IDEALS IN A FOREST, ONE-WAY INFINITE BINARY TREES AND THE CONTRACTION METHOD

SVANTE JANSON

Abstract. The analysis of an algorithm by Koda and Ruskey for listing ideals in a forest poset leads to a study of random binary trees and their limits as infinite random binary trees. The corresponding finite and infinite random forests are studied too. The infinite random binary trees and forests studied here have exactly one infinite path; they can be defined using suitable size-biazed Galton–Watson processs. Limit theorems are proved using a version of the contraction method.

1. Introduction

The vertices of a rooted forest may be regarded as a poset in a natural way, with the roots being the minimal elements. Consider the family of all ideals (or down-sets) of this poset. If the forest consists of trees $T_1,\dots,T_k$, then the ideals are the sets of the form $V_1 \cup \dots \cup V_k$, where each V_i is either empty or the vertex set of a rooted subtree of T_i.

Koda and Ruskey [11] described two algorithms for listing the ideals of a forest poset in a Gray code manner, i.e. such that consecutive ideals differ by exactly one element. (For background and applications, see [11]. For actual implementations, see Knuth [10].) We are here concerned only with their first algorithm, `Algorithm P` in [11]. Since the algorithm operates on ordered forests, we assume from now on that all forests and trees are rooted and ordered.

As noted in [11], the running time per ideal of `Algorithm P`, i.e. the total running time divided by the number of ideals listed, is not bounded. However, it is conjectured in [11, Section 6] that the expected running time per ideal for a randomly selected rooted tree on n vertices is bounded as $n \to \infty$.

In the present paper, we study random ordered rooted trees, and verify the conjecture of [11] in this case. (As pointed out by a referee, the conjecture in [11] is really stated for random rooted trees; the algorithm operates on ordered trees, but the probability distribution depends on whether the ordering is imposed before or after the random selection. Presumably, the result holds for random rooted trees and other families of simply generated

Date: November 2, 2001; revised April 25, 2001.

trees too.) Moreover, we show that both the expectation and the distribution of the running time per ideal converges as $n \to \infty$ (without further normalization).

The proofs use a version of the contraction method, which earlier has been used to study many other algorithms, see e.g. [16, 17, 18, 19]. The present application includes some novel features, however, which we find at least as interesting as the results themselves. Thus, although the paper exclusively studies `Algorithm P`, it should mainly be seen as an example illustrating a method that we hope may be useful for the study of other algorithms as well.

In the proofs we find it convenient to transfer the problem to an equivalent one for random binary trees, see Section 3. Note that we consider random binary trees with the uniform distribution over all binary trees of a given size (sometimes called Catalan trees), in contrast to the binary search trees that appear in connection with other applications of the contraction method (in particular, `Quicksort`). The distributions are quite different, with the uniform binary trees studied here tending to be much more unbalanced and stringy, which leads to new phenomena.

In the present case (unlike the case of binary search trees), there is a natural limit of the random binary tree as its size tends to infinity; this is a non-trivial infinite random binary tree. Similarly, there is an infinite random forest that is the limit of the random ordered rooted forest. We study these infinite trees and forests in Sections 5 and 6 and show that the cost per ideal can be defined (a.s.) for these infinite objects too, in such a way that its distribution is the limiting distribution of the cost per ideal for finite random forests. This enables us to deduce some properties of the limiting distribution. For example, we show that the distribution is continuous (Theorem 5.10). It is, however, an open problem whether it is absolutely continuous.

The infinite random forests and binary trees studied here have exactly one infinite branch. They can be defined using a size-biased Galton–Watson branching process, see Section 5. We include some further comments on the structure of these infinite objects in Sections 5 and 6.

Acknowledgement. I thank Donald Knuth for suggesting this problem.

2. Preliminaries

We let $|F|$ denote the number of vertices in a forest, or tree, F. If T is a tree, we let T^* be the forest obtained by deleting the root, letting the children of the old root be the new roots. Thus $|T^*| = |T| - 1$.

Denote the number of ideals of a forest F by $N(F)$. If F consists of the trees $T_1, \dots, T_m$, then

$$N(F) = \prod_{i=1}^{m} N(T_i). \tag{2.1}$$

In particular, $N(\emptyset) = 1$. Moreover, it is easily seen that if T is a tree, then

$$N(T) = 1 + N(T^*). \tag{2.2}$$

Note that (2.1) and (2.2) together determine N recursively. It is easily seen by induction, or directly, considering the ideals consisting of paths from the root and the empty ideal, that

$$N(F) \geq |F| + 1. \tag{2.3}$$

We let $W(F)$ denote a measure of the running time (work) of `Algorithm P` on a forest F. Of course, the actual running time depends on details in the implementation, but we make a precise definition as follows, using the descriptions in the proof of Lemma 3.1 in [11]:

If F is empty, we let $W(F) = 0$.

If F consists of a single tree T, then `Algorithm P` lists first $\emptyset$ and then all ideals of T^*, in the order given by `Algorithm P` on T^*, with the root of T added to each. The work required by the algorithm on T is therefore the same as for T^*, but with one extra unit for each ideal of T^* (because of the added root) and two extra units for the additional ideal. Hence

$$W(T) = W(T^*) + N(T^*) + 2 = W(T^*) + N(T) + 1. \tag{2.4}$$

If F consists of several trees $T_1, \dots, T_k$, $k \geq 2$, let $F' = F \setminus T_1 = T_2 \cup \dots \cup T_k$. Then `Algorithm P` first lists all ideals of T_1, ignoring F', then acts in F' as if running on F', then lists all ideals of T_1 in reverse order with the first nonempty ideal of F' added to each, then acts in F' again, then lists the ideals of T_1 in order with the second nonempty ideal of F' added to each, and so on. Hence the ideals of T_1 are run through $N(F')$ times (in alternating directions) with a work $W(T_1)$ each time, while the remaining steps together are equivalent to running the algorithm on F', which requires $W(F')$. Hence

$$W(F) = N(F')W(T_1) + W(F'). \tag{2.5}$$

This completes our (recursive) definition of W. (This definition of W by (2.4) and (2.5) was given, in an equivalent form, by Knuth [personal communication].)

Remark 2.1. There is some arbitrariness in the definition; in particular, (2.4) might be modified to $W(T) = W(T^*) + aN(T^*) + b$ for some other positive constants a and b. This would not cause any important differences to the results of this paper (although numerical values will differ); we can assume that $a = 1$ by dividing W by a, and a value of b different from 2 would cause only routine changes below.

Note that $N(F)$ and $W(F)$ vary wildly among forests of the same size. The extreme cases are, as is easily verified by induction:

(i) n isolated roots; $N = 2^n$, $W = 3(2^n - 1)$.
(ii) n vertices in a path; $N = n + 1$, $W = \binom{n+1}{2} + 2n = (n^2 + 5n)/2$.

We therefore study $Q(F) := W(F)/N(F)$, the work per ideal. Note that the path in (ii) shows that Q is unbounded (on trees as well as on forests). ($N(F)$ and $N(T)$ are studied in [8, 20], but we do not use the results there.)

There are C_n ordered forests with n vertices, where

$$C_n = \frac{1}{n+1}\binom{2n}{n} = \frac{(2n)!}{n!\,(n+1)!} \tag{2.6}$$

is the n:th Catalan number, and thus there are C_{n-1} ordered trees with n vertices [9, 2.3.4.4].

Let F_n denote a random ordered rooted forest with n vertices, uniformly selected among the C_n possibilities; let similarly T_n be a uniformly selected random ordered rooted tree with n vertices. We can now state the main results of the paper, proved in Section 4.

Theorem 2.2. *There exists a positive random variable* $\mathbf{Q}$ *with finite mean such that, as* $n \to \infty$, $Q(F_n) \xrightarrow{\mathrm{d}} \mathbf{Q}$ *and* $\mathbb{E}\, Q(F_n) \to \mathbb{E}\,\mathbf{Q}$.

Corollary 2.3. *As* $n \to \infty$, *with* $\mathbf{Q}$ *as in Theorem 2.2,* $Q(T_n) \xrightarrow{\mathrm{d}} \mathbf{Q}+1$ *and* $\mathbb{E}\, Q(T_n) \to \mathbb{E}\,\mathbf{Q}+1$.

Proof. By (2.2) and (2.4),

$$Q(T_n) = \frac{W(T_n)}{N(T_n)} = \frac{W(T_n^*) + N(T_n^*) + 2}{N(T_n^*) + 1} = Q(T_n^*) + 1 + \frac{1 - Q(T_n^*)}{N(T_n^*) + 1}.$$

Since T_n^* is distributed as F_{n-1}, and $N(T_n^*) \geq n$ by (2.3), the results follow from Theorem 2.2. □

We have no explicit description of the limit distribution $\mathcal{L}(\mathbf{Q})$, but it is characterized by a fixed point equation. This fixed point equation is more complicated than in many other similar cases, so we postpone it to Section 5, see Theorems 5.7 and 5.8. In Sections 5 and 6 we further show that $\mathbf{Q}$ may be interpreted as an extension of Q to random infinite forests.

