recursive tree in order to treat Pólya urns with an arbitrary inital configuration, see Section 3.1. The SBARW Pólya urns correspond to branching random walks on the (weighted) random recursive tree, and thus Theorem 1.1 is equivalent to a.s. convergence of the empirical distribution, suitably normalized, for a branching random walk on a (weighted) random recursive tree. Furthermore, as is also well-known, the random recursive tree can be embedded in the continuous time Yule tree, and thus the result can be interpreted as a.s. convergence of the normalized empirical distribution of a

branching random walk on a Yule tree. (This extends to the weighted random recursive tree and a weighted Yule tree defined in Section 3.2.) Such a.s. convergence results for branching processes have been shown, in much greater generality, by e.g. Uchiyama [33, Theorem 4], and thus Theorem 1.1 essentially follows from known results in branching process theory.

One purpose of the present paper is to make this connection explicit, by stating results for branching random walks on random recursive trees and Yule trees in a form corresponding to the Pólya urn theorem above, including the weighted cases. We prove these results for random trees using the standard method of showing convergence of a suitable martingale of functions, used also by Uchiyama [33], Biggins [8, 9] and others. (For this, we use a Sobolev space of functions; see Remark 7.11.) We give complete proofs, both for completeness and because we want to show how the proofs work in this simple case where we can give explicit expressions instead of estimates, and without the distractions caused by the greater generality in [33], and also because we have not been able to find published results with precisely the formulations used here, including the weighted case. Furthermore, we give proofs with explicit calculations both for the (weighted) Yule trees and the random recursive trees; as said above, the results for these trees are equivalent, so it suffices to prove one of the cases. Nevertheless, it is possible to prove the result directly, with explicit calculations, for both cases, and we find it interesting and instructive to do so and see the similarities and differences between the two cases.

The Pólya urns, random trees and branching random walks that we consider are defined in Sections 2–4, and the connection between them is given in Section 5. The results for random recursive trees and Yule trees are stated in Section 6 and proved in Sections 7–8. Theorem 1.1 above is proved in Section 9. Section 10 gives analoguous results for binary search trees and binary Yule trees. Section 11 contains some open problems.

2. Pólya urns

As said in the introduction, we consider a general version of Pólya urns, where we have a measurable space $\mathcal{S}$ of *colours* (i.e., types), and the state, or *composition*, of the urn is given by a finite measure μ on $\mathcal{S}$. This version of Pólya urns has been introduced in a special case by Blackwell and MacQueen [10], and in general by Bandyopadhyay and Thacker [6], see also [4; 5], and by Mailler and Marckert [25]. Although our main theorem is only for Pólya urns of the special types associated with random walks, we give the definition of Pólya urns in a general form as in [6; 4; 5; 25]. Furthermore, we allow also random replacements, see also [20].

The interpretation of the measure μ describing the state of the urn is that if $A \subseteq \mathcal{S}$, then $\mu(A)$ is the total mass of the colours in A. The classical case with a finite number of discrete balls of assorted colours, can be treated by representing each ball of colour x by a point mass δ_x; in other words, if the urn contains m balls with colours $x_1, \ldots, x_m$, then it is represented by the measure

$$\mu = \sum_i \delta_{x_i}, \tag{2.1}$$

and thus μ is a discrete measure where $\mu\{x\}$ is the number of balls of colour x. It has often been remarked that the classical case easily generalizes to non-integer "numbers of balls" of each colour; in the measure formulation considered here, this means that μ is an arbitrary discrete finite measure. The general measure version is a further generalization, where $\mathcal{S}$ may be infinite and μ may be, e.g., a continuous measure.

To define a measure-valued Pólya urn process, we assume that we are given a colour space $\mathcal{S}$ (a measurable space). We let $\mathcal{M}(\mathcal{S})$ denote the space of finite measures on $\mathcal{S}$, let $\mathcal{M}_*(\mathcal{S}) := \mathcal{M}(\mathcal{S}) \setminus \{0\}$, where 0 denotes the zero measure, and define for each $\mu \in \mathcal{M}_*(\mathcal{S})$ its normalization $\widetilde{\mu}$ as as the probability measure (1.2). We assume also that we are given a *replacement rule*, which may be deterministic or random. In the deterministic case it is a (measurable) function $x \mapsto R_x$ mapping $\mathcal{S}$ into $\mathcal{M}(\mathcal{S})$; in other words, R_x is a kernel from $\mathcal{S}$ to itself [23, p. 20]. In the random case, each R_x is a random element of $\mathcal{M}(\mathcal{S})$; formally the replacement rule is a (measurable) mapping $x \mapsto \mathcal{R}_x$ mapping $\mathcal{S}$ into the space $\mathcal{P}(\mathcal{M}(\mathcal{S}))$ of probability measures on $\mathcal{M}(\mathcal{S})$, i.e., a probability kernel from $\mathcal{S}$ to $\mathcal{M}(\mathcal{S})$, but it is convenient to represent each $\mathcal{R}_x$ by a random $R_x \in \mathcal{M}(\mathcal{S})$ having distribution $\mathcal{R}_x$.

The Pólya urn starts with a given initial composition $\mu_0 \in \mathcal{M}_*(\mathcal{S})$. In each step, we "draw a ball from the urn"; this means that, given everything that has happened so far, if the current composition of the urn is μ_n, then we randomly select a colour X_n with distribution $\widetilde{\mu}_n$. We then "return the ball together with the replacement R_{X_n}", which means that we update the state of the urn to

$$\mu_{n+1} := \mu_n + R_{X_n}. \tag{2.2}$$

In the case when the replacements R_x is random, (2.2) should be interpreted to mean that given X_n and everything that has happened earlier, we take a fresh random R_{X_n} with the distribution $\mathcal{R}_{X_n}$. Thus, given X_n, R_{X_n} is independent of the history of the process. (It is shown in [20] that an urn with random replacements is equivalent to an urn with deterministic replacements on the larger colour space $\mathcal{S} \times [0, 1]$; we will not use this.)

The update (2.2) is repeated an infinite number of times; this defines the Pólya urn process as a Markov process. The process is well-defined, with every $\mu_n \in \mathcal{M}_*(\mathcal{S})$, since we have assumed that each R_x is a finite measure. (Thus R_x is non-negative; there are no subtractions of balls in this version.)

Remark 2.1. We use the name "replacement" to conform with [4; 5; 6; 25], although R_x really is an addition to the urn rather than a replacement, since we also return the drawn balls. (The real replacement is $\delta_x + R_x$.) A version with a true replacement, without replacement of the drawn ball, is studied in [25], but will not be considered here.

One special case, which we call *single ball addition* is when each replacement consists of a single ball of a random colour (with distribution $\mathcal{R}_x$ depending on the colour x of the drawn ball as above). In other words, for each $x \in \mathcal{S}$, R_x is a random measure of the type $R_x = \delta_{r_x}$ for a random variable $r_x \in \mathcal{S}$. In this case, let $\mathbf{r}_x := \mathcal{L}(r_x) \in \mathcal{P}(\mathcal{S})$ be the distribution of r_x and define $\overline{\mu}_n \in \mathcal{M}(S)$ as the composition $\mu_n \cdot \mathbf{r}$ of μ_n and the kernel

$x \mapsto \mathbf{r}_x$ (from $\mathcal{S}$ to itself), see [23, p. 21], i.e.

$$\overline{\mu}_n(A) := \int_{\mathcal{S}} \mathbf{r}_x(A)\, \mathrm{d}\mu_n(x) = \int_{\mathcal{S}} \mathbb{P}(r_x \in A)\, \mathrm{d}\mu_n(x). \qquad (2.3)$$

Note that the total mass $\overline{\mu}_n(\mathcal{S}) = \mu_n(\mathcal{S})$ and that therefore $\widetilde{\overline{\mu}}_n = \widetilde{\mu}_n \cdot \mathbf{r}$. Hence, in the Pólya process above, given the present state μ_n, the distribution of the colour of the next ball added to the urn is $\widetilde{\overline{\mu}}_n$. Furthermore, if Y_n is this colour, i.e., $Y_n := r_{X_n} \in \mathcal{S}$, then $\mu_{n+1} = \mu_n + \delta_{Y_n}$, and thus, by (2.3),

$$\overline{\mu}_{n+1} = \mu_{n+1} \cdot \mathbf{r} = \mu_n \cdot \mathbf{r} + \delta_{Y_n} \cdot \mathbf{r} = \overline{\mu}_n + \mathbf{r}_{Y_n}. \qquad (2.4)$$

This means that $\overline{\mu}_n$ also is a Pólya urn process, as defined above, with deterministic replacements $\mathbf{r}_x$. We state this formally.

Lemma 2.2. *Let (μ_n) be a single ball addition Pólya urn process, with random replacements $R_x = \delta_{r_x}$. Then, with $\mathbf{r}_x := \mathcal{L}(r_x)$ and $\overline{\mu}_n := \mu_n \cdot \mathbf{r}$, the sequence $(\overline{\mu}_n)$ is a Pólya urn process with deterministic replacements $\mathbf{r}_x \in \mathcal{P}(\mathcal{S})$.* □

The Pólya urns studied in [4; 5; 6; 25] have deterministic replacements that furthermore are probability measures; hence these urns are of the type (2.4), and Lemma 2.2 shows that, provided the initial value is of the type $\mu_0 \cdot \mathbf{r}$, these urns correspond to urns μ_n with random single ball additions; more precisely they are given by $\mu_n \cdot \mathbf{r}$.

Example 2.3. The urns studied in Blackwell and MacQueen [10] have the special form $R_x = \delta_x$; hence they are single ball addition Pólya urns where the added ball has the same colour as the drawn one, just as for the original (two-colour) urns in [27; 16; 32]. In this case $r_x = x$ and $\overline{\mu}_n = \mu_n$, so there is no difference between the two Pólya urns in Lemma 2.2.

Example 2.4. The *SBARW Pólya urns* discussed in Section 1 are a special case of the single ball addition case, where $\mathcal{S} = \mathbb{R}^d$ and the replacements are translation invariant, i.e., $r_x \stackrel{\mathrm{d}}{=} x + r_0$ for all $x \in \mathbb{R}^d$. In other words, with $\eta := r_0$, if we draw a ball of colour X_n, it is replaced together with a ball of colour $X_n + \eta_n$, where (η_n) are independent copies of the random variable $\eta \in \mathbb{R}^d$ (with η_n independent of X_n).

By Lemma 2.2, an SBARW Pólya urn corresponds to an urn with colour space $\mathbb{R}^d$ and deterministic replacements $\mathbf{r}_x = \mathcal{L}(x + \eta)$. This is the type of urns studied in [4; 5]; they are also studied in [6; 25] together with more general ones. We call such urns *DARW Pólya urns*.

Note that for a DARW Pólya urn, the translation invariance of $\mathbf{r}$ shows that the relation $\overline{\mu}_n = \mu_n \cdot \mathbf{r}$ in Lemma 2.2 can be written as a convolution

$$\overline{\mu}_n = \mu_n * \mathbf{r}_0 = \mu_n * \nu. \qquad (2.5)$$

3. Random trees

The random trees that we study are (mostly) well-known; see for example [14] and [3; 34]. For convenience, we collect their definitions here.

The trees that we are interested in grow (randomly) in either discrete or continuous time; we thus consider either an increasing sequence of random trees T_n with an integer parameter $n \geqslant 0$, or an increasing family $\mathcal{T}_t$ of

random trees with a real parameter $t \geqslant 0$. (We use different fonts for the two cases; this will be convenient to distinguish them in e.g. the proof of Theorem 6.1 where we consider trees of both types simultaneously, but has otherwise no significance.)

For a tree T, we let $|T|$ denote its number of nodes; however, when we consider weighted trees, we instead let $|T|$ denote the total weight, i.e., the sum of the weights of the nodes.

3.1. **Trees growing in discrete time.** The *random recursive tree* T_n is constructed recursively. T_0 is just a root. Given T_n, we obtain T_{n+1} by adding a new node and choosing its parent uniformly at random from the already existing nodes. (We have chosen a notation where T_n has $n + 1$ nodes; this is of course irrelevant for our asymptotic results.)

We consider also a generalization of the random recursive tree that we call a *weighted random recursive tree*; this is characterized by a parameter $\rho > 0$ (the *weight*). The definition is as for the random recursive tree, but we give the root weight ρ and every other node weight 1, and when we add a node, its parent is chosen with probability proportional to its weight. In other words, when adding a new node to T_n, its parent is chosen to be the root o with probability $\rho/(n + \rho)$, and to be v with probability $1/(n + \rho)$, for each of the n existing nodes $v \neq o$. Note that taking the weight $\rho = 1$ gives the random recursive tree.

The *binary search tree* is defined by a similar recursive procedure, but we now have two types of nodes, *internal* and *external*. T_0 consists of a single external node (the root). The tree evolves by choosing an external node uniformly at random, and then converting it to an internal node and adding two new external nodes as children to it. (One child is labelled left and the other right.) Thus T_n has n internal nodes and $n + 1$ external nodes; an internal node has 2 children, and an external node has 0. Depending on the circumstances, one might either be interested in the complete tree T_n with $2n + 1$ nodes, or just the internal subtree T_n^{i} with the n internal nodes.

3.2. **Trees growing in continuous time.** The *Yule tree* $\mathcal{T}_t$, $t \geqslant 0$, is the family tree of the *Yule process*, which is a simple Markov continuous-time branching process starting with a single node (= individual) at time 0 and such that every node lives for ever and gets children according to a Poisson process with intensity 1.

