The Annals of Probability
1991, Vol. 19, No. 1, 1–28

THE CONTINUUM RANDOM TREE. I[1]

By David Aldous

University of California, Berkeley

Exact and asymptotic results for the uniform random labelled tree on n vertices have been studied extensively by combinatorialists. Here we treat asymptotics from a modern stochastic process viewpoint. There are three limit processes. One is an infinite discrete tree. The other two are most naturally represented as continuous two-dimensional fractal tree-like subsets of the infinite-dimensional space l_1. One is compact; the other is unbounded and self-similar.

The proofs are based upon a simple algorithm for generating the finite random tree and upon weak convergence arguments. Distributional properties of these limit processes will be discussed in a sequel.

1. Introduction. There are many ways to construct different models of random trees with n vertices. One way is via the simple Galton–Watson branching process. Given an offspring distribution θ, start with 1 individual in generation 0, and write V for the total population size. Then the "family tree" of this branching process can be considered as a random rooted unlabelled tree $\mathscr{B}$, and conditioning on $V = n$ gives a tree on n vertices.

Another way is combinatorial. Given a set of trees on n vertices and, implicitly, a notion of when two trees are to be considered the same, we can consider the uniform random tree in this set. Obviously, any question about an exact probability involving such random trees is an enumeration question; a now-classical reference for enumeration problems for trees is [10].

Our object of study is the *uniform random labelled tree* $\mathscr{T}_n$, that is the tree picked uniformly from the n^{n-2} trees on vertices $1, \dots, n$. By choosing vertex 1 as the root and removing labels, we shall often regard $\mathscr{T}_n$ as a (nonuniform) random unlabelled rooted tree. It is easy to verify this has the same distribution as the tree produced from the branching process with Poisson(1) offspring, conditioned on $V = n$. Section 2 gives more information on the formalities.

This is not an isolated fact: It turns out that there is a certain class of random tree models which can be described in both ways. Combinatorialists have defined families (such as "simply generated trees" [9]) of rooted trees for which there are nice generating function identities relating the numbers of trees of different sizes. Asymptotics for uniform random trees from such families are treated in, e.g., [9, 5]. These turn out to be essentially just the random trees which can be obtained from the branching process construction. This connection has been described, in some settings, by the Russian school of discrete probabilists: Chapter 2 of Kolchin [7] presents their results. Both

Received June 1989; revised January 1990.
[1]Research supported by NSF Grant MCS-87-01426.
AMS 1980 *subject classifications*. 60C05, 05C80.
Key words and phrases. Random tree, random fractal, critical branching process.

groups have studied asymptotic questions using different language but essentially similar generating function arguments. These techniques are especially suitable for "height profile" questions, that is the study of rescaled limits of $(X^n_m: 0 \leq m < \infty)$, where

(1) $$X^n_m = \text{number of vertices at distance } m \text{ from the root.}$$

In this paper we take a completely different approach. It is shown in [2] that the following algorithm produces a random rooted unlabelled tree which is distributed as $\mathscr{T}_n$ (regarded as a rooted unlabelled tree).

ALGORITHM 1. Fix $n \geq 2$.
Take a root vertex 1.
For $2 \leq i \leq n$, connect vertex i to vertex $V_i = \min(U_i, i-1)$, where $U_2, \dots, U_n$ are independent and uniform on $1, \dots, n$.
Remove the labels.

There are several standard ways of listing labelled trees, e.g., via Prufer code, which can be used to generate $\mathscr{T}_n$. It is not known whether Algorithm 1 can be derived from any known enumeration scheme (it certainly has not been used explicitly for the study of asymptotics). It is worth pointing out that there is no known natural way to construct $\mathscr{T}_{n+1}$ by adding a random edge to $\mathscr{T}_n$. Indeed, trees grown by simple Markovian models of adding edges tend to be quite different from our trees (e.g., having diameter of order $\log n$ instead of $n^{1/2}$). The advantage of Algorithm 1 is that it makes the asymptotic behavior of these random trees intuitively rather clear. Moreover, it leads to descriptions of an entire infinite-dimensional limit process, rather than just the one-dimensional projection processes implicit in previous studies of limits of (1).

We now describe informally three processes. The theme of the paper is that these are the limits of $\mathscr{T}_n$ as $n \to \infty$; the first when we do not rescale space, the third when we rescale distance by a factor $n^{-1/2}$, and the second when we rescale distance by a factor $n^{-\alpha}$, $0 < \alpha < \frac{1}{2}$. Process 1, and convergence to that limit, has been discussed by Grimmett [6] (see Section 3).

PROCESS 1. For each $k = 0, 1, 2, \dots$, create independent Poisson(1) branching processes, regarded as trees with root i_k and other vertices unlabelled. Then connect $i_0, i_1, i_2, \dots$ as a path, deem i_0 the root and delete labels.

PROCESS 2. At time 0 there is an infinite continuous line $[0, \infty)$. At time $0 < t < \infty$ there is a tree composed of the original line and finite line segments attached to each other, only a finite number of such segments connecting with each finite interval of the original line. The process grows according to the following rules:

(i) in each time increment $(t, t + dt)$, in each segment $(x, x + dx)$ of the tree constructed at time t, there is chance $dt\,dx$ of a "birth";

(ii) if a birth occurs at time t and place x, then a new branch with random exponential (rate t) length is instantly attached at x.

The process is the closure of the tree at time infinity.

This description is incomplete, because we did not specify the orientations of branches. To draw the process on a piece of paper it is natural to make each branch horizontal or vertical. Unfortunately, as the figure later shows, this makes different branches intersect, thus destroying the "tree" property. For the formal description of Processes 2 and 3 we take each branch to be orthogonal to all the previous branches, in some infinite-dimensional space. The space l_1 turns out to be natural, and *closure* refers to that space.

PROCESS 3. Take a half-line $[0, \infty)$, but now we cut and paste. Let $C_1, C_2, \dots$ be the times of a nonhomogeneous Poisson process of rate $r(t) = t$. Cut the half-line into intervals $[C_i, C_{i+1})$. Starting with the line segment $[0, C_1)$, grow a tree inductively by adding $[C_i, C_{i+1})$ as a branch attached to a random point B_i, chosen uniformly over the existing tree. The process is the closure of the tree at time infinity.

There is also a different description of Process 2 as a σ-finite process of (mostly small) rescaled copies of Process 3 attached to an infinite baseline.

It is perhaps easiest to understand these limits by analogy with a more familiar object. Consider for each n the combinatorial set of sequences $\mathbf{x} = (x_1, \dots, x_n)$ with $x_i \in \{-1, 1\}$ and such that $\sum_1^m x_i \geq 0$ for all $m \leq n$. Consider the uniform distribution on this set. A probabilist would regard this as the process $(S_m;\ 0 \leq m \leq n)$ which is "simple random walk conditioned to be nonnegative." As $n \to \infty$ there are three limit processes, the first when we do not rescale space, the third when we rescale distance by a factor $n^{-1/2}$, and the second when we rescale distance by a factor $n^{-\alpha}$, $0 < \alpha < 1/2$.

1. $(S_m;\ 0 \leq m < \infty)$, "random walk conditioned to be nonnegative for all time," a certain birth-and-death process.
2. The Bessel(3) process on the time interval $[0, \infty)$; that is, Brownian motion conditioned to stay positive for all time.
3. The "Brownian meander" $(W_s;\ 0 \leq s \leq 1)$, a nonhomogeneous diffusion which is Brownian motion conditioned to be nonnegative during $[0, 1]$.

The analogy is sharp in many ways. For instance, each Process 2 is self-similar and can be obtained from Process 1 as the rescaled limit at infinity, and from Process 3 as the rescaled limit at 0. Another example of this "threefold way" of taking limits, in the context of a queueing/storage model, is in [1].

This paper is devoted to formalizing the descriptions of these processes and to establishing that they are indeed limits of the uniform random finite trees $\mathscr{T}_n$. The author's natural inclination would be to give a few paragraphs of prose and assert that the mathematical details are routine. Instead, arguments are written out at a moderate level of detail. Except for a few technicalities, we do *not* use the connection with critical branching processes in this paper.

There are several directions in which this work may be extended and related to existing results. Let us briefly mention a few, which may be treated in detail elsewhere.

(a) *Distributional properties of the limit trees*. Combinatorialists have studied asymptotic distributions associated with uniform random trees, using enumeration and generating function methods. Probabilists have studied conditioned branching processes and the limiting diffusion-type process obtained in (1), which turns out to be a time-change of Brownian excursion. Our explicit constructions give yet another technique for studying explicit distributions—it is interesting to compare these methods.

(b) *Robustness*. Conditioned critical Galton–Watson processes with more general distributions lead to the same continuum trees: the offspring s.d. acts as a scale parameter. More generally, we believe that any natural model of n-vertex random trees which has height of order $n^{1/2}$ will rescale to Process 2.

(c) *Superprocesses*. Le Gall [8] gives a construction of "the superprocess" (measure-valued branching process) based upon a certain tree constructed out of Brownian excursions. One may similarly build variants of superprocesses from our Processes 2 and 3 and interpret them as "the superprocess, conditioned on a given total population size," and as "the superprocess, conditioned on nonextinction," respectively. This may be a productive way of studying superprocesses.