3. Binary trees

We find it convenient to consider binary trees instead of forests, using the well-known correspondence in [9, Section 2.3.2], which can be defined recursively as follows: If $F = \emptyset$, then $B(F) = \emptyset$. If F is a forest consisting of trees $T_1, \dots, T_k$, $k \geq 1$, then $B(F)$ is the binary tree with a root, a left subtree $B(T_1^*)$ and a right subtree $B(T_2 \cup \dots \cup T_k)$. Note that $|B(F)| = |F|$.

We define N, W and Q for binary trees by this correspondence, setting $N(B(F)) = N(F)$ and so on.

It is easily seen that (2.1), (2.2), (2.4), (2.5) translate as follows: If B is a nonempty binary tree with left and right subtrees L and R, then

$$N(B) = \bigl(N(L) + 1\bigr)N(R) \tag{3.1}$$

$$W(B) = N(R)\bigl(W(L) + N(L) + 2\bigr) + W(R). \tag{3.2}$$

Together with $N(\emptyset) = 1$ and $W(\emptyset) = 0$, (3.1) and (3.2) define N and W directly on binary trees by recursion. Taking the quotient, we further obtain

$$\begin{aligned} Q(B) &= \frac{W(L) + N(L) + 2}{N(L) + 1} + \frac{W(R)}{N(R)(N(L) + 1)} \\ &= Q(L) + 1 + \frac{1 - Q(L) + Q(R)}{N(L) + 1}. \end{aligned} \tag{3.3}$$

Let B_n denote a (uniformly selected) random binary tree with n vertices ($n \geq 0$). Then Theorem 2.2 is equivalent to the following.

Theorem 3.1. *As $n \to \infty$, $Q(B_n) \xrightarrow{\mathrm{d}} \mathbf{Q}$ and $\mathbb{E}\, Q(B_n) \to \mathbb{E}\, \mathbf{Q}$, with $\mathbf{Q}$ as in Theorem 2.2.*

Let, for $n \geq 1$, L_n and R_n denote the left and right subtrees of B_n. Note that

$$|L_n| + |R_n| = n - 1. \tag{3.4}$$

Let $(p_{n,k})_{k=0}^{n-1}$ be the probability distribution of the size of the left (or, by symmetry, the right) subtree of B_n, i.e.

$$p_{n,k} := \mathbb{P}(|L_n| = k) = \mathbb{P}(|R_n| = k). \tag{3.5}$$

By (3.4),

$$p_{n,k} = \mathbb{P}(|L_n| = k) = \mathbb{P}(|R_n| = n - 1 - k) = p_{n,n-1-k}. \tag{3.6}$$

There are C_n binary trees with n vertices, where again C_n is the Catalan number (2.6). Hence, the number of binary trees with k vertices in the left subtree and $n - 1 - k$ in the right is $C_k C_{n-1-k}$, and

$$p_{n,k} = \frac{C_k C_{n-1-k}}{C_n}.$$

Stirling's formula easily yields

$$C_n = \frac{1}{n+1}\binom{2n}{n} = \pi^{-1/2} n^{-3/2} 2^{2n} \bigl(1 + O(n^{-1})\bigr).$$

Hence, uniformly for $0 \leq k \leq n/2$,

$$\begin{aligned} p_{n,k} &= C_k (n-1-k)^{-3/2} n^{3/2} 2^{2(n-1-k)-2n} \bigl(1 + O(n^{-1})\bigr) \\ &= C_k 4^{-k-1} \Bigl(1 + O\Bigl(\frac{k+1}{n}\Bigr)\Bigr) \\ &= \pi_k \Bigl(1 + O\Bigl(\frac{k+1}{n}\Bigr)\Bigr), \end{aligned}$$

where

$$\pi_k = \lim_{n\to\infty} p_{n,k} = C_k 4^{-k-1}, \qquad k \geq 0.$$

By the generating function for Catalan numbers

$$B(z) := \sum_{k=0}^{\infty} C_k z^k = \frac{1 - \sqrt{1 - 4z}}{2z}, \tag{3.7}$$

see e.g. [9, (2.3.4.4-13)], we have

$$\sum_{k=0}^{\infty} \pi_k = \sum_{k=0}^{\infty} C_k 4^{-k-1} = \frac{1}{4} B\Bigl(\frac{1}{4}\Bigr) = \frac{1}{2}. \tag{3.8}$$

Hence $(\pi_k)_0^\infty$ is not a probability distribution. This reflects the symmetry of the left and right sides; roughly speaking, for n large, with probability 1/2 $|L_n|$ is small, and with probability 1/2 $|R_n|$ is small and then $|L_n| \approx n$. In particular, a large random binary tree is extremely unbalanced. We state this more precisely.

Lemma 3.2. *For each $\varepsilon > 0$ there exists M such that, for every n,*

$$\mathbb{P}(|L_n| < M) > \tfrac{1}{2} - \varepsilon \tag{3.9}$$

$$\mathbb{P}(|L_n| \geq n - M) > \tfrac{1}{2} - \varepsilon \tag{3.10}$$

$$\mathbb{P}(M \leq |L_n| < n - M) < 2\varepsilon. \tag{3.11}$$

Proof. Choose M_1 such that $\sum_{k=M_1}^{\infty} \pi_k < \varepsilon$. Then, as $n \to \infty$,

$$\mathbb{P}(|L_n| < M_1) \to \sum_{k=0}^{M_1-1} \pi_k > \tfrac{1}{2} - \varepsilon,$$

so (3.9) holds with $M = M_1$ for sufficiently large n, say $n \geq n_0$. Taking $M := \max(M_1, n_0)$, (3.9) holds for all n. Furthermore, (3.10) holds by (3.6), while (3.11) is an immediate consequence of (3.9) and (3.10). ☐

We can modify (π_k) to make it into a probability distribution in two ways, both of which will be used below. First, we can allow the value $+\infty$, giving it the probability 1/2 because of (3.8). We let ξ^* be a random variable with values in $N^* := \{0, 1, \dots, \infty\}$ having this distribution, i.e.

$$\mathbb{P}(\xi^* = k) = \begin{cases} \pi_k, & 0 \leq k < \infty, \\ \frac{1}{2}, & k = \infty. \end{cases}$$

Alternatively, we can renormalize (π_k) and consider the probability distribution $(2\pi_k)_{0 \leq k < \infty}$. We let ξ be a random variable with this distribution, i.e.

$$\mathbb{P}(\xi = k) = 2\pi_k = 2^{-2k-1} C_k, \qquad k \geq 0. \tag{3.12}$$

Note that ξ can be defined as ξ^* conditioned on $\xi^* < \infty$.

With this notation, the following lemmas are immediate consequences of the results above.

Lemma 3.3. *Let $n \to \infty$. Then $|L_n| \xrightarrow{\mathrm{d}} \xi^*$, as random variables in N^*.* ☐

Lemma 3.4. *Let $n \to \infty$. Then $|L_n|$, conditioned on $|L_n| < n/2$, converges in distribution to ξ.* ☐

Of course, the same results hold for $|R_n|$.

Remark 3.5. These results for the uniform random binary tree studied here should be compared with the corresponding results for random binary search trees, which have a different distribution and for which $|L_n|$ is uniformly distributed on $\{0,\dots,n-1\}$. The stronger imbalance in our case is a source of phenomena quite different from the binary search tree case.

Finally we record a simple but important observation. We let $\widetilde{B}_n$ be another (uniform) random binary tree, independent of $\{B_k\}_{k=0}^{\infty}$.

Lemma 3.6. *Let $0 \le k < n$. The conditional joint distribution of L_n and R_n given $|L_n| = k$ equals the distribution of $(B_k, \widetilde{B}_{n-1-k})$.* □

In other words, conditioned on the sizes of the subtrees L_n and R_n, they are two independent random binary trees.

4. Proof of the limit theorems

We begin with a preliminary estimate, which verifies the conjecture that the expected running time is bounded.

Lemma 4.1. $\sup_{n\ge 0} \mathbb{E}\, Q(B_n) < \infty$.