By symmetry and lack of memory, it is obvious that if τ_n is the stopping time

$$\tau_n := \min\{t : |\mathcal{T}_t| = n + 1\}, \tag{3.1}$$

then the sequence $T_n := \mathcal{T}_{\tau_n}$ is a sequence of random recursive trees.

Corresponding to the weighted random recursive tree above, we define also a *weighted Yule tree*, where the root (the initial node) has weight $\rho > 0$ and every other node has weight 1, and each node gets children with intensity equal to its weight. (Thus, only the initial node is modified.) For the weighted Yule tree, we modify (3.1) and define τ_n as the first time that the total weight is $n + \rho$; then T_{τ_n} is a weighted random recursive tree with the same weight ρ. Note that if ρ is an integer, then the weighted Yule tree can be obtained by taking ρ independent Yule trees and merging their roots.

Many authors use a different version of the Yule tree, which we call the *binary Yule tree* to distinguish the two versions. The difference is that each individual lives a random time with an exponential distribution $\mathrm{Exp}(1)$ with rate 1, and that each individual gets 2 children when she dies. (We do not define any weighted version.) It is obvious that the number of living individuals follows the same branching process (the Yule process) for both versions, but that the trees $\mathcal{T}_t$, which contain both the living and dead individuals, will be different. In fact, if we now let τ_n be the first time that the tree has n dead individuals, and thus $n+1$ living ones, it is easy to see, again because of the lack of memory, that the sequence T_{τ_n} defines a binary search tree, where the dead individuals are internal nodes and the living individuals are external nodes.

We shall use some simple facts from branching process theory.

First, it is a well-known fact [3, Theorems III.7.1–2] that for the Yule tree (and much more generally),

$$|\mathcal{T}_t|/e^t \xrightarrow{\text{a.s.}} W > 0 \tag{3.2}$$

for some random variable W. (In fact, for the Yule tree, $W \sim \mathrm{Exp}(1)$, but we do not need this.) For the weighted Yule tree, every child of the root starts an independent Yule tree, and it follows easily that (3.2) holds in this case too. (Furthermore, W then has the Gamma distribution $W \sim \Gamma(\rho)$.)

Taking $t = \tau_n$ in (3.2) yields, for a general weight ρ,

$$\frac{n+\rho}{e^{\tau_n}} \xrightarrow{\text{a.s.}} W \tag{3.3}$$

and thus, a.s.,

$$\tau_n = \log(n+\rho) - \log W + o(1) = \log n + O(1). \tag{3.4}$$

(Here and below, the implicit constant in $O(1)$ may be random.)

We note a standard fact.

Lemma 3.1. *If $\mathcal{T}_t$ is the weighted Yule tree, for any $\rho > 0$, then for every $t < \infty$ and $r < \infty$, $\mathbb{E}\,|\mathcal{T}_t|^r < \infty$.*

Proof. It is well-known that $\mathbb{E}\,|\mathcal{T}_t|^r < \infty$ for the standard Yule tree with $\rho = 1$, see [3, Corollary III.6.1].

For general ρ, we may by monotonicity (in ρ) assume that ρ is an integer, and the result then follows by regarding the tree as a union of ρ independent Yule trees. $\qquad\square$

4. Branching random walks on trees

Given a rooted tree T and a probability distribution ν on $\mathbb{R}^d$, a *branched random walk* on T, with *offset distribution* ν, is a stochastic process $(X_v)_{v\in T}$ indexed by the nodes of T that is defined recursively as follows:

(BW1) Let η_v, $v \in T$, be i.i.d. random variables with $\eta_i \sim \nu$.

(BW2) $X_o := 0$, where o is the root of T.

(BW3) If w is a child of v, then $X_w := X_v + \eta_w$.

In other words, if $v \prec w$ means that v is an ancestor of w,

$$X_v := \sum_{o \prec u \preceq v} \eta_u. \tag{4.1}$$

The tree T is usually random; we then tacitly assume that the random variables η_v are independent of the tree T.

Remark 4.1. Alternatively (and equivalently), we may start with a rooted tree T and random variable η, by taking $\nu := \mathcal{L}(\eta)$, the distribution of η; (BW1) then says that η_i are independent copies of η. In this setting, η is called the *offset*. In the sequel, we use η in this sense, to denote a generic random variable with distribution ν.

Remark 4.2. We never use η_o, and may thus ignore it. For $v \neq o$, we may think of η_v as associated to the edge leading to v from its parent; then X_v is the sum of these values for all edges on the path between o and v. (Alternatively, we could change the definition and let $X_o := v_o$; this would not affect our asymptotical results.)

We are interested in the empirical distribution of the variables X_v; this is by definition the random probability measure on $\mathbb{R}$ (or $\mathbb{R}^d$) defined by

$$\widetilde{\mu} := \frac{1}{|T|} \sum_{v \in T} \delta_{X_v}, \tag{4.2}$$

where δ_x (the Dirac delta) denotes a point mass at x, and $|T|$ is the number of nodes in T. In other words, given $(X_v)_{v \in T}$, $\widetilde{\mu}$ is the distribution $\mathcal{L}(X_V)$ of the value X_V seen at a uniformly randomly chosen node $V \in T$.

For a weighted tree, we modify the definition (4.2) by counting each node v according to its weight ω_v. In our cases, $\omega_v = 1$ for every $v \neq o$, and thus, recalling that $|T|$ denotes the total weight,

$$\widetilde{\mu} := \frac{1}{|T|} \sum_{v \in T} \omega_v \delta_{X_v} = \frac{1}{|T|} \Big(\rho \delta_{X_o} + \sum_{v \neq o} \delta_{X_v} \Big). \tag{4.3}$$

Similarly, the random node $V \in T$ is chosen with probability proportional to its weight, and then still $\widetilde{\mu} = \mathcal{L}(X_V)$. Since only the root has a weight different from 1, it is obvious that the asymptotic results below are not affected be these modifications, and the results hold for both definitions (4.2) and (4.3). However, the modifications are natural, and convenient in the proofs below, so we shall use (4.3) in the weighted case.

In analogy with the Pólya urns defined earlier, we also define the unnormalized measure

$$\mu := |T|\widetilde{\mu} = \sum_{v \in T} \omega_v \delta_{X_v}. \tag{4.4}$$

We consider an increasing sequence or family of (random) trees T_n with an integer parameter $n \geqslant 0$, or alternatively an increasing family $\mathcal{T}_t$ of random trees with a real parameter $t \geqslant 0$. We consider also an i.i.d. family $(\eta_v)_v$ of offsets, defined for all $v \in T_\infty := \bigcup_n T_n$ [or $\bigcup_t \mathcal{T}_t$]. (Thus η_v is defined for all $v \in T_n$ [$\mathcal{T}_t$], but does not depend on the parameter n [t].) We denote the empirical distribution by $\widetilde{\mu}_n$ [$\widetilde{\mu}_t$], and our goal is to show that it, suitably rescaled, a.s. converges to a normal distribution as $n \to \infty$ or $t \to \infty$; see Section 6 for precise statements. As in Theorem 1.1, this is a question of convergence of a random probability measure in the space $\mathcal{P}(\mathbb{R}^d)$ of probability measures on $\mathbb{R}^d$ with the standard (weak) topology, cf. Remark 1.3.

We give a simple lemma that will be used later.

Lemma 4.3. *If $\mathcal{T}_t$ is a weighted Yule tree and $\mathbb{E}\,|\eta|^2 < \infty$, then for every $t < \infty$, $\mathbb{E}\big(\sum_{v \in \mathcal{T}_t} |X_v|\big)^2 < \infty$.*

Proof. Trivially, $|X_v| \leqslant \sum_{w \in \mathcal{T}_t} |\eta_w|$ for every $v \in \mathcal{T}_t$ and thus

$$\sum_{v \in \mathcal{T}_t} |X_v| \leqslant |\mathcal{T}_t| \sum_{w \in \mathcal{T}_t} |\eta_w|. \tag{4.5}$$

It follows, using the Cauchy–Schwarz inequality, that

$$\mathbb{E}\Big(\Big(\sum_{v \in \mathcal{T}_t} |X_v|\Big)^2 \mid \mathcal{T}_t\Big) \leqslant |\mathcal{T}_t|^2 \sum_{u,w \in \mathcal{T}_t} \mathbb{E}(|\eta_u|\,|\eta_w|) \leqslant |\mathcal{T}_t|^4 \,\mathbb{E}\,|\eta|^2 \tag{4.6}$$

and thus, using Lemma 3.1,

$$\mathbb{E}\Big(\sum_{v \in \mathcal{T}_t} |X_v|\Big)^2 \leqslant \mathbb{E}\,|\eta|^2 \,\mathbb{E}\,|\mathcal{T}_t|^4 < \infty. \tag{4.7}$$

$\square$

5. Pólya urns and trees

The proofs by Bandyopadhyay and Thacker [4, 5, 6] and Mailler and Marckert [25] are based on a natural connection between Pólya urns and branching Markov chains on random recursive trees, and in particular between DARW Pólya urns and branching random walks on random recursive trees. In our setting, we use two versions, one for SBARW urns and one for DARW urns, and we include also the weighted case (when $\mu_0(\mathcal{S}) \neq 1$).

5.1. **SBARW Pólya urns.** Consider a Pólya urn of the single ball addition type, see Section 2; we let the initial composition μ_0 have arbitrary mass $\rho > 0$, but assume that it is concentrated at 0; thus, $\mu_0 = \rho\delta_0$. Regard this initial mass as a ball with weight ρ and colour 0; let all balls added later to the urn have weight 1. Regard the balls in the urn as nodes in a tree, where the initial ball is the root and each new ball added after drawing a ball becomes a child of the drawn ball. It is obvious that the resulting random tree process is the weighted random recursive tree defined in Section 3.1. Furthermore, if the Pólya urn is of the SBARW type, then the labels on the ball form a branching random walk (4.1), with the same offset η; note that the measures μ_n and $\widetilde{\mu}_n$ defined earlier (Sections 2 and 4) are the same for the Pólya urn and the branching random walk.

This means that, at least when $\mu_0 = \rho\delta_0$, the first part of Theorem 1.1 is equivalent to a result for the random recursive tree, see Theorem 6.1 below and Section 9.

5.2. **DARW Pólya urns.** For the DARW Pólya urns, we describe the connection used by [4; 5; 6; 25] as follows. Consider first an arbitrary measure-valued Pólya urn with deterministic replacements $R_x \in \mathcal{M}(\mathcal{S})$. Denote the successive additions to the urn by, see (2.2),

$$\Delta\mu_{n+1} := \mu_{n+1} - \mu_n = R_{X_n}, \tag{5.1}$$

and let $\Delta\mu_0 := \mu_0$, the initial composition. We may pretend that the different additions $\Delta\mu_k$, $k \leqslant n$, are identifiable parts of μ_n. Hence when we draw

a ball, we can do it in two steps; we first select an index $k \leqslant n$, with probability $\Delta \mu_k(\mathcal{S})/\mu_n(\mathcal{S})$, and then, given k, select X_n with distribution $\widetilde{\Delta \mu_k}$. This defines a growing family T_n of trees, where T_n has node set $\{0, \ldots, n\}$, and T_{n+1} is obtained from T_n by adding $n+1$ as a new node with mother k, the index selected when choosing X_n in the construction of the Pólya urn.

From now on, we assume that all replacements R_x are probability measures, i.e., have mass $R_x(\mathcal{S}) = 1$. Then it is obvious that this random family of trees T_n is the weighted random recursive tree with weight $\rho = \mu_0(\mathcal{S})$. We mark each node n in this tree by the measure $\Delta \mu_n$, and also, for $n > 0$, by the colour X_{n-1} of the ball drawn to find this addition; we write $Z_n := X_{n-1}$. Then, given the trees T_n, $n \geqslant 0$, these marks, and thus the Pólya urn, are defined recursively as follows, with $o = 0$, the root of the tree,

(P1) $\Delta \mu_o := \mu_0$

(P2) If a node $v \neq o$ has mother u, then Z_v is drawn with the distribution $\Delta \mu_u$, and then $\Delta \mu_v := R_{Z_v}$.

If we further specialize to a DARW Pólya urn, then, see Example 2.4, for $v \neq o$, assuming η to be independent of everything else,

$$\Delta \mu_v := R_{Z_v} = \mathbf{r}_{Z_v} = \mathcal{L}(\eta + Z_v \mid Z_v). \tag{5.2}$$

Remark 5.1. It is easily seen that the relations in Sections 5.1 and 5.2 are connected by the correspondence in Lemma 2.2; if we take the tree in Section 5.1 and mark each node with colour x by R_x (ρR_0 for the root), then we obtain the corresponding process in Section 5.2.

6. Results for trees

We state here the results for the (weighted) random recursive tree and Yule tree; proofs are given in Sections 7–8. As said in the introduction, the results for Yule trees are essentially proved by Uchiyama [33, Theorem 4]; the weighted Yule trees considered here are not quite included in his conditions (which otherwise are very general), but his result is easily extended to the present case. See also Section 10, where similar results for the binary search tree and the binary Yule tree are given.

Recall the definitions of the random trees in Section 3, the empirical measures $\widetilde{\mu}_n$ or $\widetilde{\mu}_t$ in Section 4, and $\Theta_{a,b}$ in (1.3).

We assume that $\mathbb{E}\,|\eta|^2 < \infty$, and let as above $m := \mathbb{E}\,\eta \in \mathbb{R}^d$. Convergence in the space $\mathcal{P}(\mathbb{R}^d)$ of probability measures is always in the usual weak topology.