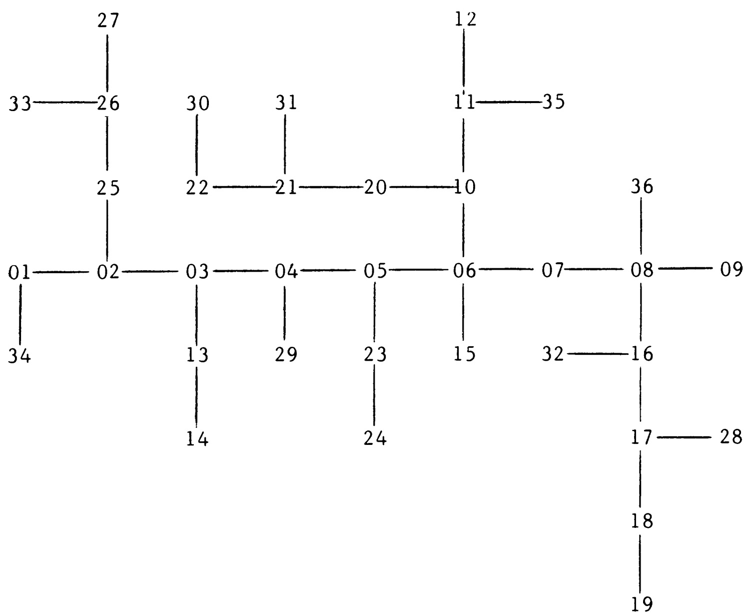

FIG. 1.

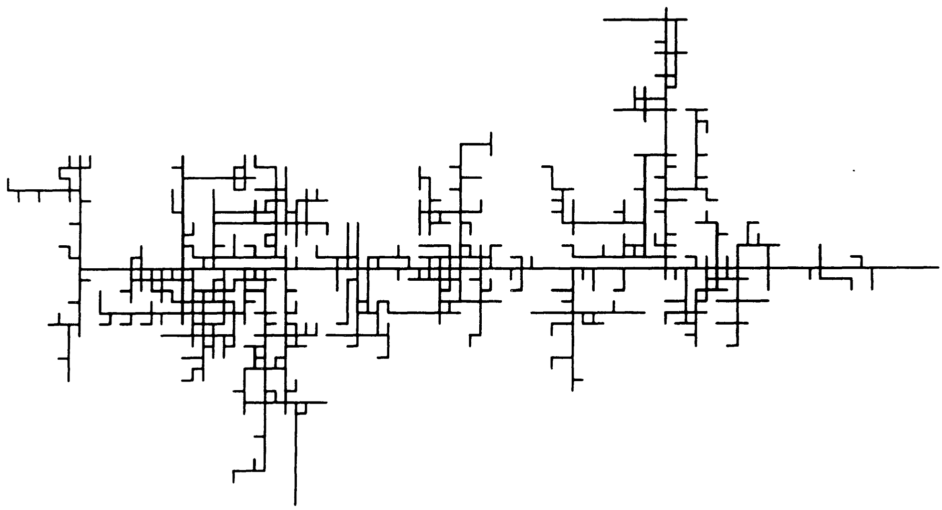

FIG. 2.

Figures 1 and 2 show trees drawn by Algorithm 1 for $n = 36$ and $n = 1000$. Thus, we may regard the second picture as an approximation to Process 2. Of course, the pictures differ from the formal construction specified later, in that successive branches are drawn horizontally and vertically in two dimensions, instead of orthogonally in infinite dimensions. The "loops" in the $n = 1000$ figure are artifacts of this two-dimensional drawing.

2. Labels, roots, embeddings and convergence. Since the paper is addressed to probabilists, let us explain some elementary combinatorial ideas. Initially, a tree is an "abstract tree:" a set of vertices, with an edge being formally an unordered pair of vertices. For a labelled tree with n vertices, the vertices are labelled by distinct integers $1, 2, \dots, n$. Labelled trees t and t^* are isomorphic ("the same") if, for each pair (i, j) of labels, (i, j) is an edge of t iff it is an edge of t^*. Cayley's formula says there are n^{n-2} labelled trees (formally, nonisomorphic labelled trees) on $n \geq 2$ vertices. Two *unlabelled* trees on n vertices are isomorphic if there exist labellings making them isomorphic as labelled trees. A *rooted* tree is merely a tree with one vertex distinguished and called the root. Isomorphisms between rooted trees are required to preserve the root. If we insist that the root be vertex 1, then there are n^{n-2} rooted labelled trees on n vertices; if not, there are n^{n-1}.

Given a rooted unlabelled tree t on n vertices, let $L(t)$ be the number of distinct ways to label t. Then $L(t) \leq n!$, but may be strictly less because superficially different labellings may produce isomorphic labelled trees: loosely, $L(t)$ is smaller for symmetric trees than for asymmetric trees.

$\mathscr{T}_n$, the uniform random labelled tree, can be regarded as a rooted unlabelled tree by making 1 the root (or equivalently, picking vertex i or a uniform random vertex as the root) and removing labels. Then, for rooted unlabelled trees t,

(2) $$P(\mathscr{T}_n = t) \text{ is proportional to } L(t),$$

and so $\mathscr{T}_n$ is not uniform on unlabelled trees. On the other hand, an unlabelled tree may be "randomly labelled" (precisely, take an arbitrary labelling and then apply a uniform random permutation to the labels). Starting with $\mathscr{T}_n$, if we first unlabel and then randomly relabel, we get back to a random labelled tree distributed as $\mathscr{T}_n$. In this sense, there is no loss of structure in unlabelling $\mathscr{T}_n$, since if we wish, we can get back the labelled tree by randomly relabelling.

For vertices v, w in a tree, the *distance* $d(v, w)$ between v and w is the number of edges along the unique path connecting them. Write $|t|$ for the number of vertices in a tree t. The *height* of t is

$$h(t) = \max_v d(\text{root}, v).$$

Now consider the Poisson(1) branching process as a random tree $\mathscr{P}$. We quote some standard asymptotics for its height and size (see, e.g., [8]):

(3) $$P(h(\mathscr{P}) > h) \sim 2/h \quad \text{as } h \to \infty;$$

(4) $$P(|\mathscr{P}| = n) \sim (2\pi)^{-1/2} n^{-3/2} \quad \text{as } n \to \infty.$$

Now make $\mathscr{P}$ a labelled tree by randomly labelling vertices $1, \ldots, |\mathscr{P}|$. It is easy to check that, for each finite rooted labelled tree t with $|t|$ vertices,

$$P(\mathscr{P} = t) = e^{-|t|}/|t|!.$$

Thus, conditional on $|\mathscr{P}| = n$, we see that $\mathscr{P}$ is uniform on labelled trees with n vertices. If instead we regard $\mathscr{P}$ as a rooted unlabelled tree, then

(5) $$P(\mathscr{P} = t) = e^{-|t|} L(t)/|t|!.$$

The limit theorems giving Processes 2 and 3 involve rescaling trees, which of course does not make sense for abstract trees. Instead, we embed trees into a linear space. A natural space is l_1, the Banach space of sequences $x = (x_1, x_2, \ldots)$ such that $\|x\| = \sum |x_i| < \infty$. Let $z_i = (0, \ldots, 0, 1, 0, 0, \ldots)$ be the ith unit vector in the natural basis of l_1. Call the directions of the z_i *orthogonal*, although of course we are not in Hilbert space. A *set-representation* of a tree t is a labelling of the vertices as a subset $S = \{v, w, \ldots\} \subset l_1$ in such a way that $d(v, w) \equiv \|v - w\|$. To see that such a representation exists, pick one vertex of t and label it 0. Order the edges arbitrarily as $e_1, \ldots, e_{n-1}$. Label each vertex v as

(6) $$v = \sum_i z_i J(v, e_i),$$

where $J(v, e) = 1$ if e is in the path from v to 0, and $J(v, e) = 0$ if not. Then

$$\|v - w\| = \sum_i |J(v, e_i) - J(w, e_i)|,$$

and it is easy to see this agrees with the graph-distance $d(v, w)$. Note that in general we cannot represent vertices in *Euclidean* space in this distance-preserving way.

Thus a set-representation of t is a subset $S = \{v_1, v_2, \dots, v_n\}$ of l_1. Clearly, S determines t as an unlabelled tree, whereas given t there will be many different set-representations S (see *equivalent* below.) A *measure-representation* μ of t is a probability measure on l_1 which is uniform on some set-representation S of t.

Write σ_a for scalar multiplication in l_1, $\sigma_a(x) = ax$. Then σ_a acts on sets and probability measures in the natural way.

We now turn to notions of convergence. For closed subsets of l_1, *convergence* shall mean convergence in the Hausdorff metric. For probability measures on l_1, *convergence* shall mean weak convergence. We are concerned with convergence of rescaled trees. In the deterministic setting, consider a sequence t_n of trees on n vertices with set-representations S_n and corresponding measure-representations μ_n. Consider the assertions

$$\sigma_{n^{-1/2}}(S_n) \to S, \quad \text{say,} \tag{7}$$

$$\sigma_{n^{-1/2}}(\mu_n) \to \mu, \quad \text{say.} \tag{8}$$

Neither implies the other. [Because adding a path of length $n^{3/4}$ would affect (7) but not (8); conversely, adding $n/2$ leaves to a single vertex would affect (8) but not (7).] But informally, one might hope that when one holds so does the other and that S is the support of μ; and we shall see that this happens in our setting.

For the random trees $\mathscr{T}_n$ there exist random set-representations S_n and corresponding random measure-representations μ_n, and so we can talk about convergence in distribution,

$$\sigma_{n^{-1/2}}(S_n) \to_d S, \quad \text{say,}$$

$$\sigma_{n^{-1/2}}(\mu_n) \to_d \mu, \quad \text{say,}$$

where now the limits are random sets and random measures. All our spaces are Polish, so in place of "convergence in distribution" we could say "there exist versions which converge a.s."

Clearly, a naive representation such as (6) will not yield convergence. The point of this paper is that Algorithm 1 leads to a representation which does.