Proof. Define, for $n \ge 0$,

$$a_n := \mathbb{E}\, Q(B_n) + 1,$$
$$b_n := \mathbb{E}\, \frac{1}{N(B_n)+1} \le \frac{1}{2}.$$

From (3.3), conditioning on $|L_n|$ and using Lemma 3.6,

$$\begin{aligned} a_n &\le \mathbb{E}\Big(Q(L_n) + 1 + (1+Q(R_n))\frac{1}{N(L_n)+1}\Big) + 1 \\ &= \sum_{l=0}^{n-1} p_{n,l} a_l + \sum_{r=0}^{n-1} p_{n,r} a_r b_{n-1-r} + 1 \\ &= \sum_{l=0}^{n-1} p_{n,l} a_l (1 + b_{n-1-l}) + 1 \\ &\le \tfrac{3}{2} \sum_{l=0}^{n-1} p_{n,l} a_l + 1. \end{aligned} \tag{4.1}$$

Let $a_n^* := \max_{0\le k\le n} a_k$. By (4.1), for any $M \ge 0$,

$$\begin{aligned} a_n &\le \tfrac{3}{2} a_M^*\, \mathbb{P}(|L_n| \le M) + \tfrac{3}{2} a_{n-1}^*\, \mathbb{P}(|L_n| > M) + 1 \\ &\le 1 + \tfrac{3}{2} a_M^* + \tfrac{3}{2} a_{n-1}^* \big(1 - \mathbb{P}(|L_n| \le M)\big). \end{aligned}$$

We choose M as in Lemma 3.2 with $\varepsilon = 0.1$. Then $\mathbb{P}(|L_n| \le M) > 0.4$ and thus, for all $n \ge 1$,

$$a_n \le 1 + \tfrac{3}{2} a_M^* + \tfrac{3}{2} \cdot 0.6 a_{n-1}^* = 0.9 a_{n-1}^* + 1 + \tfrac{3}{2} a_M^*.$$

An easy induction yields

$$a_n \le 10(1 + \tfrac{3}{2} a_M^*), \qquad n \ge 0.$$

The lemma follows because $\mathbb{E}\, Q(B_n) \le a_n$. □

We prove Theorem 3.1, and thus Theorem 2.2 and Corollary 2.3, using the Mallows metric d_1 for probability distributions with finite expectations. (This metric is also known under many other names, such as the Dudley, Fortet-Mourier, Kantorovich or Wasserstein distance.) It has several equivalent definitions, see e.g. [15]; for us the following is convenient.

If f is a real (or complex) function on $\mathbb{R}$, let

$$\|f\|_{\mathrm{Lip}} := \sup_{x \ne y} \frac{|f(x) - f(y)|}{|x - y|}.$$

If μ and ν are probability measures on $\mathbb{R}$ with finite expectations, then

$$d_1(\mu, \nu) := \sup\Bigl\{ \Bigl| \int f\, d\mu - \int f\, d\nu \Bigr| : \|f\|_{\mathrm{Lip}} \le 1 \Bigr\}. \tag{4.2}$$

In other words, we take the supremum in (4.2) over all functions f satisfying the Lipschitz condition $|f(x) - f(y)| \le |x - y|$. (It does not matter whether we consider real or complex functions.)

If X and Y are random variables with finite expectations, we will for simplicity write $d_1(X, Y)$ for the d_1 distance between their distributions. Thus

$$d_1(X, Y) := \sup\bigl\{ \bigl| \mathbb{E}\, f(X) - \mathbb{E}\, f(Y) \bigr| : \|f\|_{\mathrm{Lip}} \le 1 \bigr\}. \tag{4.3}$$

It is easily seen that $d_1(X_n, X) \to 0$ implies $X_n \xrightarrow{\mathrm{d}} X$ and $\mathbb{E}\, X_n \to \mathbb{E}\, X$. (Take $f(x) = |t|^{-1} e^{itx}$, $t \ne 0$, and $f(x) = x$.) We will show that $d_1(Q(B_n), \mathbf{Q}) \to 0$ for some random variable $\mathbf{Q}$; this thus proves Theorem 3.1. (Indeed, Theorem 3.1 is equivalent to $d_1(Q(B_n), \mathbf{Q}) \to 0$, using the fact that $d_1(X_n, X) \to 0$ is equivalent to $X_1 \xrightarrow{\mathrm{d}} X$ and $\mathbb{E}\, |X_n| \to \mathbb{E}\, |X|$ for any random variables with finite expectations.)

Remark 4.2. That $\mathbb{E}\, Q(B_n)$ converges could also be shown directly using a simplified version of the proof below, taking $f(x) = x$.

Note first that replacing f by $f - f(0)$ does not change $\mathbb{E}\, f(X) - \mathbb{E}\, f(Y)$. Hence we may in (4.3) further impose $f(0) = 0$. Since then $|f(x)| = |f(x) - f(0)| \le |x|$, we have the bound

$$d_1(X, Y) \le \mathbb{E}\, |X| + \mathbb{E}\, |Y|. \tag{4.4}$$

We now consider $Q(B_n)$. For notational convenience we write $X_n = Q(B_n)$, $Y_n = (N(B_n) + 1)^{-1}$ and $\widetilde{X}_n = Q(\widetilde{B}_n)$. Thus $\widetilde{X}_n$ has the same distribution as X_n but is independent of all X_k and Y_k. Note that, by (2.3),

$$Y_n \le \frac{1}{n+2}.$$

We further define

$$\delta_N := \sup\{d_1(X_n, X_m) : n, m \geq N\}. \tag{4.5}$$

Let $c_1 := \sup_n \mathbb{E}\, X_n$, which is finite by Lemma 4.1. By (4.4), $\delta_N \leq 2c_1 < \infty$.

Fix a function f with $\|f\|_{\mathrm{Lip}} \leq 1$ and $f(0) = 0$. By (3.3) and Lemma 3.6, conditioning on $|L_n|$, for $n \geq 1$,

$$\begin{aligned}
\mathbb{E}\, f(X_n) &= \mathbb{E}\, f(Q(B_n)) = \mathbb{E}\, f\Bigl(Q(L_n) + 1 + \frac{1 - Q(L_n) + Q(R_n)}{N(L_n) + 1}\Bigr) \\
&= \sum_k p_{n,k}\, \mathbb{E}\, f\Bigl(Q(B_k) + 1 + \frac{1 - Q(B_k) + Q(\widetilde{B}_{n-1-k})}{N(B_k) + 1}\Bigr) \\
&= \sum_k p_{n,k}\, \mathbb{E}\, f\bigl(X_k + 1 + Y_k(1 - X_k + \widetilde{X}_{n-1-k})\bigr) \\
&= \sum_k p_{n,k}\, \mathbb{E}\, f(U_{n,k})
\end{aligned} \tag{4.6}$$

where

$$U_{n,k} := X_k + 1 + Y_k(1 - X_k + \widetilde{X}_{n-1-k}). \tag{4.7}$$

We have, for $0 \leq k \leq n$,

$$|\,\mathbb{E}\, f(U_{n,k})| \leq \mathbb{E}\, |U_{n,k}| \leq \mathbb{E}(X_k + 2 + \widetilde{X}_{n-1-k}) \leq c_2 := 2c_1 + 2 \tag{4.8}$$

and

$$\begin{aligned}
|\,\mathbb{E}\, f(U_{n,k}) - \mathbb{E}\, f(X_k + 1)| &\leq \mathbb{E}\, |f(U_{n,k}) - f(X_k + 1)| \\
&\leq \mathbb{E}\, |U_{n,k} - X_k - 1| = \mathbb{E}\, |Y_k(1 - X_k + \widetilde{X}_{n-1-k})| \\
&\leq \mathbb{E}\, \frac{1}{k+2} |1 - X_k + \widetilde{X}_{n-1-k}| \leq \frac{c_2}{k+2}.
\end{aligned} \tag{4.9}$$

Let $\varepsilon > 0$ and let M be as in Lemma 3.2. By (4.8),

$$\Bigl|\sum_{k=0}^{M} p_{n,k}\, \mathbb{E}\, f(U_{n,k}) - \sum_{k=0}^{M} \pi_k\, \mathbb{E}\, f(U_{n,k})\Bigr| \leq \sum_{k=0}^{M} |p_{n,k} - \pi_k| c_2 \tag{4.10}$$

and

$$\begin{aligned}
\Bigl|\sum_{k=M+1}^{n-M-2} p_{n,k}\, \mathbb{E}\, f(U_{n,k})\Bigr| &\leq c_2 \sum_{k=M+1}^{n-M-2} p_{n,k} = c_2\, \mathbb{P}(M < |L_n| < n - 1 - M) \\
&< 2c_2\varepsilon.
\end{aligned} \tag{4.11}$$