Theorem 6.1. *Let T_n be the random recursive tree and suppose that $\mathbb{E}\,|\eta|^2 < \infty$. Then, as $n \to \infty$, in $\mathcal{P}(\mathbb{R}^d)$,*

$$\Theta_{\sqrt{\log n},\, m \log n}(\widetilde{\mu}_n) \xrightarrow{\text{a.s.}} \mathcal{L}\left(N\left(0, \mathbb{E}[\eta \eta^{\mathsf{t}}]\right)\right). \tag{6.1}$$

More generally, the same result holds for the weighted random recursive tree with an arbitrary weight $\rho > 0$ defined in Section 3.1.

Theorem 6.2 (essentially Uchiyama [33])**.** *Let $\mathcal{T}_t$ be the Yule tree and suppose that $\mathbb{E}\,|\eta|^2 < \infty$. Then, as $t \to \infty$, in $\mathcal{P}(\mathbb{R}^d)$,*

$$\Theta_{\sqrt{t},\, mt}(\widetilde{\mu}_t) \xrightarrow{\text{a.s.}} \mathcal{L}\left(N\left(0, \mathbb{E}[\eta \eta^{\mathsf{t}}]\right)\right). \tag{6.2}$$

More generally, the same result holds for the weighted version with an arbitrary weight $\rho > 0$ defined in Section 3.2.

Remark 6.3. As for Theorem 1.1, the results can be stated as conditional convergence in distribution, see Remark 1.2. Let V_n be a random node in the random recursive tree T_n, as in Section 4 chosen with probability proportional to its weight (and thus uniformly when $\rho = 1$). Then (6.1) is equivalent to

$$\mathcal{L}\Big(\frac{X_{V_n} - m\log n}{\sqrt{\log n}} \,\Big|\, T_n, \{\eta_v\}_v\Big) \xrightarrow{\text{a.s.}} \mathcal{L}\big(N\big(0, \mathbb{E}[\eta\eta^{\mathsf{t}}]\big)\big), \qquad (6.3)$$

which can be written

Conditioned on $\{T_n\}$ and $\{\eta_v\}$, a.s, $\quad \dfrac{X_{V_n} - m\log n}{\sqrt{\log n}} \xrightarrow{\text{d}} N\big(0, \mathbb{E}[\eta\eta^{\mathsf{t}}]\big).$
$$(6.4)$$

The same applies to Theorem 6.2, and the binary trees in Section 10; we leave the details to the reader.

Note also that by unconditioning in (6.3)–(6.4), we obtain a (simpler) annealed version, cf. Remark 1.3.

Example 6.4. As a special case of the results above, let $\eta \equiv 1$ (deterministically). Then X_v is the depth of v, and thus Theorems 6.1 and 6.2 show that the distribution of node depths in a (weighted) random recursive tree or a Yule tree a.s. is asymptotically normal. Note that in this case, with the normalizations above, the limit is $N(0,1)$. (This is known, at least in the unweighted case, for example from [33] or [14, Remark 6.19 and (6.25)]; see also [14, Theorem 6.17] on the insertion depth, which is related to the annealed version, and the corresponding result for binary search trees and binary Yule trees in [12] and [13].)

This special case implies that in the asymptotic variance $\mathbb{E}[\eta\eta^{\mathsf{t}}]$ in the theorems above, we can interpret mm^{t} as coming from the random fluctuations of the depths; thus the contribution coming from the fluctuations of the offsets is $\mathbb{E}[\eta\eta^{\mathsf{t}}] - mm^{\mathsf{t}}$, which is the covariance matrix of η.

Remark 6.5. The same problem for conditioned Galton–Watson trees, which includes for example uniformly random plane trees and binary trees, has been studied by Aldous [2]; see also, e.g., [21]. The results for those random trees are very different from the present ones, with convergence in distribution to a non-random limit (known as ISE).

7. Proof of Theorem 6.2

In this section we prove Theorem 6.2; we then show in Section 8 that Theorem 6.1 is a simple consequence of Theorem 6.2. On the other hand, it is also easy to prove Theorem 6.1 directly using same arguments as for Theorem 6.2 with only minor modifications; we give in Section 8 also this, alternative, proof for comparison. (This is the method used by Mailler and Marckert [25], under somewhat stronger conditions.)

The basic idea is the same as in Uchiyama [33] (and in many other papers), although the details are different from the very general case in [33];

we construct a martingale from the characteristic function (Fourier transform) of the empirical distribution, and then use martingale theory to obtain a.s. uniform convergence of this martingale, leading to convergence of the characteristic function of suitable rescaled $\Theta_{a,b}(\widetilde{\mu})$. See further Remark 7.11.

We consider a branching random walk on the Yule tree $\mathcal{T}_t$ with offsets η_v as described in Sections 3–4. In this section, we assume until further notice (at the end of the section) that $d = 1$. (Actually, all formulas except (7.16) extend with at most notational differences to $d > 1$, but we do not use this; see also Remark 7.11.) We assume also, for simplicity, that $\rho = 1$, so that we consider the standard Yule tree; the minor modifications for a general ρ are discussed in Remark 7.9 after the lemmas.

Let $\mathcal{F}_t$ be the σ-field generated by all events (births and offsets) up to time t.

C denotes positive constants that may vary from one occurrence to the next. They may depend on the offset distribution, but not on n, t or other variables.

Denote the characteristic function of the offset distribution by

$$\varphi(s) := \mathbb{E}\, e^{\mathrm{i}s\eta}. \tag{7.1}$$

Fix throughout the proof $\delta > 0$ such that $\operatorname{Re}\varphi(s) \geqslant \frac{3}{4}$ when $|s| \leqslant \delta$, and let J be the interval $[-\delta, \delta]$.

Define the Fourier transform of a measure $\mu \in \mathcal{M}(\mathbb{R})$ by

$$\widehat{\mu}(s) = \int_{\mathbb{R}} e^{\mathrm{i}sx}\, \mathrm{d}\mu; \tag{7.2}$$

recall that this is the characteristic function if μ is a probability measure. Define the complex-valued random function, with μ_t given by (4.4) for $\mathcal{T}_t$,

$$F_t(s) := \widehat{\mu}_t(s) = \sum_{v \in \mathcal{T}_t} e^{\mathrm{i}sX_v}, \qquad s \in \mathbb{R}. \tag{7.3}$$

Note that $F_t(0) = |\mathcal{T}_t|$, and that $F_t(s)/F_t(0)$ is the characteristic function of the probability measure $\widetilde{\mu}_t$, see (7.31) below.

Although $F_t(s)$ is defined for all real s, we shall mainly consider F_t as a function on J. We begin by computing the first and second moments of $F_t(s)$.

Lemma 7.1. (i) *For every $t \geqslant 0$ and $s \in \mathbb{R}$,*

$$\mathbb{E}\, F_t(s) = e^{t\varphi(s)}. \tag{7.4}$$

(ii) *For every $t \geqslant 0$ and $s_1, s_2 \in J$,*

$$\mathbb{E}\big(F_t(s_1)F_t(s_2)\big) = \frac{\varphi(s_1) + \varphi(s_2)}{\varphi(s_1) + \varphi(s_2) - \varphi(s_1 + s_2)} e^{t(\varphi(s_1)+\varphi(s_2))}$$
$$- \frac{\varphi(s_1 + s_2)}{\varphi(s_1) + \varphi(s_2) - \varphi(s_1 + s_2)} e^{t\varphi(s_1+s_2)}. \tag{7.5}$$

Proof. (i): Each existing node v gets a new child, w say, with intensity 1, independently of the past. If this happens, then $F_t(s)$ increases by $e^{\mathrm{i}sX_w} = e^{\mathrm{i}s(X_v+\eta_w)}$. Since η_w independent of $\mathcal{F}_t$, the conditional expectation given $\mathcal{F}_t$ of this possible jump of $F_t(s)$ is $\mathbb{E}\big(e^{\mathrm{i}s(X_v+\eta_w)} \mid \mathcal{F}_t\big) = e^{\mathrm{i}sX_v}\varphi(s)$. It follows by standard Poisson process theory that, for any fixed $s \in \mathbb{R}$, there exists

a martingale M'_t (depending on s) such that, in the notation of stochastic calculus,

$$\mathrm{d}F_t(s) = \sum_{v \in \mathcal{T}_t} e^{\mathrm{i}sX_v} \varphi(s) \, \mathrm{d}t + \mathrm{d}M'_t = \varphi(s)F_t(s) \, \mathrm{d}t + \mathrm{d}M'_t. \qquad (7.6)$$

In particular, taking the expectation, we obtain

$$\frac{\partial}{\partial t} \mathbb{E} \, F_t(s) = \varphi(s) \, \mathbb{E} \, F_t(s). \qquad (7.7)$$

This differential equation, with the initial value $\mathbb{E} \, F_0(s) = 1$, has the solution (7.4).

(ii): Arguing as for part (i), we obtain for every real s_1, s_2

$$\frac{\partial}{\partial t} \mathbb{E}\big(F_t(s_1)F_t(s_2)\big)$$

$$= \mathbb{E} \sum_{v \in \mathcal{T}_t} \Big(\big(F_t(s_1) + e^{\mathrm{i}s_1(X_v + \eta_w)}\big)\big(F_t(s_2) + e^{\mathrm{i}s_2(X_v + \eta_w)}\big) - F_t(s_1)F_t(s_2) \Big)$$

$$= \mathbb{E}\Big(F_t(s_1)F_t(s_2)e^{\mathrm{i}s_2\eta_w} + F_t(s_1)F_t(s_2)e^{\mathrm{i}s_1\eta_w} + F_t(s_1 + s_2)e^{\mathrm{i}(s_1+s_2)\eta_w} \Big)$$

$$= \big(\varphi(s_1) + \varphi(s_2)\big) \mathbb{E}\big(F_t(s_1)F_t(s_2)\big) + \varphi(s_1 + s_2) \mathbb{E} \, F_t(s_1 + s_2). \qquad (7.8)$$

For $s_1, s_2 \in J$, we have

$$\mathrm{Re}\big(\varphi(s_1) + \varphi(s_2) - \varphi(s_1 + s_2)\big) \geqslant \frac{3}{4} + \frac{3}{4} - 1 = \frac{1}{2} > 0 \qquad (7.9)$$

and thus (7.8) has the solution (7.5), again recalling the initial valure $F_0(s) = 1$. $\qquad \square$

The proof is based on the following martingale.

Lemma 7.2. *Let*

$$M_t(s) := \frac{F_t(s)}{\mathbb{E} \, F_t(s)} = e^{-t\varphi(s)} F_t(s). \qquad (7.10)$$

Then $M_t(s)$, $t \geqslant 0$, is a (complex) martingale for every fixed $s \in \mathbb{R}$. Furthermore, for $s \in J$, this martingale is uniformly L^2-bounded:

$$\mathbb{E} \, |M_t(s)|^2 \leqslant C, \qquad t \geqslant 0, \, s \in J. \qquad (7.11)$$

Proof. The first part is standard: it follows from (7.6) above that $e^{-\varphi(s)t}F_t(s)$ is a martingale.

Furthermore, (7.10) and (7.5) yield, for $s_1, s_2 \in J$,

$$\mathbb{E}\big(M_t(s_1)M_t(s_2)\big) = \frac{\varphi(s_1) + \varphi(s_2)}{\varphi(s_1) + \varphi(s_2) - \varphi(s_1 + s_2)}$$

$$- \frac{\varphi(s_1 + s_2)}{\varphi(s_1) + \varphi(s_2) - \varphi(s_1 + s_2)} e^{t(\varphi(s_1+s_2)-\varphi(s_1)-\varphi(s_2))}.$$
$$(7.12)$$

Using (7.9), this yields the estimate, still for $s_1, s_2 \in J$ and all $t \geqslant 0$,

$$\big|\mathbb{E}\big(M_t(s_1)M_t(s_2)\big)\big| \leqslant 2\big|\varphi(s_1) + \varphi(s_2)\big| + 2\big|\varphi(s_1 + s_2)\big| \leqslant 6. \qquad (7.13)$$

Furthermore, $F_t(-s) = \overline{F_t(s)}$ and thus $M_t(-s) = \overline{M_t(s)}$; hence (7.11) follows by taking $s_1 = -s_2 = s$ in (7.13). $\qquad \square$

Remark 7.3. The proof shows also that, typically, $M_t(s)$ is not L^2-bounded for every real s, see (7.12), which holds as soon as the denominator $\varphi(s_1) + \varphi(s_2) - \varphi(s_1 + s_2) \neq 0$. Hence it is necessary to restrict to some interval J. (Our choice of J is not the largest possible, but it is convenient for our purposes.)

Lemma 7.2 implies that for every fixed $s \in J$, the martingale $M_t(s)$ converges a.s. A crucial step is to improve this to uniform convergence for all $s \in J$, i.e., a.s. convergence of M_t as an element of the Banach space $C(J)$. However, we shall not work in $C(J)$, since we find it difficult to estimate the first or second moment of $\|M_t\|_{C(J)} = \sup_{s \in J} |M_t(s)|$ directly; another technical problem is that $C(J)$ does not have the Radon–Nikodým property (see below). Instead we use the Sobolev space $W_1^2(J)$ defined by

$$W_1^2(J) := \{f \in L^2(J) : f' \in L^2(J)\}, \tag{7.14}$$

with the norm

$$\|f\|_{W_1^2(J)}^2 := \|f\|_{L^2(J)}^2 + \|f'\|_{L^2(J)}^2. \tag{7.15}$$

Remark 7.4. The definition (7.14) has to be interpreted with some care, since f in general is not differentiable everywhere. The general definition of Sobolev spaces in several variables [1; 28] uses distributional (weak) derivatives. In the present one-variable case, we can just require that f is absolutely continuous on J, so that f' exists a.e. in J in the usual sense, and then assume $f, f' \in L^2(J)$. Equivalently, $W_1^2(J) = \{f : f(x) = f(0) + \int_0^x g(y)\,\mathrm{d}y$ for some $g \in L^2(J)\}$.