Call subsets K_1, K_2 of l_1 containing 0 *equivalent*, and write $K_1 \approx K_2$, if there exists a (nonlinear) isometry $K_1 \leftrightarrow K_2$ which fixes 0. If K_1, K_2 are set-representations of finite rooted unlabelled trees, this is exactly the condition for the trees to be isomorphic. So "equivalence" is intended to convey the idea of "isomorphism" for continuum trees. Call random subsets $\mathscr{R}_1$, $\mathscr{R}_2$ *equivalent in distribution* ($\approx_d$) if there exist $\mathscr{R}_1^*, \mathscr{R}_2^*$ such that $\mathscr{R}_1^* =_d \mathscr{R}_2^*$ and $\mathscr{R}_i^*(\omega) \approx \mathscr{R}_i(\omega)$ a.s. for $i = 1, 2$. Write $\mathscr{R}_m \sim_d \mathscr{R}_\infty$ if there exist $\mathscr{R}_m^* \to_d \mathscr{R}_\infty^*$ such that $\mathscr{R}_m^*(\omega) \approx \mathscr{R}_m(\omega)$ a.s. for $m = 1, \dots, \infty$. Make analogous definitions for measure-representations. The point is that we are typically interested only in functionals of finite trees which are invariant under isomorphism; so we should only be interested in properties of continuum trees which are invariant under equivalence. These are "intrinsic" properties, rather than artifacts of our construction.

3. The discrete infinite tree. Here we shall formalize the idea that, as $n \to \infty$, the uniform random tree $\mathscr{T}_n$ looks locally like the infinite tree of Process 1. This idea has been discussed by Grimmett [6], using a superficially different description of Process 1. He proves the convergence result (our Theorem 2) for *fixed* k and observes that this "local" limit result can be used to derive asymptotics of quantities associated with local properties of $\mathscr{T}_n$, these results having previously been proved combinatorially. For our purposes it is important that Theorem 2 holds for $k = o(n^{1/2})$. This implies, for example, that Process 2 can be regarded as a rescaled limit of Process 1 as well as of $\mathscr{T}_n$ [see Theorem 11(ii)].

The key idea is that, for $k = o(n^{1/2})$, there is a unique path from the root of length k which can be extended to length $\Omega(n^{1/2})$. Theorem 2 concerns the subtrees $\mathscr{T}_{n,k} \subseteq \mathscr{T}_n$ which branch off from this segment. The standard combinatorial arguments around (11) are similar to those in [6].

Let T be the set of rooted unlabelled trees. Let T^k be the product set $T \times T \times \cdots \times T$. An element $\mathbf{t} = (t_0, \dots, t_{k-1})$ of T^k may be regarded as an unlabelled rooted tree with a distinguished path of length $k - 1$ attached to the root. To be precise, suppose we are given a sequence $(t_0, \dots, t_{k-1})$ of rooted unlabelled trees. Temporarily label the roots $(v_0, \dots, v_{k-1})$. Join these trees into one tree $\mathbf{t}$ by adding a distinguished path $v_0, v_1, v_2, \dots, v_{k-1}$. Then deem v_0 the root and remove labels.

Let $\mathscr{P}$ be the Poisson(1) branching process, considered as a random element of T. Let $\mathscr{P}^k = (\mathscr{P}_1, \mathscr{P}_2, \dots, \mathscr{P}_k)$ be a vector of i.i.d. copies of $\mathscr{P}$, considered as a random element of T^k. Informally, Process 1 is $\mathscr{P}^\infty$; for the formalities, let us consider only finite trees.

Now fix $k \geq 1$ and $t \in T$. We shall define, under certain side conditions, a decomposition

$$t = (\mathbf{t}, t^*), \qquad \mathbf{t} \in T^k, t^* \in T. \tag{9}$$

The conditions are as follows:

1. t has height $h(t) \geq k$.
2. Each path of length $h(t)$ from the root v_0 has the same initial segment of length k, say $(v_0, v_1, \dots, v_k)$.

Then we can define $\mathbf{t}$ to be the subtree of vertices v such that the path from v to the root does not meet v_k and define t^* to be the remaining vertices. Then regard t^* as a tree rooted at v_k and $\mathbf{t}$ as a rooted tree with distinguished path $(v_0, \dots, v_{k-1})$. Applying this decomposition to the uniform random tree $\mathscr{T}_n$ gives (as first component) a random element $\mathscr{T}_{n,k}$ of T^k, which is undefined when conditions 1 and 2 are not met.

THEOREM 2. *If $n \to \infty$ and $k = k(n) = o(n^{1/2})$,*

$$\sup_{A \subset T^k} \left| P(\mathscr{T}_{n,k} \in A) - P(\mathscr{P}^k \in A) \right| \to 0.$$

In words, the total variation distance between the distributions of $\mathscr{T}_{n,k}$ and $\mathscr{P}^k$ converges to 0.

PROOF. Fix $n > k$ and $\mathbf{t} \in T^k$. Using (5), we have

$$P(\mathscr{P}^k = \mathbf{t}) = e^{-|\mathbf{t}|} \prod_{i=0}^{k-1} \frac{L(t_i)}{|t_i|!}. \tag{10}$$

We shall show that, for $|\mathbf{t}| \le n$,

$$\begin{aligned} n^{n-1} P(\mathscr{T}_{n,k} = \mathbf{t}) = \binom{n}{|t_0| \cdots |t_{k-1}|} (n - |\mathbf{t}|)^{n-|\mathbf{t}|-1} \\ \times q(n - |\mathbf{t}|, h(\mathbf{t}) - k) \prod_{i=0}^{k-1} L(t_i), \end{aligned} \tag{11}$$

where q is defined in (12).

There are n^{n-1} labelled rooted trees t on n vertices; we count how many of them have a decomposition of the form $t = (\mathbf{t}, t^*)$ for fixed $\mathbf{t}$. Plainly, we need $|t^*| = n - |\mathbf{t}|$. To label $\mathbf{t}$, we first label the distinguished roots and choose the set of labels for each tree t_i and then assign labels to vertices in t_i; the number of ways to do this is

$$\binom{n}{|t_0| \cdots |t_{k-1}|} \prod_{i=0}^{k-1} L(t_i).$$

Next, the number of rooted labelled trees t^* is

$$(n - |\mathbf{t}|)^{n-|\mathbf{t}|-1}.$$

Finally, we must ensure that $t = (\mathbf{t}, t^*)$ satisfies conditions 1 and 2. This happens iff $h(t^*) + k > h(\mathbf{t})$. Thus, we define

$$q(m, j) = P(h(\mathscr{T}_m) > j) \tag{12}$$

and then the proportion of labelled trees t^* which satisfy the condition is $q(n - |\mathbf{t}|, h(\mathbf{t}) - k)$.

Putting together these equalities gives (11).

Dividing (10) by (11) and simplifying,

$$\begin{aligned} &\frac{P(\mathscr{P}^k = \mathbf{t}) q(n - v, h(\mathbf{t}) - k)}{P(\mathscr{T}_{n,k} = \mathbf{t})} \\ &= e^{-v} \frac{(n - v)!}{n!} \left(1 - \frac{v}{n}\right)^{1-n} (n - v)^v, \end{aligned} \tag{13}$$

where $v = |\mathbf{t}|$.

Using Stirling's formula, the right-hand side tends to 1 as $n \to \infty$, $v = o(n)$. Now, to show that the total variation distance between measures Q_n, R_n tends to 0, it suffices to show that

$$Q_n\left\{x: \left|\frac{Q_n(x)}{R_n(x)} - 1\right| > \varepsilon\right\} \to 0$$

for each $\varepsilon > 0$ (provided Q_n is a probability measure). Thus, to prove the theorem, it now suffices to show

(i) $|\mathscr{P}^k| = o(n)$ in probability,
(ii) $q(n - |\mathscr{P}^k|, h(\mathscr{P}^k)) \to 1$ in probability.

But a simple fact about the random labelled tree (which follows from Algorithm 1 by considering the first i such that $V_i < i - 1$) is

$$q(m, j) \to 1 \quad \text{as } m \to \infty,\ j = o(m^{1/2}).$$

Thus (ii) will follow from (i) and

(iii) $h(\mathscr{P}^k) = o(n^{1/2})$ in probability.

But (3) and (4) imply

$$h(\mathscr{P}^k) = O(k) \quad \text{and} \quad |\mathscr{P}^k| = O(k^2) \quad \text{in probability,}$$

and so (i) and (iii) follow from the hypothesis $k = o(n^{1/2})$. □

4. The compact continuum tree. Recall the verbal description of Process 3 in the introduction. By specifying that the successive branches should be parallel to the successive unit vectors z_i in the natural basis of l_1, we obtain a rigorous construction.

Let $C_0 = B_0 = 0$. Let $(C_1, C_2, \dots)$ be the times of events in a Poisson process of rate $r(t) = t$. Let $B_i = \xi_i C_i$, where (ξ_i) are i.i.d. uniform on $(0, 1)$, independent of the C's. These are the "random" ingredients; now regard ω as fixed and proceed deterministically. Define ρ: $[0, \infty) \to l_1$ by $\rho(0) = 0$,

$$\rho(x) = \rho(B_i) + (x - C_i) z_{i+1} \quad \text{on } C_i < x \le C_{i+1},\ i \ge 0.$$

Write

$$\mathscr{S}_t = \rho([0, t]),$$
$$\mathscr{S}_\infty = \rho([0, \infty)),$$
$$\mathscr{S} = \overline{\mathscr{S}_\infty}.$$

The fundamental result is that $\mathscr{S}$ is a.s. compact. This immediately implies that $\mathscr{S}_t \to \mathscr{S}$ a.s. in the Hausdorff metric. Think of $\mathscr{S}$ as the set-representation of some (Platonic) continuum random tree. More information is contained in its measure-representation μ specified by the following theorem. The uniform probability distribution on $[0, t]$ induces a distribution on $\mathscr{S}_t$ and hence on l_1; call this random probability measure μ_t.

THEOREM 3. *With probability* 1,

(i) $\mathscr{S}$ *is compact*,
(ii) $\mu_t \to \mu$ *as* $t \to \infty$, *for a certain* μ,
(iii) $\mathscr{S}$ *is the support of* μ,
(iv) $\mu(\mathscr{S}_\infty) = 0$.