Furthermore,

$$\sum_{k=n-1-M}^{n-1} p_{n,k}\, \mathbb{E}\, f(U_{n,k}) = \sum_{j=0}^{M} p_{n,n-j-1}\, \mathbb{E}\, f(U_{n,n-1-j}) = \sum_{j=0}^{M} p_{n,j}\, \mathbb{E}\, f(U_{n,n-1-j})$$

and thus by (4.8) and (4.9)

$$\begin{aligned}
&\left|\sum_{k=n-1-M}^{n-1} p_{n,k}\,\mathbb{E}\, f(U_{n,k}) - \sum_{j=0}^{M} \pi_j\,\mathbb{E}\, f(X_{n-1-j}+1)\right| \\
&\le \sum_{j=0}^{M} |p_{n,j}-\pi_j|\,|\,\mathbb{E}\, f(U_{n,n-1-j})| + \sum_{j=0}^{M} \pi_j |\,\mathbb{E}\, f(U_{n,n-1-j}) - \mathbb{E}\, f(X_{n-1-j}+1)| \\
&\le c_2 \sum_{j=0}^{M} |p_{n,j}-\pi_j| + \frac{c_2}{n-M}\sum_{j=0}^{M}\pi_j.
\end{aligned} \tag{4.12}$$

We define R_n by

$$\mathbb{E}\, f(X_n) = \sum_{k=0}^{M} \pi_k\,\mathbb{E}\, f(U_{n,k}) + \sum_{k=0}^{M} \pi_k\,\mathbb{E}\, f(X_{n-1-k}+1) + R_n \tag{4.13}$$

and obtain by (4.6) and (4.10)–(4.12), for $n \ge 2M$,

$$|R_n| \le 2c_2 \sum_{k=0}^{M} |p_{n,k}-\pi_k| + 2c_2\varepsilon + \frac{c_2}{n-M}. \tag{4.14}$$

Let N_0 be so large that $N_0 > 2M$, $N_0 > M+1/\varepsilon$ and $|p_{n,k}-\pi_k| < \varepsilon/(M+1)$ when $n \ge N_0$ for $k \le M$. Then (4.14) yields

$$|R_n| \le 5c_2\varepsilon, \qquad n \ge N_0. \tag{4.15}$$

Now suppose $N \ge N_0$ and $m, n \ge N$. Using (4.13) and (4.15) we have

$$\begin{aligned}
&|\,\mathbb{E}\, f(X_n) - \mathbb{E}\, f(X_m)| \\
&\le \sum_{k=0}^{M} \pi_k |\,\mathbb{E}\, f(U_{n,k}) - \mathbb{E}\, f(U_{m,k})| \\
&\qquad + \sum_{k=0}^{M} \pi_k |\,\mathbb{E}\, f(X_{n-1-k}+1) - \mathbb{E}\, f(X_{m-1-k}+1)| + 10c_2\varepsilon.
\end{aligned} \tag{4.16}$$

Since $f_1(x) := f(x+1)$ is a function with $\|f_1\|_{\mathrm{Lip}} = \|f\|_{\mathrm{Lip}} \le 1$,

$$|\,\mathbb{E}\, f(X_{n-1-k}+1) - \mathbb{E}\, f(X_{m-1-k}+1)| \le d_1(X_{n-1-k}, X_{m-1-k}). \tag{4.17}$$

Similarly, for any given X_k and Y_k, the function

$$g(x) := f\bigl(X_k + 1 + Y_k(1 - X_k + x)\bigr)$$

has Lipschitz norm $\|g\|_{\mathrm{Lip}} \le Y_k$, and thus by (4.7)

$$\begin{aligned}
|\,\mathbb{E}\bigl(f(U_{n,k}) - f(U_{m,k}) \mid X_k, Y_k\bigr)| &= |\,\mathbb{E}\, g(\widetilde{X}_{n-1-k}) - \mathbb{E}\, g(\widetilde{X}_{m-1-k})| \\
&\le \|g\|_{\mathrm{Lip}} d_1(\widetilde{X}_{n-1-k}, \widetilde{X}_{m-1-k}) \le Y_k d_1(X_{n-1-k}, X_{m-1-k}).
\end{aligned}$$

Using the crude bound $Y_k \le 1/2$ and taking the expectation we have

$$|\mathbb{E} f(U_{n,k}) - \mathbb{E} f(U_{m,k})| \le \mathbb{E}\,\big|\mathbb{E}\big(f(U_{n,k}) - f(U_{m,k}) \mid X_k, Y_k\big)\big| \le \tfrac{1}{2} d_1(X_{n-1-k}, X_{m-1-k}). \tag{4.18}$$

Consequently, by (4.16), (4.18) and (4.5), for $m, n \ge N \ge N_0$,

$$\begin{aligned}|\mathbb{E} f(X_n) - \mathbb{E} f(X_m)| &\le \sum_{k=0}^{M} \pi_k \tfrac{1}{2}\delta_{N-1-k} + \sum_{k=0}^{M} \pi_k \delta_{N-1-k} + 10c_2\varepsilon \\ &\le \tfrac{3}{2}\delta_{N-1-M} \sum_{k=0}^{M} \pi_k + 10c_2\varepsilon \\ &\le \tfrac{3}{4}\delta_{N-1-M} + 10c_2\varepsilon.\end{aligned}$$

Taking the supremum over all f with $\|f\|_{\mathrm{Lip}} \le 1$ and $f(0) = 0$ we find, for $n, m \ge N \ge N_0$,

$$d_1(X_n, X_m) \le \tfrac{3}{4}\delta_{N-1-M} + 10c_2\varepsilon$$

and thus

$$\delta_N \le \tfrac{3}{4}\delta_{N-1-M} + 10c_2\varepsilon, \qquad N \ge N_0.$$

Letting $N \to \infty$ we obtain

$$\limsup_{N\to\infty} \delta_N \le \tfrac{3}{4} \limsup_{N\to\infty} \delta_N + 10c_2\varepsilon$$

and thus, since $\limsup_{N\to\infty} \delta_N \le 2c_1 < \infty$,

$$\limsup_{N\to\infty} \delta_N \le 40c_2\varepsilon.$$

Finally, letting $\varepsilon \to 0$, we obtain $\delta_N \to 0$ as $N \to \infty$.

By the definition (4.5), this shows that $(X_n)_n$, or rather the corresponding sequence of distributions, is a Cauchy sequence in the d_1 metric. It is easily seen that the space of all probability measures on $\mathbb{R}$ with finite expectation is complete with the metric d_1 [15]. Hence there exists a limit distribution, and thus a random variable $\mathbf{Q}$ such that $d_1(X_n, \mathbf{Q}) \to 0$, which completes the proof. □

Remark 4.3. The proof above shows that the distributions of $(X_n)_n$ form a Cauchy sequence, and thus converge to some limit. The limit will in the next section be characterized by fixed point equations. An anonymous referee has pointed out that, alternatively, it is possible to first define the limit distribution by the fixed point equation in Theorem 5.7 and then use arguments similar to the proof above to show that $d_1(X_n, \mathbf{Q}) \to 0$. This is the usual procedure in applications of the contraction method; it has some advantages in the current setting too but also some disadvantages, and we do not find the differences decisive. Anyone interested in extending the present results should consider both versions of the method.

5. More on binary trees

We begin with some more or less well-known (folk-lore?) observations on random binary trees.

Define a random binary tree B_* with random size by the following construction: Flip a fair coin. If it comes up tails, let B_* be empty, otherwise begin with a root. In the latter case, flip the coin again twice and add a left child of the root if the first flip results in heads and a right child if the second flip results in heads. Continue in this way, flipping the coin twice for every new vertex, as long as new vertices are added. (Equivalently, do site percolation on the complete infinite binary tree by flipping a fair coin for each vertex and removing the vertices that get tails, and let B_* be the component of the root, if any.)

We can regard B_* as the family tree of a Galton–Watson branching process with offspring distribution $\mathrm{Bi}(2, 1/2)$ (and children labelled as left or right), starting with $\mathrm{Bi}(1, 1/2)$ individuals. We thus call B_* the *Galton–Watson binary tree*. Since this Galton–Watson process is critical, it a.s. dies out, and thus B_* is finite.

The probability that B_* equals a given binary tree with n vertices is 2^{-2n-1}, since the vertices have to be chosen by n specified coin flips coming up heads, while $n+1$ other coin flips have to yield tails. Hence

$$\mathbb{P}(|B_*| = n) = C_n 2^{-2n-1} = 2\pi_n. \tag{5.1}$$

In other words, $|B_*|$ has the same distribution as ξ defined in (3.12). Moreover, the conditional distribution of B_* given $|B_*| = n$ is uniform, and thus equals the distribution of B_n. This yields yet another possibility of defining B_*: select its size by (5.1) and then select uniformly a binary tree with this size. Equivalently, if ξ is independent of $(B_n)_{n=0}^{\infty}$, we can take $B_* = B_\xi$.

Lemmas 3.4 and 3.6 imply the following:

Lemma 5.1. *Let $n \to \infty$. Then L_n, conditioned on $|L_n| < n/2$, converges in distribution to B_*.* □

Consequently, a large random binary tree has one branch at the root distributed (asymptotically) as B_*, while the other is large. We may continue recursively with the large branch, which suggests the following construction.