Then $W_1^2(J)$ is an Hilbert space. Furthermore, there is a continuous inclusion $W_1^2(J) \subset C(J)$, and thus an estimate

$$\|f\|_{C(J)} \leqslant C\|f\|_{W_1^2(J)}, \qquad f \in W_1^2(J). \tag{7.16}$$

This is a special case of the Sobolev embedding theorem [1, Theorem 5.4]; in the present (one-variable) case, it is an easy consequence of the Cauchy–Schwarz inequality, which implies that if $f \in W_1^2(J)$ and $[a, b] \subseteq J$, then

$$|f(b) - f(a)| \leqslant \int_a^b |f'(x)|\,\mathrm{d}x \leqslant (b-a)^{1/2}\|f'\|_{L^2[a,b]} \leqslant (b-a)^{1/2}\|f\|_{W_1^2(J)}. \tag{7.17}$$

Hence $f \in C(J)$, and (7.16) follows easily from (7.17).

The (random) function F_t defined in (7.3) is an infinitely differentiable function of s. Furthermore, since we assume $\mathbb{E}\,|\eta|^2 < \infty$, the characteristic function $\varphi(s)$ is twice continuously differentiable; hence so is $\mathbb{E}\,F_t(s)$ by (7.4) and $M_t(s)$ by (7.10). In particular, $M_t \in W_1^2(J)$ for every $t \geqslant 0$.

Lemma 7.5. $(M_t)_{t \geqslant 0}$ is a right-continuous L^2-bounded martingale in $W_1^2(J)$.

Proof. We begin by estimating the norm. By (7.15) and Fubini's theorem,

$$\mathbb{E}\,\|M_t\|_{W_1^2(J)}^2 = \mathbb{E}\,\|M_t\|_{L^2(J)}^2 + \mathbb{E}\,\|M_t'\|_{L^2(J)}^2$$

$$= \mathbb{E}\int_J |M_t(s)|^2\,\mathrm{d}s + \mathbb{E}\int_J \left|\frac{\partial}{\partial s}M_t(s)\right|^2\,\mathrm{d}s$$

$$= \int_J \mathbb{E}\,|M_t(s)|^2\,\mathrm{d}s + \int_J \mathbb{E}\left|\frac{\partial}{\partial s}M_t(s)\right|^2\,\mathrm{d}s. \tag{7.18}$$

We know that $\mathbb{E}\,|M_t(s)|^2$ is bounded by (7.11), but it remains to estimate the last integral.

Denote the right-hand side of (7.12) by $h(s_1, s_2; t)$, so $\mathbb{E}\big(M_t(s_1)M_t(s_2)\big) = h(s_1, s_2; t)$ when $s_1, s_2 \in J$. Taking partial derivatives with respect to both s_1 and s_2, we obtain, for $s_1, s_2 \in J^\circ$, the interior of J,

$$\frac{\partial^2}{\partial s_1 \partial s_2}\,\mathbb{E}\big(M_t(s_1)M_t(s_2)\big) = \frac{\partial^2}{\partial s_1 \partial s_2}h(s_1, s_2; t), \qquad (7.19)$$

where the right-hand side exists because (7.9) holds and, as said above, φ is twice continuously differentiable; furthermore, this implies, using the explicit form of $h(s_1, s_2; t)$ in (7.12),

$$\frac{\partial^2}{\partial s_1 \partial s_2}h(s_1, s_2; t) = O\Big(1 + (1+t^2)e^{t\,\mathrm{Re}(\varphi(s_1+s_2)-\varphi(s_1)-\varphi(s_2))}\Big)$$

$$= O\Big(1 + (1+t^2)e^{-t/2}\Big) = O(1), \qquad s_1, s_2 \in J^\circ. \quad (7.20)$$

In the left-hand side of (7.19) we interchange the order of differentiation and expectation. To justify this, we note first that by (7.3), $|F_t(s)| \leqslant |\mathcal{T}_t|$ and $|\frac{\partial}{\partial s}F_t(s)| \leqslant \sum_{v\in\mathcal{T}_t}|X_v|$. Hence, using (7.10) and Lemmas 3.1 and 4.3, and letting C_t denote constants that may depend on t but not on s_1, s_2,

$$\mathbb{E}\Big|\frac{\partial^2}{\partial s_1 \partial s_2}\big(M_t(s_1)M_t(s_2)\big)\Big| = \mathbb{E}\Big|\frac{\partial M_t(s_1)}{\partial s_1}\frac{\partial M_t(s_2)}{\partial s_2}\Big|$$

$$\leqslant C_t\,\mathbb{E}\Big(\Big(|F_t(s_1)| + \Big|\frac{\partial}{\partial s}F_t(s_1)\Big|\Big)\Big(|F_t(s_2)| + \Big|\frac{\partial}{\partial s}F_t(s_2)\Big|\Big)\Big)$$

$$\leqslant C_t\,\mathbb{E}\Big(|\mathcal{T}_t| + \sum_{v\in\mathcal{T}_t}|X_v|\Big)^2 \leqslant C_t\,\mathbb{E}\,|\mathcal{T}_t|^2 + C_t\,\mathbb{E}\Big(\sum_{v\in\mathcal{T}_t}|X_v|\Big)^2 \leqslant C_t < \infty.$$

$$(7.21)$$

Consequently, if $[a, b]$ and $[c, d]$ are any two subintervals of J°, then Fubini's theorem yields

$$\int_a^b \int_c^d \mathbb{E}\Big(\frac{\partial M_t(s_1)}{\partial s_1}\frac{\partial M_t(s_2)}{\partial s_2}\Big)\,\mathrm{d}s_1\,\mathrm{d}s_2 = \mathbb{E}\int_a^b\int_c^d \frac{\partial M_t(s_1)}{\partial s_1}\frac{\partial M_t(s_2)}{\partial s_2}\,\mathrm{d}s_1\,\mathrm{d}s_2$$

$$= \mathbb{E}\big(\big(M_t(b) - M_t(a)\big)\big(M_t(d) - M_t(c)\big)\big)$$

$$= h(b, d; t) - h(a, d; t) - h(b, c; t) + h(a, c; t) \qquad (7.22)$$

and differentiation (with respect to b and d) yields the desired formula

$$\mathbb{E}\Big(\frac{\partial M_t(s_1)}{\partial s_1}\frac{\partial M_t(s_2)}{\partial s_2}\Big) = \frac{\partial^2 h(s_1, s_2; t)}{\partial s_1 \partial s_2}, \qquad s_1, s_2 \in J^\circ. \qquad (7.23)$$

Together with (7.20), this shows, for $s \in J^\circ$,

$$\mathbb{E}\Big|\frac{\partial M_t(s)}{\partial s}\Big|^2 = -\mathbb{E}\Big(\frac{\partial M_t}{\partial s}(s)\frac{\partial M_t}{\partial s}(-s)\Big) = -\frac{\partial^2 h(s_1, s_2; t)}{\partial s_1 \partial s_2}\Big|_{s_1=s, s_2=-s}$$

$$= O(1). \qquad (7.24)$$

Finally, we use (7.24) together with (7.11) in (7.18), and find

$$\mathbb{E}\,\|M_t\|_{W_1^2(J)}^2 \leqslant C, \qquad t \geqslant 0. \qquad (7.25)$$

In other words, $\{M_t\}$ is an L^2-bounded family of random variables in $W_1^2(J)$.

In particular, each M_t is integrable, and thus the conditional expectation $\mathbb{E}(M_t \mid \mathcal{F}_u)$ is defined for every $u \leqslant t$. Point evaluations are continuous linear functionals on $W_1^2(J)$ by (7.16). Hence, if $0 \leqslant u \leqslant t$ and $s \in J$, then, using also that $M_t(s)$ is a martingale by Lemma 7.2,

$$\mathbb{E}(M_t \mid \mathcal{F}_u)(s) = \mathbb{E}(M_t(s) \mid \mathcal{F}_u) = M_u(s). \tag{7.26}$$

Consequently, $\mathbb{E}(M_t \mid \mathcal{F}_u) = M_u$, and thus M_t, $t \geqslant 0$, is a martingale with values in $W_1^2(J)$. We have shown L^2-boundedness in (7.25). Finally, $t \mapsto M_t$ is right-continuous by the definition (7.10), since F_t is a right-continuous step function. $\qquad \square$

Lemma 7.6. *There exists a random function $M_\infty \in W_1^2(J) \subset C(J)$ such that $M_t \xrightarrow{\text{a.s.}} M_\infty$ in $W_1^2(J)$ as $t \to \infty$.*

Proof. This follows from Lemma 7.5 since $W_1^2(J)$ is a Hilbert space and thus has the Radon–Nikodým property, see [30, Theorem 2.9 and Corollary 2.15], using [30, Theorem 1.49] to extend the result from the discrete-parameter martingale $(M_n)_{n \in \mathbb{N}}$ to the continuous-parameter $(M_t)_{t \geqslant 0}$. $\qquad \square$

Lemma 7.7. *As $t \to \infty$, $M_t \xrightarrow{\text{a.s.}} M_\infty$ in $C(J)$, i.e., a.s. $M_t(s) \to M_\infty(s)$ uniformly in $s \in J$.*

Proof. An immediate consequence of Lemma 7.6 and (7.16). $\qquad \square$

Lemma 7.8. $M_\infty(0) > 0$ *a.s.*

Proof. By (7.10) and (7.3),

$$M_t(0) = e^{-t} F_t(0) = e^{-t} |\mathcal{T}_t|, \tag{7.27}$$

which a.s. converges to the strictly positive limit W by (3.2). $\qquad \square$

Remark 7.9. We have so far assumed that $\rho = 1$. The results easily extend to general $\rho > 0$, provided we, as in Sections 3–4, count the nodes according to their weights ω_v and thus change (7.3) to

$$F_t(s) := \sum_{v \in \mathcal{T}_t} \omega_v e^{\mathrm{i} s X_v} = \rho + \sum_{v \neq o} e^{\mathrm{i} s X_v}. \tag{7.28}$$

Recall that $|\mathcal{T}_t|$ now is the total weight; thus $F_t(0) = |\mathcal{T}_t|$ still holds. Similarly, recalling (4.3), the characteristic function of $\widetilde{\mu}_t$ is still $F_t(s)/F_t(0)$. In the proof of Lemma 7.1, we then obtain the same differential equations (7.7) and (7.8), but the initial condition is now $F_0(s) = \rho$, giving

$$\mathbb{E}\, F_t(s) = \rho e^{t\varphi(s)} \tag{7.29}$$

and

$$\mathbb{E}\big(F_t(s_1)F_t(s_2)\big) = \rho^2 e^{t(\varphi(s_1)+\varphi(s_2))} + \rho \frac{\varphi(s_1 + s_2)}{\varphi(s_1) + \varphi(s_2) - \varphi(s_1 + s_2)} \times \\ \big(e^{t(\varphi(s_1)+\varphi(s_2))} - e^{t\varphi(s_1+s_2)}\big). \tag{7.30}$$

We define $M_t(s) := F_t(s)/\mathbb{E}\, F_t(s) = \rho^{-1} e^{-t\varphi(s)} F_t(s)$ and obtain again the estimate (7.11). The rest of the proofs above holds without changes; in particular, Lemmas 7.5–7.8 hold as stated for any $\rho > 0$.

Proof of Theorem 6.2. We first continue to assume $d = 1$, but allow $\rho > 0$ to be arbitrary, see Remark 7.9.

Let $\widetilde{\mu}_t$ be the empirical distribution (4.2) for the Yule tree $\mathcal{T}_t$ at time $t \geqslant 0$. Denote the characteristic function of $\widetilde{\mu}_t$ by $\widehat{\widetilde{\mu}}_t$. Then, by (4.2), (7.3) and (7.10) when $\rho = 1$, and their modifications (4.3) and (7.28) in general, using $\varphi(0) = 1$,

$$\widehat{\widetilde{\mu}}_t(s) = \frac{1}{|\mathcal{T}_t|} \sum_{v \in \mathcal{T}_t} \omega_v e^{\mathrm{i}sX_v} = \frac{F_t(s)}{F_t(0)} = \frac{M_t(s)}{M_t(0)} e^{t(\varphi(s)-1)}. \qquad (7.31)$$

If $s(t)$ is a function of t such that $s(t) \to 0$ as $t \to \infty$, then the uniform convergence in Lemma 7.7 together with the continuity of M_∞ implies that $M_t(s(t)) \xrightarrow{\text{a.s.}} M_\infty(0)$, and thus, using also Lemma 7.8, $M_t(s(t))/M_t(0) \xrightarrow{\text{a.s.}} M_\infty(0)/M_\infty(0) = 1$. Consequently, (7.31) shows that a.s., as $t \to \infty$,

$$\widehat{\widetilde{\mu}}_t(s(t)) = \big(1 + o(1)\big) e^{t(\varphi(s(t))-1)} = e^{t(\varphi(s(t))-1)+o(1)}. \qquad (7.32)$$

Now consider the rescaled measure $\Theta_{a,b}(\widetilde{\mu}_t)$, where we take $a = a(t) := \sqrt{t}$ and $b = b(t) := mt$. By (1.4), its characteristic function is given by

$$\widehat{\Theta_{a,b}(\widetilde{\mu}_t)}(s) = \frac{1}{|\mathcal{T}_t|} \sum_{v \in \mathcal{T}_t} e^{\mathrm{i}s(X_v-b)/a} = e^{-\mathrm{i}sb/a} \widehat{\widetilde{\mu}}_t(s/a). \qquad (7.33)$$

For any fixed $s \in \mathbb{R}$, $s/a = s/\sqrt{t} \to 0$ as $t \to \infty$, and thus (7.32) applies with $s(t) := s/a = s/\sqrt{t}$; hence (7.33) yields, a.s.,

$$\widehat{\Theta_{a,b}(\widetilde{\mu}_t)}(s) = e^{-\mathrm{i}sb/a} e^{t(\varphi(s/a)-1)+o(1)} = e^{t(\varphi(s/\sqrt{t})-1-\mathrm{i}ms/\sqrt{t})+o(1)}. \qquad (7.34)$$

Furthermore, $\varphi(s/\sqrt{t}) = 1 + \mathrm{i}ms/\sqrt{t} - \frac{1}{2} \mathbb{E}[\eta^2]s^2/t + o(t^{-1})$, and thus a.s., as $t \to \infty$,

$$\widehat{\Theta_{a,b}(\widetilde{\mu}_t)}(s) = e^{-\frac{1}{2} \mathbb{E}[\eta^2]s^2+o(1)} \to e^{-\frac{1}{2} \mathbb{E}[\eta^2]s^2}. \qquad (7.35)$$

This shows that for each fixed s, the characteristic function of $\Theta_{a,b}(\widetilde{\mu}_t)$ converges a.s. to the characteristic function of $N(0, m^2 + \sigma^2)$.