It is easy to see that $\mathscr{S}$ is "tree-like" in several senses. Call $\mathscr{S}_\infty \setminus \bigcup_{i=0}^{\infty}\{\rho(C_i)\}$ the *skeleton* of $\mathscr{S}$, and call the remainder of $\mathscr{S}$ the *leaves*. For each pair of points x, y in $\mathscr{S}$ there is a unique path (continuous, non-self-intersecting) in $\mathscr{S}$ from x to y, and this path lies within the skeleton except at x and y. Removing a point x in the skeleton will disconnect $\mathscr{S}$, whereas removing a leaf x will not.

The rest of this section contains the Proof of Theorem 3. The hardest part is compactness, which certainly follows from the more quantitative result stated next. For $A \subset l_1$ let $N(A, \delta)$ be the cardinality of the smallest set (x_i) such that

$$\sup_{x \in A} \min_i \|x - x_i\| \le \delta.$$

PROPOSITION 4. *There exists* $K < \infty$ *such that, a.s.*,

$$N(\mathscr{S}, \delta) \le K\delta^{-2}\log^2(1/\delta) \quad \textit{for all sufficiently small } \delta.$$

We conjecture that a.s.

$$N(S, \delta) \sim C\delta^{-2}\log(1/\delta),$$

for some constant C. It is easy to see (Lemma 7) that $N(\mathscr{S}, \delta)$ is at least order δ^{-2}. Thus $\mathscr{S}$ has "dimension 2", in the sense of metric entropy. Later results [Theorem 11(iii) and (53)] imply that it is also two-dimensional in the sense of "measures of small balls," i.e., $\mu\{x\colon \|x\| \le \delta\}$ is of order δ^2 as $\delta \to 0$. Presumably any natural notion of "fractal dimension" will give dimension 2 to $\mathscr{S}$.

To start the proof of Proposition 4, write

$$D(s,t) = \inf_{0 \le r \le s} \|\rho(t) - \rho(r)\|, \qquad 0 < s < t, \tag{14}$$

for the distance (in l_1, or equivalently along the skeleton $\mathscr{S}_\infty$) between the point labelled t and the partial skeleton $\mathscr{S}_s$.

LEMMA 5. (i) $P(D(s,t) > b) \le \exp(-bs)$, $0 < s < t$, $b > 0$.
(ii) $D(s,t) \to_d \beta_s$, *say as* $t \to \infty$, *where* $s\beta_s \to_d$ exponential(1) *as* $s \to \infty$.

PROOF. Part (i) says $D(s,t)$ is stochastically bounded by the exponential(s) distribution. Lemma 9 gives a simple probabilistic argument for a corresponding property in the discrete setting, and the present result can be proved similarly. To prove (ii), fix θ and s. Write $g(t) = E\exp(\theta D(s,t))$. By considering whether or not some cut-point C_i occurs in $[t, t + dt]$, standard arguments give the equation

$$g'(t) = \theta g(t) - tg(t) + \int_0^t g(u)\,du, \qquad t > s, \tag{15}$$

with boundary conditions

$$g = 1 \quad \text{on } [0, s], \qquad g'(s+) = \theta.$$

Fortunately, we can solve this explicitly. Writing $h(t) = g'(t)$, we have

$$h'(t) = \theta h(t) - th(t),$$

and one solution of this equation is

$$h(t) = \exp\left(\theta t - \tfrac{1}{2}t^2\right).$$

Fitting initial conditions gives the solution

$$g(t) = 1 + \frac{\theta}{h(s)} \int_s^t h(u)\, du, \qquad t \geq s,$$

and hence

$$E \exp(\theta s D(s,t)) = 1 + \theta s \int_s^t \exp\left\{\theta s(u - s) - \tfrac{1}{2}(u^2 - s^2)\right\} du.$$

The obvious calculus estimates give the desired result. □

From the marginal bound of Lemma 5 we proceed to a maximal inequality.

LEMMA 6. $P(\sup_{e^j \leq t \leq e^{j+1}} D(e^j, t) \geq 3je^{-j}) \leq e^{2-j}$.

PROOF. Write $s = e^j$, $u = e^{j+1}$. Fix $b > 1/s$. Consider

$$\tau = \inf\{t \geq s\colon D(s,t) = b\}.$$

Recall the Poisson process $(C_1, C_2, \dots)$ of cut-points. Let $C_t = \min\{C_i\colon C_i \geq t\}$ be the time of the first cut-point after t. Then $D(s,t) \geq b$ on $\tau \leq t \leq C_\tau$ and so

$$(16) \qquad \lambda\{t\colon s \leq t \leq u + 1, D(s,t) \geq b\} \geq \left[C_\tau \wedge (u+1) - \tau\right] 1_{(\tau \leq u)},$$

where λ is Lebesgue measure. Now take expectations and bound each side. The left-hand side has expectation

$$\int_s^{u+1} P(D(s,t) \geq b)\, dt \leq (u + 1 - s)\exp(-bs)$$

by Lemma 5(i). For the right-hand side, it is easy to see that

$$(17) \qquad E(C_t \wedge (u+1) - t)1_{(t \leq u)} \geq \frac{1 - \exp(-u-1)}{u+1} 1_{(t \leq u)}.$$

Now, the conditional distribution of C_τ given $\{\tau = t\}$ is the distribution of C_t, by the independence structure of the Poisson process (C_i) of cuts. So, using (17), the right-hand side of (16) has expectation at least

$$\frac{1 - \exp(-u-1)}{u+1} P(\tau \leq u).$$

Now, the process $t \to D(s,t)$ cannot make upward jumps, so

$$P\Big(\sup_{s \le t \le u} D(s,t) \ge b\Big) = P(\tau \le u).$$

Rearranging the preceding inequalities, and doing some algebra (noting that we can take $j \ge 2$ in the lemma), we obtain

$$P\Big(\sup_{s \le t \le u} D(s,t) \ge b\Big) \le e^2 s^2 \exp(-bs).$$

Putting $b = 3s^{-1}\log(s)$ establishes the lemma. □

Now note that, for k sufficiently large,

$$\sum_{j \ge k} je^{-j} \le 2ke^{-k}.$$

Since

$$\sup_{t \ge e^k} D(e^k, t) \le \sum_{j \ge k} \sup_{e^j \le t \le e^{j+1}} D(e^j, t),$$

Lemma 6 implies that, for k sufficiently large,

$$P\Big(\sup_{t \ge e^k} D(e^k, t) \ge 6ke^{-k}\Big) \le (1 - 1/e)^{-1} e^{2-k}. \tag{18}$$

The Borel–Cantelli lemma and some manipulation show that a.s.

$$\sup_{t \ge s(\delta)} D(s(\delta), t) \le \delta/2 \quad \text{for all sufficiently small } \delta, \tag{19}$$

where $s(\delta) = 13e\delta^{-1}\log(1/\delta)$.

Now the covering number N in Proposition 4 satisfies

$$\begin{aligned} N(\mathscr{S}, \delta) &= N(\mathscr{S}_\infty, \delta) \\ &\le N(\mathscr{S}_{s(\delta)}, \delta/2) \quad \text{for sufficiently small } \delta \end{aligned} \tag{20}$$

a.s., by (19). But a $\delta/2$-cover of $\mathscr{S}_{s(\delta)}$ is given by

$$\rho(0), \rho(\delta/2), \rho(\delta), \rho(3\delta/2), \dots, \rho(s(\delta)), \rho(B_1), \rho(B_2), \dots, \rho(B_{Q(s(\delta))}),$$

where $Q(s)$ is the number of cut-points in $[0, s]$. So

$$N(\mathscr{S}_{s(\delta)}, \delta/2) \le 2 + 2\delta^{-1}s(\delta) + Q(s(\delta)). \tag{21}$$

But $Q(s)$ has Poisson$(s^2/2)$ distribution and, by the law of large numbers (for the homogeneous Poisson process), a.s.

$$Q(s) \le \tfrac{3}{4}s^2 \quad \text{for sufficiently large } s.$$

Now (20) and (21) imply that a.s.

$$N(\mathscr{S}, \delta) \le s^2(\delta) \quad \text{for sufficiently small } \delta.$$

This establishes the proposition.

We digress to give

LEMMA 7. *$N(\mathscr{S}, \delta) \geq 1/(20\delta^2)$ for all sufficiently small δ, a.s.*

PROOF. Fix $\delta > 0$. Let $A(t, \delta)$ be the event that no cut-point C_i falls in the interval $[t - 2\delta, t]$. Let

$$M_\delta = \{\rho(2\delta n): n \geq 1,\ A(2\delta n, \delta) \text{ occurs}\}.$$

The points of M_δ are 2δ-separated, so $N(\mathscr{S}_\infty, \delta) \geq |M_\delta|$. The events in the definition of M_δ are independent, so it is easy to calculate

$$E|M_\delta| \sim \operatorname{var}|M_\delta| \sim \int_0^\infty (2\delta)^{-1} \exp(-2\delta t)\, dt \sim (2\delta)^{-2} \quad \text{as } \delta \to 0.$$

Then Chebyshev's inequality and the Borel–Cantelli lemma give

$$|M_{\delta(k)}| \sim (2\delta(k))^{-2} \quad \text{a.s., for } \delta(k) = 2^{-k}.$$

Hence $N(\mathscr{S}_\infty, \delta(k)) \geq 1/(5\delta^2(k))$ ultimately a.s., and the result follows by monotonicity in δ. □

PROOF OF THEOREM 3. Proposition 4 gives (i). For (ii), let $\pi_k: l_1 \to l_1^k$ be the projection $\pi_k(x) = (x_1, \dots, x_k)$. Then π_k acts on measures in the natural way. The key fact (shown later) is that, for fixed k,

(22) $\{\pi_k \mu_{C_m(\cdot)};\ m \geq k\}$ is a measure-valued martingale.

