RANDOM CUTTING AND RECORDS IN DETERMINISTIC AND RANDOM TREES

SVANTE JANSON

ABSTRACT. We study random cutting down of a rooted tree and show that the number of cuts is equal (in distribution) to the number of records in the tree when edges (or vertices) are assigned random labels.

Limit theorems are given for this number, in particular when the tree is a random conditioned Galton–Watson tree. We consider both the distribution when both the tree and the cutting (or labels) are random, and the case when we condition on the tree.

The proofs are based on Aldous' theory of the continuum random tree.

1. INTRODUCTION

We consider random cutting down of rooted trees, defined as follows [31]. If T is a rooted tree with number of vertices $|T| \geq 2$, we make a random cut by choosing one edge uniformly at random. Delete this edge so that the tree separates into two parts, and keep only the part containing the root. Continue recusively until only the root is left. We let $X(T)$ denote the (random) number of cuts that are performed until the tree is gone.

The same random variable appears when we consider records in a tree. Let each edge e have a random value λ_e attached to it, and assume that these values are i.i.d. with a continuous distribution. Say that a value λ_e is a *record* if it is the largest value in the path from the root to e. Then the number of records is again given by $X(T)$.

To see this, generate first the values λ_e and then cut the tree, each time choosing the edge with the largest λ_e among the remaining ones. By symmetry, this gives the cutting procedure above, and an edge is cut at some time if and only if its value is a record. Hence the number of records equals the number of cuts.

Remark 1.1. When we say that cutting and records give the same random variable, we really mean that they give random variables with the same distribution. (The proof just given gives a natural coupling where the two variables really coincide.)

Date: January 27, 2004; revised May 17, 2005.

Remark 1.2. As is well-known, and seen by the argument above, the distribution of λ_e does not matter, because only the order relations are important. (We assume the distribution to be continuous to avoid ties.) For the same reason, we could alternatively let the values λ_e be a random permutation of $1, \dots, |T| - 1$.

Remark 1.3. An alternative way to see the equivalence between the number of cuts and the number of records is to chop up the tree completely by cutting all edges in random order. Label the edges by $1, \dots, |T|-1$ in the order they are cut. If we count only the cuts where the cut edge still is connected to the root, we recover $X(T)$. These edges are the edges with minimal labels on the path to the root, i.e. the records for the reversed order.

There are also vertex versions of cuttings and records. For cuttings, choose a vertex at random and destroy it together with all its descendants. Continue until the root is chosen and thus the whole tree is destroyed. We let $X_v(T)$ denote the random number of vertex deletions that are needed.

For records, we assign i.i.d. values λ_v (or a random permutation) to the vertices, and define a record as above. The equivalence between cuttings and records is seen as above.

The edge and vertex versions are closely related. Indeed, let $\tilde{T}$ be the tree obtained by adding a new root to T, with the old root as its only child. Then there is a natural correspondence between edges of $\tilde{T}$ and vertices of T (each edge corresponds to the endpoint of it most distant from the root), and this correspondence preserves the cutting and record operations defined above. Consequently, $X_v(T) = X(\tilde{T})$.

Conversely, if T' is the rooted forest obtained from T by deleting the root, letting its neighbours be the new roots, then $X(T) = X_v(T')$, with the obvious extension of the definition above to rooted forests. This extension is trivial, since if F is a rooted forest with tree components $T_1, \dots, T_k$, then $X_v(F) = \sum_j X_v(T_j)$ (and similarly $X(F) = \sum_j X(T_j)$) with the summands independent, because cuttings and records in the different components are independent. (This is easiest seen with records, since the cuttings appear in a jumbled order.)

We will mainly study the edge version, which is traditional for cuttings (although the vertex version seems more natural for records). In Section 6 we show that the results transfer to the vertex version.

Our main results concern the asymptotical behaviour of $X(T)$ for a class of random trees T (i.e. for a class of distributions of T). Let us, however, first remark that it also is of interest to study $X(T)$ for deterministic trees T. We give one example here, and two others in Section 8.

Example 1.4. Take $T = P_n$, a path with n edges, with the root at an end. $X(P_n)$ (or, equivalently, $X_v(P_{n-1})$) is the number of records in a sequence of n i.i.d. values $\lambda_1, \dots, \lambda_n$, or in a random permutation of $1, \dots, n$. This is the classical record problem, which has been much studied, see for example [36]. Let $I_j = 1$ if λ_j is a record, and $I_j = 0$ otherwise, $j = 1, \dots, n$. It is easily seen that $\mathbb{P}(I_j = 1) = 1/j$, so $I_j \sim \mathrm{Be}(1/j)$. Moreover, the random variables I_j are independent [36]. Since $X(P_n) = \sum_{j=1}^n I_j$, we have

$$\mathbb{E}\, X(P_n) = \sum_{j=1}^{n} \mathbb{E}\, I_j = \sum_{j=1}^{n} \frac{1}{j} \sim \ln n. \tag{1.1}$$

The representation $X(P_n) = \sum_{j=1}^n I_j$ further yields easily, by the central limit theorem with Liapounov's condition [23, Exercise 5.20] or via an approximation by a Poisson distribution $\mathrm{Po}(\mathbb{E}\, X)$ or $\mathrm{Po}(\ln n)$ [4, Theorem 2.M], asymptotic normality:

$$(\ln n)^{-1/2}\big(X(P_n) - \ln n\big) \xrightarrow{\mathrm{d}} N(0,1) \quad \text{as } n \to \infty.$$

We can write $X(T)$ as a sum of indicators as in Example 1.4 for any tree T, see the proof of Lemma 4.3 below, but paths are very special; it is essentially only for paths that these indicators are independent. (More precisely, for T such that T' is a collection of paths rooted at one end; for $X_v(T)$ the condition is that T is a path rooted at one end.) For general trees we therefore need other methods.

Example 1.5. The simplest example where the indicators are dependent is $X_v(T)$ where T is a tree with three vertices: one root 0 attached to two leaves 1 and 2. We have $X_v(T) = I_0 + I_1 + I_2$ with $\mathbb{P}(I_0 = 1) = 1$ and $\mathbb{P}(I_1 = 1) = \mathbb{P}(I_2 = 1) = 1/2$, but $\mathbb{P}(I_1 = I_2 = 1) = 1/3$. In fact, $X_v(T)$ has in this case a uniform distribution on $\{1, 2, 3\}$.

The classes of random trees that we consider are the conditioned Galton–Watson trees, obtained as the family tree of a Galton–Watson process conditioned on a given total size. (Other classes of random trees will presumably yield other interesting results with different normalizations. Random recursive trees and binary search trees would be interesting examples.) More precisely, let ξ be a non-negative integer valued random variable, and consider the Galton–Watson process with offspring distribution ξ. Let $\mathcal{T}_n$ be the family tree, conditioned on its number of edges being n. (We consider only n such that n edges is possible.) Note that the order of $\mathcal{T}_n$ thus is $n+1$; a more common notation is to let $\mathcal{T}_n$ have order n, but our choice will be more convenient in the proofs because we consider edge cuttings and records. For the limit results, it does not matter whether n denotes the number of edges or vertices.

We let ξ (or rather its distribution) be fixed throughout the paper. We assume always

$$\mathbb{E}\,\xi = 1 \quad \text{(the Galton–Watson process is critical)}, \tag{1.2}$$

$$0 < \sigma^2 = \operatorname{Var} \xi < \infty, \tag{1.3}$$

(In papers on conditioned Galton–Watson trees, it is often assumed that ξ has an exponential moment, $\mathbb{E}\, e^{\alpha\xi} < \infty$ for some $\alpha > 0$. This is sometimes a technically useful assumption, but we will in this paper only assume finite variance (1.3), and sometimes finite higher moments.)

It is well known [1] that the families of random trees obtained in this way are the same as the simply generated families [32]. Many combinatorially interesting families are of this type; some examples to which our results apply are the following, for further examples see e.g. [1, 12].

(i) Ordered (=plane) trees. $\mathbb{P}(\xi = k) = 2^{-k-1}$; $\sigma^2 = 2$.
(ii) Unordered labelled trees (Cayley trees). $\xi \sim \mathrm{Po}(1)$; $\sigma^2 = 1$.
(iii) Binary trees. $\xi \sim \mathrm{Bi}(2, 1/2)$; $\sigma^2 = 1/2$.
(iv) Strict binary trees. $\mathbb{P}(\xi = 0) = \mathbb{P}(\xi = 2) = 1/2$; $\sigma^2 = 1$.
(v) d-ary trees. $\xi \sim \mathrm{Bi}(d, 1/d)$; $\sigma^2 = 1 - 1/d$.

We will thus study $X(\mathcal{T}_n)$ where $\mathcal{T}_n$ is as above. Since both the cutting (or records) and the tree are random, this can be regarded in (at least) two ways.

First, we can regard $X(\mathcal{T}_n)$ as a random variable, obtained by picking a random tree $\mathcal{T}_n$ and then a random cutting of it. This point of view has been taken by Meir and Moon [31] (mean and variance for Cayley trees), Chassaing and Marchand [9] (asymptotic distribution for Cayley trees), Panholzer [33, 34] (asymptotic distribution for some special families of simply generated trees, and for non-crossing trees). One of the main results of this paper is to extend these results to all conditioned Galton–Watson trees. All unspecified limits in this paper are as $n \to \infty$.

Theorem 1.6. *Let $\mathcal{T}_n$ be a conditioned Galton–Watson tree of size n, defined by an offspring distribution ξ satisfying* (1.2)–(1.3)*. Then,*

$$\frac{X(\mathcal{T}_n)}{\sigma n^{1/2}} \xrightarrow{\mathrm{d}} Z, \tag{1.4}$$

where Z has a Rayleigh distribution with density $xe^{-x^2/2}$, $x > 0$. Moreover, if $\mathbb{E}\,\xi^m < \infty$ for every $m > 0$, then all moments converge in (1.4)*, and thus, for every $r > 0$,*

$$\mathbb{E}\,X(\mathcal{T}_n)^r \sim \sigma^r n^{r/2}\,\mathbb{E}\,Z^r = 2^{r/2}\sigma^r\Gamma\bigl(\tfrac{r}{2}+1\bigr)n^{r/2}. \tag{1.5}$$

Remark 1.7. The proofs of special cases of Theorem 1.6 by Chassaing and Marchand [9] (using an equivalence with hash tables) and Panholzer [33, 34] (using generating functions) are quite different from our proof.

Remark 1.8. The proof shows that (1.5) holds provided $\mathbb{E}\,\xi^{\lfloor r\rfloor+2} < \infty$; this is presumably not sharp. For $r = 1$, we can show that $\mathbb{E}\,X(\mathcal{T}_n) \sim \sigma\sqrt{\pi n/2}$ holds assuming only (1.3), see Appendix A; we do not know if moment conditions on ξ really are needed for the higher moments. Similarly, $\mathbb{E}\,\xi^{\lfloor rk\rfloor+2} < \infty$ is sufficient for (1.11) below, and $\mathbb{E}\,\xi^4 < \infty$ is sufficient for Theorem 1.12; we doubt that these conditions are sharp.

The other point of view is to study $X(\mathcal{T}_n)$ as a random variable conditioned on $\mathcal{T}_n$. In other words, we consider the random procedure in two steps: First we choose a random tree $T = \mathcal{T}_n$. Then we keep this tree fixed and consider random cuttings of it; this gives a random variable $X(T)$ with a distribution that depends on T. Normalizing as in Theorem 1.6, we let μ_T denote the distribution of $\sigma^{-1}n^{-1/2}X(T)$; thus $\mu_{\mathcal{T}_n}$ is a random probability distribution, viz. the distribution of $\sigma^{-1}n^{-1/2}X(\mathcal{T}_n)$ given $\mathcal{T}_n$.

The reader who is not comfortable with a random probability distribution can instead consider the moments $m_k(T) := \mathbb{E}\,X(T)^k$, $k = 1, 2, \dots$. For any tree T, these are some numbers; taking T to be the random tree $\mathcal{T}_n$, we obtain the random variables

$$m_k(\mathcal{T}_n) = \mathbb{E}\big(X(\mathcal{T}_n)^k \mid \mathcal{T}_n\big). \tag{1.6}$$

The moments of $\mu_{\mathcal{T}_n}$ are thus $\sigma^{-k}n^{-k/2}m_k(\mathcal{T}_n)$.

We define, for a function f defined on an interval J and $t_1, \dots, t_k \in J$, with $k \geq 1$ is arbitrary,

$$L_f(t_1, \dots, t_k) := \sum_{i=1}^{k} f(t_{(i)}) - \sum_{i=1}^{k-1} \inf_{[t_{(i)}, t_{(i+1)}]} f, \tag{1.7}$$

where $t_{(1)}, \dots, t_{(k)}$ are $t_1, \dots, t_k$ arranged in nondecreasing order. (Hence, $t_{(i)} = t_i$ if $t_1 \leq t_2 \leq \dots \leq t_k$.) $L_f(t_1, \dots, t_k)$ is thus symmetric in $t_1, \dots, t_k$. Note that $L_f(t) = f(t)$.

We are mainly interested in non-negative functions defined on $[0, 1]$ and then further define, for $k \geq 1$,

$$m_k(f) := k! \int_0^1 \dots \int_0^1 \frac{dt_1 \cdots dt_k}{L_f(t_1)L_f(t_1, t_2) \cdots L_f(t_1, t_2, \dots, t_k)}\,. \tag{1.8}$$

We also let $m_0(f) := 1$.

We will give background and motivation for these definitions in Sections 3 and 4. Let $C[0, 1]_+$ denote the set of non-negative, continuous functions on $[0, 1]$.

Theorem 1.9. *If $f \in C[0,1]_+$ is such that $\int_0^1 dt/f(t) < \infty$, then there exists a unique probability measure ν_f on $[0, \infty)$ with (finite) moments*

$$\int x^k \, d\nu_f(x) = m_k(f)$$

given by (1.8).

We will see in Section 9 that this theorem extends to discontinuous f too.

Let B_{ex} denote the normalized Brownian excursion. Recall that this is a random function in $C[0, 1]_+$, see e.g. [8] or [37]. It is well-known, see Remark 5.2 below, that $\int_0^1 dt/B_{\text{ex}}(t) < \infty$ a.s.; hence $\nu_{cB_{\text{ex}}}$ exists a.s. for every constant $c > 0$. ($\nu_{cB_{\text{ex}}}$ is thus a random probability measure.)

Theorem 1.10. *If $\mathcal{T}_n$ is a conditioned Galton–Watson tree as above, then*

$$\mu_{\mathcal{T}_n} \xrightarrow{\text{d}} \nu_{2B_{\text{ex}}} \tag{1.9}$$

in the space of probability measures on $\mathbb{R}$*. Moreover, moment convergence holds in* (1.9)*, that is, for every* $k \geq 1$*, using the notation* (1.6)*,*

$$\sigma^{-k} n^{-k/2} m_k(\mathcal{T}_n) \xrightarrow{\mathrm{d}} \int x^k \, d\nu_{2B_{\mathrm{ex}}}(x) = m_k(2B_{\mathrm{ex}}), \tag{1.10}$$

with the right hand side given by (1.8)*. Further, if* $\mathbb{E}\,\xi^m < \infty$ *for every* $m > 0$*, then moment convergence holds in* (1.10) *too; for* $k \geq 1$ *and* $r > 0$*,*

$$\mathbb{E}\, m_k(\mathcal{T}_n)^r \sim \sigma^{kr} n^{kr/2}\, \mathbb{E}\, m_k(2B_{\mathrm{ex}})^r. \tag{1.11}$$

Joint convergence holds in (1.9)*,* (1.10) *for all* $k \geq 1$*, and* (3.4) *below.*

Remark 1.11. It ought to be possible to define a random variable with the distribution $\nu_{2B_{\mathrm{ex}}}$ by some construction that can be interpreted as continuous cutting on the Brownian continuum random tree defined by Aldous [1, 2]. We have, however, not had enough imagination to construct such a variable.

We can use these results to see how much of the variance of $X(\mathcal{T}_n)$ that comes from the random choice of tree and how much that comes from the cutting. We have, as always in such cases, the decomposition

$$X(\mathcal{T}_n) = \Big(X(\mathcal{T}_n) - \mathbb{E}\big(X(\mathcal{T}_n) \mid \mathcal{T}_n\big)\Big) + \mathbb{E}\big(X(\mathcal{T}_n) \mid \mathcal{T}_n\big)$$

and the corresponding analysis of variance

$$\begin{aligned} \operatorname{Var} X(\mathcal{T}_n) &= \mathbb{E}\Big(X(\mathcal{T}_n) - \mathbb{E}\big(X(\mathcal{T}_n) \mid \mathcal{T}_n\big)\Big)^2 + \operatorname{Var}\Big(\mathbb{E}\big(X(\mathcal{T}_n) \mid \mathcal{T}_n\big)\Big) \\ &= \mathbb{E}\big(\operatorname{Var}\big(X(\mathcal{T}_n) \mid \mathcal{T}_n\big)\big) + \operatorname{Var}\big(\mathbb{E}\big(X(\mathcal{T}_n) \mid \mathcal{T}_n\big)\big). \end{aligned} \tag{1.12}$$

Theorem 1.12. *For large* n*, at least provided* $\mathbb{E}\,\xi^r < \infty$ *for all* $r > 0$*,*

$$\begin{aligned} \operatorname{Var} X(\mathcal{T}_n) &= \mathbb{E}\, m_2(\mathcal{T}_n) - \big(\mathbb{E}\, m_1(\mathcal{T}_n)\big)^2 \sim \big(2 - \tfrac{\pi}{2}\big)\sigma^2 n, \\ \mathbb{E}\big(\operatorname{Var}\big(X(\mathcal{T}_n) \mid \mathcal{T}_n\big)\big) &= \mathbb{E}\big(m_2(\mathcal{T}_n) - m_1(\mathcal{T}_n)^2\big) \sim \big(2 - \tfrac{\pi^2}{6}\big)\sigma^2 n, \\ \operatorname{Var}\big(\mathbb{E}\big(X(\mathcal{T}_n) \mid \mathcal{T}_n\big)\big) &= \operatorname{Var}\big(m_1(\mathcal{T}_n)\big) \sim \big(\tfrac{\pi^2}{6} - \tfrac{\pi}{2}\big)\sigma^2 n. \end{aligned}$$

Hence, asymptotically, the first term in (1.12) is $(2 - \pi^2/6)/(2 - \pi/2) \approx 0.827$ of the total. Thus, for a conditioned Galton–Watson tree, for large n, about 83% of the variance of $X(\mathcal{T}_n)$ comes from the random choice of cutting, and 17% from the random choice of tree.

In the proofs we will use an estimate that might be of independent interest. Let $w_k(T)$ be the number of vertices of depth k in a rooted tree T. As above, let $\mathcal{T}_n$ be a conditioned Galton–Watson tree of size n, defined by an offspring distribution ξ satisfying (1.2)–(1.3).