Consider now an arbitrary $d \geqslant 1$. For any fixed $u \in \mathbb{R}^d$, consider the linear projections $u^{\mathrm{t}}X_v \in \mathbb{R}$, which are obtained as in (4.1) from the offsets $u^{\mathrm{t}}\eta_v$. Applying (7.35) (with $s = 1$) to these variables yields

$$\widehat{\Theta_{a,b}(\widetilde{\mu}_t)}(u) \xrightarrow{\text{a.s.}} e^{-\frac{1}{2} \mathbb{E}[(u^{\mathrm{t}}\eta)^2]} = e^{-\frac{1}{2} u^{\mathrm{t}} \mathbb{E}[\eta\eta^{\mathrm{t}}]u} \qquad (7.36)$$

for every fixed $u \in \mathbb{R}^d$. This implies that $\Theta_{a,b}(\widetilde{\mu}_t) \xrightarrow{\text{d}} \mathcal{L}\big(N\big(0, \mathbb{E}[\eta\eta^{\mathrm{t}}]\big)\big)$ a.s., see [7]. $\qquad\square$

Remark 7.10. The final part of the proof, from (7.32), is very similar to standard proofs of the central limit theorem, and we can interpret (7.32) as showing that $\widetilde{\mu}_t$ asymptotically is like the distribution of a sum of independent copies of η; note that (7.32) says that $\widehat{\widetilde{\mu}}_t(s) \approx e^{t(\varphi(s)-1)}$, which is the characteristic function of a sum of a random $\mathrm{Po}(t)$ number of independent copies of η.

Remark 7.11. The main idea in the proof above, as in many other related works, including [25], is to obtain uniform convergence of certain random functions in some interval $J \ni 0$, i.e., convergence in $C(J)$, since this allows us to obtain convergence for an argument $s = s(t)$ depending on t.

(Lemma 7.7.) Pointwise convergence a.s. to a random function follows in our cases, and in many related problems, from the martingale limit theorem. One method to improve this to uniform convergence goes back to Joffe, Le Cam and Neveu [22]; the idea is to use the Kolmogorov continuity criterion [23, Theorem 3.23] to show that the limit can be taken as a continuous function; then uniform convergence follows by the martingale convergence theorem in a space of continuous functions. We have here chosen a slightly different method; we use a Sobolev space $W_1^2(J)$ and show that the martingale is bounded there; then both existence of the limit and uniform convergence follows. Nevertheless the methods are quite similar; the first requires estimates of moments of differences while the second requires estimates of moments of derivatives, and the required estimates are similar. Hence, for practical applications, the two methods seem to be essentially equivalent.

Uchiyama [33] uses pointwise convergence of random functions; he does not explicitly show uniform convergence, but he too uses estimates on moments of differences, in a way which seems related.

Another method to obtain uniform continuity, see Biggins [8, 9], assumes that the random functions are analytic functions in an open domain in the complex plane. Then pointwise estimates of moments yield automatically (by Cauchy's estimates) uniform estimates of the functions and their derivatives on compact sets, and thus uniform convergence on compact subsets. (It may be convenient to use the Bergman space of square integrable analytic functions in a suitable domain, cf. [15].) This is the method used by Mailler and Marckert [25] for the Pólya urns and random recursive trees discussed in the present paper; it is elegant but it requires in our case exponential moments of the offset distribution so that its characteristic function can be extended to an analytic functions in a complex domain.

Chauvin, Drmota and Jabbour-Hattab [12] (for binary trees) use a combination of both the Kolmogorov criterion and properties of analytic functions.

Note that also when the offset is vector-valued and takes values in $\mathbb{R}^d$ with $d > 1$, we consider one-dimensional projections and use a one-dimensional Sobolev space. We may define $M_t(s)$ as above for $s \in \mathbb{R}^d$ and show, by the same arguments, that $M_t(s)$ is a L^2-bounded martingale in the Sobolev space $W_1^2(B)$, for a small ball $B \subset \mathbb{R}^d$. However, we cannot use this to claim convergence in $C(B)$ (or in a smaller ball), since the Sobolev imbedding theorem in higher dimensions require more derivatives, see [1, Theorem 5.4]. (One might use a Sobolev space $W_k^2(B)$ with more derivatives, but that would require more moments for the offset distribution, apart from complicating the proof.)

8. Proof of Theorem 6.1

Proof of Theorem 6.1. Since the (weighted) random recursive tree can be realized as the (weighted) Yule tree at the stopping times τ_n defined by (3.1), we obtain by taking $t = \tau_n$ in (6.2), where $\widetilde{\mu}_n$ now refers to the random recursive tree,

$$\Theta_{\sqrt{\tau_n},\, m\tau_n}(\widetilde{\mu}_n) \xrightarrow{\text{a.s.}} \mathcal{L}\left(N\left(0, \mathbb{E}[\eta\eta^{\mathsf{t}}]\right)\right). \tag{8.1}$$

Combined with (3.4), this implies (6.1). To see this in detail, we can write (8.1) in the form (6.4): conditioned on $\{\mathcal{T}_t\}$ and $\{\eta_v\}$, a.s,

$$\frac{X_{V_n} - m\log\tau_n}{\sqrt{\log\tau_n}} \xrightarrow{\mathrm{d}} N\big(0, \mathbb{E}[\eta\eta^{\mathsf{t}}]\big). \tag{8.2}$$

This implies (6.4) by (3.4) and the Cramér–Slutsky theorem (still conditioning on $\{\mathcal{T}_t\}$ and $\{\eta_v\}$). □

We have chosen to prove Theorem 6.1 using the continuous-time Yule tree. However, it is also possible to argue directly in discrete time in the same way. We find it interesting to sketch this version of the argument too, for comparison, leaving some details to the reader; see also the proof of [25, Theorem 1.6], where the result is proved under stronger assumptions using similar and partly the same arguments. We consider an arbitrary $\rho > 0$.

Define $F_n(s)$ as in (7.28). Then, labelling the nodes in order of appearance, $F_{n+1}(s) = F_n(s) + e^{\mathrm{i}sX_{n+1}}$ and thus, cf. (7.6),

$$\mathbb{E}\big(F_{n+1}(s)\mid\mathcal{F}_n\big) = F_n(s) + \frac{1}{n+\rho}\sum_{v\in T_n}\omega_v e^{\mathrm{i}sX_v}\varphi(s) = F_n(s) + \frac{\varphi(s)}{n+\rho}F_n(s)$$

$$= \frac{n+\rho+\varphi(s)}{n+\rho}F_n(s). \tag{8.3}$$

In particular,

$$\mathbb{E}\,F_{n+1}(s) = \frac{n+\rho+\varphi(s)}{n+\rho}\,\mathbb{E}\,F_n(s) \tag{8.4}$$

and thus by induction, since $F_0(s) = \rho$,

$$\mathbb{E}\,F_n(s) = \frac{\Gamma(\rho+1)}{\Gamma(\rho+\varphi(s))}\frac{\Gamma(n+\rho+\varphi(s))}{\Gamma(n+\rho)}. \tag{8.5}$$

While this formula is more complicated than (7.29), it is still easy to get asymptotics. As a well-known consequence of Stirling's formula, see [29, (5.11.12)], for any complex a, b,

$$\frac{\Gamma(n+a)}{\Gamma(n+b)} = \big(1+o(1)\big)n^{a-b}, \qquad n\to\infty. \tag{8.6}$$

Hence,

$$\mathbb{E}\,F_n(s) = \big(1+o(1)\big)\frac{\Gamma(\rho+1)}{\Gamma(\rho+\varphi(s))}n^{\varphi(s)}, \qquad n\to\infty. \tag{8.7}$$

For the second moment, we similarly obtain, cf. (7.8),

$$\mathbb{E}\big(F_{n+1}(s_1)F_{n+1}(s_2)\mid\mathcal{F}_n\big) = F_n(s_1)F_n(s_2) + \frac{1}{n+\rho}\Big(\varphi(s_1)F_n(s_1)F_n(s_2)$$

$$+ \varphi(s_2)F_n(s_1)F_n(s_2) + \varphi(s_1+s_2)F_n(s_1+s_2)\Big)$$

$$= \frac{n+\rho+\varphi(s_1)+\varphi(s_2)}{n+\rho}F_n(s_1)F_n(s_2) + \frac{\varphi(s_1+s_2)}{n+\rho}F_n(s_1+s_2) \tag{8.8}$$

and thus, using induction and (8.5),

$$\mathbb{E}\big(F_n(s_1)F_n(s_2)\big) = \frac{\Gamma(n+\rho+\varphi(s_1)+\varphi(s_2)}{\Gamma(n+\rho)}\left(\frac{\rho^2\Gamma(\rho)}{\Gamma(\rho+\varphi(s_1)+\varphi(s_2))}\right.$$

$$+ \sum_{k=1}^{n} \frac{\Gamma(k + \rho + \varphi(s_1 + s_2) - 1)}{\Gamma(k + \rho + \varphi(s_1) + \varphi(s_2))} \frac{\Gamma(\rho + 1)\varphi(s_1 + s_2)}{\Gamma(\rho + \varphi(s_1 + s_2))} \Big). \quad (8.9)$$

We define, similarly to (7.10), $M_n(s) := F_n(s)/\mathbb{E}\, F_n(s)$ (at least for $s \in J$, which implies $F_n(s) \neq 0$ by (8.5)). It follows from (8.3) that $M_n(s)$ is a martingale for each s. Furthermore, it follows from (8.9), (8.6) and (8.7) that, uniformly for $s_1, s_2 \in J$, recalling (7.9),

$$\mathbb{E}\big(M_n(s_1)M_n(s_2)\big) = O\Big(1 + \sum_{k=1}^{n} k^{-3/2}\Big) = O(1). \quad (8.10)$$

Moreover, (8.6) holds uniformly for a and b in any fixed bounded sets in $\mathbb{C}$, and thus (using Cauchy's estimates), we can differentiate (8.6) with respect to a and b, arbitrarily many times. It thus follows from (8.9) and (8.5) that also, for $s_1, s_2 \in J^\circ$,

$$\mathbb{E}\Big(\frac{\partial M_n(s_1)}{\partial s_1} \frac{\partial M_n(s_2)}{\partial s_2}\Big) = O\Big(1 + \sum_{k=1}^{n} \frac{\log^2 k}{k^{3/2}}\Big) = O(1). \quad (8.11)$$

Hence, cf. (7.23)–(7.25), M_n is an L^2-bounded martingale in $W_1^2(J)$. We can now argue as in Section 7, and obtain for any $s(n) \to 0$, cf. (7.32),

$$\widehat{\widetilde{\mu}}_n(s(n)) = \big(1 + o(1)\big) n^{\varphi(s(n))-1} = e^{\log n(\varphi(s(n))-1)+o(1)}, \quad (8.12)$$

and the proof is completed as in Section 7. (A minor simplification is that now $F_n(0) = n + \rho$ and thus $M_n(0) = 1$ deterministically, so Lemma 7.8 is not needed.)

We thus see that there is no essential difference between using the discrete time random recursive tree or the continuous time Yule tree in our arguments. However, the formulas for the first and, in particular, second moments are much simpler in the continuous time case, which leads to simpler calculations in order to obtain the desired estimates.

9. Proof of Theorem 1.1

We extend the notations (7.2) and (7.1) to $\mathbb{R}^d$. Let $a_n := \sqrt{\log n}$ and $b_n := m \log n$.

Proof of Theorem 1.1. *Step 1:* Consider first an SBARW Pólya urn, with initial composition $\rho\delta_0$, for some $\rho > 0$. Then, as shown in Section 5.1, the sequence μ_n is the same as for a weighted random recursive tree, and thus (1.6) follows from Theorem 6.1.