It is well known that probability-measure-valued martingales converge a.s. (use the convergence theorem for positive martingales and the existence of countable convergence-determining classes of sets). Thus, it easily follows that, a.s.,

$$\lim_{t \to \infty} \pi_k \mu_t \quad \text{exists, for each } k.$$

But (i) implies that a.s. $\mathscr{S}$ is compact and hence

$$\{\mu_t;\ t \geq 0\} \quad \text{is tight.}$$

Since a probability distribution ν on l_1 is determined by $(\pi_k \nu;\ k \geq 1)$, it follows that $\mu = \lim_t \mu_t$ exists a.s., which is part (ii) of the theorem.

To prove (iii), we shall show later that, for fixed m,

(23) $$\operatorname{support}(\pi_m \mu) = \mathscr{S}_{C_m} \quad \text{a.s.}$$

Recall that the mth branch added to the tree is $\rho(C_{m-1}, C_m]$ and has length $C_m - C_{m-1}$. Consider a point $x \in \mathscr{S}_\infty$ such that

$$\|\pi_m(x) - \rho(C_m)\| \leq \varepsilon < C_m - C_{m-1}.$$

Then the path from 0 to x must enter the branch $\rho(C_{m-1}, C_m]$, by the orthogonal property of the branches. Thus, $\|x - \pi_m(x)\| \leq h_m$, where h_m is the Hausdorff distance between $\mathscr{S}_{C_m}$ and $\mathscr{S}$. So for a point x as above,

$$\|x - \rho(C_m)\| \leq \varepsilon + h_m.$$

But (23) says that $\rho(C_m)$ is a.s. in the support of $\pi_m(\mu)$, and so, by taking ε sufficiently small,

$$\mu\{x: \|x - \rho(C_m)\| \le 2h_m\} > 0 \quad \text{a.s.} \tag{24}$$

Now $C_m \sim \sqrt{2m}$ a.s. and B_m is uniform on $[0, C_m]$, and it follows easily that a.s. (B_m) is dense in $[0, \infty)$. Also, we know that $C_{m+1} - C_m \to 0$ a.s. and $h_m \downarrow 0$ a.s. Fix a tree $\mathscr{S}(\omega)$ satisfying these a.s. properties. For each $y \in \mathscr{S}_\infty$ there is a subsequence $\rho(B_{m_j}) \to y$ and then $\rho(C_{m_j+1}) \to y$. Applying (24), we deduce that y is in the support of μ. Thus, we have shown $\mathscr{S}_\infty \subseteq \text{support}(\mu)$ a.s., which implies $\mathscr{S} \subseteq \text{support}(\mu)$ a.s. But on the other hand, $\mu_t(\mathscr{S}) = 1$ for all t, so (ii) implies $\mu(\mathscr{S}) = 1$ a.s. Thus, $\mathscr{S} = \text{support}(\mu)$ a.s., which is part (iii) of the theorem.

To prove part (iv), write $g_s(x) = \min_{y \in S_s} \|x - y\|$, $x \in l_1$. By averaging over t in part (ii) of Lemma 5,

$$E\mu_t\{x: g_s(x) \in \cdot\} \to_d P(\beta_s \in \cdot) \quad \text{as } t \to \infty.$$

So, by part (ii),

$$E\mu\{x: g_s(x) \in \cdot\} = P(\beta_s \in \cdot).$$

Therefore,

$$\begin{aligned} E\mu(S_\infty) &= \lim_{s \to \infty} E\mu(S_s) \\ &= \lim_{s \to \infty} E\mu\{x: g_s(x) = 0\} \\ &= \lim_{s \to \infty} P(\beta_s = 0) \\ &= 0 \end{aligned}$$

because Lemma 5 says $s\beta_s \to_d$ exponential(1).

The urn model. It remains to prove (22) and (23). These are most easily viewed as facts about generalized urn models. Informally, let U_m, V_m the numbers of black and white balls in an urn at time m. Choose a color by touching a uniform random ball in the urn; add Δ_m balls of that color, where Δ_m is random with arbitrary distribution (varying with m), but is independent of the choice of color. Formally, let the r.v.'s U_m, V_m, Δ_m and events A_m be adapted to $\mathscr{F}_m$. Let $U_0, V_0, \Delta_m > 0$. Let

$$(U_{m+1}, V_{m+1}) = \begin{cases} (U_m + \Delta_{m+1}, V_m) & \text{on } A_m, \\ (U_m, V_m + \Delta_{m+1}) & \text{on } A_m^c, \end{cases}$$

$$P(A_{m+1} | \mathscr{F}_m) = U_m / (U_m + V_m),$$

A_{m+1} and Δ_{m+1} be conditionally independent given $\mathscr{F}_m$.

Note that we allow U_m, V_m to be nonintegral. It is elementary that

$$R_m = \frac{U_m}{U_m + V_m} \quad \text{is a martingale;} \tag{25}$$

$$\operatorname{var}(R_{m+1}) - \operatorname{var}(R_m) \le E\left(\frac{\Delta_{m+1}}{U_m + V_m}\right)^2. \tag{26}$$

Now $R_m \to R_\infty$ a.s., say. A more subtle result, contained essentially in Pemantle [11], is

$$(27) \qquad P(0 < R_\infty < 1) = 1 \quad \text{provided } \sup_m \Delta_m < \infty \text{ a.s.}$$

To apply in our setting, fix k, condition on the partial tree $\mathscr{S}_{C_k} = \rho[0, C_k]$ and let (G_1, G_2) be a partition of that partial tree into "black" and "white" subsets of positive size (induced Lebesgue measure). As the tree is grown by adding branches $\rho(C_m, C_{m+1}]$, color each branch black or white according to the color of the join-point $\rho(B_m)$. For $m > k$, let U_m, V_m be the sizes of the black and white subsets of $\mathscr{S}_{C_m}$, and let $\Delta_m = C_m - C_{m-1}$. Now the process "proportion of $\mathscr{S}_{C_m}$ colored black" can be written as

$$\pi_k \mu_{C_m}(G_1), \qquad m > k,$$

and (25) says this process is a martingale. This result (for each G_1) establishes (22). Now from the definition of C_m it is easy to see that $\Delta_m \to 0$ a.s., and then (27) says that the limit proportion of black balls is a.s. nonzero. Using (ii), for closed G_1

$$(28) \qquad \pi_k \mu(G_1) \geq \lim_m \pi_k \mu_{C_m}(J_1) > 0 \quad \text{a.s.}$$

Varying G_1 gives $\mathscr{S}_{C_k} \subseteq \text{support}(\pi_k \mu)$ a.s., establishing (23). □

5. Convergence of the uniform tree to the compact continuum tree. Algorithm 1 gives a natural set-representation (recall Section 2) of the uniform random tree $\mathscr{T}_n$. Informally, call the times i, where $U_i < i - 1$ *cut-times*; build the tree by adding unit edges in the same direction between cut-times, forming a *branch*; and at a cut-time start a new branch by adding edges in a new, orthogonal direction from a random vertex already constructed.

Formally, recall that $(z_i;\ i \geq 1)$ is the natural unit vector basis of l_1. Fix n, and let $(U_i;\ 2 \leq i \leq n)$ be independent uniform on $\{1, \dots, n\}$. Define $J_1 = 1$,

$$(29) \qquad J_i = \begin{cases} J_{i-1} + 1 & \text{if } U_i < i - 1, \\ J_{i-1} & \text{if } U_i \geq i - 1. \end{cases}$$

Then Algorithm 1 implies that a set-representation $\mathscr{S}_n = \{V_1, \dots, V_n\}$ of $\mathscr{T}_n$ can be defined as follows: $V_1 = 0$,

$$(30) \qquad V_i = V_{\min(U_i, i-1)} + z_{J_i}, \qquad i > 1.$$

Let μ_n be the corresponding measure-representation. For $t \leq n$ we can consider the *partial tree*

$$(31) \qquad \mathscr{S}_n(t) = \{V_1, \dots, V_{[t]}\}.$$

Let $\mu_n(t)$ be the uniform probability distribution on $\mathscr{S}_n(t)$. As in Section 2, the rescaling operator $\sigma_{n^{-1/2}}$ acts naturally on these objects.

This section is devoted to the proof of our main convergence result:

THEOREM 8. *Let $\mathscr{S}$, μ be the set- and measure-representations of the compact continuum tree, as in Theorem 3. Then as $n \to \infty$,*

$$(\sigma_{n^{-1/2}}(\mathscr{S}_n), \sigma_{n^{-1/2}}(\mu_n)) \to_d (\mathscr{S}, \mu).$$

PROOF. For each n, construct (C_j^n, B_j^n), $j \geq 1$ as follows:
C_j^n is the jth cut-time, that is, the jth element of $\{i: U_i < i - 1\}$.
B_j^n is the corresponding value of U, that is $B_j^n = U_{C_j^n}$.
It is straightforward to show

$$(32)\qquad \begin{aligned} &((n^{-1/2}C_1^n, n^{-1/2}B_1^n), (n^{-1/2}C_2^n, n^{-1/2}B_2^n), \dots) \\ &\qquad \to_d ((C_1, B_1), (C_2, B_2), \dots) \end{aligned}$$

for (B_i, C_i) as in Section 4. This implies convergence of the partial trees defined in (31),

$$(33)\qquad \begin{aligned} &(\sigma_{n^{-1/2}}(\mathscr{S}_n(tn^{1/2})), \sigma_{n^{-1/2}}(\mu_n(tn^{1/2}))) \\ &\qquad \to_d (\mathscr{S}_t, \mu_t) \quad \text{for fixed } 0 < t < \infty, \end{aligned}$$

where $\mathscr{S}_t$, μ_t are as in the construction of Theorem 3. We know from Theorem 3 that $\mathscr{S}_t \to \mathscr{S}$ a.s. and $\mu_t \to \mu$ a.s. So by a routine weak convergence lemma (cf. [3], Theorem 4.2) it suffices to prove

$$(34)\qquad \lim_{t\to\infty} \limsup_{n\to\infty} P\big(d_H(\sigma_{n^{-1/2}}(\mathscr{S}_n(tn^{1/2})), \sigma_{n^{-1/2}}(\mathscr{S}_n)) > \varepsilon\big) = 0,$$

$$(35)\qquad \lim_{t\to\infty} \limsup_{n\to\infty} P\big(d_W(\sigma_{n^{-1/2}}(\mu_n(tn^{1/2})), \sigma_{n^{-1/2}}(\mu_n)) > \varepsilon\big) = 0,$$

where d_H is the Hausdorff metric on compact sets and d_W is a metrization of weak convergence of probability measures.