Theorem 1.13. *Suppose that* $r \geq 1$ *is an integer such that* $\mathbb{E}\,\xi^{r+1} < \infty$*. Then, for all* n *and* $k \geq 1$*,* $\mathbb{E}\big(w_k(\mathcal{T}_n)^r\big) \leq C k^r$ *for some constant* C *depending on* r *and* ξ *only.*

For the expectation $\mathbb{E}\, w_k(\mathcal{T}_n)$, related asymptotic results are given by Meir and Moon [32].

Proofs of the theorems above are given in Sections 2–5.

In Section 6 we show that the results above are valid for the vertex versions too. We also give a generalization to a somewhat larger class of random trees, including the non-crossing trees studied by Panholzer [34].

In Section 7 we connect our results to known results about the height and width of random trees.

We end the paper with some comments and further results related to the main results. Section 8 contains two examples with deterministic trees (a path, with connections Hoare's algorithm FIND, and a binary tree); these behave quite differently than the conditioned Galton–Watson trees. Section 9 extends Theorem 1.9 to discontinuous f. Although the resulting probability distributions are not needed for our study of random cuttings and records for conditioned Galton–Watson trees, they arise as limits for other classes of trees; moreover, we find them interesting in themselves. We study a few simple examples.

Finally, we want to draw attention to the following open problems, related to Theorem 1.13; see further Section 10. As above, let $\mathcal{T}_n$ be a conditioned Galton–Watson tree of size n, defined by an offspring distribution ξ satisfying (1.2)–(1.3).

Problem 1.14. Is, for every fixed $k \geq 1$, $\mathbb{E}\, w_k(\mathcal{T}_n)$ an increasing function of n?

Problem 1.15. Is it possible to define the trees $\mathcal{T}_n$ on a common probability space so that the sequence $\mathcal{T}_n$ is increasing? In other words, does there exist a stochastic process $\mathcal{T}_n$ describing a growing tree with the right marginal distributions?

Problem 1.15 was considered for d-ary (including binary) trees by Luczak and Winkler [28], who proved that the answer is affirmative in this case. The proof is non-trivial, and there is no "natural" definition of the growing process. We do not know any similar results for other conditioned Galton–Watson trees, nor any counterexample. Intuitively, it is natural to guess that $\mathcal{T}_n$ is (stochastically) increasing in this way, but the definition by conditioning precludes any simple monotonicity argument.

A positive answer to Problem 1.15 obviously implies a positive answer to Problem 1.14, so this problem too is solved for d-ary trees. The exact formulas in [32] for labelled (Cayley) trees, plane trees and strict binary trees give a positive answer to Problem 1.14 in these cases too.

Acknowledgements. I thank several participants in the Ninth Seminar on Analysis of Algorithms in San Miniato, June 2003, for valuable discussions.

This research was partly done during a visit to Université de Versailles Saint-Quentin, Versailles, France, September 2003.

2. Proof of Theorem 1.13

We will in this section prove the estimate Theorem 1.13, which is used in the proof of the main results. The reader that is eager to see the main arguments can omit this section at the first reading.

The *span* of ξ, $\operatorname{span}(\xi)$, is the smallest positive integer d such that d divides ξ a.s. We will for simplicity assume that $\operatorname{span}(\xi) = 1$ and leave the minor modifications when $\operatorname{span}(\xi) = d > 1$ to the reader. We will in this section let C and c denote various positive constants depending on the distribution of ξ and the power r only; their values may change from one occurence to the next.

Let $S_N := \sum_1^N \xi_i$, where ξ_i are i.i.d. copies of ξ. As is well-known, see e.g. [26, Lemma 2.1.3], if $\mathcal{T}^{(i)}$ are i.i.d. copies of $\mathcal{T}$, then

$$\mathbb{P}\Bigl(\sum_1^m |\mathcal{T}^{(i)}| = n\Bigr) = \frac{m}{n}\,\mathbb{P}(S_n = n-m), \qquad n \ge m \ge 1. \tag{2.1}$$

In particular, using the local central limit theorem [26, Theorem 1.4.2],

$$\mathbb{P}\bigl(|\mathcal{T}| = n\bigr) = \frac{1}{n}\,\mathbb{P}(S_n = n-1) \sim (2\pi)^{-1/2}\sigma^{-1}n^{-3/2}. \tag{2.2}$$

We will use the following general estimate. (It can be regarded as a coarse but general version of local central limit and large deviation theorems.)

Lemma 2.1. *There exists constants C and $c > 0$ such that for all N and $k \ge 0$*

$$\mathbb{P}(S_N = N-k) \le CN^{-1/2}e^{-ck^2/N}.$$

Proof. We may assume $0 \le k \le N$. Let $F(z) := \mathbb{E}\, z^\xi$ be the probability generating function of ξ. Then

$$\mathbb{P}(S_N = N-k) = \frac{1}{2\pi i}\oint z^{k-N}F(z)^N\,\frac{dz}{z},$$

where we choose to integrate around the circle $|z| = r$ with radius $r := e^{-\delta k/N}$, for some small δ to be chosen later. We therefore let $G(z) := F(z)/z$, and have

$$\mathbb{P}(S_N = N-k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-\delta k^2/N + ikt}G(re^{it})^N\,dt. \tag{2.3}$$

Since $\mathbb{E}\,\xi = 1$ and $\mathbb{E}\,\xi(\xi-1) = \sigma^2$, we have the Taylor expansion

$$F(z) = 1 + (z-1) + \tfrac{\sigma^2}{2}(z-1)^2 + o(|z-1|^2), \qquad |z| \le 1,$$

and thus

$$\begin{aligned}
G(z) &= 1 + \tfrac{\sigma^2}{2}(z-1)^2 + o(|z-1|^2), \qquad |z| \le 1,\\
G(e^w) &= 1 + \tfrac{\sigma^2}{2}w^2 + o(|w|^2), \qquad \operatorname{Re} w \le 0,\\
\ln G(e^w) &= \tfrac{\sigma^2}{2}w^2 + o(|w|^2), \qquad \operatorname{Re} w \le 0.
\end{aligned}$$

Hence, if $0 < \delta \le \delta_0$ and $|t| \le t_0$ for sufficiently small positive δ_0 and t_0,

$$\begin{aligned}\ln|G(re^{it})| &= \operatorname{Re}\ln G(e^{-\delta k/N+it}) = \tfrac{\sigma^2}{2}(\delta^2k^2/N^2 - t^2) + o(\delta^2k^2/N^2 + t^2)\\ &\le \sigma^2\delta^2k^2/N^2 - \sigma^2t^2/4. \end{aligned} \tag{2.4}$$

Since $|F(z)| < 1$ for $|z| \le 1$ with $z \ne 1$ (when $\operatorname{span}(\xi) = 1$), continuity and compactness shows that $|F(re^{it})| \le 1 - \varepsilon < e^{-\varepsilon}$ for some $\varepsilon > 0$ when $e^{-\delta_0} \le r \le 1$ and $t_0 \le |t| \le \pi$. Hence, for $t_0 \le |t| \le \pi$ and $0 \le \delta \le \delta_1 := \min(\delta_0, \varepsilon/2)$,

$$|G(re^{it})| = e^{\delta k/N}|F(re^{it})| \le e^{\delta}e^{-\varepsilon} \le e^{-\varepsilon/2}. \tag{2.5}$$

Combining (2.4) and (2.5), we see that if $\delta \le \delta_1$ and $|t| \le \pi$, then

$$|G(re^{it})| \le e^{\sigma^2\delta^2k^2/N^2 - c_1t^2},$$

with $c_1 := \min(\sigma^2/4,\, \varepsilon/2\pi^2) > 0$. Using this in (2.3) we obtain

$$\mathbb{P}(S_N = N-k) \le e^{\sigma^2\delta^2k^2/N - \delta k^2/N}\int_{-\infty}^{\infty} e^{-c_1Nt^2}\,dt, \qquad 0 \le \delta \le \delta_1,$$

and the result follows by choosing $\delta \le 1/2\sigma^2$. □

If T is a tree, let T^k denote T pruned at height k, i.e. the subtree consisting of all vertices of depth $\le k$. As $n \to \infty$, the conditioned Galton–Watson tree $\mathcal{T}_n$ converges in distribution to a random infinite tree $\mathcal{T}_\infty$, in the sense that $\mathcal{T}_n^k \xrightarrow{\mathrm{d}} \mathcal{T}_\infty^k$ for every fixed k, see [1]. (This follows easily from the argument in (9.11) below. Actually, we will not use this fact, except as a motivation.) The tree $\mathcal{T}_\infty$ can be described in several ways, see e.g. [1] and [27]; we will use the fact that it is a size-biased version of the (a.s. finite) random Galton–Watson tree $\mathcal{T}$; more precisely, for every tree T (with height k),

$$\mathbb{P}(\mathcal{T}_\infty^k = T) = w_k(T)\,\mathbb{P}(\mathcal{T}^k = T). \tag{2.6}$$

(Note that the sum over T of the right hand side equals $\mathbb{E}\,w_k(\mathcal{T}^k) = \mathbb{E}\,w_k(\mathcal{T}) = (\mathbb{E}\,\xi)^k = 1$.)

Let T be a tree of height k, with $w_k(T) = m$. If the Galton–Watson tree $\mathcal{T}$ has $\mathcal{T}^k = T$, then the part above T_k consists of m independent copies of $\mathcal{T}$. The total order of these subtrees is $|\mathcal{T}| - |T| + m$, and thus (2.1), (2.2), Lemma 2.1 and (2.6) yield, if $N = n + 1 - |T| + m$,

$$\begin{aligned}\mathbb{P}(\mathcal{T}_n^k = T) &= \frac{\mathbb{P}(\mathcal{T}^k = T,\ |\mathcal{T}| = n+1)}{\mathbb{P}(|\mathcal{T}| = n+1)} = \frac{\mathbb{P}(\mathcal{T}^k = T)\frac{m}{N}\,\mathbb{P}(S_N = N-m)}{\mathbb{P}(|\mathcal{T}| = n+1)}\\ &\le Cn^{3/2}\frac{m}{N^{3/2}}e^{-cm^2/N}\,\mathbb{P}(\mathcal{T}^k = T)\\ &= C\Big(\frac{n}{N}\Big)^{3/2}e^{-cm^2/N}\,\mathbb{P}(\mathcal{T}_\infty^k = T). \end{aligned} \tag{2.7}$$

Lemma 2.2. *If $r \ge 1$ is an integer and $\mathbb{E}\,\xi^r < \infty$, then $\mathbb{E}\big(w_k(\mathcal{T})^r\big)$ is a polynomial in k of degree $r-1$.*

Proof. Recall that $w_k(\mathcal{T})$ is the size of the k:th generation in a critical Galton–Watson process. Thus, conditioned on $w_k(\mathcal{T}) = M$, $w_{k+1}(\mathcal{T})$ is distributed as S_M.

First, for $r = 1$, we have $\mathbb{E}\, w_k(\mathcal{T}) = (\mathbb{E}\,\xi)^k = 1$.

Next, $w_k(\mathcal{T})^2$ is the number of pairs (v_1, v_2) in the k:th level (generation). Distinguishing between the cases when their fathers are different or the same, we see that

$$\mathbb{E}\big(w_{k+1}(\mathcal{T})^2 \mid w_k(\mathcal{T}) = M\big) = M(M-1)(\mathbb{E}\,\xi)^2 + M\,\mathbb{E}\,\xi^2 = M^2 + M\sigma^2$$

and thus

$$\mathbb{E}\, w_{k+1}(\mathcal{T})^2 = \mathbb{E}\, w_k(\mathcal{T})^2 + \sigma^2\,\mathbb{E}\, w_k(\mathcal{T}) = \mathbb{E}\, w_k(\mathcal{T})^2 + \sigma^2.$$

By induction,

$$\mathbb{E}\, w_k(\mathcal{T})^2 = 1 + k\sigma^2. \tag{2.8}$$

For $r > 2$ we argue in the same way. We consider all sequences of r vertices $v_1, \dots, v_r$ at level k, and separate them according to the partition of $\{1, \dots, r\}$ formed by the sets of siblings. This yields

$$\mathbb{E}\big(w_{k+1}(\mathcal{T})^r \mid w_k(\mathcal{T}) = M\big) = M^r + q_r(M),$$

where q_r is a polynomial of degree $r-1$, and thus

$$\mathbb{E}\, w_{k+1}(\mathcal{T})^r = \mathbb{E}\, w_k(\mathcal{T})^r + \mathbb{E}\, q_r(w_k(\mathcal{T})). \tag{2.9}$$

By induction on r, $\mathbb{E}\, q_r(w_k(\mathcal{T}))$ is a polynomial in k of degree $r-2$, and (2.9) implies the result. ◻

Lemma 2.3. *If $r \geq 1$ is an integer with $\mathbb{E}\,\xi^{r+1} < \infty$, then $\mathbb{E}\big(w_k(\mathcal{T}_\infty)^r\big)$ is a polynomial in k of degree r.*

Proof. By (2.6), $\mathbb{E}\big(w_k(\mathcal{T}_\infty)^r\big) = \sum_T w_k(T)^r\,\mathbb{P}(\mathcal{T}_\infty^k = T) = \mathbb{E}\big(w_k(\mathcal{T})^{r+1}\big)$ and the result follows by Lemma 2.2. ◻

Let $\mathcal{E}_k$ be the event $\{\sum_{j=0}^{k-1} w_j(\mathcal{T}_n) \leq n/2\}$ and define $\tilde{w}_k(\mathcal{T}_n) := w_k(\mathcal{T}_n)\mathbf{1}[\mathcal{E}_k]$, where $\mathbf{1}[\mathcal{E}]$ denotes the indicator of $\mathcal{E}$. ($\tilde{w}_k(\mathcal{T}_n)$ is a truncated version of w_k, roughly speaking we ignore vertices with depth larger than the median.)

Fix $r \geq 1$ with $\mathbb{E}\,\xi^{r+1} < \infty$. If $\mathcal{T}_n^k = T$ and $\tilde{w}_k(\mathcal{T}_n) > 0$, then $\mathcal{E}_k$ occurs and thus $n + 1 - (|T| - w_k(T)) = n - \sum_{j<k} w_j(T) \geq n/2$; hence $\mathbb{P}(\mathcal{T}_n^k = T) \leq C\,\mathbb{P}(\mathcal{T}_\infty^k = T)$ by (2.7). Summing over all T with $w_k(T) = j$, we see that, for every $j \geq 1$, $\mathbb{P}(\tilde{w}_k(\mathcal{T}_n) = j) \leq C\,\mathbb{P}(w_k(\mathcal{T}_\infty) = j)$. Hence,

$$\mathbb{E}\big(\tilde{w}_k(\mathcal{T}_n)^r\big) = \sum_j j^r\,\mathbb{P}(\tilde{w}_k(\mathcal{T}_n) = j) \leq C\,\mathbb{E}\big(w_k(\mathcal{T}_\infty)^r\big),$$

which by Lemma 2.3 yields

$$\mathbb{E}\big(\tilde{w}_k(\mathcal{T}_n)^r\big) \leq Ck^r, \qquad k \geq 1. \tag{2.10}$$

We use the notation $\|X\|_r = (\mathbb{E}\,|X|^r)^{1/r}$ and rewrite (2.10) as $\|\tilde{w}_k(\mathcal{T}_n)\|_r \leq Ck$. By Minkowski's inequality, thus

$$\Big\|\sum_{j=0}^{k} \tilde{w}_j(\mathcal{T}_n)\Big\|_r \leq \sum_{j=0}^{k} \|\tilde{w}_j(\mathcal{T}_n)\|_r \leq Ck^2, \qquad k \geq 1. \tag{2.11}$$

If $\mathcal{E}_k$ does not occur, let l be the smallest integer such that $\sum_{j<l} w_j(\mathcal{T}_n) > n/2$. Then $l \leq k$ and $\tilde{w}_j(\mathcal{T}_n) = w_j(\mathcal{T}_n)$ for $j < l$, and thus $\sum_{j<k} \tilde{w}_j(\mathcal{T}_n) = \sum_{j<l} w_j(\mathcal{T}_n) > n/2$. Hence, by Markov's inequality and (2.11),

$$\mathbb{P}(\mathcal{E}_k^c) \leq \Big(\frac{2}{n}\Big)^r \mathbb{E}\Big(\sum_{j=0}^{k} \tilde{w}_j(\mathcal{T}_n)\Big)^r \leq Cn^{-r}k^{2r}. \tag{2.12}$$

In particular, $\mathbb{P}(\mathcal{E}_k^c) \leq \mathbb{P}(\mathcal{E}_k^c)^{1/2} \leq Ck^r n^{-r/2}$ and

$$\mathbb{E}\big(w_k(\mathcal{T}_n)^r \mathbf{1}[\mathcal{E}_k^c]\mathbf{1}[w_k(\mathcal{T}_n) \leq n^{1/2}]\big) \leq n^{r/2}\,\mathbb{P}(\mathcal{E}_k^c) \leq Ck^r. \tag{2.13}$$

Finally, if $m = w_k(T) \geq n^{1/2}$, then $e^{-cm^2/N} \leq e^{-cn/N} \leq C(N/n)^{3/2}$ and (2.7) shows that $\mathbb{P}(\mathcal{T}_n^k = T) \leq C\,\mathbb{P}(\mathcal{T}_\infty^k = T)$. Consequently, using Lemma 2.3,

$$\mathbb{E}\big(w_k(\mathcal{T}_n)^r \mathbf{1}[w_k(\mathcal{T}_n) > n^{1/2}]\big) \leq C\,\mathbb{E}\big(w_k(\mathcal{T}_\infty)^r\big) \leq Ck^r. \tag{2.14}$$

Summing (2.10), (2.13) and (2.14), we obtain Theorem 1.13.

3. Depth-first search and walk

We will use the idea of coding trees by walks [1, 2, 35]. We will denote the root of a tree by o. The depth $d(v)$ of a vertex v in a rooted tree is the distance from o to v.

Let T be an ordered tree with root o and $n = |T| - 1$ edges. The *depth-first search* of T is the function ψ from $\{0, 1, \dots, 2n\}$ to the set of vertices of T such that $\psi(0) = o$ and, for $0 \leq i < 2n$, if $\psi(i) = v$, then $\psi(i+1)$ is the first child of v that has not already been visited, if such a child exists, and the parent of v otherwise. Note that $\psi(i)$ and $\psi(i+1)$ always are neigbours; we extend ψ to $[0, 2n]$ by letting, for $0 \leq i < t < i+1 \leq 2n$, $\psi(t)$ to be the one of $\psi(i)$ and $\psi(i+1)$ that has largest depth. Then each non-root vertex in T is $\psi(t)$ for t in exactly two (possibly adjacent) intervals of unit lengths, which proves the following, cf. [2, Lemma 12].