Step 2: Consider now a DARW Pólya urn μ_n with initial composition $\mu_0 = \rho\nu = \rho\delta_0 * \nu$ for some $\rho \geqslant 0$. Then Lemma 2.2 and (2.5) show that the urn is given by $\mu_n = \mu'_n * \nu$ for an SBARW Polya urn μ'_n, with the same offset distribution ν and initial composition $\mu'_0 = \rho\delta_0$. Hence, $\widetilde{\mu}_n = \widetilde{\mu}'_n * \nu$, and thus, cf. (7.33), for every $s \in \mathbb{R}^d$,

$$\widehat{\Theta_{a_n,b_n}(\widetilde{\mu}_n)}(s) = e^{-\mathrm{i}s\cdot b_n/a_n} \widehat{\widetilde{\mu}}_n\Big(\frac{s}{a_n}\Big) = e^{-\mathrm{i}s\cdot b_n/a_n} \widehat{\widetilde{\mu}'_n}\Big(\frac{s}{a_n}\Big) \widehat{\nu}\Big(\frac{s}{a_n}\Big)$$

$$= \widehat{\Theta_{a_n,b_n}(\widetilde{\mu}'_n)}(s)\, \widehat{\nu}\Big(\frac{s}{a_n}\Big). \quad (9.1)$$

Here, $\widehat{\nu}(s/a_n) \to 1$ as $n \to \infty$ since $a_n \to \infty$, and Step 1 shows that a.s., for all s,

$$\widehat{\Theta_{a_n,b_n}(\widetilde{\mu}'_n)}(s) \to e^{-\frac{1}{2}\mathbb{E}[\eta^2]s^2}. \tag{9.2}$$

Hence, (9.1) yields a.s., for all $s \in \mathbb{R}^d$,

$$\widehat{\Theta_{a_n,b_n}(\widetilde{\mu}_n)}(s) \to e^{-\frac{1}{2}\mathbb{E}[\eta^2]s^2}, \tag{9.3}$$

which proves (1.6) in this case.

Step 3: Consider now a DARW Pólya urn μ_n, with arbitrary initial composition $\mu_0 \in \mathcal{M}_*(\mathbb{R}^d)$. Let $\rho := \mu_0(\mathbb{R}^d)$, and consider a Pólya urn with the same offset distribution ν but initial composition $\mu'_0 := \rho\nu$. Then Step 2 applies to the second urn, which shows that a.s., for all $s \in \mathbb{R}^d$,

$$\widehat{\Theta_{a_n,b_n}(\widetilde{\mu}'_n)}(s) \to e^{-\frac{1}{2}\mathbb{E}[\eta^2]s^2}. \tag{9.4}$$

By Lemma 9.1 below, this implies

$$\widehat{\Theta_{a_n,b_n}(\widetilde{\mu}_n)}(s) \xrightarrow{\text{a.s.}} e^{-\frac{1}{2}\mathbb{E}[\eta^2]s^2}, \tag{9.5}$$

for every $s \in \mathbb{R}^d$, which implies (1.6). (We may again use [7], or note that the proof of Lemma 9.1 shows that a.s. (9.7) holds for all s simultaneously.)

Step 4: Finally, consider an SBARW Pólya urn μ'_n with arbitrary initial composition. We use Lemma 2.2 and construct a corresponding DARW Pólya urn $\mu_n = \mu'_n * \nu$. Step 3 applies to the latter urn, which shows (9.3) a.s., and then (9.1) shows (9.2) which proves the result in this case.

This completes the proof of Theorem 1.1, assuming the lemma below. $\square$

Lemma 9.1. *Let μ_n and μ'_n be two DARW Pólya urn processes, with the same offset distribution $\nu \in \mathcal{P}(\mathbb{R}^d)$ but (possibly) different initial distributions μ_0 and μ'_0. Assume that $\mu_0(\mathbb{R}^d) = \mu'_0(\mathbb{R}^d)$. Then, for any sequence $s_n \in \mathbb{R}^d$ with $s_n \to 0$, as $n \to \infty$,*

$$\left|\widehat{\widetilde{\mu}_n}(s_n) - \widehat{\widetilde{\mu}'_n}(s_n)\right| \xrightarrow{\text{a.s.}} 0. \tag{9.6}$$

Hence, for every $s \in \mathbb{R}^d$,

$$\left|\widehat{\Theta_{a_n,b_n}(\widetilde{\mu}_n)}(s) - \widehat{\Theta_{a_n,b_n}(\widetilde{\mu}'_n)}(s)\right| \xrightarrow{\text{a.s.}} 0. \tag{9.7}$$

Proof. We use the coupling with a weighted random recursive tree T_n in Section 5.2, and may assume that both Pólya urns are defined by (P1)–(P2) from the same trees T_n. We use the notations $\Delta\mu'_v$ and Z'_v for the second urn. Let $\Delta Z_v := Z'_v - Z_v \in \mathbb{R}^d$.

Recall that by (5.2), for any $v \neq o$ and conditioned on Z_v, we have

$$\Delta\mu_v = \mathcal{L}(Z_v + \eta). \tag{9.8}$$

Hence, by (P2), if v has a mother $u \neq o$, then, conditioned on the tree process (T_n) and on Z_u and Z'_u, Z_v has the distribution $\mathcal{L}(\eta + Z_u)$, i.e., $Z_v \overset{\text{d}}{=} \eta + Z_u$, and

$$Z'_v \overset{\text{d}}{=} \eta + Z'_u = \eta + Z_u + \Delta Z_u \overset{\text{d}}{=} Z_v + \Delta Z_u. \tag{9.9}$$

Hence, we may couple the two urn processes such that when v has a mother $u \neq o$, then

$$\Delta Z_v := Z'_v - Z_v = \Delta Z_u. \tag{9.10}$$

If v is a daughter of o, we may, for example, choose Z_v and Z_v' independent.

Note also that (5.2) implies,

$$\widehat{\Delta\mu_v}(s) = \mathbb{E}\big(e^{\mathrm{i}s\cdot(Z_v+\eta)} \mid Z_v\big) = e^{\mathrm{i}s\cdot Z_v}\,\mathbb{E}\,e^{\mathrm{i}s\cdot\eta} = e^{\mathrm{i}s\cdot Z_v}\varphi(s). \tag{9.11}$$

For any node $v \neq o$, let $u(v)$ be the ancestor of v that is a daughter of o. (With $u(v) = v$ if v is a daughter of o.) Then, repeated application of (9.10) shows that $\Delta Z_v = \Delta Z_{u(v)}$. Consequently, for $v \neq o$,

$$\begin{aligned}
\big|\widehat{\Delta\mu_v}(s) - \widehat{\Delta\mu_v'}(s)\big| &= \big|e^{\mathrm{i}s\cdot Z_v}\varphi(s) - e^{\mathrm{i}s\cdot Z_v'}\varphi(s)\big| = \big|e^{\mathrm{i}s\cdot Z_v} - e^{\mathrm{i}s\cdot Z_v'}\big||\varphi(s)| \\
&\leqslant \big|e^{\mathrm{i}s\cdot Z_v} - e^{\mathrm{i}s\cdot Z_v'}\big| = \big|e^{\mathrm{i}s\cdot\Delta Z_v} - 1\big| = \big|e^{\mathrm{i}s\cdot\Delta Z_{u(v)}} - 1\big|. \tag{9.12}
\end{aligned}$$

We sum (9.12) over all $v \in T_n$ with $v \neq o$. Let D be the set of nodes in T_∞ that are daughters of o, and for each $u \in D$, let $N_n(u)$ be the number of descendants of u in T_n. Then, with $\rho = \mu_0(\mathbb{R}) = \mu_0'(\mathbb{R})$,

$$\begin{aligned}
\big|\widehat{\mu_n}(s) - \widehat{\mu_n'}(s)\big| &\leqslant \sum_{v\in T_n}\big|\widehat{\Delta\mu_v}(s) - \widehat{\Delta\mu_v'}(s)\big| \\
&\leqslant \big|\widehat{\Delta\mu_0}(s) - \widehat{\Delta\mu_0'}(s)\big| + \sum_{o\neq v\in T_n}\big|e^{\mathrm{i}s\cdot\Delta Z_{u(v)}} - 1\big| \\
&\leqslant 2\rho + \sum_{u\in D} N_n(u)\big|e^{\mathrm{i}s\cdot\Delta Z_u} - 1\big|. \tag{9.13}
\end{aligned}$$

Dividing by $|T_n| = n + \rho$ we find for the normalized distributions

$$\big|\widehat{\widetilde{\mu}_n}(s) - \widehat{\widetilde{\mu}_n'}(s)\big| \leqslant \frac{2\rho}{n} + \sum_{u\in D}\frac{N_n(u)}{n}\big|e^{\mathrm{i}s\cdot\Delta Z_u} - 1\big|. \tag{9.14}$$

Number the elements of D as $u_1, u_2, \ldots$ (in order of appearance). It is well known, at least in the unweighted case, that the relative sizes $N_n(u_k)/n$ of the branches converge a.s. as $n \to \infty$, say

$$N_n(u_k)/n \xrightarrow{\text{a.s.}} V_k, \qquad k \geqslant 1. \tag{9.15}$$

In fact, this is an instance of the Chinese restaurant process with parameters (seating plan) $(0, \rho)$, see [31, Section 3.2]; take one table for each $u \in D$ and all its descendants, and then the probability that node $n+1$ joins a table with n_k nodes is $n_k/(n + \rho)$, and the probability that it starts a new table (and thus belongs to D) is $\rho/(n+\rho)$. Hence, (9.15) follows from [31, Theorem 3.2], which furthermore shows that $V_k = W_k \prod_{j=1}^{k-1}(1 - W_j)$ for an i.i.d. sequence $W_j \sim \text{Beta}(1, \rho)$. (This result can also be proved easily using a sequence of classical two-colour Pólya urns, cf. e.g. [19, Appendix A].) It follows that, a.s.,

$$\sum_{k=1}^{\infty} V_k = 1 - \prod_{j=1}^{\infty}(1 - W_j) = 1, \tag{9.16}$$

and thus the limits V_k in (9.15) form a random probability distribution on $\mathbb{N}$. (The random distribution $\text{GEM}(0, \rho)$ [31].) For a discrete distributions, pointwise convergence of the probabilities to a limiting distribution

is equivalent to convergence in total variation, see [18, Theorem 5.6.4], and thus (9.15) and (9.16) imply

$$\sum_{k=1}^{\infty} \left| \frac{N_n(u_k)}{n} - V_k \right| \xrightarrow{\text{a.s.}} 0. \tag{9.17}$$

We now return to (9.14) and find, for any sequence $s_n \in \mathbb{R}^d$,

$$\left| \widehat{\widetilde{\mu}_n}(s_n) - \widehat{\widetilde{\mu}'_n}(s_n) \right| \leqslant \frac{2\rho}{n} + \sum_{k=1}^{\infty} V_k \left| e^{\mathrm{i} s_n \cdot \Delta Z_{u_k}} - 1 \right| + \sum_{k=1}^{\infty} 2 \left| \frac{N_n(u_k)}{n} - V_k \right|. \tag{9.18}$$

Here the last sum converges to 0 a.s. by (9.17). Furthermore, note that ΔZ_{u_k} does not depend on n. Hence, if $s_n \to 0$, then the first sum converges to 0 a.s., by (9.16) and dominated convergence. Consequently, (9.6) follows from (9.18). Finally, (9.7) follows because $\widehat{\Theta_{a_n,b_n}(\widetilde{\mu}_n)}(s) = e^{-\mathrm{i} s \cdot b_n/a_n} \widehat{\widetilde{\mu}_n}(s/a_n)$ and similarly for $\widetilde{\mu}'_n$. $\qquad\square$

10. Binary trees

Theorems 6.1–6.2 have analogues for binary search trees and binary Yule trees. (In fact, the unweighted cases of Theorems 6.1–6.2 can be seen as special cases of the results for binary trees, see Example 10.4 below.) We let in the present section T_n and $\mathcal{T}_t$ denote these random binary trees. Recall from Section 3 that in the these trees, nodes are either internal (dead) or external (living), and that each internal node has two children, labelled left and right. If v is an internal node, we denote its left and right child by v_{L} and v_{R}, respectively.

In this context, it is natural to generalize the definitions in Section 4, and allow the offsets to have different distributions for left and right children; furthermore, we may allow a dependency between the offsets of the two children of a node. Hence, in this section we assume, instead of (BW1) in Section 4 (see also Remark 4.1), that we are given a random variable $\eta^* = (\eta_{\mathsf{L}}, \eta_{\mathsf{R}}) \in \mathbb{R}^d \times \mathbb{R}^d$ with distribution $\nu^* := \mathcal{L}(\eta^*)$, and that $\eta_v^* = (\eta_{v_{\mathsf{L}}}, \eta_{v_{\mathsf{R}}})$, $v \in T$, are i.i.d. copies of η^*. This defines η_w for every $w \neq o$, which is enough to define X_v as before; recall (4.1) and Remark 4.2.

This setting (for the binary Yule tree) is a special case of the one in Uchiyama [33], where also, more generally, the number of children may vary. Theorem 10.2 below (at least the external case) is thus a special case of [33, Theorem 4], (and Theorem 10.1 an easy consequence), but as in Sections 7–8 we give (rather) complete proofs with explicit calculations for both binary Yule trees and binary search trees, for comparison with the proofs of Theorems 6.1–6.2 and other similar proofs in the literature.

This setting has also been used by Fekete [17], who showed (among other results) a weaker version of Theorem 10.1 below with convergence in probability; we improve this to convergence a.s.