Analogously to (14), write

$$D_n(i, j) = \min_{1 \leq u \leq i} \|V_j - V_u\|.$$

Then (34) can be rephrased as

$$(36)\qquad \lim_{t\to\infty} \limsup_{n\to\infty} P\Big(\sup_j D_n([tn^{1/2}], j) > \varepsilon n^{1/2}\Big) = 0.$$

But this can be shown by following closely the arguments for Proposition 4. Here is the analog of Lemma 5.

LEMMA 9. *For integers $1 \leq i < j \leq n$, $b \geq 1$,*

$$P(D_n(i, j) > b) \leq (1 - i/n)^b.$$

PROOF. Let $\xi \ge 1$ be an integer-valued r.v. Let (A, B, C) be a partition of events, independent of ξ, and let $P(A) = q$. Consider a r.v. Ψ of the form

$$\Psi = \begin{cases} 1 & \text{on } A, \\ \xi + 1 & \text{on } B, \end{cases}$$

$$\text{dist}(\Psi | C) = \text{dist}(\eta + 1),$$

for some η. If ξ and η have geometric(q) distributions [that is, $p(m) = q(1-q)^m$, $m \ge 1$], then it is easy to verify that Ψ also has geometric(q) distribution. It follows that

(37) if ξ and η are stochastically bounded by the geometric(q) distribution, then so is Ψ.

The lemma asserts that $D_n(i, j)$ is stochastically bounded by the geometric (i/n) distribution. We argue by induction on j, for fixed n, i. Certainly $D_n(i, i+1) = 1$. Writing $\xi = D_n(i, j)$, we see that $\Psi = D_n(i, j+1)$ is of the preceding form, for

$$A = \{U_{j+1} \le i\}, \qquad B = \{U_{j+1} \ge j\}, \qquad C = \{i + 1 \le U_{j+1} \le j - 1\}.$$

Here dist$(\Psi | C)$ is a mixture over $i + 1 \le u \le j - 1$ of $1 + D_n(i, u)$. So (37) carries the induction forward from j to $j + 1$, establishing the lemma. □

With only minor changes to allow for discreteness, we can now follow the argument for Lemma 6 and then establish the analog of (18): There exist $k_0 < \infty$ and $A < \infty$ such that for all $k \ge k_0$ and all n,

$$P\Big(\max_{u \le n} D_n([n^{1/2} e^k], u) \ge 6ke^{-k} n^{1/2}\Big) \le Ae^{-k}.$$

This yields (36).

To prove (35), let d_W be the Lévy–Prohorov (dual bounded Lipschitz) metric ([4], Section 11.3) on probability measures on l_1:

$$d_W(\theta_1, \theta_2) = \sup\left\{\left|\int f d\theta_1 - \int f d\theta_2\right| : \|f\|_{BL} \le 1\right\}.$$

We record a simple fact.

LEMMA 10. *Let θ_1, θ_2 be probability measures. Let $K_1, \dots, K_N$ be disjoint sets, of diameter at most d_0, which cover the support of θ_1. Then,*

$$d_W(\theta_1, \theta_2) \le d_0 + \sum_i |\theta_1(K_i) - \theta_2(K_i)|.$$

Now fix n, M and condition on the partial tree $\mathscr{S}_n(C_M^n)$ consisting of the first M branches. Let K be a subset of l_1^M. For each $j \ge C_M^n$, $\pi_M(\mu_n(j))$ is a probability measure on l_1^M, and $\pi_M \mu_n(j)(K)$ denotes the mass it assigns to K.

We shall establish a bound of the form

$$\max_K P\Big(\max_{j\geq C_M^n}|\pi_M\mu_n(j)(K)-\mu_n(C_M^n)(K)|>\delta|\mathscr{S}_n(C_M^n)\Big) \leq \delta^{-2}A(n,M), \tag{38}$$

where $A(n,M)$ are random variables such that

$$\lim_{M\to\infty}\limsup_{n\to\infty} EA(n,M)=0. \tag{39}$$

Then let $N=N(\mathscr{S}_n(C_M^n),\varepsilon n^{1/2})$ be the size of the smallest $\varepsilon n^{1/2}$-cover $\{x_1,\dots,x_N\}$ of $\mathscr{S}_n(C_M^n)$. Associate with $\{x_i\}$ a partition $\{K_i\}$ of $\mathscr{S}_n(C_M^n)$ into subsets of diameter at most $2\varepsilon n^{1/2}$. Applying (38) and Lemma 10, and rescaling space, we obtain

$$P\Big(\max_{j\geq C_M^n} d_W(\sigma_{n^{-1/2}}\pi_M\mu_n(j),\sigma_{n^{-1/2}}\mu_n(C_M^n))\geq 2\varepsilon+\delta N|\mathscr{S}_n(C_M^n)\Big) \leq \delta^{-2}NA(n,M). \tag{40}$$

Now

$$N=N(\sigma_{n^{-1/2}}\mathscr{S}_n(C_M^n),\varepsilon)\leq N(\sigma_{n^{-1/2}}\mathscr{S}_n(n),\varepsilon),$$

the equality being just scaling. We know $\sigma_{n^{-1/2}}(\mathscr{S}_n(n))\to_d\mathscr{S}$, and so we may take versions such that $\sigma_{n^{-1/2}}(\mathscr{S}_n(n))\to\mathscr{S}$ a.s. So,

$$N^*(\varepsilon)\equiv\max_n N(\sigma_{n^{-1/2}}\mathscr{S}_n(n),\varepsilon)<\infty \quad \text{a.s.} \tag{41}$$

Thus in (40) we may replace N by $N^*(\varepsilon)$. Then, putting $\delta=\varepsilon/N^*(\varepsilon)$ and considering both $j=[n^{1/2}t]$ and $j=n$, (40) gives

$$P\big(d_W(\sigma_{n^{-1/2}}\pi_M\mu_n,\sigma_{n^{-1/2}}\pi_M\mu_n(n^{1/2}t))>6\varepsilon|\mathscr{S}_n(C_M^n)\big) \leq \varepsilon^{-2}(N^*(\varepsilon))^3A(n,M) \quad \text{on } \{n^{1/2}t\geq C_M^n\}. \tag{42}$$

Now,

$$\begin{aligned} d_W(\sigma_{n^{-1/2}}\pi_M\mu_n,\sigma_{n^{-1/2}}\mu_n) &\leq \sup_{x\in\mathscr{S}_n}\|\sigma_{n^{-1/2}}\pi_M x-\sigma_{n^{-1/2}}x\| \\ &= d_H(\sigma_{n^{-1/2}}\mathscr{S}_n,\sigma_{n^{-1/2}}\mathscr{S}_n(C_M^n)) \end{aligned}$$

and the same for $\mu_n(n^{1/2}t)$. Thus, taking expectations in (42), for any $b>0$,

$$\begin{aligned} &P\big(d_W(\sigma_{n^{-1/2}}(\mu_n(tn^{1/2})),\sigma_{n^{-1/2}}(\mu_n))>8\varepsilon\big) \\ &\quad\leq \varepsilon^{-2}b^3EA(n,M)+P(N^*(\varepsilon)>b)+P(t<n^{-1/2}C_M^n) \\ &\qquad+P\big(d_H(\sigma_{n^{-1/2}}(\mathscr{S}_n(tn^{1/2})),\sigma_{n^{-1/2}}(\mathscr{S}_n))>\varepsilon\big). \end{aligned} \tag{43}$$

For any $M=M(t)\to\infty$ as $t\to\infty$, (39) and (34) now imply

$$\begin{aligned} &\limsup_{t\to\infty}\limsup_{n\to\infty}P\big(d_W(\sigma_{n^{-1/2}}(\mu_n(tn^{1/2})),\sigma_{n^{-1/2}}(\mu_n))>8\varepsilon\big) \\ &\quad\leq P(N^*(\varepsilon)>b)+\limsup_{t\to\infty}\limsup_{n\to\infty}P(t<n^{-1/2}C_M^n). \end{aligned} \tag{44}$$

Since b is arbitrary, the first term in the bound may be made arbitrarily small. And for fixed M, (32) says $n^{-1/2}C_M^n \to_d C_M$, so by taking $M(t) = o(t^{1/2})$ the final term of (44) is zero. This establishes (35).