Lemma 3.1. *If we choose t in $(0, 2n)$ uniformly at random, then $\psi(t)$ will have a uniform distribution over all non-root vertices in T.* □

We further define

$$V(i) = V_T(i) := d(\psi(i)), \qquad i = 0, \dots, 2n,$$

and extend, as is customary, V to $[0, 2n]$ by linear interpolation; thus $V \in C[0, 2n]$. Note that

$$d(\psi(t)) = \lceil V(t) \rceil, \qquad t \in [0, 2n]. \tag{3.1}$$

We rescale V (by constants adapted to the families of trees we are interested in) and define

$$\widetilde{V}(t) := n^{-1/2} V(2nt), \tag{3.2}$$

$$\widehat{V}(t) := n^{-1/2} \lceil V(2nt) \rceil. \tag{3.3}$$

Hence $\widetilde{V} \in C[0,1]$, and $\widetilde{V} \le \widehat{V} \le \widetilde{V} + n^{-1/2}$. We use the name *depth-first walk* for V, and call $\widetilde{V}$ and $\widehat{V}$ *rescaled depth-first walks.* Since $|\widetilde{V} - \widehat{V}| \le n^{-1/2}$, it often does not matter whether we use $\widetilde{V}$ or $\widehat{V}$ in our asymptotic results, and we then usually prefer $\widetilde{V}$ which is traditional. However, $\int_0^1 dt/\widetilde{V}(t)$ always diverges, which forces us to use $\widehat{V}$ in e.g. (4.9) below.

The definitions so far apply to any tree, deterministic or not. If T is a random conditioned Galton–Watson tree as in Section 1, then $\widetilde{V}$ becomes a random function in $C[0,1]$, and Aldous [2, Theorem 23 with Remark 2] has shown the deep result that, in $C[0,1]$ with its usual topology, as $n \to \infty$,

$$\widetilde{V} \xrightarrow{\mathrm{d}} 2\sigma^{-1} B_{\mathrm{ex}}. \tag{3.4}$$

(See also [29].) This will be the basis of our proofs.

4. Proof of Theorems 1.9 and 1.10

We begin by showing uniqueness in Theorem 1.9.

Lemma 4.1. *If f is defined on an interval J and $t_1, \dots, t_k \in J$, then $L_f(t_1, \dots, t_k) \ge \max_{1 \le i \le k} f(t_i)$. Consequently,*

$$\frac{1}{L_f(t_1) \cdots L_f(t_1, \dots, t_k)} \le \frac{1}{f(t_1) \cdots f(t_k)}. \tag{4.1}$$

Proof. Since L_f is a symmetric function, we may for the first part assume that $0 \le t_1 \le \dots \le t_k \le 1$, so $t_{(i)} = t_i$. If $1 \le j \le k$, we use $\inf_{[t_i, t_{i+1}]} f \le f(t_i)$ for $i < j$ and $\inf_{[t_i, t_{i+1}]} f \le f(t_{i+1})$ for $i \ge j$; hence, by (1.7),

$$L_f(t_1, \dots, t_k) = \sum_{i=1}^{k} f(t_i) - \sum_{i=1}^{k-1} \inf_{[t_i, t_{i+1}]} f \ge f(t_j).$$

This yields the first inequality; (4.1) follows immediately. □

Lemma 4.2. *If $f \ge 0$ on $[0,1]$, and $A = \int_0^1 dt/f(t) < \infty$, then*

$$0 \le m_k(f) \le k!\, A^k, \qquad k \ge 1. \tag{4.2}$$

Hence, for $0 \le x < 1/A$,

$$\sum_{k=0}^{\infty} m_k(f) \frac{x^k}{k!} < \infty. \tag{4.3}$$

In particular, each $m_k(f)$ is finite and there exists at most one probability measure on $\mathbb{R}$ with moments $m_k(f)$.

Proof. By (1.8) and Lemma 4.1,

$$0 \le m_k(f) \le k! \int_0^1 \cdots \int_0^1 \frac{dt_1 \cdots dt_k}{f(t_1)\cdots f(t_k)} = k!\Bigl(\int_0^1 \frac{dt}{f(t)}\Bigr)^k.$$

This proves (4.2), and thus (4.3). A probability measure with moments $m_k(f)$ thus has finite moment generating function in a neighborhood of 0; it is well known that this implies that the measure is unique, see e.g. [19, Section 4.10]. □

We continue by computing the moments of $X(T)$ for a fixed tree T. We denote falling factorials by $x^{\underline{k}} := x(x-1)\cdots(x-k+1)$. If $v_1,\dots,v_k$ are vertices in a rooted tree T, let $L_T(v_1,\dots,v_k)$ be the number of edges in the subtree of T spanned by $v_1,\dots,v_k$ and the root, i.e. in the union of the paths from $v_1,\dots,v_k$ to the root. In particular, for $k=1$, $L_T(v)=d(v)$.

Lemma 4.3. *For any tree T with root o, the factorial moments of $X(T)$ are given by, for $k \ge 1$,*

$$\mathbb{E}\, X(T)^{\underline{k}} = k! \sum_{v_1,\dots,v_k}^{**} \frac{1}{L_T(v_1)\cdot L_T(v_1,v_2)\cdots L_T(v_1,\dots,v_k)}, \tag{4.4}$$

with $\sum^{**}$ denoting summation over $v_1,\dots,v_k$ that are distinct, $\neq o$, and such that v_i is not a descendant of v_j when $i<j$. In particular,

$$m_1(T) = \mathbb{E}\, X(T) = \sum_{v\neq o} \frac{1}{d(v)}. \tag{4.5}$$

Proof. Using the equivalence $X(T) = X_v(T')$ and the record formulation we have, as in Example 1.4, $X(T) = \sum_{v\neq o} I_v$, where I_v is the indicator that there is a record in T' at the vertex v. Hence, letting $\sum^*$ denote the sum over distinct vertices $\neq o$,

$$X(T)^{\underline{k}} = \sum_{v_1,\dots,v_k}^{*} I_{v_1}\cdots I_{v_k}. \tag{4.6}$$

In this sum, each product $I_{v_1}\cdots I_{v_k}$ occurs $k!$ times, with the indices permuted. For exactly one of these permutations we have $\lambda_{v_1} < \cdots < \lambda_{v_k}$. Consequently,

$$X(T)^{\underline{k}} = k! \sum_{v_1,\dots,v_k}^{*} \mathbf{1}[\mathcal{E}(v_1,\dots,v_k)], \tag{4.7}$$

where $\mathcal{E}(v_1,\dots,v_k)$ is the event

$$\begin{aligned}&\{\lambda_{v_1} < \cdots < \lambda_{v_k} \text{ and all are records in } T'\}\\ &\qquad = \{\lambda_{v_j} \text{ is the largest value in } T(v_1,\dots,v_j)' \text{ for every } j=1,\dots,k\}.\end{aligned}$$

The event $\mathcal{E}(v_1,\dots,v_k)$ is impossible if v_i is a descendant of v_j for some i and j with $i<j$. For any other sequence $v_1,\dots,v_k$, the probability that λ_{v_k} is the largest value in $T(v_1,\dots,v_k)'$ is, by symmetry, 1 divided by the number of vertices in $T(v_1,\dots,v_k)'$, i.e. $1/L_T(v_1,\dots,v_k)$. Moreover, conditioned on

this happening, the values in $T(v_1,\dots,v_{k-1})'$ are exchangeable, again by symmetry, and thus it follows by induction that, for such $v_1,\dots,v_k$,

$$\mathbb{P}[\mathcal{E}(v_1,\dots,v_k)] = \prod_{j=1}^{k} \frac{1}{L_T(v_1,\dots,v_j)}.$$

Taking expectations in (4.7) we thus obtain (4.4), and (4.5) follows because $L_T(v) = d(v)$. ☐

We next connect the subtree size L_T to L_f in (1.7) using the depth-first walks in Section 3, cf. [1, 2].

Lemma 4.4. *Let T be a tree with depth-first search and walk ψ and V. If $t_1,\dots,t_k \in [0,1]$, then*

$$L_T\big(\psi(t_1),\dots,\psi(t_k)\big) = L_{\lceil V\rceil}(t_1,\dots,t_k).$$

Proof. Since, by definition, both L_T and $L_{\lceil V\rceil}$ are symmetric, we may assume that $0 \le t_1 \le \dots \le t_k \le 1$. Let $v_i = \psi(t_i)$, $i = 1,\dots,k$.

First, if $k = 1$, we have by (3.1)

$$L_T(v_1) = d(v_1) = \lceil V(t_1)\rceil = L_{\lceil V\rceil}(t_1).$$

Next, if $k = 2$, let w be the last common ancestor of v_1 and v_2. It is easily seen that $d(w) = \inf_{[t_1,t_2]}\lceil V(t)\rceil$, cf. [2], and thus

$$L_T(v_1,v_2) = d(v_1) + d(v_2) - d(w) = L_{\lceil V\rceil}(t_1,t_2).$$

The general case follows similarly, using induction on k. ☐

Lemma 4.5. *Suppose that T_n, $n = 1,2,\dots$, is a sequence of ordered trees with $|T_n| = n+1$, and denote the corresponding depth-first walks by V_n, rescaled to $\widetilde{V}_n$ and $\widehat{V}_n$. Suppose further that $f \in C[0,1]$ is a function such that*

$$\widetilde{V}_n(t) \to f(t) \quad \textit{in } C[0,1] \tag{4.8}$$

and

$$\int_0^1 \frac{dt}{\widehat{V}_n(t)} \to \int_0^1 \frac{dt}{f(t)} < \infty. \tag{4.9}$$

Then, for each $k \ge 1$,

$$n^{-k/2} m_k(T_n) = n^{-k/2}\,\mathbb{E}\,X(T_n)^k \to m_k(f)$$

given by (1.8), and $n^{-1/2}X(T_n) \xrightarrow{\mathrm{d}} \nu_f$ given by Theorem 1.9.

Proof. Consider first the mean. Let ψ_n be the depth-first search for T_n. By (4.5), Lemma 3.1, and (3.1),

$$m_1(T_n) = \mathbb{E}\,X(T_n) = \sum_{v\in T_n'} \frac{1}{d(v)} = \frac{1}{2}\int_0^{2n} \frac{dt}{d(\psi_n(t))} = \frac{1}{2}\int_0^{2n} \frac{dt}{\lceil V_n(t)\rceil}.$$

A change of variables yields, see (3.3),

$$m_1(T_n) = n\int_0^1 \frac{dt}{\lceil V_n(2nt)\rceil} = n^{1/2}\int_0^1 \frac{dt}{\widehat{V}_n(t)}. \tag{4.10}$$

By (4.9), the latter integral converges to $\int_0^1 dt/f(t) = m_1(f)$.

Recall now that a sequence (g_n) of functions on a measure space (Ω,μ) with total mass 1 is *uniformly integrable* if $\sup_n \int_\Omega |g_n|\,d\mu < \infty$ and

$$\sup_{\mu(A)\le\delta}\sup_n \int_A |g_n|\,d\mu \to 0 \qquad \text{as } \delta\to 0.$$

If all $g_n \ge 0$ and $g_n \to g$ a.e., we have the useful equivalence, see e.g. [23, Proposition 4.12],

$$\{g_n\}_1^\infty \text{ is uniformly integrable} \qquad \Longleftrightarrow \qquad \int g_n \to \int g < \infty. \tag{4.11}$$

Since (4.8) implies $\widetilde{V}_n(t) \to f(t)$ for every $t\in[0,1]$, and thus $\widehat{V}_n(t)\to f(t)$ and $1/\widehat{V}_n(t) \to 1/f(t)$, (4.9) implies that $\{1/\widehat{V}_n(t)\}$ is uniformly integrable. More generally, for every fixed $k \ge 1$,

$$\int_0^1\cdots\int_0^1 \frac{dt_1\cdots dt_k}{\widehat{V}_n(t_1)\cdots\widehat{V}_n(t_k)} = \left(\int_0^1 \frac{dt}{\widehat{V}_n(t)}\right)^k \to \left(\int_0^1 \frac{dt}{f(t)}\right)^k = \int_0^1\cdots\int_0^1 \frac{dt_1\cdots dt_k}{f(t_1)\cdots f(t_k)},$$

and thus, by (4.11), $\bigl\{1/\bigl(\widehat{V}_n(t_1)\cdots\widehat{V}_n(t_k)\bigr)\bigr\}$ is uniformly integrable on $[0,1]^k$. By (4.1), this implies that

$$\left\{\frac{1}{L_{\widehat{V}_n}(t_1)\cdots L_{\widehat{V}_n}(t_1,\dots,t_k)}\right\}_{n=1}^\infty \quad \text{is uniformly integrable on } [0,1]^k. \tag{4.12}$$

Let $\mathcal{D}$ be the set of pairs (v,w) of non-root vertices in T_n such that $v=w$ or v is a descendant of w. Then the sum in (4.4) is over all non-roots $(v_1,\dots,v_k)$ such that $(v_i,v_j)\notin\mathcal{D}$ for $1\le i<j\le k$. Fix $k\ge1$ and let

$$\mathcal{E} = \bigcup_{1\le i<j\le k}\bigl\{(x_1,\dots,x_k)\in[0,2n]^k : (\psi(x_i),\psi(x_j))\in\mathcal{D}\bigr\}$$

$$\widehat{\mathcal{E}} = \bigcup_{1\le i<j\le k}\bigl\{(t_1,\dots,t_k)\in[0,1]^k : (\psi(2nt_i),\psi(2nt_j))\in\mathcal{D}\bigr\}.$$

For each w, $\mathcal{D}$ contains $d(w)$ pairs (v,w). Hence,

$$|\mathcal{D}| = \sum_w d(w) \le n\max_w d(w) = n\max V_n$$

and, using Lemma 3.1,

$$|\widehat{\mathcal{E}}| \le \binom{k}{2} n^{-2}|\mathcal{D}| \le k^2 n^{-1}\max V_n = k^2 n^{-1/2}\max\widetilde{V}_n. \tag{4.13}$$

We now take $v_i = \psi_n(x_i)$ in (4.4), and obtain by Lemmas 3.1 and 4.4,

$$\begin{aligned}\mathbb{E}\, X(T_n)^{\underline{k}} &= k!\, 2^{-k} \int \cdots \int_{[0,2n]^k \setminus \mathcal{E}} \frac{dx_1 \cdots dx_k}{L_{\lceil V_n \rceil}(x_1) \cdots L_{\lceil V_n \rceil}(x_1, \dots, x_k)} \\ &= k!\, n^{k/2} \int \cdots \int_{[0,1]^k \setminus \widehat{\mathcal{E}}} \frac{dt_1 \cdots dt_k}{L_{\widehat{V}_n}(t_1) \cdots L_{\widehat{V}_n}(t_1, \dots, t_k)}.\end{aligned}$$

Since $\max \widetilde{V}_n \to \max f < \infty$ by (4.8), we have by (4.13) $|\widehat{\mathcal{E}}| \to 0$ as $n \to \infty$. The uniform integrability (4.12) thus implies that the integral over $\widehat{\mathcal{E}}$ tends to 0. Hence,

$$n^{-k/2}\, \mathbb{E}\, X(T_n)^{\underline{k}} = k! \int \cdots \int_{[0,1]^k} \frac{dt_1 \cdots dt_k}{L_{\widehat{V}_n}(t_1) \cdots L_{\widehat{V}_n}(t_1, \dots, t_k)} + o(1). \tag{4.14}$$

Moreover, (4.8) implies that also $\widehat{V}_n \to f$ uniformly on $[0,1]$. Hence, whenever $0 \le t_1 \le t_2 \le 1$, $\inf_{[t_1,t_2]} \widehat{V}_n \to \inf_{[t_1,t_2]} f$. Thus, by (1.7),

$$L_{\widehat{V}_n}(t_1, \dots, t_k) \to L_f(t_1, \dots, t_k), \qquad t_1, \dots, t_k \in [0,1].$$

It now follows from (4.14), (4.12) and (4.11) that

$$n^{-k/2}\, \mathbb{E}\, X(T_n)^{\underline{k}} \to k! \int \cdots \int_{[0,1]^k} \frac{dt_1 \cdots dt_k}{L_f(t_1) \cdots L_f(t_1, \dots, t_k)} = m_k(f).$$

In particular, $\mathbb{E}\, X(T_n)^{\underline{k}} = O(n^{k/2})$ for every fixed k. The relation between ordinary and factorial moments now shows that

$$n^{-k/2}\, \mathbb{E}\, X(T_n)^k \to m_k(f), \qquad k \ge 1,$$

as asserted. Lemma 4.2 shows that the method of moment applies, so $n^{-1/2} X(T_n)$ converges in distribution to a limit with moments $m_k(f)$. Thus ν_f in Theorem 1.9 exists, and $n^{-1/2} X(T_n) \xrightarrow{\mathrm{d}} \nu_f$. ∎

Remark 4.6. The assumption (4.8) may be relaxed. For example, it is enough (by the same proof) to assume that $\sup_n \sup_t \widetilde{V}_n(t) < \infty$ and that $\widetilde{V}_n \to f$ uniformly on each subinterval $[a,b]$ with $0 < a < b < 1$. See also Section 9.

Lemma 4.7. *Let $\mathcal{T}_n$ be a conditioned Galton–Watson tree as in Section 1, and let $F = 2\sigma^{-1} B_{\mathrm{ex}}$. Then*

$$\left(\widetilde{V}_n, \int_0^1 \frac{dt}{\widehat{V}_n(t)} \right) \xrightarrow{\mathrm{d}} \left(F, \int_0^1 \frac{dt}{F(t)} \right)$$

in $C[0,1] \times \mathbb{R}$.

Proof. Of course, this is based on (3.4). The only problem is that $f \mapsto \int_0^1 dt/f(t)$ is not a continuous functional on $C[0,1]$. We therefore use a truncated version.

Let ϕ_ε be the function with $\phi_\varepsilon = 0$ on $[0,\varepsilon]$, $\phi_\varepsilon = 1$ on $[2\varepsilon,\infty)$, and ϕ_ε linear on $[\varepsilon, 2\varepsilon]$. Define

$$
\begin{aligned}
Y_n &:= \int_0^1 \frac{1}{\widehat{V}_n(t)}\,dt, & Y &:= \int_0^1 \frac{1}{F(t)}\,dt,\\
Y_n^\varepsilon &:= \int_0^1 \frac{\phi_\varepsilon\bigl(\widehat{V}_n(t)\bigr)}{\widehat{V}_n(t)}\,dt, & Y^\varepsilon &:= \int_0^1 \frac{\phi_\varepsilon\bigl(F(t)\bigr)}{F(t)}\,dt.
\end{aligned}
\tag{4.15}
$$

By (3.4), $\widetilde{V}_n \xrightarrow{\mathrm{d}} F$ in $C[0,1]$. Using the Skorohod coupling theorem, see e.g. [23, Theorem 4.30], we may pretend that $\widetilde{V}_n \xrightarrow{\text{a.s.}} F$, i.e. a.s. $\widetilde{V}_n \to F$ uniformly. Then $\widehat{V}_n \to F$ uniformly too, and since $x \mapsto \phi_\varepsilon(x)/x$ is uniformly continuous, it follows that $Y_n^\varepsilon \to Y^\varepsilon$. Consequently (or by [7, Theorem 5.5]), for every fixed $\varepsilon > 0$, (3.4) implies

$$
\bigl(\widetilde{V}_n, Y_n^\varepsilon\bigr) \xrightarrow{\mathrm{d}} \bigl(F, Y^\varepsilon\bigr) \qquad \text{as } n \to \infty. \tag{4.16}
$$

Further it is clear, by monotone convergence, that $Y^\varepsilon \to Y$ as $\varepsilon \to 0$, for every fixed n.