Define a (noncomplete) *infinite* random binary tree B_∞ as follows. Begin with the root and create an infinite path from it by randomly adding, an infinite number of times, either a left or a right child to the last added vertex. Finally, add independent copies of B_* at the free sites of the vertices in the path, i.e. as left or right subtrees depending on which side is not already occupied by the remainder of the infinite path. Note that B_∞ has exactly one infinite path from the root; we call this path the *trunk*.

It is easily seen that B_∞ can be defined by the following modification of the branching process above creating B_*. Consider a Galton–Watson process with two types of individuals, *mortals* and *immortals*. Let a mortal have $\mathrm{Bi}(2, 1/2)$ children, all mortal, and let an immortal have exactly one immortal child and $\mathrm{Bi}(1, 1/2)$ mortal children. Moreover, label each child as left or right, at random but ensuring that two siblings get different labels. The resulting family tree is B_∞.

It is now easy to see that Lemma 5.1 implies the following description of the asymptotic shape of large random binary trees. Let, for any tree B, $B^{(M)}$ be the first M levels of B, i.e. the tree with all branches pruned at height M.

Lemma 5.2. *As $n \to \infty$, $B_n \xrightarrow{\mathrm{d}} B_\infty$ in the sense of finite-dimensional distributions, i.e., $B_n^{(M)} \xrightarrow{\mathrm{d}} B_\infty^{(M)}$ for every finite M.* □

Remark 5.3. If we regard the space $\mathcal{B}$ of all finite or infinite binary trees as a subset of the power set of the vertex set of the complete infinite binary tree, with the natural product space topology on the power set, $\mathcal{B}$ is a metrizable compact space. A metric can be defined e.g. by $d(B, B') = 1/(k+1)$ if B and B' differ in the k-th level but not below it. The conclusion of Lemma 5.2 is equivalent to $B_n \xrightarrow{\mathrm{d}} B_\infty$ in this compact metric space $\mathcal{B}$.

Remark 5.4. The construction of B_∞ is a special case of the following general construction of the *size-biased* Galton–Watson process (regarded as a family tree); see e.g. [1] and [13]. Starting from a Galton–Watson process with an offspring distribution μ having finite, positive mean, the size-biased process can be obtained by considering a branching process with two types: mortals with an offspring distribution μ and all children mortals, and immortals with the size-biased offspring distribution $\widehat{\mu}$ and exactly one immortal child (in a random position among its siblings). The process starts with a single immortal. In the critical case studied here (and in the subcritical case), the size-biased process is the same as the *Q-process* studied in [3, Section I.14]. It is shown there that this process arises as the limit (in the sense of finite-dimensional distributions) as $t \to \infty$ of the original process conditioned on extinction occuring after time t (see also [5]). Informally (for critical and subcritical processes), it is the process conditioned on living forever. Similarly, it is easily shown that for a critical Galton–Watson process with finite offspring variance, the size-biased process is the limit as $n \to \infty$ of the process conditioned on the total progeny being n [7, 1]. In the case of random binary trees, this conditioning yields B_n, and we recover Lemma 5.2.

Having proved that both the trees B_n and the functional $Q(B_n)$ defined on them converge in distribution, it is natural to try to interpret the limit in Theorem 3.1 as $Q(B_\infty)$ for an extension of Q to infinite trees. Unfortunately, we cannot define this extension by continuity on the space $\mathcal{B}$ in Remark 5.3. Indeed, it is easily seen that for any infinite binary tree b, there is a sequence b_n of finite binary trees such that $b_n \to b$ in $\mathcal{B}$ but $Q(b_n) \to \infty$; for example, construct b_n by pruning b at height n and adding a sufficiently large complete binary tree at one of the cuts. (We leave the verification to the reader.) Hence, Q has no continuous extension to $\mathcal{B}$.

However, we can extend Q in the following, somewhat weaker, way. We let $N(B)=\infty$ for any infinite tree B. We further let L_∞ and R_∞ denote the left and right subtree of the root of B_∞. Note that exactly one of L_∞ and R_∞ is finite.

Theorem 5.5. *There exists an extension of Q to infinite binary trees such that $\mathbb{E}\,|Q(B_\infty^{(M)})-Q(B_\infty)|\to 0$ as $M\to\infty$. This extension satisfies a.s. the equation*

$$Q(B_\infty)=Q(L_\infty)+1+\frac{1}{N(L_\infty)+1}\bigl(1-Q(L_\infty)+Q(R_\infty)\bigr). \tag{5.2}$$

Moreover, the limit random variable $\mathbf{Q}$ in Theorems 2.2 and 3.1 can be taken as $Q(B_\infty)$, i.e. $Q(B_n)\xrightarrow{\mathrm{d}}Q(B_\infty)$ as $n\to\infty$.

We do not know whether $Q(B_\infty^{(M)})\to Q(B_\infty)$ a.s. as $M\to\infty$.

We begin with a lemma on truncations of finite trees.

Lemma 5.6. *Let, for $M\ge 1$,*

$$\delta^{(M)}:=\sup_n\mathbb{E}\,|Q(B_n^{(M)})-Q(B_n)|.$$

Then $\delta^{(M)}\to 0$ as $M\to\infty$.

Proof. Note first that $\delta^{(M)}<\infty$ by Lemma 4.1 because $Q(B_n^{(M)})$ attains only a finite number of values for each M.

For any n and $M\ge 1$, the left and right subtrees of $B_n^{(M+1)}$ are $L_n^{(M)}$ and $R_n^{(M)}$, and thus (3.3) yields

$$\begin{aligned}Q(B_n^{(M+1)})-Q(B_n)&=\frac{N(L_n^{(M)})}{N(L_n^{(M)})+1}\bigl(Q(L_n^{(M)})-Q(L_n)\bigr)\\&\quad+\frac{1}{N(L_n^{(M)})+1}\bigl(Q(R_n^{(M)})-Q(R_n)\bigr)\\&\quad+\Bigl(\frac{1}{N(L_n^{(M)})+1}-\frac{1}{N(L_n)+1}\Bigr)\bigl(1-Q(L_n)+Q(R_n)\bigr).\end{aligned}$$

Since either $L_n^{(M)}=L_n$ or $|L_n|>|L_n^{(M)}|>M$, this implies

$$\begin{aligned}|Q(B_n^{(M+1)})-Q(B_n)|&\le|Q(L_n^{(M)})-Q(L_n)|+\tfrac12|Q(R_n^{(M)})-Q(R_n)|\\&\quad+\frac{1}{M}|1-Q(L_n)+Q(R_n)|.\end{aligned}\tag{5.3}$$

For any $k\ge 0$, by Lemma 3.6,

$$\mathbb{E}\bigl(|Q(L_n^{(M)})-Q(L_n)|\bigm| |L_n|=k\bigr)=\mathbb{E}\,|Q(B_k^{(M)})-Q(B_k)|\le\delta^{(M)}.$$

Moreover, if $|L_n|=k\le M$, then $L_n^{(M)}=L_n$. Hence,

$$\mathbb{E}\,|Q(L_n^{(M)})-Q(L_n)|\le\delta^{(M)}\,\mathbb{P}(|L_n|>M).$$

The same estimate holds for $\mathbb{E}\,|Q(R_n^{(M)}) - Q(R_n)|$. We thus obtain from (5.3), again letting $c_1 := \sup_n \mathbb{E}\,Q(B_n) < \infty$, see Lemma 4.1,

$$\mathbb{E}\,|Q(B_n^{(M+1)}) - Q(B_n)| \le \tfrac{3}{2}\delta^{(M)}\,\mathbb{P}(|L_n| > M) + \frac{1+2c_1}{M}. \tag{5.4}$$

Let M_0 be as in Lemma 3.2 with $\varepsilon = 0.1$. Then, for every $M \ge M_0$, we have $\mathbb{P}(|L_n| > M) \le \mathbb{P}(|L_n| > M_0) < 0.6$, and thus by (5.4)

$$\mathbb{E}\,|Q(B_n^{(M+1)}) - Q(B_n)| \le 0.9\delta^{(M)} + \frac{c_2}{M}.$$

for every n and thus

$$\delta^{(M+1)} \le 0.9\delta^{(M)} + \frac{c_2}{M}, \qquad M \ge M_0. \tag{5.5}$$

It follows by induction that $\delta^{(M)} \le \delta^{(M_0)} + 10c_2$, $M \ge M_0$, and thus $\delta := \limsup_{M\to\infty} \delta^{(M)} < \infty$. Furthermore, (5.5) implies $\delta \le 0.9\delta$, and consequently $\delta = 0$. ☐

Proof of Theorem 5.5. For any $M, N \ge 1$, it follows from Lemma 5.2 that $Q(B_n^{(M)}) - Q(B_n^{(N)}) \xrightarrow{\mathrm{d}} Q(B_\infty^{(M)}) - Q(B_\infty^{(N)})$ as $n \to \infty$, and thus by Fatou's lemma

$$\mathbb{E}\,|Q(B_\infty^{(M)}) - Q(B_\infty^{(N)})| \le \liminf_{n\to\infty} \mathbb{E}\,|Q(B_n^{(M)}) - Q(B_n^{(N)})| \le \delta^{(M)} + \delta^{(N)}.$$

It follows from Lemma 5.6 that $B_\infty^{(M)}$, $M \ge 1$, is a Cauchy sequence in L^1, and thus this sequence converges to a limit, which can be written $Q(B_\infty)$.