It remains to prove (38) and (39). Fix n, M and condition on the partial tree $\mathscr{S}_n(C_M^n)$ consisting of the first M branches. Let K be a subset of l_1^M. As in the previous section, there is an "urn model" description of how the algorithm constructs the tree. Color black the vertices of $\mathscr{S}(C_M^n - 1)$ which are in K, and color the other vertices white. The algorithm grows the tree in segments $[C_m^n, C_{m+1}^n - 1]$. Color each new vertex according to the color of the vertex it is attached to. Then each branch is either all black or all white, according to whether its first vertex C_m^n was attached to a black or white vertex of $\mathscr{S}_n(C_m^n - 1)$, and the color probabilities are just the color proportions of $\mathscr{S}_n(C_m^n - 1)$. Thus, if we look at the tree only at "times" (number of vertices) $\mathscr{S}_n(C_m^n - 1)$, then the proportions of black vertices

$$\pi_M \mu_n(C_m^n - 1)(K), \qquad m \geq M,$$

form a martingale: we are in the setting of the urn model, with

$$\Delta_{m+1} = C_{m+1}^n - C_m^n, \qquad U_{m+1} + V_{m+1} = C_{m+1}^n - 1.$$

Now appeal to the variance bound (26) and the L^2 martingale maximal inequality. We see that (38) holds for

$$A(n, M) = \sum_{m \geq M} E\left(\left(\frac{C_{m+1}^n - C_m^n}{C_m^n - 1}\right)^2 \middle| \mathscr{S}(C_M^n - 1)\right).$$

But conditional on C_m^n the distribution of $C_{m+1}^n - C_m^n$ is stochastically bounded by the geometric $(n^{-1}(C_m^n - 1))$ distribution, and so

$$E\left((C_{m+1}^n - C_m^n)^2 \middle| C_m^n\right) \leq 2\left(\frac{n}{C_m^n - 1}\right)^2.$$

Thus, to prove (39), we must prove

$$\lim_{M \to \infty} \limsup_{n \to \infty} \sum_{m \geq M} n^2 E(C_m^n - 1)^{-4} = 0. \tag{45}$$

This is heuristically clear, because $n^{-1/2}C_m^n \to C_m \approx \sqrt{2m}$, so the summands are of order m^{-2}. To make a formal proof, we need to bound $P(C_m^n \leq k)$ for k smaller than $\sqrt{2mn}$. But this probability is $P(J_k \geq m)$ for J_k defined at (29). And J_k is the sum of independent indicators, so routine large deviation upper bounds can be produced, and these lead to a proof of (45). □

REMARKS. To see why we treat both set- and measure-representations, consider the height profile process at (1). To describe its limit process in terms of the continuum random tree, we need the measure-representation μ. On the other hand, to show by our approach that the rescaled heights $h(\mathscr{T}_n)$ of uniform random labelled trees satisfy

$$n^{-1/2}h(\mathcal{T}_n) \to_d h(\mathcal{S}),$$

we want Hausdorff convergence of set-representation. Such issues, and convergence of moments, are deferred to future work.

The existence of a limit distribution for $n^{-1/2}h(\mathcal{T}_n)$ goes back to Rényi and Szekeres [12]. A technical fact, needed later is that

$$Eh(\mathcal{S}) < \infty. \tag{46}$$

We could prove this from our construction, but it is easier to quote the fact from [12] that

$$\lim_n E\left(n^{-1/2}h(\mathcal{T}_n)\right)^2 < \infty. \tag{47}$$

6. The self-similar continuum tree. A verbal description of Process 2 was given in the introduction. We also mentioned an alternative construction via Process 3. It is technically convenient to start with the latter, deferring the formalization of Process 2 to the next section.

Let $\mathcal{S}$ be the compact continuum tree of Section 4. Write ψ_t for the distribution of $\sigma_t\mathcal{S}$, the compact tree rescaled by a factor t. So ψ_t is a probability measure on the space $\mathcal{K}$ of compact subsets K of l_1. Define a σ-finite measure ψ on $\mathcal{K}$ by

$$\psi(\cdot) = \sqrt{2/\pi}\int_0^\infty t^{-2}\psi_t(\cdot)\,dt. \tag{48}$$

Write $h(K) = \max\{\|x\|: x \in K\}$ for the radius of a compact set K. Then,

$$\begin{aligned}\sqrt{\pi/2}\,\psi\{K: h(K) > \delta\} &= \int_0^\infty t^{-2}\psi_t\{K: h(K) > \delta\}\,dt\\ &= \int_0^\infty t^{-2}\psi_1\{K: h(K) > \delta/t\}\,dt\\ &= \int_0^\infty \delta^{-1}\psi_1\{K: h(K) > s\}\,ds\\ &= \delta^{-1}Eh(\mathcal{S})\\ &< \infty \quad \text{by (46)}.\end{aligned}$$

Thus ψ is finite "off small sets." Regard ψ as analogous to the σ-finite excursion measure for Brownian motion. We know how to build up Brownian motion on $[0,\infty)$ from its excursions: The following construction builds up an infinite continuum random tree from compact random trees with distributions controlled by ψ.

Let $\tilde{l}_1 = \{\mathbf{x} \in l_1: x_1 = 0\}$. Let $J = (j_i)$ be an infinite subset of $\{2,3,\dots\}$, which for the moment we may take to be all of $\{2,3,\dots\}$. Let $\pi_J: l_1 \to \tilde{l}_1$ be the map $\sum_i a_i z_i \to \sum_i a_i z_{j_i}$, which acts naturally on subsets and measures on l_1. Let $\tilde{\mathcal{K}}$ be the space of compact subsets of $\tilde{l}_1$. Consider a Poisson point process on $[0, \infty) \times \tilde{\mathscr{K}}$ with intensity Leb $\times\, \psi$. Let $\mathscr{R}^0$ be the random subset of l_1 defined as the union of $\{uz_1: 0 \le u < \infty\}$ and the sets $\{t^* z_1\} \times \pi_J K^*$, where (t^*, K^*) are the points of the Poisson point process. We picture $\mathscr{R}^0$ as i.i.d. compact *bushes* attached to the *baseline* $\{uz_1: 0 \le u < \infty\}$. Clearly, $\mathscr{R}^0$ is connected, closed and locally compact.

By construction ψ is self-similar,

$$\sigma_t \psi = t\psi,$$

and it follows that $\mathscr{R}^0$ is self-similar,

$$\sigma_t \mathscr{R}^0 =_d \mathscr{R}^0. \tag{49}$$

Note that σ_t is acting both on the baseline and on the attached bushes.

Of course, $\mathscr{R}^0$ is the *set*-representation of a certain random tree. We can construct the corresponding measure-representation μ^0 similarly. Let μ be the measure-representation of the compact continuum tree of Section 4. Write Θ_t for the distribution of $t^2 \sigma_t \mu$. So Θ_t is a probability measure on the space $\mathscr{M}$ of positive measures on l_1. Write

$$\Theta(\cdot) = \sqrt{2/\pi} \int_0^\infty t^{-2} \Theta_t(\cdot)\, dt. \tag{50}$$

Informally, in (48) and (50) the factor t^{-2} gives the relative frequencies of trees of relative linear size t. The "measure" in our measure-representations is indicating relative numbers of vertices. Scaling by t changes this number by t^2 (our trees are "two-dimensional"), which accounts for the t^2 term in the definition of Θ_t.

The formal construction proceeds as before. Let $\tilde{\mathscr{M}}$ be the space of measures on $\tilde{l}_1$. Take a Poisson point process on $[0, \infty) \times \tilde{\mathscr{M}}$ with intensity Leb $\times\, \Theta$. Then μ^0 is defined to be the sum, over points (t^*, μ^*) of the Poisson point process, of $\delta_{t^*} \times \pi_J \mu^*$. It is easy to see from (50) that μ^0 is locally a.s. finite. We may take $\mathscr{R}^0(\omega) = \text{support}(\mu^0(\omega))$. In terms of this measure-representation, the self-similarity property becomes

$$\sigma_t \mu^0 =_d t^{-2} \mu^0. \tag{51}$$

This $(\mathscr{R}^0, \mu^0)$ is almost the random tree we want, but there is one difficulty. As discussed in Section 2, the whole point of representing trees as subsets of l_1 is to have the l_1 distance $\|x - y\|$ agree with graph-distance $d(x, y)$. But with the construction given with i.i.d. bushes K^* at points t^*, our $\mathscr{R}^0$ will contain elements $(t_1^* z_1, uz_2)$, $(t_2^* z_1, uz_2)$, for small $u > 0$. The graph distance between them is $|t_1^* - t_2^*| + 2u$ but the l_1 distance is $|t_1^* - t_2^*|$. We need to modify the construction to make different bushes orthogonal. There seems to be no elegant way to do this: Here is a simple inelegant way. Let $(J_{i,j}: i, j \ge 1)$ be a partition of $\{2, 3, \dots, \}$ into infinite components. Take the points (t^*, μ^*) of the Poisson process with $t^* \le 1$ and order them as (t_1^*, μ_1^*), $(t_2^*, \mu_2^*), \dots$ in decreasing order of total mass $\mu^*(\tilde{l}_1)$. Then, instead of using $(t_j^*, \pi_J \mu_j^*)$ in the construction of μ^0, we use $(t_j^*, \pi_{J_{1,j}} \mu_j^*)$. Repeat separately on the ith unit segment of the baseline, using $J_{i,j}$. Summing over these bush-measures gives as before a random measure; call this ν. Define $\mathscr{R}(\omega) = \text{support}(\nu(\omega))$. Call this Process $\hat{2}$.

This modified construction destroys the "equal in distribution" formulation of self-similarity. Instead we need the notion of equivalence in distribution ($\approx_d$), which with the related convergence notion ($\sim_d$) was discussed in Section 2. We get

$$\sigma_t \mathscr{R} \approx_d \mathscr{R}, \tag{52}$$

$$\sigma_t \nu \approx_d t^{-2}\nu. \tag{53}$$

Here is the main result of this section.

THEOREM 11. (i) *Let* $(\mathscr{P}^\infty, \phi)$ *be set- and counting-measure-representations of Process* 1, *the discrete infinite tree. Then,*

$$\sigma_{1/n}\mathscr{P}^\infty \sim_d \mathscr{R}, \qquad n^{-2}\sigma_{1/n}\phi \sim_d \nu \quad \text{as } n \to \infty.$$

(ii) *Let* $(\mathscr{S}_n, \mu_n)$ *be set- and measure-representations of* $\mathscr{T}_n$, *the uniform random tree on n vertices. Let* $k = k(n)$ *satisfy* $k(n) \to \infty$, $k(n) = o(n^{1/2})$. *Then,*

$$\sigma_{1/k}\mathscr{S}_n \sim_d \mathscr{R}, \qquad k^{-2}\sigma_{1/k}\mu_n \sim_d \nu \quad \text{as } n \to \infty.$$

(iii) *Let* $(\mathscr{S}, \mu)$ *be set- and measure-representations of Process* 3, *the compact continuum tree. As* $\varepsilon \to 0$,

$$\sigma_{1/\varepsilon}\mathscr{S} \sim_d \mathscr{R}, \qquad \varepsilon^{-2}\sigma_{1/\varepsilon}\mu \sim_d \nu.$$

TECHNICAL REMARK. The convergence assertions involve unbounded closed sets and infinite (locally finite) measures. Of course, we intend the *localized* versions of Hausdorff convergence of sets and weak convergence of measures.