Arguing as for (4.10) (backwards),

$$
0 \le Y_n - Y_n^\varepsilon \le n^{-1/2} \sum_{d(v)\le 2\varepsilon n^{1/2}} \frac{1}{d(v)} = n^{-1/2} \sum_{k=1}^{2\varepsilon n^{1/2}} \frac{w_k(\mathcal{T}_n)}{k}
$$

and thus, by Theorem 1.13,

$$
\mathbb{E}\,|Y_n - Y_n^\varepsilon| \le n^{-1/2} \sum_{k=1}^{2\varepsilon n^{1/2}} \frac{\mathbb{E}\,w_k(\mathcal{T}_n)}{k} \le 2C\varepsilon. \tag{4.17}
$$

Consequently,

$$
\lim_{\varepsilon\to 0} \limsup_{n\to\infty} \mathbb{E}\,|(\widetilde{V}_n, Y_n^\varepsilon) - (\widetilde{V}_n, Y_n)| = \lim_{\varepsilon\to 0} \limsup_{n\to\infty} \mathbb{E}\,|Y_n^\varepsilon - Y_n| = 0. \tag{4.18}
$$

By [7, Theorem 4.2], we thus can let $\varepsilon \to 0$ in (4.16) (interchanging the order of the limits) and obtain $(\widetilde{V}_n, Y_n) \xrightarrow{\mathrm{d}} (F, Y)$. $\square$

Lemma 4.8. *Let $\mathcal{T}_n$ be a conditioned Galton–Watson tree. If r is an integer such that $\mathbb{E}\,\xi^{r+1} < \infty$, then $\mathbb{E}\,m_1(\mathcal{T}_n)^r = O(n^{r/2})$.*

Proof. By (4.5),

$$
m_1(\mathcal{T}_n) = \sum_{k=1}^{\infty} \frac{w_k(\mathcal{T}_n)}{k} \le \sum_{k=1}^{n^{1/2}} \frac{w_k(\mathcal{T}_n)}{k} + \frac{n}{n^{1/2}}.
$$

Hence, by Minkowski's inequality and Theorem 1.13,

$$
\|m_1(\mathcal{T}_n)\|_r \le \sum_{k=1}^{n^{1/2}} \frac{\|w_k(\mathcal{T}_n)\|_r}{k} + n^{1/2} \le C n^{1/2}. \qquad \square
$$

Lemma 4.9. *Let $\mathcal{T}_n$ be a conditioned Galton–Watson tree. For every fixed integer $k \geq 1$ such that $\mathbb{E}\,\xi^{k+1} < \infty$, $\mathbb{E}\,X(\mathcal{T}_n)^k = O(n^{k/2})$.*

Proof. For any tree T, $L_T(v_1, \dots, v_j) \geq L_T(v_j) = d(v_j)$. Lemma 4.3 thus implies

$$\mathbb{E}\,X(T)^{\underline{k}} \leq k! \sum_{v_1,\dots,v_k \neq o} \frac{1}{d(v_1)\cdots d(v_k)} = k!\,\bigl(\mathbb{E}\,X(T)\bigr)^k = k!\,m_1(T)^k. \tag{4.19}$$

Consequently, $\mathbb{E}\,X(\mathcal{T}_n)^{\underline{k}} = O(n^{k/2})$ by Lemma 4.8, and the result follows by expressing X^k in falling factorials. $\square$

Proof of Theorem 1.10. By Lemma 4.7 and the Skorohod coupling theorem, see e.g. [23, Theorem 4.30], we may assume that the trees $\mathcal{T}_n$ are defined on a common probability space and that

$$\left(\widetilde{V}_n, \int_0^1 \frac{dt}{\widehat{V}_n(t)}\right) \xrightarrow{\text{a.s.}} \left(F, \int_0^1 \frac{dt}{F(t)}\right),$$

with $F = 2\sigma^{-1}B_{\mathrm{ex}}$. Lemma 4.5 now shows that a.s., for every $k \geq 1$,

$$\sigma^{-k} n^{-k/2} m_k(\mathcal{T}_n) \to \sigma^{-k} m_k(F) = m_k(2B_{\mathrm{ex}}),$$

and thus $\mu_{\mathcal{T}_n} \to \nu_{2B_{\mathrm{ex}}}$. This proves (1.9) and (1.10), jointly with (3.4).

Finally, assume $\mathbb{E}\,\xi^m < \infty$ for all m. By Jensen's inequality, for integers $k, r \geq 1$,

$$m_k(\mathcal{T}_n)^r = \mathbb{E}\bigl(X(\mathcal{T}_n)^k \mid \mathcal{T}_n\bigr)^r \leq \mathbb{E}\bigl(X(\mathcal{T}_n)^{rk} \mid \mathcal{T}_n\bigr) = m_{rk}(\mathcal{T}_n)$$

and thus $\mathbb{E}\,m_k(\mathcal{T}_n)^r \leq \mathbb{E}\,X(\mathcal{T}_n)^{rk} = O\bigl(n^{rk/2}\bigr)$ by Lemma 4.9. Hence, every moment of the left hand side of (1.10) stays bounded as $n \to \infty$. This implies moment convergence in (1.10), which clearly is equivalent to (1.11). $\square$

Finally, we prove existence in Theorem 1.9. We do this in three steps.

Step 1: $\min f > 0$ *and* f *is Lipschitz:* $|f(x) - f(y)| \leq C|x - y|$ *for some* C *and all* $x, y \in [0, 1]$. Define

$$g_n(2k) := 2\lceil \tfrac{1}{2}\sqrt{n} f(k/n)\rceil$$

for even integers $2k = 0, 2, \dots, 2n$. Assume that $n > C^2$; then the Lipschitz assumption yields $|f((k+1)/n) - f(k/n)| \leq C/n < 1/\sqrt{n}$ and thus $g_n(2k+2) - g_n(2k) \in \{-2, 0, 2\}$ for every $k = 0, \dots, n-1$. Define $g_n(2k+1) := 1 + \min\bigl(g(2k), g(2k+2)\bigr)$; then $g_n(j) - g_n(j-1) = \pm 1$ for every integer $j = 1, \dots, 2n$. Hence, g_n is a simple walk on $\{0, 1, \dots, 2n\}$, but it is not 0 at the endpoints. We thus define $V_n(j) := \min\bigl(g_n(j),\, j,\, 2n - j\bigr)$, and observe that V_n is a simple walk that is the depth-first walk of some tree T_n with n edges.

Extend g_n to $[0, 2n]$ by linear interpolation and let, cf. (3.2) and (3.3), $\tilde{g}_n(t) := n^{-1/2} g_n(2nt)$ and $\hat{g}_n(t) := n^{-1/2}\lceil g_n(2nt)\rceil$. Then $|\tilde{g}_n(k/n) - f(k/n)| < 2n^{-1/2}$ for each $k = 0, \dots, n$, and it follows easily that $\tilde{g}_n \to f$ and $\hat{g}_n \to f$

uniformly on $[0,1]$. Further, $\hat{g}_n \geq \min f > 0$, so by dominated convergence, $\int dt/\hat{g}_n(t) \to \int dt/f(t)$.

If $A := \max f$, then $g_n \leq An^{1/2}+3$, and thus $V_n(j) = g_n(j)$ whenever $An^{1/2}+3 \leq j \leq 2n - An^{1/2} - 3$; hence $\widetilde{V}_n(t) = \tilde{g}_n(t)$ and $\widehat{V}_n(t) = \hat{g}_n(t)$ on $[(A+4)n^{-1/2}, 2n-(A+4)n^{-1/2}]$. Consequently, $\widetilde{V}_n(t) \to f(t)$ uniformly on every interval $[a,b]$ with $0<a<b<1$. Moreover, $V_n(t) = \min(g_n(t),\, t,\, 2n-t)$ for non-integer $t \in [0,2n]$ too, and thus

$$\frac{1}{\widehat{V}_n(t)} = \max\Big(\frac{1}{\hat{g}_n(t)}, \frac{n^{1/2}}{\lceil 2nt\rceil}, \frac{n^{1/2}}{\lceil 2n(1-t)\rceil}\Big) \leq \frac{1}{\hat{g}_n(t)} + \frac{n^{1/2}}{\lceil 2nt\rceil} + \frac{n^{1/2}}{\lceil 2n(1-t)\rceil}.$$

Consequently,

$$0 \leq \int_0^1 \frac{dt}{\widehat{V}_n(t)} - \int_0^1 \frac{dt}{\hat{g}_n(t)} \leq 2n^{1/2}\int_0^1 \frac{dt}{\lceil 2nt\rceil} = n^{-1/2}\sum_{j=1}^{2n}\frac{1}{j} = o(1).$$

The trees T_n thus satisfy the assumptions of Lemma 4.5 as modified in Remark 4.6. Consequently, $n^{-1/2}X(T_n) \xrightarrow{\mathrm{d}} \nu_f$, which shows that ν_f exists.

Step 2: $f \in C[0,1]_+$ with $\min f > 0$. There exist strictly positive Lipschitz functions f_N such that $f_N \to f$ uniformly on $[0,1]$ as $N \to \infty$. ν_{f_N} exists for every N by Step 1. It follows easily that $m_k(f_N) \to m_k(f)$ for every $k \geq 1$, and thus ν_{f_N} converges by the method of moments to a distribution ν_f. (See also Lemma 9.2 below.)

Step 3: $f \in C[0,1]_+$ with $\int_0^1 dt/f(t) < \infty$. Define $f_N(t) := f(t) + 1/N$. The method of moment applies again, and shows the existence of ν_f.

5. Proofs of Theorems 1.6 and 1.12

Proof of Theorem 1.6. By the definition of $\mu_{\mathcal{T}_n}$ and Theorem 1.10, for any bounded continuous function $f : \mathbb{R} \to \mathbb{R}$,

$$\mathbb{E}\big(f(\sigma^{-1}n^{-1/2}X(\mathcal{T}_n)) \mid \mathcal{T}_n\big) = \int f\, d\mu_{\mathcal{T}_n} \xrightarrow{\mathrm{d}} \int f\, \nu_{2B_{\mathrm{ex}}}.$$

Taking expectations we find, by dominated convergence,

$$\mathbb{E}\big(f(\sigma^{-1}n^{-1/2}X(\mathcal{T}_n))\big) \to \mathbb{E}\int f\, d\nu_{2B_{\mathrm{ex}}} = \int f\, d\nu,$$

where $\nu = \mathbb{E}\,\nu_{2B_{\mathrm{ex}}}$. This shows convergence of $\sigma^{-1}n^{-1/2}X(\mathcal{T}_n)$ in distribution to some limit ν, i.e. (1.4) holds for some Z.

By Lemma 4.9, every moment on $n^{-1/2}X(\mathcal{T}_n)$ stays bounded as $n \to \infty$, which together with (1.4) implies moment convergence in (1.4).

It remains to identify the limit ν as the Rayleigh distribution. Note that ν does not depend on the distribution of ξ. We have thus proved an invariance principle, so in order to identify the limit we can appeal to the special cases proved by Chassaing and Marchand [9] and Panholzer [33].

We can also identify ν directly as follows. We have

$$\int x^k \, d\nu(x) = \mathbb{E} \int x^k \, d\nu_{2B_{\mathrm{ex}}}(x) = \mathbb{E}\, m_k(2B_{\mathrm{ex}}).$$

The following lemma computes these moments. A simple integration shows that Z has the same moments, and the proof is complete. $\square$

Lemma 5.1. $\mathbb{E}\, m_k(2B_{\mathrm{ex}}) = 2^{k/2}\Gamma(k/2+1)$, *for every* $k \geq 1$.

Proof. In this proof, the edges of trees may have arbitrary positive real lengths.

The continuum random tree is a metric space constructed by Aldous [2, §4.3] in several different ways. One construction represents the continuum random tree by the random function $2B_{\mathrm{ex}}$, such that each $t \in [0,1]$ corresponds to a vertex (point) $\psi(t)$ in the continuum tree and the subtree spanned by the root and $\psi(t_1), \dots, \psi(t_k)$ has total edge length $L_{2B_{\mathrm{ex}}}(t_1, \dots, t_k)$, cf. [2, Theorem 13]. Another construction says that if $U_1, \dots, U_k$ are random numbers in $[0,1]$, uniformly distributed and independent, then the random subtree of the continuum random tree spanned by the corresponding vertices and the root has the same distribution as the following tree: Let $Y_1, \dots, Y_k$ be the first k points in a Poisson process on $(0, \infty)$ with intensity $x\,dx$. Let T_1 be a single edge of length Y_1 from the root to v_1. T_i for $i \geq 2$ is defined inductively by choosing a new branch-point uniformly on the edges of T_{i-1}, and attaching v_i to this point by an edge of length $Y_i - Y_{i-1}$. It follows that $L_{2B_{\mathrm{ex}}}(U_1, \dots, U_i) \overset{\mathrm{d}}{=} Y_i$ for $i = 1, \dots, k$ (jointly). Since $Y_1, \dots, Y_k$ have the joint density function $y_1 \cdots y_k e^{-y_k^2/2}$ on $0 < y_1 < \cdots < y_k$ by standard properties of Poisson processes [2], (1.8) yields

$$\begin{aligned}
\mathbb{E}\, m_k(2B_{\mathrm{ex}}) &= \mathbb{E} \frac{k!}{L_{2B_{\mathrm{ex}}}(U_1) \cdots L_{2B_{\mathrm{ex}}}(U_1, \dots, U_k)} = \mathbb{E} \frac{k!}{Y_1 \cdots Y_k} \\
&= \int \cdots \int_{0<y_1<\cdots<y_k} \frac{k!}{y_1 \cdots y_k} y_1 \cdots y_k e^{-y_k^2/2} \, dy_1 \cdots dy_k \\
&= k! \int \cdots \int_{0<y_1<\cdots<y_k} e^{-y_k^2/2} \, dy_1 \cdots dy_k = k! \int_0^\infty \frac{y_k^{k-1}}{(k-1)!} e^{-y_k^2/2} \, dy_k \\
&= k \int_0^\infty (2x)^{k/2-1} e^{-x} \, dx = k 2^{k/2-1} \Gamma(k/2) = 2^{k/2}\Gamma(k/2+1). \qquad \square
\end{aligned}$$

Proof of Theorem 1.12. By (1.6) and (1.11), it remains only to show

$$\mathbb{E}\, m_2(2B_{\mathrm{ex}}) = 2, \qquad \big(\mathbb{E}\, m_1(2B_{\mathrm{ex}})\big)^2 = \pi/2, \qquad \mathbb{E}\big(m_1(2B_{\mathrm{ex}})\big)^2 = \pi^2/6.$$

The two first follow by taking $k = 2$ and 1 in Lemma 5.1. The third follows from the identity in law $m_1(B_{\mathrm{ex}}) \overset{\mathrm{d}}{=} \max B_{\mathrm{ex}}$, see (7.4) below, and known expressions for its moments (following from (7.5)), see e.g. [5].

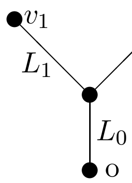

FIGURE 1. The tree T_2 with two leaves.

We can also use the same method as in the proof of Lemma 5.1. Using the notations there,

$$\begin{aligned}\mathbb{E}\big(m_1(2B_{\mathrm{ex}})\big)^2 &= \mathbb{E}\int_0^1\int_0^1 \frac{dt_1\,dt_2}{2B_{\mathrm{ex}}(t_1)\cdot 2B_{\mathrm{ex}}(t_2)}\\ &= \mathbb{E}\,\frac{1}{2B_{\mathrm{ex}}(U_1)\cdot 2B_{\mathrm{ex}}(U_2)} = \mathbb{E}\,\frac{1}{d(v_1)\cdot d(v_2)},\end{aligned}$$

where v_i is the vertex in the continuum random tree corresponding to U_i. The tree T_2 spanned by v_1, v_2 and the root o contains also a branchpoint and three edges of lengths L_0, L_1 and L_2, say, see Figure 1. By the construction above, $d(v_1) = L_0 + L_1 = Y_1$, $L_2 = Y_2 - Y_1$, and L_0 is chosen uniformly in $(0, Y_1)$. Hence, $\mathbb{E}\big(m_1(2B_{\mathrm{ex}})\big)^2$ equals

$$\begin{aligned}\mathbb{E}\,&\frac{1}{(L_0+L_1)(L_0+L_2)} = \mathbb{E}\,\frac{1}{Y_1(Y_2-L_1)} = \mathbb{E}\,\frac{1}{Y_1}\int_0^{Y_1}\frac{dl}{Y_1(Y_2-l)}\\ &= \mathbb{E}\,Y_1^{-2}\big(\ln Y_2 - \ln(Y_2-Y_1)\big)\\ &= \iint_{0<y_1<y_2} y_1^{-2}\big(-\ln(1-y_1/y_2)\big)\,y_1y_2e^{-y_2^2/2}\,dy_1\,dy_2\\ &= \iint_{0<y_1<y_2}\sum_{k=1}^\infty \frac{1}{k}y_1^{k-1}y_2^{1-k}e^{-y_2^2/2}\,dy_1\,dy_2\\ &= \sum_{k=1}^\infty\frac{1}{k^2}\int_0^\infty y_2e^{-y_2^2/2}\,dy_2 = \frac{\pi^2}{6}.\end{aligned}$$

□

Remark 5.2. The case $k = 1$ of Lemma 5.1, $\mathbb{E}\,m_1(2B_{\mathrm{ex}}) = \sqrt{\pi/2}$, can by (1.8) be written $\mathbb{E}\int_0^1 dt/B_{\mathrm{ex}}(t) = \sqrt{2\pi}$. This well-known fact [37, Exercise XI.(3.9)] can be proved in several other ways too. One, straightforward, way is to compute $\mathbb{E}(1/B_{\mathrm{ex}}(t))$ for each t from the density function of $B_{\mathrm{ex}}(t)$ [8, II.(1.4)], and then integrate. Another way is to use the identity in distribution (7.4) below together with (7.5).

6. VERTEX CUTTINGS AND RECORDS

Now consider the vertex version $X_v(T)$. We couple $X_v(T)$ and $X(T)$ by using the vertex record formulation and $X(T) = X_v(T')$, where, as in the introduction, T' is T with the root deleted.