This proves the first assertion and the third follows from this and Lemmas 5.2 and 5.6 by a standard 3ε argument, see e.g. [4, Theorem 4.2].

For (5.2), we observe again that by (3.3)

$$Q(B_\infty^{(M+1)}) = Q(L_\infty^{(M)}) + 1 + \frac{1 - Q(L_\infty^{(M)}) + Q(R_\infty^{(M)})}{N(L_\infty^{(M)}) + 1}. \tag{5.6}$$

As $M \to \infty$, the left hand side converges to $Q(B_\infty)$ in L^1 and thus in probability by the first part of the theorem. Similarly, conditioned on $|L_\infty| = \infty$, $Q(L_\infty^{(M)}) \to Q(L_\infty)$ in L^1 and thus in probability, since the conditional distribution of L_∞ given that it is infinite equals the distribution of B_∞. On the other hand, conditioned on $|L_\infty| < \infty$, obviously $Q(L_\infty^{(M)}) \to Q(L_\infty)$ a.s.. Combining the two cases, $Q(L_\infty^{(M)}) \xrightarrow{\mathrm{p}} Q(L_\infty)$. Similarly $Q(R_\infty^{(M)}) \xrightarrow{\mathrm{p}} Q(R_\infty)$, while $N(L_\infty^{(M)}) \xrightarrow{\mathrm{p}} N(L_\infty) \le \infty$ is evident. Letting $M \to \infty$ in (5.6) thus yields (5.2). ☐

We can develop (5.2) further. First, L_∞ is infinite with probability 1/2. In this case, $N(L_\infty) = \infty$ and (5.2) reduces to

$$Q(B_\infty) = Q(L_\infty) + 1. \tag{5.7}$$

Moreover, the conditional distribution of L_∞ given $|L_\infty| = \infty$ equals the unconditional distribution of B_∞.

The other possibility is L_∞ finite; in this case R_∞ is infinite and its (conditional) distribution equals the unconditional distribution of B_∞, while L_∞ has the same distribution as B_*. We rewrite (5.2) as

$$Q(B_\infty) = \beta(L_\infty) + \alpha(L_\infty)Q(R_\infty) \tag{5.8}$$

where, for a finite binary tree B,

$$\alpha(B) := \frac{1}{N(B)+1}$$
$$\beta(B) := \frac{N(B)Q(B)+N(B)+2}{N(B)+1} = \frac{W(B)+N(B)+2}{N(B)+1}.$$

We can combine (5.7) and (5.8) into the following fixed point equation.

Theorem 5.7. *The limit random variable* $\mathbf{Q} = Q(B_\infty)$ *in Theorems 2.2 and 3.1 satifies the fixed point equation* $\mathbf{Q} \overset{\mathrm{d}}{=} A\mathbf{Q} + B$, *where* (A,B) *is independent of* $\mathbf{Q}$ *and has the distribution given by*

$$(A,B) \overset{\mathrm{d}}{=} \begin{cases} (1,1), & \eta = 0, \\ \big(\alpha(B_*), \beta(B_*)\big), & \eta = 1, \end{cases}$$

where $\eta \sim \mathrm{Bi}(1,1/2)$ *and* B_* *are independent.* □

We can obtain a slightly simpler fixed point equation if we follow the leftmost branch of B_∞ until we find a vertex v with a finite left subtree, i.e. until the infinite path makes its first right turn. (In the branching process construction above, we continue until the left child is mortal.) Let $\zeta \geq 0$ be the height of v, and denote its left and right subtrees by L and R. Then, ζ, L and R are independent; ζ has a geometric distribution $\mathrm{Ge}(1/2)$; L is finite and $L \overset{\mathrm{d}}{=} B_*$; and R is infinite and $R \overset{\mathrm{d}}{=} B_\infty$. Applying (5.7) ζ times followed by (5.8), we find

$$Q(B_\infty) = \beta(L) + \alpha(L)Q(R) + \zeta.$$

This yields the following alternative fixed point equation.

Theorem 5.8. *The limit random variable* $\mathbf{Q}$ *in Theorems 2.2 and 3.1 satifies the fixed point equation* $\mathbf{Q} \overset{\mathrm{d}}{=} A'\mathbf{Q} + B'$, *where* (A',B') *is independent of* $\mathbf{Q}$ *and has the distribution given by*

$$(A',B') \overset{\mathrm{d}}{=} \big(\alpha(B_*), \beta(B_*) + \zeta\big),$$

where $\zeta \sim \mathrm{Ge}(1/2)$ *and* B_* *are independent.* □

Corollary 5.9. *The limit of* $\mathbb{E}\, Q(F_n)$ *and* $\mathbb{E}\, Q(B_n)$ *is given by*

$$\mathbb{E}\,\mathbf{Q} = \mathbb{E}\, Q(B_\infty) = \frac{\mathbb{E}\, B'}{1-\mathbb{E}\, A'} = \frac{\mathbb{E}\,\beta(B_*)+1}{1-\mathbb{E}\,\alpha(B_*)}$$
$$= \frac{(2-\mathbb{E}\,\alpha(B_*))\,\mathbb{E}\, Q(B_*)+1}{1-\mathbb{E}\,\alpha(B_*)} = \mathbb{E}\, Q(B_*) + \frac{\mathbb{E}\, Q(B_*)+1}{1-\mathbb{E}\,\alpha(B_*)}.$$

Proof. Taking expectations in Theorem 5.8 we find $\mathbb{E}\,\mathbf{Q} = \mathbb{E}\,A'\,\mathbb{E}\,\mathbf{Q} + \mathbb{E}\,B'$, which yields the second inequality, and the third follows by the definitions of A' and B', since $\mathbb{E}\,\zeta = 1$. (Theorem 5.7 leads to the same result.)

Next, we argue as above for the finite random tree B_* too. In this case, the tree is empty and $Q(B_*) = 0$ with probability 1/2, and otherwise $Q(B_*) = \beta(L_*) + \alpha(L_*)Q(R_*)$, where L_* and R_* are independent with the same distribution as B_*. This can be written, in analogy with Theorem 5.7, $Q(B_*) \overset{\mathrm{d}}{=} A_0 Q(B_*) + B_0$, where (A_0, B_0) is independent of $Q(B_*)$ and has the distribution given by

$$(A_0, B_0) \overset{\mathrm{d}}{=} \begin{cases} (0,0), & \eta = 0, \\ \big(\alpha(B_*), \beta(B_*)\big), & \eta = 1, \end{cases}$$

where $\eta \sim \mathrm{Bi}(1, 1/2)$ as above is independent of B_*. Taking expectations we find

$$\mathbb{E}\,Q(B_*) = \frac{\mathbb{E}\,B_0}{1 - \mathbb{E}\,A_0} = \frac{\mathbb{E}\,\beta(B_*)}{2 - \mathbb{E}\,\alpha(B_*)}$$

or $\mathbb{E}\,\beta(B_*) = (2 - \mathbb{E}\,\alpha(B_*))\,\mathbb{E}\,Q(B_*)$, and the result follows. □

Note that the variables A, B, A' and B' are discrete and take only rational values; for example, A' takes the values $\{1/k\}_{k=2}^{\infty}$, while B and B' are unbounded. (We do not know whether the range B and B' is the set of all nonnegative rational numbers.) Since B_* and the auxiliary variables η and ζ only take countably many values, with explicitly given probabilities, the distributions of these variables, and in particular their expectations, can in principle be determined numerically with arbitrary accuracy. In practice, the slow convergence of $\mathbb{P}(|B_*| > n)$ to zero together with the exponential growth of the number of trees of a given size may make it difficult to attain high precision.

We have found, using `Maple`, the estimates $\mathbb{E}\,\alpha(B_*) = \mathbb{E}(1+N(B_*))^{-1}) \doteq 0.318$, $\mathbb{E}\,\beta(B_*) \doteq 2.9$, and $\mathbb{E}\,Q(B_*) \doteq 1.7$, which yields $\mathbb{E}\,\mathbf{Q} \doteq 5.7$; we have no sharp rigorous error bounds, however, so these values should not be taken as absolute truths.

The fixed point equations imply further some qualitative properties of $\mathbf{Q}$.

Theorem 5.10. *The limit random variable* $\mathbf{Q}$ *has a continuous distribution with support* $[3, \infty)$.