Since we have two types of nodes, we define besides $\widetilde{\mu}$ given by (4.2), where we use all nodes, also the internal and external versions. Let T^{i} and

T^{e} be the sets of internal and external nodes in T, respectively, and let

$$\widetilde{\mu}^{\mathsf{i}} := \frac{1}{|T^{\mathsf{i}}|} \sum_{v \in T^{\mathsf{i}}} \delta_{X_v}, \qquad \widetilde{\mu}^{\mathsf{e}} := \frac{1}{|T^{\mathsf{e}}|} \sum_{v \in T^{\mathsf{e}}} \delta_{X_v}. \qquad (10.1)$$

Note that, since the trees are binary, $|T^{\mathsf{e}}| = |T^{\mathsf{i}}| + 1$, and thus $|T| = |T^{\mathsf{i}}| + |T^{\mathsf{e}}| = 2|T^{\mathsf{i}}| + 1$. In particular, for the binary search tree, $|T_n^{\mathsf{i}}| = n$, $|T_n^{\mathsf{e}}| = n + 1$ and $|T_n| = 2n + 1$.

Let

$$m := \mathbb{E}\,\eta_{\mathsf{L}} + \mathbb{E}\,\eta_{\mathsf{R}} \in \mathbb{R}^d. \qquad (10.2)$$

Theorem 10.1. *Let T_n be the binary search tree and suppose that $\mathbb{E}\,|\eta^*|^2 < \infty$. Then, as $n \to \infty$, in $\mathcal{P}(\mathbb{R}^d)$,*

$$\Theta_{\sqrt{\log n},\, m \log n}(\widetilde{\mu}_n) \xrightarrow{\text{a.s.}} \mathcal{L}\left(N\left(0, \mathbb{E}[\eta_{\mathsf{L}}\eta_{\mathsf{L}}^{\mathsf{t}}] + \mathbb{E}[\eta_{\mathsf{R}}\eta_{\mathsf{R}}^{\mathsf{t}}]\right)\right). \qquad (10.3)$$

Furthermore, the same result holds for the internal and external empirical distributions $\widetilde{\mu}_n^{\mathsf{i}}$ and $\widetilde{\mu}_n^{\mathsf{e}}$.

Theorem 10.2 (Uchiyama [33]). *Let $\mathcal{T}_t$ be the binary Yule tree and suppose that $\mathbb{E}\,|\eta^*|^2 < \infty$. Then, as $t \to \infty$, in $\mathcal{P}(\mathbb{R}^d)$,*

$$\Theta_{\sqrt{t},\, mt}(\widetilde{\mu}_t) \xrightarrow{\text{a.s.}} \mathcal{L}\left(N\left(0, \mathbb{E}[\eta_{\mathsf{L}}\eta_{\mathsf{L}}^{\mathsf{t}}] + \mathbb{E}[\eta_{\mathsf{R}}\eta_{\mathsf{R}}^{\mathsf{t}}]\right)\right). \qquad (10.4)$$

Furthermore, the same result holds for the internal and external empirical distributions $\widetilde{\mu}_t^{\mathsf{i}}$ and $\widetilde{\mu}_t^{\mathsf{e}}$.

Remark 10.3. Note that in Theorems 10.1–10.2 it does not matter whether there is a dependency between η_{L} and η_{R} or not; m and the asymptotic variance $\mathbb{E}[\eta_{\mathsf{L}}\eta_{\mathsf{L}}^{\mathsf{t}}] + \mathbb{E}[\eta_{\mathsf{R}}\eta_{\mathsf{R}}^{\mathsf{t}}]$ are obtained by summing separate contributions from η_{L} and η_{R}. This is not surprising, since if we consider X_v for a fixed node v, it is a sum of either $\eta_{u_{\mathsf{L}}}$ or $\eta_{u_{\mathsf{R}}}$ for all $u \prec v$, and these are all independent. Furthermore, there are typically about as many left and right steps in the path to a node v, and it is easy to see that the distribution of X_V for a uniformly random node V, after normalizing as above, has the normal limit in (10.3) and (10.4), cf. Remark 1.3.

Example 10.4. Let $\eta_{\mathsf{R}} = 0$. Then each right child can be identified with its mother; this reduces the binary Yule tree to the Yule tree, and the binary search tree to the random recursive tree, and the theorems above (for external nodes) reduce to Theorems 6.1–6.2. (This reduction of the binary search tree to the random recursive tree identifies the nodes in the random recursive tree with the external nodes in the binary search tree. It is related to the well-known rotation correspondence [14, p. 72], which, however, identifies the nodes in the random recursive tree with the internal nodes in the binary search tree.)

Proof of Theorems 10.1 and 10.2. The arguments in Sections 7–8 require only minor modifications. We therefore omit many details. We assume again $d = 1$. We denote the characteristic function of η^* by

$$\varphi^*(s_1, s_2) := \mathbb{E}\,e^{\mathrm{i}(s_1\eta_{\mathsf{L}} + s_2\eta_{\mathsf{R}})}, \qquad s_1, s_2 \in \mathbb{R}^d, \qquad (10.5)$$

and the characteristic functions of the marginal distributions η_{L} and η_{R} by

$$\varphi_{\mathsf{L}}(s) := \mathbb{E}\,e^{\mathrm{i}s\eta_{\mathsf{L}}} = \varphi^*(s, 0), \qquad \varphi_{\mathsf{R}}(s) := \mathbb{E}\,e^{\mathrm{i}s\eta_{\mathsf{R}}} = \varphi^*(0, s). \qquad (10.6)$$

We consider the binary Yule tree $\mathcal{T}_t$ and define, in addition to $F_t(s)$,

$$F_t^{\mathrm{i}}(s) := \sum_{v \in \mathcal{T}_t^{\mathrm{i}}} e^{\mathrm{i}sX_v}, \qquad\qquad F_t^{\mathrm{e}}(s) := \sum_{v \in \mathcal{T}_t^{\mathrm{e}}} e^{\mathrm{i}sX_v}. \qquad (10.7)$$

For the exterior version F_t^{e} we can argue as in Section 7. We assume again first $d = 1$. Each external node v dies with intensity 1, and then gets two children. This changes $F_t^{\mathrm{e}}(s)$ by

$$-e^{\mathrm{i}sX_v} + e^{\mathrm{i}sX_{v_{\mathsf{L}}}} + e^{\mathrm{i}sX_{v_{\mathsf{R}}}} = e^{\mathrm{i}sX_v}\bigl(-1 + e^{\mathrm{i}s\eta_{v_{\mathsf{L}}}} + e^{\mathrm{i}s\eta_{v_{\mathsf{R}}}}\bigr) \qquad (10.8)$$

For convenience, define

$$\tilde{\varphi}(s) := \mathbb{E}\bigl(e^{\mathrm{i}s\eta_{v_{\mathsf{L}}}} + e^{\mathrm{i}s\eta_{v_{\mathsf{R}}}} - 1\bigr) = \varphi_{\mathsf{L}}(s) + \varphi_{\mathsf{R}}(s) - 1 \qquad (10.9)$$

and

$$\begin{aligned}
\psi(s_1, s_2) &:= \mathbb{E}\bigl[\bigl(e^{\mathrm{i}s_1\eta_{v_{\mathsf{L}}}} + e^{\mathrm{i}s_1\eta_{v_{\mathsf{R}}}} - 1\bigr)\bigl(e^{\mathrm{i}s_2\eta_{v_{\mathsf{L}}}} + e^{\mathrm{i}s_2\eta_{v_{\mathsf{R}}}} - 1\bigr)\bigr] \\
&= \varphi^*(s_1, s_2) + \varphi^*(s_2, s_1) + \varphi_{\mathsf{L}}(s_1 + s_2) + \varphi_{\mathsf{R}}(s_1 + s_2) \\
&\qquad -\varphi_{\mathsf{L}}(s_1) - \varphi_{\mathsf{L}}(s_2) - \varphi_{\mathsf{R}}(s_1) - \varphi_{\mathsf{R}}(s_2) + 1 \\
&= \varphi^*(s_1, s_2) + \varphi^*(s_2, s_1) + \tilde{\varphi}(s_1 + s_2) - \tilde{\varphi}(s_1) - \tilde{\varphi}(s_2). \quad (10.10)
\end{aligned}$$

It follows from (10.8) that, cf. (7.7),

$$\frac{\partial}{\partial t}\mathbb{E}\,F_t^{\mathrm{e}}(s) = \bigl(\varphi_{\mathsf{L}}(s) + \varphi_{\mathsf{R}}(s) - 1\bigr)\mathbb{E}\,F_t^{\mathrm{e}}(s) = \tilde{\varphi}(s)\,\mathbb{E}\,F_t^{\mathrm{e}}(s) \qquad (10.11)$$

and thus

$$\mathbb{E}\,F_t^{\mathrm{e}}(s) = e^{t\tilde{\varphi}(s)}. \qquad (10.12)$$

Similarly, cf. (7.8) and (7.5),

$$\begin{aligned}
\frac{\partial}{\partial t}&\mathbb{E}\bigl(F_t^{\mathrm{e}}(s_1)F_t^{\mathrm{e}}(s_2)\bigr) \\
&= \bigl(\tilde{\varphi}(s_1) + \tilde{\varphi}(s_2)\bigr)\mathbb{E}\bigl(F_t^{\mathrm{e}}(s_1)F_t^{\mathrm{e}}(s_2)\bigr) + \psi(s_1, s_2)\,\mathbb{E}\,F_t^{\mathrm{e}}(s_1 + s_2)
\end{aligned}$$

and hence, at least for s_1, s_2 in a suitably small interval $J = (-\delta, \delta)$,

$$\begin{aligned}
\mathbb{E}\bigl(F_t^{\mathrm{e}}(s_1)F_t^{\mathrm{e}}(s_2)\bigr) &= \frac{\varphi^*(s_1, s_2) + \varphi^*(s_2, s_1)}{\tilde{\varphi}(s_1) + \tilde{\varphi}(s_2) - \tilde{\varphi}(s_1 + s_2)}e^{t(\tilde{\varphi}(s_1)+\tilde{\varphi}(s_2))} \\
&\qquad - \frac{\psi(s_1, s_2)}{\tilde{\varphi}(s_1) + \tilde{\varphi}(s_2) - \tilde{\varphi}(s_1 + s_2)}e^{t\tilde{\varphi}(s_1+s_2)}. \quad (10.13)
\end{aligned}$$

The rest of the proof for $\widetilde{\mu}^{\mathrm{e}}$ is as before. It follows that $M_t(s) := F^{\mathrm{e}}(s)/\mathbb{E}\,F^{\mathrm{e}}(s)$ is an L^2-bounded martingale in $W_1^2(J)$ (if J is small enough) and we obtain, cf. (7.32),

$$\widehat{\widetilde{\mu}_t^{\mathrm{e}}}(s/\sqrt{t}) = \frac{F_t^{\mathrm{e}}(s/\sqrt{t})}{|\mathcal{T}_t^{\mathrm{e}}|} = \frac{F_t^{\mathrm{e}}(s/\sqrt{t})}{F_t^{\mathrm{e}}(0)} = e^{t(\tilde{\varphi}(s/\sqrt{t})-1)+o(1)}. \qquad (10.14)$$

Furthermore, (10.9) implies

$$\tilde{\varphi}(s) = 1 + \mathrm{i}s(\mathbb{E}\,\eta_{\mathsf{L}} + \mathbb{E}\,\eta_{\mathsf{R}}) - \tfrac{s^2}{2}\bigl(\mathbb{E}[\eta_{\mathsf{L}}\eta_{\mathsf{L}}^{\mathsf{t}}] + \mathbb{E}[\eta_{\mathsf{R}}\eta_{\mathsf{R}}^{\mathsf{t}}]\bigr) + o(s^2). \qquad (10.15)$$

Consequently, cf. (7.34), it follows from (10.14) that, again with $a := \sqrt{t}$ and $b := mt$ and for any fixed $s \in \mathbb{R}$, a.s.,

$$\widehat{\Theta_{a,b}(\widetilde{\mu}_t^{\mathrm{e}})}(s) = e^{t(\tilde{\varphi}(s/\sqrt{t})-1-\mathrm{i}ms/\sqrt{t})+o(1)} \to e^{-\frac{s^2}{2}(\mathbb{E}[\eta_{\mathsf{L}}\eta_{\mathsf{L}}^{\mathsf{t}}]+\mathbb{E}[\eta_{\mathsf{R}}\eta_{\mathsf{R}}^{\mathsf{t}}])}. \qquad (10.16)$$

The proofs of Theorem 10.2 and 10.1 for $\widetilde{\mu}^{\mathrm{e}}$ are completed as before.