The root is always a record, and $X_v(T) - 1$ counts the number of other vertices that are records, while $X_v(T')$ counts the number of other vertices that are records if we ignore the root. Hence $X_v(T) - 1 \le X_v(T') = X(T)$. Moreover, the probability that a vertex v with depth $d(v) = k \ge 1$ is a record in T is $1/(k+1)$, while it is $1/k$ if we ignore the root (i.e. in T'). Hence

$$\mathbb{E}\big(X_v(T') - (X_v(T) - 1)\big) = \sum_{k=1}^{\infty} w_k(T)\Big(\frac{1}{k} - \frac{1}{k+1}\Big) = \sum_{k=1}^{\infty} \frac{w_k(T)}{k(k+1)}.$$

We have shown the following.

Lemma 6.1. *For any rooted tree T, it is possible to couple $X(T)$ and $X_v(T)$ such that $X_v(T) \le X(T) + 1$ and*

$$\mathbb{E}\,|X(T) - X_v(T)| \le 1 + \sum_{k=1}^{\infty} \frac{w_k(T)}{k(k+1)}. \qquad \square$$

Theorem 6.2. *Theorems 1.6, 1.10 and 1.12 hold for the vertex version $X_v(\mathcal{T}_n)$ too.*

Proof. Let $\mathcal{T}_n$ be a conditioned Galton–Watson tree as in Section 1, and use the coupling in Lemma 6.1. Since $w_k(\mathcal{T}_n) = 0$ for $k > n$, we have by Theorem 1.13

$$\begin{aligned} n^{-1/2}\,\mathbb{E}\,|X(\mathcal{T}_n) - X_v(\mathcal{T}_n)| &\le n^{-1/2} + n^{-1/2} \sum_{k=1}^{n} \frac{\mathbb{E}\, w_k(\mathcal{T}_n)}{k(k+1)} \\ &\le n^{-1/2} + n^{-1/2} C \sum_{k=1}^{n} \frac{1}{k} \to 0. \end{aligned}$$

It follows that (1.4) holds with X_v too.

Further, conditioning on $\mathcal{T}_n$, we see that $n^{-1/2}\,\mathbb{E}\big(|X_v(\mathcal{T}_n) - X(\mathcal{T}_n)| \bigm| \mathcal{T}_n\big) \xrightarrow{\text{p}} 0$. By the Skorohod coupling theorem [23, Theorem 4.30], we may assume that this and (1.9) hold together a.s., which implies that (1.9) holds for X_v too. Using $X_v(\mathcal{T}_n) \le X(\mathcal{T}_n) + 1$, it can similarly be shown that (1.10) and (1.11) (when $\mathbb{E}\,\xi^m < \infty$, $m \ge 1$) hold for X_v too. Theorem 1.12 for X_v then follows as before. We omit the details, since the result also follows by Theorem 6.5 below, see Example 6.3. $\square$

Let us generalize the Galton–Watson tree by assuming that the root may have a different offspring distribution than the other vertices; say that the number of children of the root is η. Each of the η children then grows (independently) into a tree as before, with offspring distribution ξ in all following generations. Let $\mathcal{T}^\eta$ be the resulting tree, and let $\mathcal{T}_n^\eta$ be $\mathcal{T}^\eta$ conditioned to have order $n+1$.

Example 6.3. If $\eta = 1$, then $\mathcal{T}_n^\eta$ is just $\widetilde{\mathcal{T}_{n-1}}$, i.e. $\mathcal{T}_{n-1}$ with a new root attached. Thus $X(\mathcal{T}_n^\eta) = X(\widetilde{\mathcal{T}_{n-1}}) = X_v(\mathcal{T}_{n-1})$, so this is another way of looking at X_v and the results just proved for it.

Example 6.4. The non-crossing trees were shown by Marckert and Panholzer [30] to be of this type, with $\eta \sim \mathrm{Ge}(2/3)$ and $\xi \sim \mathrm{NegBin}(2, 2/3)$. (Thus ξ is distributed as the sum of two independent copies of η; this corresponds to the fact that at all vertices except the root, we have two sides and may distinguish between children to the left and to the right [30].) As a consequence, it is shown in [30] that (3.4) holds for random non-crossing trees too, with $\sigma^2 = \operatorname{Var} \xi = 3/2$.

Random cutting of random non-crossing trees was studied by Panholzer [34]; his result is a version of our Theorem 1.6 for non-crossing trees. We generalize this result.

Theorem 6.5. *Let $\mathcal{T}_n^\eta$ be as above, with $0 < \mathbb{E}\,\eta < \infty$, $\mathbb{E}\,\xi = 1$ and $0 < \sigma^2 = \operatorname{Var} \xi < \infty$. Then Theorems 1.6, 1.10, 1.12 and 1.13 hold for $\mathcal{T}_n^\eta$ too, provided the assumptions on existence of higher moments of ξ now include η too. (C in Theorem 1.13 may depend on η too.)*

Proof. We begin with a lemma. We write $Y_n = O_p(1)$ for a family $\{Y_n\}$ of random variables if $\sup_n \mathbb{P}(|Y_n| > M) \to 0$ as $M \to \infty$; this is also known as *stochastically bounded* or *tight*.

Let $d_0(T)$ denote the degree of the root of T. Deleting the root of $\mathcal{T}_n^\eta$, we obtain $d_0(\mathcal{T}_n^\eta)$ branches; we order them $B_1, \dots, B_{d_0}$ such that $|B_1| \ge |B_2| \ge \dots$. We show first that all but a few vertices belong to the largest branch.

Lemma 6.6. $|B_1| = n - O_p(1)$.

Proof. First, note that by (2.1), for any $n \ge 1$,

$$\mathbb{P}(|\mathcal{T}^\eta| = n+1) = \sum_{m=1}^{\infty} \mathbb{P}(\eta = m) \frac{m}{n} \mathbb{P}(S_n = n - m). \tag{6.1}$$

Assume now, for simplicity, that $\operatorname{span}(\xi) = 1$; we leave the minor differences in the general case to the reader. We will in this section let C and c denote various positive constants that depend on ξ and η. Fixing some $m > 0$ with $\mathbb{P}(\eta = m) > 0$, we see from (6.1) and the local central limit theorem, cf. (2.2), that (for large n)

$$\mathbb{P}(|\mathcal{T}^\eta| = n+1) \ge \mathbb{P}(\eta = m) \frac{m}{n} \mathbb{P}(S_n = n - m) \ge c n^{-3/2}.$$

Conversely, by (6.1) and Lemma 2.1,

$$\mathbb{P}(|\mathcal{T}^\eta| = n+1) \le \sum_{m=1}^{\infty} \mathbb{P}(\eta = m) \frac{m}{n} C n^{-1/2} = C n^{-3/2}.$$

Consequently, for n large,

$$c n^{-3/2} \le \mathbb{P}(|\mathcal{T}^\eta| = n+1) \le C n^{-3/2}. \tag{6.2}$$

We see in the same way by (6.1) and Lemma 2.1, that for any $M \geq 0$,

$$\begin{aligned}\mathbb{P}(|\mathcal{T}^\eta| = n+1,\, d_0(\mathcal{T}^\eta) \geq M) &\leq \sum_{m=M}^{\infty} \mathbb{P}(\eta = m)\frac{m}{n} C n^{-1/2}\\ &= C n^{-3/2} \sum_{M}^{\infty} m\,\mathbb{P}(\eta = m)\end{aligned}$$

and thus,

$$\mathbb{P}(d_0(\mathcal{T}_n^\eta) \geq M) = \frac{\mathbb{P}(|\mathcal{T}^\eta| = n+1,\, d_0(\mathcal{T}^\eta) \geq M)}{\mathbb{P}(|\mathcal{T}^\eta| = n+1)} \leq C \sum_{M}^{\infty} m\,\mathbb{P}(\eta = m),$$

which tends to 0 as $M \to \infty$. Hence $d_0(\mathcal{T}_n^\eta) = O_p(1)$.

To prove the lemma, it is therefore sufficient to prove it conditioned on $d_0(\mathcal{T}_n^\eta) \leq M$, for every fixed M. By further conditioning, it is sufficient to prove it conditioned on $d_0(\mathcal{T}_n^\eta) = m$ for every $m \geq 1$, i.e. to prove the lemma in the case when $\eta = m$ is constant.

Hence assume $\eta = m$, so there are m branches. Then $\mathcal{T}^\eta$ consists of a root and branches $\mathcal{T}^{(i)}$, $i = 1\dots,m$, which are independent copies of $\mathcal{T}$. Given $\mathcal{T}^{(2)},\dots,\mathcal{T}^{(m)}$, let $N = n - \sum_2^m |\mathcal{T}^{(i)}|$; then the (conditional) probability that $\sum_1^m |\mathcal{T}^{(i)}| = n$ and $|\mathcal{T}^{(1)}| \geq |\mathcal{T}^{(2)}| \geq \dots$ is either 0 or $\mathbb{P}(|\mathcal{T}^{(1)}| = N) \leq CN^{-3/2} \leq C_m n^{-3/2}$, since the event is possible only if $N \geq n/m$. (C_m denotes constants that depend on m.) It follows by (6.2) that, returning to $\mathcal{T}_n^\eta$, $\mathbb{P}(|B_2| = k) \leq C_m\,\mathbb{P}(|\mathcal{T}| = k)$. Hence $|B_2| = O_p(1)$ and $n - |B_1| \leq (m-1)|B_2| = O_p(1)$. $\square$

It follows easily that the difference between the rescaled depth-first walks $\widetilde{V}$ for $\mathcal{T}_n^\eta$ and for B_1 tends to 0 uniformly, in probability. Given $|B_1| = n-k$, B_1 is a conditioned Galton–Watson tree of order $n-k$, so (3.4) holds for B_1. It follows that (3.4) holds for $\mathcal{T}_n^\eta$ too.

It is now easy to check that the proofs in Sections 2, 4, 5 hold for $\mathcal{T}_n^\eta$ too (with a few trivial modifications). $\square$

7. Height and width

The sequence $\{w_k(T)\}_{k=0}^\infty$ is called the *profile* of T, and $W(T) := \max_k w_k(T)$ is called the *width* of T. Further, the *height* of T is

$$H(T) := \max_{v \in T} d(v) = \max\{k : w_k(T) > 0\}.$$

The asymptotics of these for conditioned Galton–Watson trees are well-known, see e.g. [1], [10] and the further references there. First [1], using the depth-first walks in Section 3, since $H(T) = \max_t V_T(t) = n^{1/2} \max_t \widetilde{V}(t)$, (3.4) implies that

$$n^{-1/2} H(\mathcal{T}_n) \xrightarrow{\mathrm{d}} 2\sigma^{-1} \max_t B_{\mathrm{ex}}(t). \tag{7.1}$$

By Theorem 1.10, this extends to joint convergence with $m_1(\mathcal{T}_n)$, the expected number of cuts or records in the tree which is given by (4.5), in the form

$$\begin{aligned} n^{-1/2}\bigl(H(\mathcal{T}_n),\ m_1(\mathcal{T}_n)\bigr) &\xrightarrow{\mathrm{d}} \bigl(2\sigma^{-1}\max_t B_{\mathrm{ex}}(t),\ \sigma m_1(2B_{\mathrm{ex}})\bigr) \\ &= \Bigl(\frac{2}{\sigma}\max_t B_{\mathrm{ex}}(t),\ \frac{\sigma}{2}\int_0^1 \frac{dt}{B_{\mathrm{ex}}(t)}\Bigr). \end{aligned} \tag{7.2}$$

The profile and the width can be treated similarly [1, 3, 15], but the limits will be described by the local time of the Brownian excursion; this was extended by [10] to include joint distribution with the height. Chassaing, Marckert and Yor [10] further gave a second proof using instead the *breadth-first walk* (see below), which proves

$$n^{-1/2}\bigl(H(\mathcal{T}_n), W(\mathcal{T}_n)\bigr) \xrightarrow{\mathrm{d}} \Bigl(\frac{1}{\sigma}\int_0^1 \frac{dt}{B_{\mathrm{ex}}(t)},\ \sigma\max_t B_{\mathrm{ex}}(t)\Bigr). \tag{7.3}$$

(For simplicity, they considered only binary trees, but the argument extends, see below.) Note that we have the same random variables on the right hand sides of (7.2) and (7.3) (apart from constant factors), but in different order. (For this joint distribution, see [14].) In particular, we see that we have two different descriptions of the limit of $H(\mathcal{T}_n)$, and thus [6]

$$\max_t B_{\mathrm{ex}}(t) \overset{\mathrm{d}}{=} \frac{1}{2}\int_0^1 \frac{dt}{B_{\mathrm{ex}}(t)}. \tag{7.4}$$

(Of course, this is an equality in distribution, and not for individual excursions. Informally, we have two different Brownian excursions in (7.2) and (7.3); the second is a time change of the local time of the first [10], [21].) We remark that the distributions of these random variables are known [11, 24]; see [5] for much more information:

$$\mathbb{P}(\max_t B_{\mathrm{ex}}(t) \le x) = 1 + 2\sum_{k=1}^{\infty}(1-4k^2x^2)\exp(-2k^2x^2), \qquad x>0. \tag{7.5}$$

We employ the second method used by Chassaing, Marckert and Yor [10] to prove (7.3), and extend it to include $m_1(\mathcal{T}_n)$ too:

Theorem 7.1. *Let $\mathcal{T}_n$ be a conditioned Galton–Watson tree of order n, defined by an offspring distribution ξ satisfying* (1.2)–(1.3)*. Then, jointly,*

$$\begin{aligned} n^{-1/2}H(\mathcal{T}_n) &\xrightarrow{\mathrm{d}} \sigma^{-1}\int_0^1 \frac{dt}{B_{\mathrm{ex}}(t)}, \\ n^{-1/2}W(\mathcal{T}_n) &\xrightarrow{\mathrm{d}} \sigma\max_t B_{\mathrm{ex}}(t), \\ n^{-1/2}m_1(\mathcal{T}_n) &\xrightarrow{\mathrm{d}} \sigma\int_0^1 \frac{dt}{\int_0^t ds/B_{\mathrm{ex}}(s)}. \end{aligned}$$

Remark 7.2. By (7.2) and (7.4), the second and third limits have the same distribution, which only differs by a scale factor from the first.

Proof. We follow [10] (extending the argument to arbitrary conditioned Galton–Watson trees).

Define the *breadth-first search* of an ordered tree T with $|T| = n+1$ to be the vertices ordered in a sequence $v_0, v_1, \dots, v_n$ with non-decreasing depths, such that vertices of equal depth are ordered by the ordering in T. (In particular, v_0 is the root o.) With $w_k = w_k(T)$ as above, let $z_k := \sum_{j=1}^k w_j$. (Thus $z_0 = 0$.) Then z_k is the index of the last vertex of depth k, and thus (with $z_{-1} = -1$)

$$d(v_j) = k \iff z_{k-1} < j \le z_k. \tag{7.6}$$

Further, let $S(j) = S_T(j)$ $(0 \le j \le n)$ be the number of vertices v_k with $k > j$ and parents in $\{v_0, \dots, v_j\}$. Thus $S(0)$ is the degree of the root, $S(n-1) = 1$ and $S(n) = 0$. The sequence $\{(S(j)\}_0^n$ is called the *breadth-first walk*. We define $S(t)$ for all real $t \in [0, n]$ by linear interpolation. Note that

$$S(z_k) = w_{k+1}, \qquad k \ge 0. \tag{7.7}$$

We further define $S^*(j) = S_T^*(j) := w_{d(v_j)}$. Thus $S^*(j) := w_k = S(z_{k-1})$ if $z_{k-1} < j \le z_k$, i.e.

$$S^*(j) = S(j^*), \qquad \text{where} \quad j^* := z_{d(v_j)-1}. \tag{7.8}$$

Clearly (for $n \ge 1$),

$$W(T) = \max_k w_k = \max_k S(z_k) = \max_j S^*(j), \tag{7.9}$$

and it is easily seen that $H(T) = d(v_n) = \sum_{i=1}^n \frac{1}{S^*(i)}$ and, more generally,

$$d(v_j) = \left\lceil \sum_{i=1}^j \frac{1}{S^*(i)} \right\rceil, \qquad 0 \le j \le n. \tag{7.10}$$

Consider now the conditioned Galton–Watson tree $\mathcal{T}_n$. Then the breadth-first walk is a random walk with independent increments distributed as $\xi - 1$ (started at $S(-1) = 1$), conditioned on $S(j) > 0$ for $j = 0, \dots, n-1$ but $S(n) = 0$. We normalize by

$$\widetilde{S}_n(t) := n^{-1/2} S_{\mathcal{T}_n}(nt), \qquad 0 \le t \le 1,$$

and have as is well-known, see [22],

$$\widetilde{S}_n \xrightarrow{\mathrm{d}} \sigma B_{\mathrm{ex}} \qquad \text{in } C[0,1]. \tag{7.11}$$

We further define

$$\widehat{S}_n(t) := n^{-1/2} S^*_{\mathcal{T}_n}(\lceil nt \rceil), \qquad 0 \le t \le 1. \tag{7.12}$$

Thus

$$H(\mathcal{T}_n) = \int_0^n S^*_{\mathcal{T}_n}(\lceil x \rceil)^{-1} dx = n^{1/2} \int_0^1 \frac{dt}{\widehat{S}_n(t)}. \tag{7.13}$$

By (7.11) and the Skorohod coupling theorem [23, Theorem 4.30], we may assume that all random variables are defined on the same probability space

Ω, and that for a.e. $\omega \in \Omega$, $\widetilde{S}_n \to \sigma B_{\mathrm{ex}}$ in $C[0,1]$. We fix such an ω; thus $\widetilde{S}_n \to \sigma B_{\mathrm{ex}}$ uniformly. Since, by (7.9),

$$W(\mathcal{T}_n) = \max_k S(z_k) \le \max S = n^{1/2} \max \widetilde{S}_n,$$

we have $W(\mathcal{T}_n)/n = O(n^{-1/2}) \to 0$. By (7.8) and (7.6), for $0 \le j \le n$, $|j - j^*| = |j - z_{d(v_j)-1}| \le \sup_k w_k = W$ and thus $\max_j |j/n - j^*/n| \to 0$. Hence,

$$\begin{aligned}\max_j n^{-1/2} |S^*_{\mathcal{T}_n}(j) - S_{\mathcal{T}_n}(j)| &= \max_j |\widetilde{S}_n(j^*/n) - \widetilde{S}_n(j/n)| \\ &\le 2 \sup |\widetilde{S}_n - \sigma B_{\mathrm{ex}}| + \sigma \max_j |B_{\mathrm{ex}}(j^*/n) - B_{\mathrm{ex}}(j/n)| \to 0. \end{aligned} \tag{7.14}$$

Consequently, by (7.9) again,

$$n^{-1/2} W(\mathcal{T}_n) - \max_t \widetilde{S}_n(t) = n^{-1/2} \max_j S^*_{\mathcal{T}_n}(j) - n^{-1/2} \max_j S_{\mathcal{T}_n}(j) \to 0,$$

and hence

$$n^{-1/2} W(\mathcal{T}_n) \to \sigma \max_t B_{\mathrm{ex}}(t). \tag{7.15}$$

For the height, we note first that similarly to (7.14), $\sup_t |\widehat{S}_n(t) - \widetilde{S}_n(t)| \to 0$ and thus $\sup_t |\widehat{S}_n(t) - \sigma B_{\mathrm{ex}}(t)| \to 0$. We truncate as in the proof of Lemma 4.7, using (4.15) with $\widehat{V}_n$ replaced by $\widehat{S}_n$ and F by σB_{ex}. We then have $Y^\varepsilon_n \to Y^\varepsilon$ as $n \to \infty$ for every $\varepsilon > 0$. Now, by (7.13), $Y_n = n^{-1/2} H(\mathcal{T}_n)$.