Remark 5.11. Although A and B (and A' and B') are discrete, $\mathbf{Q}$ is continuous. Indeed, this is very general, and the proof below uses only $A' \neq 0$ a.s.. However, we have not been able to resolve whether $\mathbf{Q}$ is absolutely continuous, although it seems very plausible. Note that singular distributions may occur in this type of fixed point equations. For example, $A = 1/3$ and $B \sim \mathrm{Bi}(1, 1/2)$ yields the Cantor measure (up to a scale factor).

Proof. Let $p(x) := \mathbb{P}(\mathbf{Q} = x)$ and suppose that $p(x) > 0$ for some x. Let $p_0 := \sup_x p(x) > 0$. It is easily seen that this supremum is attained, since

$\sum_x p(x) \le 1$, so we can choose x with $p(x) = p_0$. By Theorem 5.8,

$$p_0 = \mathbb{P}(A'\mathbf{Q} + B' = x) = \mathbb{E}\big(\mathbb{P}(\mathbf{Q} = (x - B')/A')\big) = \mathbb{E}\, p\big((x - B')/A'\big).$$

Since $p(y) \le p_0$, this is possible only if $p\big((x - B')/A'\big) = p_0$ for all values of A' and B', but this implies that $p(y) = p_0$ for infinitely many values of y, which contradicts $\sum_y p(y) \le 1$. Hence $p(x) = 0$ for every x, i.e., the distribution of $\mathbf{Q}$ is continuous.

Next, it is easily shown by (3.1) and (3.2) and induction that for any finite binary tree B,

$$W(B) \ge 2N(B) - 2.$$

Consequently,

$$B' \ge \beta(B_*) \ge \frac{3N(B_*)}{N(B_*) + 1}.$$

Hence, for any $\varepsilon > 0$, again using Theorem 5.8,

$$\begin{aligned}\mathbb{P}(\mathbf{Q} < 3 - \varepsilon) = \mathbb{P}(A'\mathbf{Q} + B' < 3 - \varepsilon) &\le \mathbb{P}\Big(\frac{\mathbf{Q} + 3N(B_*)}{N(B_*) + 1} < 3 - \varepsilon\Big) \\ &= \mathbb{P}\big(\mathbf{Q} < 3 - (N(B_*) + 1)\varepsilon\big) \le \mathbb{P}(\mathbf{Q} < 3 - 2\varepsilon).\end{aligned}$$

Evidently, this implies $\mathbb{P}(\mathbf{Q} < 3 - \varepsilon) = 0$ for every $\varepsilon > 0$, and thus $\mathbf{Q} \ge 3$ a.s..

Conversely, let E be the support of the distribution of $\mathbf{Q}$; by definition, E is closed. It follows from the fixed point equation that if $x \in E$ and $\mathbb{P}\big((A', B') = (a, b)\big) > 0$, then $ax + b \in E$. In particular, taking $B_* = \emptyset$ which yields $\alpha(B_*) = 1/2$ and $\beta(B_*) = 3/2$, we find

$$x \in E \implies (x + 3)/2 + n \in E \quad \text{for every integer } n \ge 0. \tag{5.9}$$

Starting with any $x \in E$, taking $n = 0$ and iterating (5.9), we find in the limit $3 \in E$. Taking $x = 3$ in (5.9), we find $3 + n \in E$ for every $n \ge 0$. Finally, again taking $n = 0$ in (5.9), we find by induction on k, that E contains every dyadic rational $3 + m2^{-k}$ with $m, k \ge 0$. Since E is closed, $E \supseteq [3, \infty)$. □

Remark 5.12. Although B_* is finite, its size has infinite expectation. Indeed, for every critical branching process, the expected size of each generation is the same, in this case 1/2; this follows also from the fact that each of the 2^k possible vertices at height k appears with probability 2^{-k-1}.

In B_∞, there is at height k one immortal and on the average 1/2 mortal in each of the k finite branches descending from the k immortals closer to the root. Hence the expected number of vertices at height k is $k/2 + 1$ and, by symmetry, each of the 2^k possible vertices appears with probability $(k+2)2^{-k-1}$. This illustrates that the infinite tree B_∞ is sparse and stringy.

As a further illustration, consider the intersection of two independent copies of B_∞; the expected size is $\sum_{k=0}^{\infty} 2^k (k + 2)^2 2^{-2k-2} = 11/2$. (This can also be seen by considering the two independent two-type branching processes generating the trees as a single branching process with 4 types representing the common vertices and the pairs of types there. We leave the details as an exercise.) Hence, two independent random large binary trees have on the average close to 5.5 vertices in common.

For the finite trees B_n, and more generally for any conditioned Galton–Watson trees with finite offspring variance, it is known that the bulk of the vertices have heights of the order $\sqrt{n}$; see e.g. [1], [2] and [14] for much more detailed results.

6. Back to the forest

The results on binary trees in Section 5 can be translated to results on forests by the correspondence discussed in Section 3, which extends to infinite forests and binary trees. Note that the number of trees in a forest equals the number of vertices in the rightmost branch of the corresponding binary tree. Again, we begin with some simple, more or less well-known observations.

We let F_* be the (finite) random forest corresponding to the random binary tree B_*. The construction of B_* in Section 5 shows that the number of vertices in the rightmost branch has the geometric distribution $\mathrm{Ge}(1/2)$. Consequently, the number of trees in F_* is $\mathrm{Ge}(1/2)$. Similarly, the number of children of any vertex is $\mathrm{Ge}(1/2)$, and all these numbers are independent. Consequently, F_* is a Galton–Watson forest obtained from a Galton–Watson process with $\mathrm{Ge}(1/2)$ initial individuals (roots) and offspring distribution $\mathrm{Ge}(1/2)$. Note that this, too, is a critical Galton–Watson process.

Equivalently, if T_* is the Galton–Watson tree with offspring distribution $\mathrm{Ge}(1/2)$, then T_* equals F_* with all components joined to a common added root; conversely, $F_* = T_*^*$.

It follows immediately that $|F_*| = |B_*| \overset{\mathrm{d}}{=} \xi$ and $|T_*| = |F_*| + 1 \overset{\mathrm{d}}{=} \xi + 1$, that F_* conditioned on $|F_*| = n$ has the distribution of F_n, and that T_* conditioned on $|T_*| = n$ has the distribution of T_n.

Similarly, let F_∞ be the random infinite forest corresponding to B_∞, and let T_∞ be the random infinite tree obtained by adding a root to F_∞; thus $F_\infty = T_\infty^*$. We can decompose the rightmost branch of B_∞ into the part belonging to the infinite path, which has $1 + \mathrm{Ge}(1/2)$ vertices, and the part after it, which is independent of the first part and has the same distribution as the rightmost branch in B_*, i.e. it has $\mathrm{Ge}(1/2)$ vertices. Hence, if $\widehat{\zeta}$ is the total number of vertices in the rightmost branch of B_∞, then $\widehat{\zeta} = 1 + \zeta + \zeta'$, where ζ and ζ' are independent and $\mathrm{Ge}(1/2)$. (In the equivalent branching process construction, the rightmost branch has $1+\zeta$ immortal and ζ' mortal vertices.) It follows that $\widehat{\zeta}$, which also is the number of components in F_∞ and the degree of the root in T_∞, has a shifted negative binomial distribution,

$$\mathbb{P}(\widehat{\zeta} = k) = k2^{-k-1}, \qquad k = 1, 2, \dots; \tag{6.1}$$

this is the size-biased distribution $\widehat{\mathrm{Ge}}(1/2)$.

Using the branching process construction of B_∞, exposing first the rightmost branch, then the rightmost branches in the left subtrees sprouting from it, and so on, it is now easily seen that T_∞ is the tree produced by the size-biased Galton–Watson process defined in Remark 5.4 with the offspring distribution $\mathrm{Ge}(1/2)$ for the mortals, and thus $\widehat{\mathrm{Ge}}(1/2)$ for the immortals. F_∞ is obtained by chopping off the root of T_∞, or by starting with $\widehat{\mathrm{Ge}}(1/2)$ individuals (roots), one of them immortal.

Note that F_∞ and T_∞ are locally finite and have exactly one infinite path (the immortals). The equation $\widehat{\zeta} = \zeta + 1 + \zeta'$ above shows that F_∞ and T_∞ also can be constructed by starting with an infinite path (the trunk) and adding to each vertex in it a $\mathrm{Ge}(1/2)$ number of branches to each side, each branch being an independent copy of T_*; for F_∞ we further add a $\mathrm{Ge}(1/2)$ number of copies of T_* on each side of the infinite component as separate components. (All random choices should be independent.)

It is easily seen that Lemma 5.2 implies the corresponding statements for forests and trees. (This is another instance of the general result given in Remark 5.4.) Note, however, that the truncation $F_\infty^{(M)}$ does not correspond to the truncation $B_\infty^{(M)}$; it corresponds to $B_\infty^{[M]}$, where we let $B^{[M]}$ denote the binary tree B with each branch truncated after M steps to the left. (Note that $B_\infty^{[M]}$ a.s. is a finite tree.) We give a formal statement.