To transfer the results to the internal version, we note that each internal nod has two children, and that these children comprise all internal and external nodes except the root. Hence,

$$F_t(s) = 1 + \sum_{v \in \mathcal{T}_t^i} \left(e^{\mathrm{i}sX_{v_\mathsf{L}}} + e^{\mathrm{i}sX_{v_\mathsf{R}}} \right) \tag{10.17}$$

and

$$F_t^{\mathsf{e}}(s) - F_t^{\mathsf{i}}(s) = F_t(s) - 2F_t^{\mathsf{i}}(s) = 1 + \sum_{v \in \mathcal{T}_t^i} \left(e^{\mathrm{i}sX_{v_\mathsf{L}}} + e^{\mathrm{i}sX_{v_\mathsf{R}}} - 2e^{\mathrm{i}sX_v} \right)$$

$$= 1 + \sum_{v \in \mathcal{T}_t^i} e^{\mathrm{i}sX_v} \left(e^{\mathrm{i}s\eta_{v_\mathsf{L}}} + e^{\mathrm{i}s\eta_{v_\mathsf{R}}} - 2 \right). \tag{10.18}$$

Thus,

$$|F_t^{\mathsf{e}}(s) - F_t^{\mathsf{i}}(s)| \leqslant 1 + \sum_{v \in \mathcal{T}_t^i} \left| e^{\mathrm{i}s\eta_{v_\mathsf{L}}} + e^{\mathrm{i}s\eta_{v_\mathsf{R}}} - 2 \right|$$

$$\leqslant 1 + \sum_{v \in \mathcal{T}_t^i} \left(|e^{\mathrm{i}s\eta_{v_\mathsf{L}}} - 1| + |e^{\mathrm{i}s\eta_{v_\mathsf{R}}} - 1| \right)$$

$$\leqslant 1 + 2 \sum_{v \in \mathcal{T}_t^i} \left(|s\eta_{v_\mathsf{L}}| \wedge 1 + |s\eta_{v_\mathsf{R}}| \wedge 1 \right). \tag{10.19}$$

Conditioned on the tree process $(\mathcal{T}_t)_t$, the internal nodes are added one by one, and thus the standard law of large numbers implies that for any fixed $s \in \mathbb{R}$, as $t \to \infty$,

$$\frac{1}{|\mathcal{T}_t^i|} \sum_{v \in \mathcal{T}_t^i} \left(|s\eta_{v_\mathsf{L}}| \wedge 1 \right) \xrightarrow{\text{a.s.}} \mathbb{E}[|s\eta_\mathsf{L}| \wedge 1]. \tag{10.20}$$

Let again $s(t) := s/\sqrt{t}$ for some fixed $s \in \mathbb{R}$. Then for any given $\varepsilon > 0$, $|s(t)| \leqslant \varepsilon$ for all large t, and thus (10.20) implies that a.s.,

$$\limsup_{t \to \infty} \frac{1}{|\mathcal{T}_t^i|} \sum_{v \in \mathcal{T}_t^i} \left(|s(t)\eta_{v_\mathsf{L}}| \wedge 1 \right) \leqslant \limsup_{t \to \infty} \frac{1}{|\mathcal{T}_t^i|} \sum_{v \in \mathcal{T}_t^i} \left(|\varepsilon\eta_{v_\mathsf{L}}| \wedge 1 \right) = \mathbb{E}[|\varepsilon\eta_\mathsf{L}| \wedge 1].$$

$$\tag{10.21}$$

As $\varepsilon \to 0$, the right-hand side of (10.21) tends to 0 by dominated convergence, and thus (10.21) implies that

$$\frac{1}{|\mathcal{T}_t^i|} \sum_{v \in \mathcal{T}_t^i} \left(|s(t)\eta_{v_\mathsf{L}}| \wedge 1 \right) \xrightarrow{\text{a.s.}} 0. \tag{10.22}$$

The same holds for η_{v_R}, and thus (10.19) implies

$$\left| \frac{F_t^{\mathsf{e}}(s(t))}{|\mathcal{T}_t^i|} - \frac{F_t^{\mathsf{i}}(s(t))}{|\mathcal{T}_t^i|} \right| \leqslant \frac{1}{|\mathcal{T}_t^i|} + \frac{2}{|\mathcal{T}_t^i|} \sum_{v \in \mathcal{T}_t^i} \left(|s(t)\eta_{v_\mathsf{L}}| \wedge 1 + |s(t)\eta_{v_\mathsf{R}}| \wedge 1 \right) \xrightarrow{\text{a.s.}} 0.$$

$$\tag{10.23}$$

Since $|\mathcal{T}_t^{\mathsf{e}}|/|\mathcal{T}_t^{\mathsf{i}}| = 1 + 1/|\mathcal{T}_t^{\mathsf{i}}| \to 1$ a.s., (10.23) and (10.14) yield, a.s.,

$$\widehat{\widetilde{\mu}}_t^{\mathsf{i}}(s/\sqrt{t}) = \frac{F_t^{\mathsf{i}}(s/\sqrt{t})}{|\mathcal{T}_t^{\mathsf{i}}|} = \frac{F_t^{\mathsf{e}}(s/\sqrt{t})}{|\mathcal{T}_t^{\mathsf{e}}|} \frac{|\mathcal{T}_t^{\mathsf{e}}|}{|\mathcal{T}_t^{\mathsf{i}}|} + o(1) = e^{t(\tilde{\varphi}(s/\sqrt{t})-1)} + o(1).$$
(10.24)

It follows that (10.16) holds also for $\widetilde{\mu}^{\mathsf{i}}$, and the results for $\widetilde{\mu}^{\mathsf{i}}$ now follow by the arguments used above for $\widetilde{\mu}^{\mathsf{e}}$ (and earlier in Sections 7–8).

Finally,

$$\widetilde{\mu}_t = \frac{|\mathcal{T}_t^{\mathsf{i}}|}{|\mathcal{T}_t|}\widetilde{\mu}_t^{\mathsf{i}} + \frac{|\mathcal{T}_t^{\mathsf{e}}|}{|\mathcal{T}_t|}\widetilde{\mu}_t^{\mathsf{e}},$$
(10.25)

and the results for $\widetilde{\mu}$ follow from the results for $\widetilde{\mu}^{\mathsf{i}}$ and $\widetilde{\mu}^{\mathsf{e}}$. $\square$

Example 10.5. Let $\eta^* = (-1, 1)$ (deterministically). Then X_v is the difference between the number of right and left steps in the path from the root to v, and $\widetilde{\mu}$ (or $\widetilde{\mu}^{\mathsf{i}}$) is known as the vertical profile of T. Theorem 10.1 shows that the vertical profile a.s. is asymptotically normal, with

$$\Theta_{\sqrt{\log n},\,0}(\widetilde{\mu}_n) \xrightarrow{\text{a.s.}} \mathcal{L}\left(N(0, 2)\right),$$
(10.26)

as proved by Kuba and Panholzer [24]. This can be compared to (completely different) results for uniformly random binary trees by Marckert [26] and Bousquet-Mélou and Janson [11].

11. Some open problems

The results above suggest some possible directions for future research.

11.1. **Stable limits?** We assume throughout the present paper that the offsets have finite second moments. If this does not hold, we may still conjecture that asymptotically, the empirical distribution $\widetilde{\mu}$ a.s. is as the distribution of a sum of independent copies of η, cf. Remark 7.10. In particular, this leads to the following problem.

Problem 11.1. Suppose that the offsets have a distribution in the domain of attraction of a stable distribution Λ with index $\alpha < 2$. Does $\widetilde{\mu}$ after suitable normalization converge a.s. to Λ?

Fekete [17] has shown such results with convergence in probability for the binary search tree.

In the proof above, the second moment is used not only in the final part (cf. Remark 7.10), but also in showing that $M_t \in W_1^2(J)$ and in the proof of Lemma 7.5. Hence some new idea is needed for this problem.

Note that it is easy to show that the annealed variable X_{V_n} has the asymptotic distribution Λ, cf. Remarks 1.3 and 10.3.

11.2. **Fluctuations?** The theorems above yield convergence to a deterministic limit. What can be said about the fluctuations? For example:

Problem 11.2. Let $d = 1$ and let $\bar{\sigma}^2 := \mathbb{E}\,\eta^2 \in (0, \infty)$. For $x \in \mathbb{R}$ and $n \geqslant 1$, let $I_{n,x} := (-\infty, m \log n + x\sqrt{\log n}]$. Then Theorem 6.1 says that a.s., for every fixed $x \in \mathbb{R}$,

$$\widetilde{\mu}_n(I_{n,x}) \to \Phi\left(x/\bar{\sigma}\right),$$
(11.1)

where Φ is the standard normal distribution function. What can be said about the distribution of the difference between the two sides in (11.1)? Is it asymptotically normal?

References

[1] Robert A. Adams, *Sobolev Spaces*. Academic Press, New York, 1975.

[2] David Aldous, Tree-based models for random distribution of mass. *J. Statist. Phys.* **73**, 625–641.

[3] Krishna B. Athreya & Peter E. Ney, *Branching Processes*. Springer-Verlag, Berlin, 1972.

[4] Antar Bandyopadhyay & Debleena Thacker, Rate of convergence and large deviation for the infinite color Pólya urn schemes. *Statist. Probab. Lett.* **92** (2014), 232–240.

[5] Antar Bandyopadhyay & Debleena Thacker, Pólya urn schemes with infinitely many colors. *Bernoulli* **23** (2017), no. 4B, 3243–3267.

[6] Antar Bandyopadhyay & Debleena Thacker, A new approach to Pólya urn schemes and its infinite color generalization. Preprint, 2016. `arXiv:1606.05317`

[7] Patrizia Berti, Luca Pratelli & Pietro Rigo, Almost sure weak convergence of random probability measures. *Stochastics* **78** (2006), no. 2, 91–97.

[8] J. D. Biggins, Uniform convergence of martingales in the one-dimensional branching random walk. *Selected Proceedings of the Sheffield Symposium on Applied Probability (Sheffield, 1989)*, 159–173, IMS Lecture Notes Monogr. Ser., 18, Inst. Math. Statist., Hayward, CA, 1991.

[9] J. D. Biggins, Uniform convergence of martingales in the branching random walk. *Ann. Probab.* **20** (1992), no. 1, 137–151.

[10] David Blackwell & James B. MacQueen, Ferguson distributions via Pólya urn schemes. *Ann. Statist.* **1** (1973), 353–355.

[11] Mireille Bousquet-Mélou & Svante Janson, The density of the ISE and local limit laws for embedded trees. *Ann. Appl. Probab.* **16** (2006), no. 3, 1597–1632.

[12] Brigitte Chauvin, Michael Drmota & Jean Jabbour-Hattab, The profile of binary search trees. *Ann. Appl. Probab.* **11** (2001), no. 4, 1042–1062.

[13] Brigitte Chauvin, Thierry Klein, Jean-François Marckert & Alain Rouault, Martingales and profile of binary search trees. *Electron. J. Probab.* **10** (2005), no. 12, 420–435.

[14] Michael Drmota, *Random Trees*. Springer-Verlag, Vienna, 2009.

[15] Michael Drmota, Svante Janson & Ralph Neininger, A functional limit theorem for the profile of search trees. *Ann. Appl. Probab.* **18** (2008), 288–333.

[16] F. Eggenberger & G. Pólya, Über die Statistik verketteter Vorgänge. *Zeitschrift Angew. Math. Mech.* **3** (1923), 279–289.

[17] Eric Fekete, Branching random walks on binary search trees: convergence of the occupation measure. *ESAIM Probab. Stat.* **14** (2010), 286–298.

[18] Allan Gut, *Probability: A Graduate Course*, 2nd ed., Springer, New York, 2013.

[19] Svante Janson, Random recursive trees and preferential attachment trees are random split trees. *Combin. Probab. Comput.* **28** (2019), no. 1, 81–99.

[20] Svante Janson, Random replacements in Pólya urns with infinitely many colours. *Electron. Commun. Probab.* **24** (2019), paper no. 23, 1–11.

[21] Svante Janson & Jean-François Marckert, Convergence of discrete snakes. *J. Theoret. Probab.* **18** (2005), no. 3, 615–647.

[22] Anatole Joffe, Lucien Le Cam & Jacques Neveu, Sur la loi des grands nombres pour des variables aléatoires de Bernoulli attachées à un arbre dyadique. *C. R. Acad. Sci. Paris Sér. A-B* **277** (1973), A963-A964.

[23] Olav Kallenberg. *Foundations of Modern Probability.* 2nd ed., Springer-Verlag, New York, 2002.

[24] Markus Kuba & Alois Panholzer, The left-right-imbalance of binary search trees. *Theoret. Comput. Sci.* **370** (2007), no. 1-3, 265–278.

[25] Cécile Mailler & Jean-François Marckert, Measure-valued Pólya processes. *Electron. J. Probab.* **22** (2017), no. 26, 1–33.

[26] Jean-François Marckert, The rotation correspondence is asymptotically a dilatation. *Random Structures Algorithms* **24** (2004), no. 2, 118–132.

[27] A. A. Markov, Sur quelques formules limites du calcul des probabilités. (Russian.) *Bulletin de l'Académie Impériale des Sciences* **11** (1917), no. 3, 177–186.

[28] Vladimir G. Maz'ja, *Sobolev Spaces.* Springer-Verlag, Berlin, 1985.

[29] *NIST Handbook of Mathematical Functions.* Edited by Frank W. J. Olver, Daniel W. Lozier, Ronald F. Boisvert and Charles W. Clark. Cambridge Univ. Press, 2010.
Also available as *NIST Digital Library of Mathematical Functions*, `http://dlmf.nist.gov/`

[30] Gilles Pisier, *Martingales in Banach Spaces.* Cambridge Univ. Press, Cambridge, UK, 2016.

[31] Jim Pitman. *Combinatorial Stochastic Processes.* École d'Été de Probabilités de Saint-Flour XXXII – 2002. Lecture Notes in Math. 1875, Springer-Verlag, Berlin, 2006.

[32] G. Pólya, Sur quelques points de la théorie des probabilités. *Ann. Inst. H. Poincaré* **1** (1930), no. 2, 117–161.

[33] Kōhei Uchiyama, Spatial growth of a branching process of particles living in R^d. *Ann. Probab.* **10** (1982), no. 4, 89–918.

[34] G. Udny Yule, A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F.R.S. *Philos. Trans. Roy. Soc. B* **213** (1925), 21–87.

DEPARTMENT OF MATHEMATICS, UPPSALA UNIVERSITY, PO BOX 480, SE-751 06 UPPSALA, SWEDEN

Email address: svante.janson@math.uu.se

URL: http://www.math.uu.se/svante-janson