Flajolet and Odlyzko [16] proved, assuming that ξ has a finite exponential moment, that $\mathbb{E}\, Y_n = \mathbb{E}\, n^{-1/2} H(\mathcal{T}_n) \to \sigma^{-1}\sqrt{2\pi}$, which equals $\mathbb{E}\, Y = \sigma^{-1}\, \mathbb{E} \int_0^1 dt/B_{\mathrm{ex}}(t)$ by Remark 5.2. Hence,

$$\lim_{n\to\infty} \mathbb{E}\, |Y^\varepsilon_n - Y_n| = \lim_{n\to\infty} \bigl(\mathbb{E}\, Y_n - \mathbb{E}\, Y^\varepsilon_n\bigr) = \mathbb{E}\, Y - \mathbb{E}\, Y^\varepsilon,$$

which by monotone convergence tends to 0 as $\varepsilon \to 0$. As in Lemma 4.7, it then follows from [7, Theorem 4.2] that

$$n^{-1/2} H(\mathcal{T}_n) = Y_n \xrightarrow{\mathrm{d}} Y = \sigma^{-1} \int_0^1 \frac{dt}{B_{\mathrm{ex}}(t)}, \tag{7.16}$$

jointly with (7.11) and (7.15).

Unfortunately, as far as we know, the convergence of $n^{-1/2}\, \mathbb{E}\, H(\mathcal{T}_n)$ to $\sigma^{-1}\sqrt{2\pi}$ [16] is not yet proved assuming only a finite second moment. Hence we give another argument for the general case. (For the case of binary trees, [10] used a similar truncation with another argument based on explicit estimates.)

We know by (7.1) and (7.11) that both $Y_n = n^{-1/2} H(\mathcal{T}_n)$ and $\widetilde{S}_n$ converge in distribution. Hence the sequence of pairs $(Y_n, \widetilde{S}_n)$ is tight (in $\mathbb{R} \times C[0,1]$) and thus relatively compact by Prohorov's theorem [7, §6], so every subsequence has a subsequence that converges in distribution. Now consider only n in such a subsequence; we thus have $(Y_n, \widetilde{S}_n) \xrightarrow{\mathrm{d}} (Z, Z')$ for some random variable Z and random function Z'. By (7.1) and (7.11), $Z \overset{\mathrm{d}}{=} 2\sigma^{-1} \max B_{\mathrm{ex}}$

and $Z' \overset{\mathrm{d}}{=} \sigma B_{\mathrm{ex}}$. We may thus assume that $Z' = \sigma B_{\mathrm{ex}}$ [23, Theorem 6.10]. We can now redo the Skorohod coupling theorem argument above and assume that not only $\widetilde{S}_n \to \sigma B_{\mathrm{ex}}$ but also $Y_n \to Z$. We have shown that then $Y_n^\varepsilon \to Y^\varepsilon$ as $n \to \infty$ for every $\varepsilon > 0$. Since $Y_n \geq Y_n^\varepsilon$, it follows that $Z \geq Y^\varepsilon$ for every ε, and thus $Z \geq Y$. However, by (7.4), $Z \overset{\mathrm{d}}{=} Y$, and thus $Z = Y$ a.s.

We have considered a convergent subsequence of $(Y_n, \widetilde{S}_n)$, and shown that the limit in distribution is $(Y, \sigma B_{\mathrm{ex}})$ for any such subsequence. Since the sequence $(Y_n, \widetilde{S}_n)$ is tight, it follows that the full sequence converges to this limit, i.e. (7.16) and (7.11) hold jointly.

Redoing the Skorohod coupling theorem argument again, we may thus assume that a.s. both $\widetilde{S}_n \to \sigma B_{\mathrm{ex}}$ and $\int_0^1 dt/\widehat{S}_n(t) \to \int_0^1 dt/\sigma B_{\mathrm{ex}}(t)$. We fix again an $\omega \in \Omega$ such that these hold, and recall that then (7.15) holds too. By (4.11), $\{1/\widehat{S}_n\}$ is uniformly integrable on $[0,1]$, and it follows that

$$\int_0^t \frac{du}{\widehat{S}_n(u)} \to \int_0^t \frac{du}{\sigma B_{\mathrm{ex}}(u)}, \qquad 0 \leq t \leq 1. \tag{7.17}$$

We now truncate $m_1(\mathcal{T}_n)$ too and define $M_n^\varepsilon := n^{-1/2} \sum_{\varepsilon n}^n 1/d(v_i)$. Then, using (7.10), with $0 \leq \delta(x) \leq 1$ (with minor modifications if εn is an integer),

$$M_n^\varepsilon = n^{-1/2} \sum_{\lceil \varepsilon n \rceil}^{n} \frac{1}{d(v_i)} = n^{-1/2} \int_{\lfloor \varepsilon n \rfloor}^{n} \frac{dx}{\int_0^x du/S^*(\lceil u \rceil) + \delta(x)}.$$

By changes of variables, (7.17), (7.12) and bounded convergence, with $\delta_n(t) := n^{-1/2}\delta(nt)$,

$$M_n^\varepsilon = \int_{\lceil \varepsilon n \rceil/n}^{1} \frac{dt}{\int_0^t du/\widehat{S}_n(u) + \delta_n(t)} \to Z^\varepsilon := \int_\varepsilon^1 \frac{dt}{\sigma^{-1}\int_0^t du/B_{\mathrm{ex}}(u)}.$$

Consequently, $\bigl(n^{-1/2}H(\mathcal{T}_n),\, n^{-1/2}W(\mathcal{T}_n),\, M_n^\varepsilon\bigr) \to (Y,\, \sigma \max B_{\mathrm{ex}},\, Z^\varepsilon)$ for every $\varepsilon > 0$. We complete the proof by another application of [7, Theorem 4.2]; to verify the condition there it suffices to show that

$$\lim_{\varepsilon \to 0} \limsup_{n \to \infty} \mathbb{E}\, |n^{-1/2} m_1(\mathcal{T}_n) - M_n^\varepsilon| = 0. \tag{7.18}$$

To prove this, note that

$$\begin{aligned} 0 \leq n^{-1/2} m_1(\mathcal{T}_n) - M_n^{\varepsilon^2} &= n^{-1/2} \sum_{i=1}^{\varepsilon^2 n} \frac{1}{d(v_i)} \\ &\leq n^{-1/2} \sum_{i=1}^{n} \frac{\mathbf{1}[d(v_i) \leq \varepsilon n^{1/2}]}{d(v_i)} + n^{-1/2} \frac{\varepsilon^2 n}{\varepsilon n^{1/2}} = n^{-1/2} \sum_{k=1}^{\varepsilon n^{1/2}} \frac{w_k(\mathcal{T}_n)}{k} + \varepsilon. \end{aligned}$$

Hence Theorem 1.13 shows that (7.18) holds, cf. (4.17), and the proof is completed. ∎

8. Examples

We give two further examples with deterministic trees.

Example 8.1. FIND is an algorithm (due to Hoare) to find the element of a given rank k in a set of n elements, see [25, Exercises 5.2.2-31 and 32]. It chooses (in one version) an element (called *pivot*) at random, compares it to all others to determine its rank and to separate the other elements into two subsets: the ones larger and the ones smaller than the pivot. If the rank of the pivot is k, stop. Otherwise, decide on basis of the rank of the pivot which of the two subsets that contains the element with rank k, and its rank there, and continue recursively.

It is easy to see that the process of choosing pivots is the same as vertex cutting in $P_{n,k}$, the path with n vertices and the root at number k. This is the tree consisting of a root and two paths, of lengths $k-1$ and $n-k$, from it. If $k=1$ or n, we recover Example 1.4.

The number of passes used by FIND, i.e. the number of pivots, is thus $X_v(P_{n,k}) = X(\widetilde{P_{n,k}})$. (Note that the random variable usually studied in connection with FIND is different; it is the total number of comparisons, which is the sum for all cuts of the number of edges in the current subtree.) Equivalently, $X_v(P_{n,k})-1$ equals the depth of node k in a random binary search tree.

For asymptotics, we note that Lemma 6.1 implies that $\mathbb{E}\,|X_v(P_{n,k}) - X(P_{n,k})| \le 3$; hence we may instead consider $X(P_{n,k})$, which is a sum of two independent parts $X_1 \overset{\mathrm{d}}{=} X(P_{k-1})$ and $X_2 \overset{\mathrm{d}}{=} X(P_{n-k})$ for the two paths from the root. By Example 1.4, each of X_1 and X_2 may be written as the sum of independent indicator variables, and thus the same is true for $X(P_{n,k})$. As in Example 1.4, this leads to Poisson approximation, now by $\mathrm{Po}\big(\ln k + \ln(n-k+1)\big)$, and to asymptotic normality.

If we assume $1 \le k \le (n+1)/2$, as we may by symmetry, then $\ln(n-k+1) = \ln n + O(1)$, and we obtain, omitting the details,

$$(\ln n + \ln k)^{-1/2}\big(X_v(P_{n,k}) - \ln n - \ln k\big) \xrightarrow{\mathrm{d}} N(0,1).$$

In particular, if $n^{1-o(1)} < k < n - n^{1-o(1)}$, we have

$$(2\ln n)^{-1/2}\big(X_v(P_{n,k}) - 2\ln n\big) \xrightarrow{\mathrm{d}} N(0,1). \tag{8.1}$$

It follows that (8.1) holds also if k is chosen at random in $\{1,\dots,n\}$.

These results have previously been obtained by Devroye and Neininger [13] and Grübel and Stefanoski [18].

Example 8.2. Let T be a complete binary tree with $n = 2^m - 1$ vertices.

It is easy to see that most vertices have depth close to m, and that the path between most pairs of vertices goes down almost to the root. More formally, if $V_1,\dots,V_k$ are independent random vertices in T, then $d(V_1) = m - O_p(1)$ and $d(V_1,V_2) = 2m - O_p(1)$, where $O_p(1)$ means bounded in probability as $m \to \infty$, and thus $L_T(V_1, \dots, V_k) = km - O_p(1)$. It follows easily from Lemma 4.3 that $\mathbb{E}\, X(T)^k \sim n^k m^{-k} \sim (n/\log_2 n)^k$. Consequently,

$$\frac{X(T)}{n/\log_2 n} \xrightarrow{\mathrm{p}} 1 \qquad \text{as } n \to \infty.$$

This example is studied in detail in the companion paper [20], where we show that another normalization yields a non-degenerate limit.

9. More on ν_f

Although our results on random cuttings and records only involve ν_f for continuous f, we find it interesting for its own sake to extend Theorem 1.9 to more general f. We also consider arbitary (finite or infinite) intervals J as domains. We then define $m_k(f)$ by modifying (1.8), now integrating over J instead of $[0,1]$.

Theorem 9.1. *If f is a non-negative Lebesgue measurable function on an interval J such that $\int_J dt/f(t) < \infty$, then there exists a unique probability measure ν_f on $[0,\infty)$ with (finite) moments*

$$\int x^k \, d\nu_f(x) = m_k(f)$$

given by (1.8) (with $[0,1]$ replaced by J). ν_f has the following invariance properties:

(i) *(Homogeneity.) If $c > 0$, then $X \sim \nu_f \implies c^{-1}X \sim \nu_{cf}$.*
(ii) *(Translation and dilation.) If $\tilde{J} = aJ + b$, where $a > 0$ and $b \in \mathbb{R}$, and $\tilde{f}$ is defined on $\tilde{J}$ by $\tilde{f}(ax+b) = f(x)$, $x \in J$, then $\nu_{\tilde{f}} = \nu_f$.*
(iii) *(Endpoints.) The value of f at an endpoint of J does not affect ν_f. Indeed, we may remove an endpoint of J, or add one with an arbitrary value of f, without changing ν_f.*

Proof. First we observe that, with the same proof as in Lemma 4.2 ,

$$m_k(f) \le k! \Bigl(\int_J \frac{dt}{f(t)} \Bigr)^k, \qquad k \ge 1,$$

which implies both finiteness of $m_k(f)$ and uniqueness of ν_f, if it exists. The properties (i), (ii) and (iii) follow from the corresponding invariances of $m_k(f)$, which are obvious.

To show existence, we use the method of monents as before; we state the central argument as a lemma.

Lemma 9.2. *Let $f, f_1, f_2, \dots$ be non-negative measurable functions on an interval J such that, as $N \to \infty$, $f_N(t) \to f(t)$ for a.e. $t \in J$, $\int_J dt/f_N(t) \to \int_J dt/f(t) < \infty$, and $\inf_{[s,t]} f_N \to \inf_{[s,t]} f$ for a.e. $(s,t) \in J^2$ with $s < t$. If ν_{f_N} exists for each N, then ν_f exists and $\nu_{f_N} \to \nu_f$.*

Proof. The same argument as in the proof of Lemma 4.5 shows that the sequence $\{1/f_N\}_N$ is uniformly integrable on J, and more generally that the sequence $\{\bigl(L_{f_N}(t_1)\cdots L_{f_N}(t_1,\dots,t_k)\bigr)^{-1}\}_N$ is uniformly integrable on J^k for each $k\ge 1$. Moreover, the assumptions imply that $L_{f_N}(t_1,\dots,t_k)\to L_f(t_1,\dots,t_k)$ for a.e. $(t_1,\dots,t_k)$. Thus, by (1.8) and (4.11), $m_k(f_N)\to m_k(f)$ for each $k\ge 1$. The lemma follows by the method of moments, using the uniqueness already proved. ∎

To complete the proof of Theorem 9.1, it remains to show that we can approximate an arbitrary f as in Lemma 9.2 by f_N for which we know that ν_{f_N} exists. We do the approximation in several steps, cf. the proof of existence in Theorem 1.9.

Step 1: $J=[0,1]$, f is a simple strictly positive function. Thus f takes a finite number of values only, say $a_1,\dots,a_r$ with $0<a_1<\dots<a_r$.

For each $N\ge 1$ and $i=1,\dots,2^N$, let I_{Ni} be the dyadic interval $[(i-1)2^{-N},i2^{-N}]$, and let $m_{Ni}:=\inf_{I_{Ni}}f$. Let g_{Ni} be a continuous function on I_{Ni} such that $\int_{I_{Ni}}|f-g_{Ni}|<4^{-N}$; we may further assume that $m_{Ni}\le g_{Ni}\le a_r=\max_{[0,1]}f$ (otherwise we replace g_{Ni} by $\min(a_r,\max(m_{Ni},g_{Ni}))$. Let $\varepsilon=\varepsilon_N<1/4$ be a small positive number chosen below, and define f_{Ni} by $f_{Ni}=g_{Ni}$ on $[(i-1+\varepsilon)2^{-N},(i-1/2-\varepsilon)2^{-N}]\cup[(i-1/2+\varepsilon)2^{-N},(i-\varepsilon)2^{-N}]$, $f_{Ni}\bigl((i-1)2^{-N}\bigr)=f_{Ni}(i2^{-N})=a_r$ and $f_{Ni}\bigl((i-1/2)2^{-N}\bigr)=m_{Ni}$, with f_{Ni} linear in the four gaps in between. We choose ε so small that $\int_{I_{Ni}}|f_{Ni}-g_{Ni}|<4^{-N}$ ($(a_r-a_1)^{-1}2^{-N-2}$ will do).

We have thus for each i constructed a continuous f_{Ni} on I_{Ni} that is a_r at the endpoints, so together they define a continuous function f_N on $[0,1]$. By the construction follows

$$\inf_{I_{Ni}} f_N = m_{Ni} = \inf_{I_{Ni}} f, \qquad i=1,\dots,2^N, \tag{9.1}$$

and

$$\int_0^1 |f_N-f| \le \sum_i \int_{I_{Ni}} \bigl(|f_{Ni}-g_{Ni}|+|g_{Ni}-f|\bigr) \le 2^{1-N}. \tag{9.2}$$

By a standard argument, (9.2) implies

$$\int_0^1 \sum_{N=1}^{\infty} |f_N(t)-f(t)|\,dt < \sum_{N=1}^{\infty} 2^{1-N} < \infty;$$

hence, for a.e. t, $\sum_{N=1}^{\infty}|f_N(t)-f(t)|<\infty$ and thus $f_N(t)-f(t)\to 0$ as $N\to\infty$. Since $f_N\ge a_1>0$, this further implies $\int dt/f_N(t)\to\int dt/f(t)$ by dominated convergence.

Next, let $E_j:=\{t\in[0,1]:f(t)=a_j\}$ and let $F_j:=\overline{E_j}$. The open set $G_j:=\mathbb{R}\setminus F_j$ consists of a countable set of open intervals. Let H_j be the set of their endpoints, and $H:=\bigcup_1^r H_j$.

Let $0<s<t<1$ and assume that $s,t\notin H$. Since f takes the values $a_1,\dots,a_r$ only, $\inf_{(s,t)}f=a_l$ for some l, and there exists $y\in(s,t)$ such that $f(y)=a_l$.

Assume now $j < l$. Then $f(x) \geq a_l > a_j$ for all $x \in (s,t)$, and thus $(s,t) \cap E_j = \emptyset$. Hence $(s,t) \cap F_j = \emptyset$, so $(s,t) \subseteq G_j$. Since we assume $s,t \notin H_j$, $(s-\delta_j, t+\delta_j) \subseteq G_j$ for some $\delta_j > 0$. In other words, $f(x) \neq a_j$ on $(s-\delta_j, t+\delta_j)$.

Consequently, with $\delta := \min\bigl(s, 1-t, \min_{j<l} \delta_j\bigr)$, if $x \in (s-\delta, t+\delta)$, then $f(x) \neq a_j$ for $j < l$, and thus $f(x) \geq a_l$. Hence

$$\inf_{(s-\delta,t+\delta)} f = \inf_{[s,t]} f = \inf_{(s,t)} f = a_l. \tag{9.3}$$

Let N be so large that $2^{-N} < \min(\delta, y-s, t-y)$. We have $y \in I_{Ni}$ for some i, and then $I_{Ni} \subset (s,t)$. Consequently, by (9.1) and (9.3),

$$\inf_{[s,t]} f_N \leq \inf_{I_{Ni}} f_N = \inf_{I_{Ni}} f = f(y) = a_l = \inf_{[s,t]} f.$$

Conversely, if $\inf_{[s,t]} f_N < a_l$, then $f_N(z) < a_l$ for some $z \in [s,t]$. We have $z \in I_{Ni}$ for some i, and then $I_{Ni} \subseteq (s-\delta, t+\delta)$. Hence

$$a_l > f_N(z) \geq \inf_{I_{Ni}} f_N = \inf_{I_{Ni}} f \geq \inf_{(s-\delta,t+\delta)} f_N = a_l,$$

a contradiction. Thus, for large N, $\inf_{[s,t]} f_N = \inf_{[s,t]} f$.