Lemma 6.1. *As $n \to \infty$, $F_n \xrightarrow{\mathrm{d}} F_\infty$ and $T_n \xrightarrow{\mathrm{d}} T_\infty$ in the sense of finite-dimensional distributions, in the sense $F_n^{(M)} \xrightarrow{\mathrm{d}} F_\infty^{(M)}$ and $T_n^{(M)} \xrightarrow{\mathrm{d}} T_\infty^{(M)}$ for every finite M.*

Proof. Fix $M \ge 0$. Lemma 5.2 implies that for each fixed finite binary tree b, $\mathbb{P}(B_n^{[M]} = b) \to \mathbb{P}(B_\infty^{[M]} = b)$, and thus $B_n^{[M]} \xrightarrow{\mathrm{d}} B_\infty^{[M]}$. $\square$

Using the correspondence between forests and binary trees, we now define Q for infinite forests too; thus $Q(F_\infty) = Q(B_\infty)$, and the limit $\mathbf{Q}$ in Theorem 2.2 can be taken as $Q(F_\infty)$. The following theorem shows that Q can be defined (a.s.) directly on infinite forests without our use of binary forests as a convenient technical tool.

Theorem 6.2. *There exists an extension of Q to infinite forests such that we have $\mathbb{E}\,|Q(F_\infty^{(M)}) - Q(F_\infty)| \to 0$ as $M \to \infty$. We have $Q(F_n) \xrightarrow{\mathrm{d}} Q(F_\infty)$ as $n \to \infty$. Furthermore, $Q(F_\infty) = Q(B_\infty)$ when B_∞ corresponds to F_∞.*

Proof. It remains only to prove that $\mathbb{E}\,|Q(F_\infty^{(M)}) - Q(F_\infty)| \to 0$, or equivalently, transfering to binary trees again, that $\mathbb{E}\,|Q(B_\infty^{[M]}) - Q(B_\infty)| \to 0$.

Let, for $M \ge 1$,

$$\delta^{[M]} := \sup_n \mathbb{E}\,|Q(B_n^{[M]}) - Q(B_n)|.$$

The proof of Lemma 5.6 shows with minor modifications that $\delta^{[M]} \to 0$; note that $B_n^{[M+1]}$ has the subtrees $L_n^{[M]}$ and $R_n^{[M+1]}$, but this causes no significant problem. Moreover, we now need a preliminary step to ensure that $\delta^{[M]} < \infty$; this is easily done using induction, since (3.3) implies $Q(B) < Q(L) + 2 + \frac{1}{2}Q(R)$, and we omit the details.

For every M we have as $n \to \infty$, see the proof of Lemma 6.1, $B_n^{[M]} \xrightarrow{\mathrm{d}} B_\infty^{[M]}$. Moreover, since $B^{(M)}$ is a truncation of $B^{[M]}$, we have joint convergence of $(B_n^{[M]}, B_n^{(M)})$ to $(B_\infty^{[M]}, B_\infty^{(M)})$, and consequently

$$Q(B_n^{[M]}) - Q(B_n^{(M)}) \xrightarrow{\mathrm{d}} Q(B_\infty^{[M]}) - Q(B_\infty^{(M)}).$$

Since $\mathbb{E}\,|Q(B_n^{[M]}) - Q(B_n^{(M)})| \le \delta^{[M]} + \delta^{(M)}$ for each n, Fatou's lemma yields

$$\mathbb{E}\,|Q(B_\infty^{[M]}) - Q(B_\infty^{(M)})| \le \delta^{[M]} + \delta^{(M)},$$

which tends to 0 as $M \to \infty$ by Lemma 5.6 and the claim above. Finally, the triangle inequality and Theorem 5.5 yields

$$\mathbb{E}\,|Q(F_\infty^{(M)}) - Q(F_\infty)| = \mathbb{E}\,|Q(B_\infty^{[M]}) - Q(B_\infty)| \to 0, \qquad M \to \infty. \qquad \square$$

Remark 6.3. We saw above that the number of components of the infinite random forest F_∞ has the shifted negative binomial distribution in (6.1); hence, by Lemma 6.1, the number of components of the random forest F_n has asymptotically this distribution. It is easy to find the exact distribution for finite n as follows. The generating function for ordered trees is $zB(z)$, with $B(z)$ given in (3.7), and thus the generating function for ordered forests with k components is $z^k B(z)^k$. It follows as an easy exercise, using e.g. [6, (5.70)], that, with $n^{\underline{k}}$ denoting the falling factorial,

$$\mathbb{P}(F_n \text{ has } k \text{ components}) = \frac{\frac{k}{2n-k}\binom{2n-k}{n}}{C_n} = k\frac{(n+1)^{\underline{k+1}}}{(2n)^{\underline{k+1}}}.$$

This evidently converges to $k2^{-k-1}$ as $n \to \infty$, as asserted above.

Remark 6.4. There is a well-known correspondence between (random) trees and (random) walks on the non-negative integers by means of the depth first walk, see e.g. [1]. In this context, several nice results are known for the random trees studied here.

The random tree T_n corresponds to a simple random walk of length $2n$ conditioned on returning to 0 at the end but not before (sometimes called Dyck paths). The random tree T_* corresponds to a simple random walk stopped at its first return to 0 [12]. For the infinite tree T_∞, the depth first walk only captures the structure on one side of the infinite trunk; the other side is described by a depth first walk running in the opposite direction. The two depth first walks are independent, and each is a biased random walk which is a discrete version of the three-dimensional Bessel process, see Le Gall [12].

References

[1] D. Aldous (1991), The continuum random tree II: an overview. In *Proc. Durham Symp. Stoch. Analysis 1990*, Cambridge Univ. Press, 23–70.
[2] D. Aldous (1993), The continuum random tree III. *Ann. Probab.* **21**, no. 1, 248–289.
[3] K.B. Athreya & P.E. Ney (1972), *Branching Processes.* Springer-Verlag, Berlin, 1972.
[4] P. Billingsley (1968), *Convergence of Probability Measures.* Wiley, New York.
[5] T.E. Harris (1951), Some mathematical models for branching processes. In *Second Berkeley Symposium on Mathematical Statistics and Probability*, Univ. California Press, Berkeley, 305–328.
[6] R.L. Graham, D.E. Knuth & O. Patashnik (1994), *Concrete Mathematics.* 2nd ed., Addison–Wesley, Reading, Mass.
[7] D.P. Kennedy (1975), The Galton–Watson process conditioned on the total progeny. *J. Appl. Probab.* **12**, 800–806.
[8] M. Klazar (1997), Twelve countings with rooted plane trees. *European J. Combin.* **18**, no. 2, 195–210. Addendum: *European J. Combin.* **18**, no. 6, 739–740.
[9] D.E. Knuth (1997), *The Art of Computer Programming. Vol. 1: Fundamental algorithms.* 3rd ed., Addison–Wesley, Reading, Mass.
[10] D.E. Knuth, `KODA-RUSKEY`. CWEB program. Available from `http://Sunburn.Stanford.EDU/~knuth/programs/koda-ruskey.w`
[11] Y. Koda & F. Ruskey (1993), A Gray code for the ideals of a forest poset. *J. Algorithms* **15**, no. 2, 324–340.
[12] J.-F. Le Gall (1986), Une approche élémentaire des théorèmes de décomposition de Williams. In *Séminaire de Probabilités, XX, 1984/85*, Lecture Notes in Math., 1204, Springer, Berlin, 447–464.
[13] R. Lyons, R. Pemantle & Y. Peres (1995), Conceptual proofs of $L \log L$ criteria for mean behavior of branching processes. *Ann. Probab.* **23**, no. 3, 1125–1138.
[14] J.-F. Marckert & A. Mokkadem (2001), The depth first processes of Galton–Watson trees converge to the same Brownian excursion. To appear in *Ann. Probab.*
[15] S.T. Rachev (1991), *Probability metrics and the stability of stochastic models.* Wiley, Chichester, U.K.
[16] S.T. Rachev & L. Rüschendorf (1995), Probability metrics and recursive algorithms, *Adv. Appl. Probab.* **27**, 770–799.
[17] U. Rösler(1991), A limit theorem for "Quicksort". *RAIRO Inform. Théor. Appl.* **25**, no. 1, 85–100.
[18] U. Rösler(1992), A fixed point theorem for distributions. *Stochastic Process. Appl.* **42**, no. 2, 195–214.
[19] U. Rösler & L. Rüschendorf (2001), The contraction method for recursive algorithms. *Algorithmica* **29**, 3-33.
[20] F. Ruskey (1981), Listing and counting subtrees of a tree. *SIAM J. Comput.* **10**, no. 1, 141–150.

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/`