We have shown that for $s,t \notin H$, $\inf_{[s,t]} f_N \to \inf_{[s,t]} f$ as $N \to \infty$, where H is countable and thus a null set. Since each ν_{f_N} exists by Theorem 1.9, Lemma 9.2 shows that ν_f exists.

Step 2: $J = [0,1]$, $a \leq f \leq b$ where $0 < a < b$. Let $f_N(t) := \lceil N f(t) \rceil / N$. f_N is simple and strictly positive, so ν_{f_N} exists by Step 1. $f_N(t) \to f(t)$ uniformly for $t \in J$; hence also $\inf_{[s,t]} f_N \to \inf_{[s,t]} f$ when $s < t$. Since $f_N(t) \geq a$, we further have $\int dt/f_N(t) \to \int dt/f(t)$ by dominated convergence. Consequently, Lemma 9.2 shows that ν_f exists.

Step 3: $J = [0,1]$, $f \geq a > 0$. Let $f_N(t) := \min\bigl(f(t), N\bigr)$. Then ν_{f_N} exists by Step 2, $f_N(t) \to f(t)$ for every t and $\inf_{[s,t]} f_N \to \inf_{[s,t]} f$ when $s < t$. Further, $\int dt/f_N(t) \to \int dt/f(t)$ by dominated convergence. Consequently, Lemma 9.2 shows that ν_f exists.

Step 4: $J = [0,1]$, $f \geq 0$ and $\int dt/f(t) < \infty$. Let $f_N(t) := f(t) + 1/N$. Then ν_{f_N} exists by Step 3, and Lemma 9.2 applies again.

Step 5: J arbitrary finite, $\int_J dt/f(t) < \infty$. Follows by Step 4, (ii) and (iii).

Step 6: J infinite, $\int_J dt/f(t) < \infty$. Let $J_N := J \cap [-N,N]$ and let f_N be the restriction of f to J_N. Then $m_k(f_N) \to m_k(f)$ by monotone convergence, and thus $\nu_f = \lim_N \nu_{f_N}$ exists by the method of moments. $\square$

Distributions of the type ν_f arise as limit distributions of $X(T_n)$ for suitable sequences T_n of trees (with appropriate normalization factors, not necessarily $n^{-1/2}$). This is implicit in the proof above where ν_f is approximated in several steps, going back through Theorem 1.9 to Lemma 4.5 and Remark 4.6. We can also relax the assumptions of Lemma 4.5 further to show such limit result for suitable T_n. We leave the details to the reader. In the examples below, we will indicate such trees T_n, but we will usually neither give precise limit statements nor formal proofs, again leaving the details to the interested reader.

Example 9.3. If $f = c > 0$ is constant and J is finite, then $m_k(f) = k!\,|J|^k c^{-k}$, and thus $\nu_f = \operatorname{Exp}(c^{-1}|J|)$. This corresponds to trees with a long trunk with many leaves or short branches attached at the top. (The time to cut down the tree is essentially determined by the time until the trunk first is hit, since the remaining time is at most about $\ln n$ by Example 1.4.) For a precise example, let T_n consist of a trunk of height a_n with $n - a_n$ leaves added to the top of the trunk, where $a_n \to \infty$ with $a_n = o(n/\ln n)$. It is the easy to see that $(a_n/n)X(T_n) \xrightarrow{\mathrm{d}} \operatorname{Exp}(1)$.

In contrast to Theorem 9.1(iii), changing f at a single interior point may change ν_f; see the following theorem and Example 9.5 below.

Theorem 9.4. *Suppose that $f \geq 0$ on (a, b) and that $f(c) = 0$, where $-\infty \leq a < c < b \leq \infty$. Let $J_1 = (a, c)$, $J_2 = (c, b)$, and let f_1 and f_2 be the restrictions of f to J_1 and J_2. If $X_1 \sim \nu_{f_1}$ and $X_2 \sim \nu_{f_2}$ are independent, then $X_1 + X_2 \sim \nu_f$.*

This corresponds to the obvious fact that if a tree consists of two subtrees T_1 and T_2 with only the root in common, then $X(T) = X(T_1) + X(T_2)$, with $X(T_1)$ and $X(T_2)$ independent.

Proof. Suppose that $t_1, \dots, t_k \in J$. Let $t'_1, \dots, t'_j$ be the subsequence of all t_i that belong to J_1 (in order), and let $t''_1, \dots, t''_l$ be the subsequence of all t_i that belong to J_2. Suppose that every $t_i \neq c$; thus $j + l = k$.

Now keep $t'_1, \dots, t'_j$ and $t''_1, \dots, t''_l$ fixed, and let $\sum^*_{t_1,\dots,t_k}$ denote the sum over the $\binom{k}{j}$ sequences $t_1, \dots, t_k$ that consist of $t'_1, \dots, t'_j$ and $t''_1, \dots, t''_l$ merged in some order (keeping the order inside $t'_1, \dots, t'_j$ and $t''_1, \dots, t''_l$). We use the notation $M(t_1, \dots, t_k) = 1/\bigl(L_f(t_1) \cdots L_f(t_1, \dots, t_k)\bigr)$, interpreted as 1 if $k = 0$, and claim that

$$\sum_{t_1,\dots,t_k}^{*} M(t_1, \dots, t_k) = M(t'_1, \dots, t'_j) M(t''_1, \dots, t''_l). \tag{9.4}$$

This holds trivially if $j = 0$ or $l = 0$. Thus assume $0 < j < k$, and use induction on k. Note that t_k is either t'_j or t''_l. The sequences $t_1, \dots, t_k$ with $t_k = t'_j$ are obtained by merging $t'_1, \dots, t'_{j-1}$ and $t''_1, \dots, t''_l$ as above and adding $t_k = t'_j$. Summing $M(t_1, \dots, t_k)$ over these sequences we obtain by

(9.4) and induction

$$\begin{aligned}\frac{1}{L_f(t_1,\dots,t_k)}&\sum_{t_1,\dots,t_{k-1}}^{*} M(t_1,\dots,t_{k-1})\\&=\frac{1}{L_f(t_1,\dots,t_k)}M(t'_1,\dots,t'_{j-1})M(t''_1,\dots,t''_l)\\&=\frac{L_f(t'_1,\dots,t'_j)}{L_f(t_1,\dots,t_k)}M(t'_1,\dots,t'_j)M(t''_1,\dots,t''_l).\end{aligned}\tag{9.5}$$

The sum with $t_k=t''_l$ is similarly

$$\frac{L_f(t''_1,\dots,t''_l)}{L_f(t_1,\dots,t_k)}M(t'_1,\dots,t'_j)M(t''_1,\dots,t''_l).\tag{9.6}$$

Since $f(c)=0$, (1.7) implies $L_f(t_1,\dots,t_k)=L_f(t'_1,\dots,t'_j)+L_f(t''_1,\dots,t''_l)$, so summing (9.5) and (9.6) we obtain (9.4).

Integrating (9.4) over all $t'_1,\dots,t'_j\in J_1$ and $t''_1,\dots,t''_l\in J_2$, we obtain, if $\mathcal{E}_j=\{(t_1,\dots,t_k)\in J^k:|\{t_1,\dots,t_k\}\cap J_1|=j\}$,

$$\begin{aligned}\int_{\mathcal{E}_j}&M(t_1,\dots,t_k)\,dt_1\cdots dt_k\\&=\int_{J_1^j}M(t'_1,\dots,t'_j)\,dt'_1\cdots dt'_j\int_{J_2^l}M(t''_1,\dots,t''_l)\,dt''_1\cdots dt''_l\\&=\tfrac{1}{j!}m_j(f_1)\tfrac{1}{l!}m_l(f_2).\end{aligned}$$

Summing over j, we finally obtain

$$\begin{aligned}m_k(f)&=\sum_{j=0}^{k}\binom{k}{j}m_j(f_1)m_{k-j}(f_2)=\sum_{j=0}^{k}\binom{k}{j}\mathbb{E}(X_1)^j\,\mathbb{E}(X_2)^{k-j}\\&=\mathbb{E}(X_1+X_2)^k.\end{aligned}$$

□

Example 9.5. If $f=1$ on $J=(a,b)$, then ν_f is an exponential distribution by Example 9.3. If $c\in(a,b)$ and $g(x)=f(x)$ for $x\neq c$ but $g(c)=0$, then ν_g is by Theorem 9.4 and Example 9.3 the distribution of a sum of two independent exponential variables. Hence $\nu_g\neq\nu_f$.

If f has several zeroes, we may use Theorem 9.4 repeatedly. Going to extremes, we have the following result. (This vaguely corresponds to trees consisting of many branches joined at the root, or close to the root as in Example 8.2, so that the law of large numbers applies.)

Theorem 9.6. *If $f(x)=0$ for $x\in D$, where D is dense in J, then ν_f is a point mass at $\int_J dt/f(t)$.*

Proof. By (1.7) and our assumption, $L_f(t_1,t_2) = f(t_1) + f(t_2)$ whenever $t_1 < t_2$. Hence,

$$m_2(f) = 2\iint_{t_1<t_2}\left(\frac{1}{L_f(t_1)L_f(t_1,t_2)} + \frac{1}{L_f(t_2)L_f(t_1,t_2)}\right)dt_1\,dt_2$$
$$= 2\iint_{t_1<t_2}\left(\frac{f(t_1)+f(t_2)}{f(t_1)f(t_2)L_f(t_1,t_2)}\right)dt_1\,dt_2 = m_1(f)^2.$$

Hence, if $X \sim \nu_f$, then $\operatorname{Var} X = m_2(f) - m_1(f)^2 = 0$, so $X = \mathbb{E}\,X = m_1(f)$ a.s. $\square$

Example 9.7. If $J = [0,1]$ and $f(x) = 1$ for irrational x and $f(x) = 0$ for rational x, then ν_f is a point mass at 1.

Finally we study another class of examples. These arise for trees consisting of a long trunk with leaves or short branches attached along the trunk so that vertex i has relative height $f(i/n)$.

Theorem 9.8. *Suppose that f is non-decreasing on $J = (0,b)$ with $0 < b \le \infty$, and that $\int_J dx/f(x) < \infty$. Let $\rho := \inf_J f(x)/x$ and suppose $X \sim \nu_f$. Then $\rho > 0$ and the moment generating function of ν_f is*

$$\mathbb{E}\,e^{tX} = \exp\left(\int_J \frac{t\,dx}{f(x)-tx}\right), \qquad \operatorname{Re} t < \rho \tag{9.7}$$

(where the integral converges absolutely), while $\mathbb{E}\,e^{tX} = \infty$ for real $t > \rho$; in particular, the characteristic function is $\exp\bigl(it\int_J dx/(f(x)-itx)\bigr)$. The cumulants are given by

$$\varkappa_k(\nu_f) = k!\int_J \frac{x^{k-1}}{f(x)^k}\,dx, \qquad k \ge 1. \tag{9.8}$$

Further, ν_f is an infinitely divisible distribution with Lévy measure μ given by the density

$$\frac{d\mu}{dx} = \int_J \frac{f(y)}{y^2}e^{-xf(y)/y}dy, \qquad x > 0, \tag{9.9}$$

Proof. Let $A = \int_J dx/f(x) < \infty$, and note that, for every $x \in J$, $A \ge \int_0^x dy/f(y) \ge x/f(x)$, so $f(x)/x \ge A^{-1}$ and thus $\rho \ge A^{-1} > 0$. Note further that if $0 < t < \rho$, then $f(x) - tx \ge (1 - t\rho^{-1})f(x)$ and thus

$$\int_J \frac{dx}{f(x)-tx} \le (1-t\rho^{-1})^{-1}A < \infty. \tag{9.10}$$

It follows that the integral in (9.7) converges for all complex t with $\operatorname{Re} t < \rho$, and that the right hand side of (9.7) is analytic in this half-plane.

Since f is non-decreasing, (1.7) implies that

$$L_f(x_1,\dots,x_k) = f(x_{(k)}) = f(x_1 \vee \dots \vee x_k),$$

where $x \vee y$ denotes the maximum of x and y. Hence, (1.8) gives

$$m_k(f) = k!\int_{J^k} f^{-1}(x_1)f^{-1}(x_1\vee x_2)\cdots f^{-1}(x_1\vee\dots\vee x_k)\,dx_1\cdots dx_k. \tag{9.11}$$

Let B be the Banach space of all bounded measurable functions on $[0,b)$, with $\|g\|_B := \sup_{[0,b)} |g|$, and define, for non-increasing $\phi \geq 0$ with $\int_J \phi < \infty$,

$$T_\phi g(x) := \int_J \phi(x \vee y) g(x \vee y)\, dy.$$

Clearly, T_ϕ is a bounded linear operator in B, with norm $\|T_\phi\| = \int_J \phi$.

By induction, for $g \in B$,

$$T_\phi^k g(x_0) = \int_{J^k} \phi(x_0 \vee x_1) \cdots \phi(x_0 \vee \cdots \vee x_k) g(x_0 \vee \cdots \vee x_k)\, dx_k \cdots dx_1$$

and thus (9.11) yields

$$\mathbb{E}\, X^k = m_k(f) = k!\, T_{1/f}^k \mathbf{1}(0), \tag{9.12}$$

where $\mathbf{1}$ is the function identically 1. Now suppose that $0 \leq t < A^{-1}$. Then $\|tT_{1/f}\| = tA < 1$ and thus the operator $I - tT_{1/f}$ is invertible and, with convergent sums,

$$\mathbb{E}\, e^{tX} = \sum_{k=0}^{\infty} t^k \frac{\mathbb{E}\, X^k}{k!} = \sum_{k=0}^{\infty} t^k T_{1/f}^k \mathbf{1}(0) = (I - tT_{1/f})^{-1} \mathbf{1}(0).$$

In other words, $\mathbb{E}\, e^{tX} = g(0)$, where $g \in B$ satisfies $(I - tT_{1/f}) g = \mathbf{1}$, i.e.

$$g(x) - \int_0^b t f^{-1}(x \vee y) g(x \vee y)\, dy = 1, \qquad 0 \leq x < b$$

or

$$1 = g(x) - txf^{-1}(x) g(x) - \int_x^b t f^{-1}(y) g(y)\, dy, \qquad 0 \leq x < b. \tag{9.13}$$

To solve this, define $H(x) := t \int_x^b f^{-1}(y) g(y) dy$. Then (at least if f and g are continuous) (9.13) yields

$$1 = -\big(t^{-1} f(x) - x\big) H'(x) - H(x)$$

with the solution, since $H(b) = 0$,

$$H(x) = \exp\biggl(\int_x^b \frac{dy}{t^{-1} f(y) - y}\biggr) - 1$$

and thus, by (9.13) again,

$$g(x) = \frac{H(x) + 1}{1 - txf^{-1}(x)} = \big(1 - txf^{-1}(x)\big)^{-1} \exp\left(\int_x^b \frac{t\, dy}{f(y) - ty}\right). \tag{9.14}$$

Indeed, knowing the answer (9.14), it is easy to verify it by substitution back in (9.13), also if f (and thus g) is discontinuous. (Recall that $txf^{-1}(x) \leq tA < 1$ and that (9.10) holds, whichs shows $g \in B$.) For $x = 0$, $f(x)$ is not defined, but (9.13) and thus (9.14) holds with $xf^{-1}(x) = 0$, i.e.

$$g(0) = 1 + H(0) = \exp\left(\int_0^b \frac{t\, dy}{f(y) - ty}\right).$$

This verifies (9.7) for $0 \le t < A^{-1}$. Both sides of (9.7) are analytic functions in the halfplane $\operatorname{Re} t < A^{-1}$, and thus (9.7) holds there by analytic continuation.

For $|t| < A^{-1}$ we have $|tx/f(x)| < 1$, $x \in J$, and

$$\int_J \frac{t\,dx}{f(x) - tx} = t \int_J \frac{dx/f(x)}{1 - tx/f(x)} = \sum_{k=0}^{\infty} t^{k+1} \int_J \frac{x^k\,dx}{f(x)^{k+1}},$$

which by (9.7) yields the formula (9.8) for the cumulants. It follows that

$$\left|\frac{\varkappa_k(\nu_f)}{k!}\right| = \int_J \frac{x^{k-1}}{f(x)^k}\,dx \le \rho^{1-k} \int_J \frac{dx}{f(x)} = A\rho^{1-k}.$$

Hence the Taylor series of $\ln \mathbb{E}\, e^{tX} = \sum_k \varkappa_k(\nu_f) t^k/k!$ has radius of convergence at least ρ; hence $\mathbb{E}\, e^{tx}$ is finite for all real t with $0 < t < \rho$, and thus for all complex t with $\operatorname{Re} t < \rho$. By analytic continuation, (9.7) holds in this halfplane too.

Conversely, if $t > \rho$, then $f(x_0)/x_0 < t$ for some $x_0 \in J$. Hence, if $\delta > 0$ is small enough, $f(x) \le f(x_0) \le tx$ for $x \in (x_0 - \delta, x_0)$, and thus

$$\int_J \frac{x^{k-1}}{f(x)^k}\,dx \ge \frac{\delta}{x_0} t^{-k}.$$

Hence, $\sum_k \varkappa_k(\nu_f) t^k/k!$ diverges for $t > \rho$, so $\ln \mathbb{E}\, e^{tX}$ has radius of convergence exactly ρ. The Taylor coefficients are positive, and thus there is a singularity at ρ. Hence $\mathbb{E}\, e^{tX}$ too has a singularity there, which is impossible if $\mathbb{E}\, e^{tX} < \infty$ for some $t > \rho$.

To see that ν_f is infinitely divisible and to verify the formula (9.9) for the Lévy measure, note first that, by the Fubini–Tonelli theorem, $\int_0^\infty x\,d\mu(x) = \int_J dy/f(y) < \infty$. Similarly, for $t > 0$, say,

$$\begin{aligned}\int_0^\infty (e^{-tx} - 1)\,d\mu(x) &= \int_J \frac{f(y)}{y^2} \int_0^\infty \Bigl(e^{-x(f(y)/y+t)} - e^{-xf(y)/y}\Bigr)\,dx\,dy \\ &= \int_J \frac{f(y)}{y^2} \Bigl(\frac{1}{f(y)/y + t} - \frac{1}{f(y)/y}\Bigr)\,dy = \int_J \frac{-t}{f(y) + ty}\,dy.\end{aligned}$$

Thus, see e.g. [23, Theorem 15.12 and (15.8)], ν_f and the infinitely divisible distribution with Lévy measure μ have the same Laplace transform, so they are equal. □

Example 9.9. Let $J = [0, 1]$ and $f(x) = x^\alpha$, with $0 \le \alpha < 1$. By (9.8), the cumulants are

$$\varkappa_k(\nu_f) = k! \int_0^1 x^{k-1-k\alpha}\,dx = \frac{k!}{k(1-\alpha)},$$

and thus the moment generating function is

$$\exp\Bigl(\sum_{k=1}^{\infty} \frac{t^k}{k(1-\alpha)}\Bigr) = \exp\Bigl(\frac{-\ln(1-t)}{1-\alpha}\Bigr) = (1-t)^{-1/(1-\alpha)}.$$

Hence ν_f is the Gamma distribution $\Gamma(1/(1-\alpha))$.

10. On Problems 1.14 and 1.15

As said in Section 2, the conditioned Galton–Watson tree $\mathcal{T}_n$ converges in distribution to the infinite random tree $\mathcal{T}_\infty$ as $n \to \infty$. Hence, if it possible to define a growing process $\mathcal{T}_n$ as in Problem 1.15, then $\mathcal{T}_n$ has to grow to $\mathcal{T}_\infty$, so $\mathcal{T}_n$ could be constructed by a certain pruning of $\mathcal{T}_\infty$.

Further, by the proof of Lemma 2.3 and (2.8), $\mathbb{E}\, w_k(\mathcal{T}_\infty) = \mathbb{E}\, w_k(\mathcal{T})^2 = 1 + k\sigma^2$. Moreover, it can be shown from (2.7) and (2.12) that for fixed k, the sequence $w_k(\mathcal{T}_n)$ is uniformly integrable. Hence,

$$\mathbb{E}\, w_k(\mathcal{T}_n) \to \mathbb{E}\, w_k(\mathcal{T}_\infty) = 1 + k\sigma^2.$$

(This was shown, at least assuming an exponential moment, by Meir and Moon [32] using different arguments.) We now see that Problem 1.14 would imply the estimate $\mathbb{E}\, w_k(\mathcal{T}_n) \le 1 + k\sigma^2$, sharpening the estimate in Theorem 1.13.

Appendix A

The purpose of the appendix is to give a proof of the following result, stated in Remark 1.8. (A special case is given by Meir and Moon [31].) This proof was originally intended to be included in the paper, but was replaced by the more general Theorem 1.6 with only slightly stronger assumptions, proved by somewhat different methods. Nevertheless, since the result and the methods still may have some interest, it is added as an appendix.

Theorem A.1. *With assumptions and notations as above, as $n \to \infty$,*

$$\mathbb{E}\, X(\mathcal{T}_n) \sim \sigma\sqrt{\frac{\pi n}{2}}.$$

Proof. We use generating functions. Let $\mathcal{T}$ be the (unconditioned) Galton–Watson tree with offspring distribution ξ, and define

$$\begin{aligned}
\Phi(z) &:= \mathbb{E}\, z^{\xi}, \\
F(z) &:= \mathbb{E}\, z^{|\mathcal{T}|}, \\
G(z,w) &:= \mathbb{E}\Bigl(z^{|\mathcal{T}|}\sum_{v\in\mathcal{T}} w^{d(v)}\Bigr) = \mathbb{E}\Bigl(z^{|\mathcal{T}|}\sum_{k\ge0} w_k(\mathcal{T})w^k\Bigr), \\
H(z) &:= \mathbb{E}\Bigl(z^{|\mathcal{T}|}m_1(\mathcal{T})\Bigr) = \mathbb{E}\Bigl(z^{|\mathcal{T}|}\sum_{k\ge1}\frac{w_k(\mathcal{T})}{k}\Bigr).
\end{aligned}$$

These are defined and analytic at least for $|z| < 1$ and $|w| < 1$. In studies of conditioned Galton–Watson trees (or, equivalently, simply generated trees), it is, as said in the introduction, often assumed that ξ has an exponential moment, $\mathbb{E}\, e^{\alpha\xi} < \infty$ for some $\alpha > 0$; this says that Φ is analytic in a disc with radius greater than 1, which is a technical advantage. We will see, however, that with some care, we can use the standard methods also in our case where we assume only a second moment of ξ.

Note that we allow $k = 0$ in the formula for G; if we sum over $k \geq 1$ only, we get $G(z, w) - F(z)$. It follows by termwise integration that

$$H(z) = \int_0^1 \bigl(G(z, w) - F(z)\bigr) \frac{dw}{w}, \qquad |z| < 1. \tag{A.1}$$

Conditioning on the degree of the root of $\mathcal{T}$, we find the standard formula (for $|z| \leq 1$, at least)

$$F(z) = \sum_{j=0}^{\infty} \mathbb{P}(\xi = j) z F(z)^j = z\Phi(F(z)) \tag{A.2}$$

and, similarly, (for $|z|, |w| < 1$, at least)

$$G(z, w) - F(z) = \sum_{j=0}^{\infty} \mathbb{P}(\xi = j) j z w G(z, w) F(z)^{j-1} = zw\Phi'(F(z))G(z, w),$$

which gives

$$G(z, w) = \frac{F(z)}{1 - zw\Phi'(F(z))}. \tag{A.3}$$

By (A.1), we then find

$$\begin{aligned} H(z) &= F(z) \int_0^1 \frac{zw\Phi'(F(z))}{1 - zw\Phi'(F(z))} \frac{dw}{w} = F(z)\bigl[-\ln\bigl(1 - zw\Phi'(F(z))\bigr)\bigr]_{w=0}^1 \\ &= -F(z) \ln\bigl(1 - z\Phi'(F(z))\bigr). \end{aligned} \tag{A.4}$$

We will for simplicity assume that $\operatorname{span}(\xi) = 1$ and leave the usual modifications when $\operatorname{span}(\xi) = d > 1$ to the reader. (There will be two factors d that cancel in (A.10), so the result remains the same.)

Let $D := \{z : |z| < 1\}$ be the open unit disc; thus $\overline{D}$ is the closed unit disc. Further, for $0 < \beta < \pi$, let $S_\beta := \{z : |\arg(z - 1)| > \beta\}$ (a sector with tip at 1) and let $\Omega_\eta := \{z : |z| < 1 + \eta\} \cap S_{\pi/4}$, $\eta > 0$. Thus, Ω_η is a disc larger than D with an indentation at $z = 1$. (The angle $\pi/4$ could be replaced by any value in $(0, \pi/2)$.)

Lemma A.2. *If* $\operatorname{span}(\xi) = 1$*, then, for some* $\eta > 0$*,* F *extends to an analytic function in* Ω_η *with* $|F(z)| < 1$ *for* $z \in \Omega_\eta$ *and*

$$F(z) = 1 - \sqrt{2}\sigma^{-1}\sqrt{1 - z} + o\bigl(|z - 1|^{1/2}\bigr), \qquad \textit{as } z \to 1 \textit{ with } z \in \Omega_\eta. \tag{A.5}$$

Proof. This is well-known if ξ has an exponential moment. In general, we may argue as follows. Since $\mathbb{E}\,\xi^2 < \infty$, Φ and its first two derivatives have continuous extensions to $\overline{D}$, with $\Phi(1) = 1$, $\Phi'(1) = \mathbb{E}\,\xi = 1$, and $\Phi''(1) = \mathbb{E}\,\xi(\xi - 1) = \sigma^2$. A Taylor expansion shows that, as $z \to 1$ in $\overline{D}$,

$$\Phi(z) = 1 + (z - 1) + \frac{\sigma^2}{2}(z - 1)^2 + o(|z - 1|^2). \tag{A.6}$$

Let $\Psi(z) := z/\Phi(z)$, and note that (A.2) may be written $\Psi(F(z)) = z$. Let $B_\varepsilon := \{z : |z-1| < \varepsilon\}$ be a small disc around 1. If $\varepsilon > 0$ is small, then Ψ is analytic in $D \cap B_\varepsilon$, and Ψ, Ψ' and Ψ'' have continuous extensions to $\overline{D \cap B_\varepsilon}$. It follows from (A.6) that, for $z \to 1$ in $\overline{D}$,

$$\Psi(z) = \frac{z}{\Phi(z)} = 1 - \frac{\sigma^2}{2}(z-1)^2 + o(|z-1|^2). \tag{A.7}$$

Hence (or by differentiation), $\Psi(1) = 1$, $\Psi'(1) = 0$ and $\Psi''(1) = -\sigma^2$. Another Taylor expansion yields, for $z \to 1$ in $\overline{D}$,

$$\Psi'(z) = -\sigma^2(z-1) + o(|z-1|). \tag{A.8}$$

Let $U_\varepsilon := B_\varepsilon \cap S_{0.6\,\pi}$. If ε is small enough, then $U_\varepsilon \subset D$ and, by (A.8), $\operatorname{Re}\Psi'(z) > 0$ in U_ε. Hence $\Psi(z_2) - \Psi(z_1) = \int_{z_1}^{z_2} \Psi'(w)\,dw \neq 0$ if $z_1, z_2 \in U_\varepsilon$ with $z_1 \neq z_2$; thus Ψ is a bijection of U_ε onto $V_\varepsilon := \Psi(U_\varepsilon)$, with an analytic inverse $\Psi^{-1} : V_\varepsilon \to U_\varepsilon$. Close to 1, the boundary of U_ε consists of the two rays $\arg(z-1) = \pm 0.6\,\pi$; by (A.7), Ψ maps them to two curves emanating from 1 in the directions $\pm 0.2\,\pi$, and it follows that if $\delta > 0$ is small enough, then $B_\delta \cap S_{\pi/4} \subset V_\varepsilon$. We can thus define $F = \Psi^{-1}$ in $B_\delta \cap S_{\pi/4}$, with $|F| < 1$. This extends F close to 1, and (A.5) follows from (A.7).

Now let $z_0 \in \partial D$ with $z_0 \neq 1$, and let $w_0 := F(z_0)$. Since $\operatorname{span}(\xi) = 1$, $|w_0| = |F(z_0)| < F(1) = 1$ and thus $|\Phi'(w_0)| < 1$. Hence, $(\partial/\partial w)\bigl(w - z\Phi(w)\bigr) = 1 - z\Phi'(w) \neq 0$ at (z_0, w_0), and the implicit function theorem shows that F can be extended to a small disc about z_0 by a function satisfying (A.2) and with $|F| < 1$. A simple compactness argument shows that a finite number of these extensions together with the extension to $B_\delta \cap S_{\pi/4}$ combine to the desired extension in Ω_η, if η is small enough. ☐

Since $\Phi''(1) = \sigma^2$, (A.5) implies that, as $z \to 1$ in Ω_η and thus $F(z) \to 1$ in D,

$$\begin{aligned} z\Phi'(F(z)) &= z\bigl(1 + \sigma^2(F(z)-1) + o(|F(z)-1|)\bigr) \\ &= 1 - \gamma\sqrt{1-z} + o\bigl(|z-1|^{1/2}\bigr), \end{aligned} \tag{A.9}$$

with $\gamma := \sqrt{2}\sigma$. It follows that $|z\Phi'(F(z))| < 1$ for $z \in \Omega_\eta$ close to 1; since $|F(z)| < 1$ for $z \in \overline{D} \setminus \{1\}$, a simple compactness argument shows that $|z\Phi'(F(z))| < 1$ for all $z \in \Omega_\eta$ if η is small enough. We then can use (A.4) to define H in Ω_η. Moreover, when $z \in \Omega_\eta$ with $|z-1|$ small, by (A.9),

$$H(z) = -\ln\bigl(1 - z\Phi'(F(z))\bigr) = -\tfrac{1}{2}\ln(1-z) - \ln\gamma + o(1).$$

Consequently, see Flajolet and Odlyzko [17, Corollary 1],

$$[z^n]H(z) = -\tfrac{1}{2}[z^n]\ln(1-z) + o(1/n) = \frac{1+o(1)}{2n}.$$

As is well known [26, Theorem 2.1.2], see also [2, Proposition 24] or [32, Theorem 3.1], or use Lemma A.2 and [17],

$$[z^n]F(z) = \mathbb{P}(|\mathcal{T}| = n) \sim (2\pi)^{-1/2}\sigma^{-1}n^{-3/2}.$$

Consequently (recalling our convention that $\mathcal{T}_n$ has $n+1$ vertices),

$$\mathbb{E}\, X(\mathcal{T}_{n-1}) = \mathbb{E}\, m_1(\mathcal{T}_{n-1}) = \frac{[z^n]H(z)}{[z^n]F(z)} \sim (\pi/2)^{1/2}\sigma n^{1/2}, \tag{A.10}$$

which completes the proof. □

References

[1] D. Aldous, The continuum random tree II: an overview. *Stochastic Analysis (Proc., Durham, 1990)*, 23–70, London Math. Soc. Lecture Note Ser. 167, Cambridge Univ. Press, Cambridge, 1991.

[2] D. Aldous, The continuum random tree III. *Ann. Probab.* **21**, no. 1, 248–289.

[3] D. Aldous, Brownian excursion conditioned on its local time. *Electron. Comm. Probab.* **3** (1998), 79–90.

[4] A.D. Barbour, L. Holst & S. Janson, *Poisson Approximation.* Oxford University Press, Oxford, 1992.

[5] P. Biane, J. Pitman & M. Yor, Probability laws related to the Jacobi theta and Riemann zeta functions, and Brownian excursions. *Bull. Amer. Math. Soc. (N.S.)* **38** (2001), no. 4, 435–465

[6] P. Biane & M. Yor, Valeurs principales associées aux temps locaux browniens. *Bull. Sci. Math.* (2) **111** (1987), no. 1, 23–101.

[7] P. Billingsley, *Convergence of Probability Measures.* Wiley, New York, 1968.

[8] R.M. Blumenthal, *Excursions of Markov processes.* Birkhäuser, Boston, 1992.

[9] P. Chassaing & R. Marchand. In preparation.

[10] P. Chassaing, J.F. Marckert & M. Yor, The height and width of simple trees. *Mathematics and Computer Science (Versailles, 2000)*, 17–30, Trends Math., Birkhäuser, Basel, 2000.

[11] K.L. Chung, Excursions in Brownian motion. *Ark. Mat.* **14** (1976), no. 2, 155–177.

[12] L. Devroye, Branching processes and their applications in the analysis of tree structures and tree algorithms. *Probabilistic methods for algorithmic discrete mathematics*, 249–314, eds. M. Habib et al., Algorithms Combin. 16, Springer-Verlag, Berlin, 1998.

[13] L. Devroye & R. Neininger, Distances and finger search in random binary search trees. *SIAM J. Comput.* **33** (2004), no. 3, 647–658.

[14] C. Donati-Martin, Some remarks about the identity in law for the Bessel bridge $\int_0^1 \frac{ds}{r(s)} \overset{\text{(law)}}{=} 2\sup_{s\le 1} r(s)$. *Studia Sci. Math. Hungar.* **37** (2001), no. 1–2, 131–144.

[15] M. Drmota & B. Gittenberger, On the profile of random trees. *Random Struct. Alg.* **10** (1997), no. 4, 421–451.

[16] P. Flajolet & A. Odlyzko, The average height of binary trees and other simple trees. *J. Comp. Sys. Sci.* **25** (1982), 171–213.

[17] P. Flajolet & A. Odlyzko, Singularity analysis of generating functions. *SIAM J. Discrete Math.* **3** (1990), no. 2, 216–240.

[18] R. Grübel & N. Stefanoski, Mixed Poisson approximation of node depth distributions in random binary search trees. *Ann. Appl. Probab.* **15** (2005), no. 1A, 279–297.

[19] A. Gut, *Probability: A Graduate Course.* Springer, New York, 2005.

[20] S. Janson, Random records and cuttings in complete binary trees. *Mathematics and Computer Science III, Algorithms, Trees, Combinatorics and Probabilities (Vienna 2004)*, Eds. M. Drmota, P. Flajolet, D. Gardy, B. Gittenberger, Birkhäuser, Basel, 2004, pp. 241–253.

[21] T. Jeulin, *Semi-martingales et grossissement d'une filtration.* Lecture Notes in Mathematics 833, Springer-Verlag, Berlin, 1980.

[22] W.D. Kaigh, An invariance principle for random walk conditioned by a late return to zero. *Ann. Probab.* **4** (1976), no. 1, 115–121.

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

[24] D.P. Kennedy, The distribution of the maximum Brownian excursion. *J. Appl. Probab.* **13** (1976), no. 2, 371–376.

[25] D.E. Knuth, *The Art of Computer Programming. Vol. 3: Sorting and Searching.* 2nd ed., Addison-Wesley, Reading, Mass., 1998.

[26] V.F. Kolchin, *Random Mappings.* Optimization Software, New York 1986.

[27] R. Lyons, R. Pemantle & Y. Peres, A conceptual proof of the Kesten–Stigum theorem. *Ann. Probab.* **23** (1995), 1125–1138.

[28] M. Luczak & P. Winkler, Building uniform subtrees of a Cayley tree. Preprint, 2003.

[29] J.-F. Marckert & A. Mokkadem, The depth first processes of Galton–Watson trees converge to the same Brownian excursion. *Ann. Probab.* **31** (2003), no. 3, 1655–1678.

[30] J.-F. Marckert & A. Panholzer, Noncrossing trees are almost conditioned Galton–Watson trees. *Random Struct. Alg.* **20** (2002), no. 1, 115–125.

[31] A. Meir & J.W. Moon, Cutting down random trees. *J. Australian Math. Soc.* **11** (1970), 313–324.

[32] A. Meir & J.W. Moon, On the altitude of nodes in random trees. *Canad. J. Math.* **30** (1978), 997–1015.

[33] A. Panholzer, Cutting down very simple trees. Preprint, 2003.

[34] A. Panholzer, Non-crossing trees revisited: cutting down and spanning subtrees. *Proceedings, Discrete Random Walks 2003*, Cyril Banderier and Christian Krattenthaler, Eds., *Discr. Math. Theor. Comput. Sci.* **AC** (2003), 265–276.

[35] J. Pitman, *Combinatorial Stochastic Processes*, Lecture Notes for St. Flour Summer School, July 2002. Preprint, available from `http://stat-www.berkeley.edu/users/pitman/bibliog.html`

[36] A. Rényi, (1962). On the extreme elements of observations. *MTA III, Oszt. Közl.* **12** (1962) 105–121. Reprinted in *Collected Works*, Vol III, pp. 50-66, Akadémiai Kiadó, Budapest, 1976.

[37] D. Revuz & M. Yor, *Continuous Martingales and Brownian Motion.* 3^{rd} edition, Springer-Verlag, Berlin, 1999.

Department of Mathematics, Uppsala University, PO Box 480, S-751 06 Uppsala, Sweden

E-mail address: `svante.janson@math.uu.se`

URL: `http://www.math.uu.se/~svante/`