PROOF OF THEOREM 11. It will suffice to prove (i). For then (ii) is immediate from Theorem 2, and (iii) follows from (ii) and Theorem 8 by considering $k(n) = n^{1/2}\varepsilon(n)$ for $\varepsilon(n) \to 0$ slowly.

To prove (i), construct $\mathscr{P}^\infty$ as follows. Let $\mathscr{P}$ be the tree generated by the Poisson(1) branching process and $|\mathscr{P}|$ its total population size. Let $A_0, A_1, \dots$ be i.i.d. distributed as $|\mathscr{P}|$. Conditional on that sequence, let $\mathscr{S}_{A_0}, \mathscr{S}_{A_1}, \dots$ be independent set-representations of the uniform random trees on $|A_0|, |A_1|, \dots$ vertices, constructed as in (30) but using unit vectors $z_2, z_3, z_4, \dots$ instead of $z_1, z_2, z_3, \dots$. Then let $\mathscr{P}^\infty$ be the union over $m \ge 0$ of $\{mz_1\} \times \mathscr{S}_{A_m}$.

We shall show that rescaled $\mathscr{P}^\infty$ converges to $\mathscr{R}^0$. Recall the modification previously discussed which takes $\mathscr{R}^0$ to $\mathscr{R}$. The same modification takes $\mathscr{P}^\infty$ to a set-representation of Process 1. It is straightforward to deduce convergence of the modified processes from convergence of unmodified processes, giving assertion (i).

In studying the asymptotic behavior of the rescaled process $\mathscr{P}^\infty$, it is clear that we may replace the i.i.d. process (m, A_m) by the Poisson process of the same intensity. Having done so, the rescaled process $\sigma_{1/n}\mathscr{P}^\infty$ is the union of points in a Poisson point process on $[0, \infty) \times \tilde{\mathscr{K}}$ with intensity

$$n \cdot \text{Leb} \times \sigma_{1/n} \operatorname{dist}(\mathscr{S}_{|\mathscr{P}|}) = \text{Leb} \times n\sigma_{1/n} \operatorname{dist}(\mathscr{S}_{|\mathscr{P}|}).$$

Now the process $\mathscr{R}^0$ is, off the baseline, of the same form with intensity Leb $\times\ \psi$, so the only issue is proving that these intensities converge (in the natural sense of convergence for σ-finite measures which are finite off small sets). Conditioning on $|\mathscr{P}|$, we have to prove

$$(54) \qquad n \sum_k P(|\mathscr{P}| = k)\sigma_{1/n} \operatorname{dist}(\mathscr{S}_k) \to \sqrt{2/\pi} \int_0^\infty t^{-2}\psi_t \, dt \quad \text{as } n \to \infty.$$

Now Theorem 8 says

$$\sigma_{k^{-1/2}} \operatorname{dist}(\mathscr{S}_k) \to \psi_1$$

and then scaling gives

$$\sigma_{1/n} \operatorname{dist}(\mathscr{S}_k) \to \psi_t \quad \text{as } k^{1/2}/n \to t.$$

We also have (4):

$$P(|\mathscr{P}| = k) \sim (2\pi)^{-1/2} k^{-3/2} \quad \text{as } k \to \infty.$$

It follows that, for fixed $\varepsilon > 0$, the sum in (54) over $\varepsilon^2 n^2 \le k \le \varepsilon^{-2} n^2$ converges to the integral over $[\varepsilon, \varepsilon^{-1}]$.

To take care of the right-hand tail, we need

$$\lim_{\varepsilon \downarrow 0} \lim_n P(|\mathscr{P}| \ge \varepsilon^{-2} n^2) = 0$$

which again follows from (4).

To take care of the left-hand tail we need, for fixed $\delta > 0$,

$$(55) \qquad \lim_{\varepsilon \downarrow 0} \lim_n n \sum_{k \le \varepsilon^2 n^2} P(|\mathscr{P}| = k) P\big(h(\sigma_{1/n} \mathscr{S}_k) > \delta\big) = 0,$$

where h is height. But

$$P\big(h(\sigma_{1/n}\mathscr{S}_k) > \delta\big) = P(h(\mathscr{S}_k) > \delta n) \le \frac{Ck}{\delta^2 n^2}$$

for some constant C, by (47). Thus the proof of (55) reduces to the proof of

$$(56) \qquad \lim_{\varepsilon \downarrow 0} \lim_n n^{-1} E|\mathscr{P}| 1_{(|\mathscr{P}| \le \varepsilon^2 n^2)} = 0$$

and this is a consequence of (4).

This establishes convergence of set-representations in (i). By a *measure*-representation ϕ here we mean counting measure (i.e., unnormalized) on a set-representation. The proof of convergence of scaled measure-representations follows the same plan. To handle left-hand tails, we have to show that arbitrarily small rescaled bushes contribute negligible mass, and this reduces to verifying the same condition (56) as before. □

7. Construction of Process 2. Theorem 12 gives a construction of a random tree $(\mathscr{R}^*, \nu^*)$ formalizing the verbal description of Process 2 in the introduction. At the end of the section, we indicate why this process is equivalent to Process $\hat{2}$ constructed in the previous section.

We shall give the construction for the part of the tree attached to the first unit segment of the baseline. The whole tree can then be built up as i.i.d. (and orthogonal in l_1) copies of this process on successive unit segments of the baseline.

Define $\tau_0 = L_0 = Q_0 = 0$ and let $\mathscr{R}_0 = \{uz_1\colon 0 \le u \le 1\}$ be the first unit segment of the baseline. For $i \ge 1$ define inductively times τ_i, lengths L_i and positions Q_i as follows:

$$\operatorname{dist}(\tau_i - \tau_{i-1}) \text{ is exponential, rate } 1 + L_1 + \cdots + L_{i-1},$$

$$\operatorname{dist}(L_i | \mathscr{F}_{i-1}, \tau_i) \text{ is exponential, rate } \tau_i,$$

$$\operatorname{dist}(Q_i | \mathscr{F}_{i-1}, L_i) \text{ is uniform on } \mathscr{R}_{\tau_{i-1}},$$

where $\mathscr{F}_i$ is the natural filtration of (τ_i, L_i, Q_i). Then define

$$\mathscr{R}_t = \mathscr{R}_{\tau_{i-1}} \quad \text{on } \tau_{i-1} \le t < \tau_i,$$

$$\mathscr{R}_{\tau_i} = \mathscr{R}_{\tau_{i-1}} \cup \{Q_i + uz_{i+1}\colon 0 \le u \le L_i\}.$$

This corresponds to our verbal description of Process 2: τ_i is the time of the ith birth, Q_i its position and L_i the length of the branch which is instantly attached there.

It is easy to see that only a finite number of branches are created in finite time. So $\mathscr{R}_t$ is well-defined, for each $t < \infty$. Let $\nu_t(\omega)$ be the natural measure on $\mathscr{R}_t(\omega)$, that is induced Lebesgue measure on each branch (note we are not normalizing). We can now state the analog of Theorem 3.

THEOREM 12. *Write* $\mathscr{R}^* = \overline{\bigcup_t \mathscr{R}_t}$. *Then*

(i) $\mathscr{R}^*$ *is a.s. compact,*
(ii) $t^{-1}\nu_t \to \nu^*$, *say, a.s.,*
(iii) $\mathscr{R}^*(\omega) = \operatorname{support}(\nu^*(\omega))$ *a.s.*

This $(\mathscr{R}^*, \nu^*)$, or rather its extension to the whole baseline, is our formalization of Process 2. One could give a bare-hands proof of Theorem 12, analogous to the proof of Theorem 3, but it turns out that the hardest part (compactness) follows from existing results. We shall show how later, and sketch the remainder of the proof.

The next lemma shows how Process 2 arises as a limit in Algorithm 1. As at (31), let $\mathscr{S}_n(t)$ be the partial tree constructed by the first $[t]$ steps of Algorithm 1. Choose some $k = k(n)$ satisfying $k \to \infty$, $k = o(n^{1/2})$. Write $D = \{x \in l_1\colon \pi_1(x) \le 1\}$.

LEMMA 13. *For fixed* $0 < t < \infty$,

$$\sigma_{1/k}(\mathscr{S}_n(tn/k) \cap D) \rightsquigarrow_d \mathscr{R}_t \quad \text{as } n \to \infty.$$

PROOF. We shall argue informally that the "dynamics" of the rescaled $\mathscr{S}_n(\cdot)$ process converge to those of $\mathscr{R}_t$; making a formal proof is then routine. First, consider the partial tree $\mathscr{S}_n(k)$ constructed by the first k steps. Since

$$P(U_i < i - 1 \text{ for some } i \le k) \le \frac{k^2}{2n} \to 0,$$

we see that

$$\sigma_{1/k}\left(\mathscr{S}_n\left(t'_n \frac{n}{k}\right) \cap D\right) \to \mathscr{R}_0$$

for $t'_n = k^2/n \to 0$.

Next, fix $t < t_1$ with $t_1 - t$ small. Condition on the existing tree $\mathscr{S}_n(tn/k) \cap D$, which has F_t vertices, say. For $i \in [tn/k, t_1 n/k]$, each U_i has chance approximately F_t/n to hit the existing partial tree. So the chance that the partial tree does not grow during that period is

$$\left(1 - \frac{F_t}{n}\right)^{[t_1 n/k] - [tn/k]} \sim \exp\left(-\frac{F_t}{k}(t_1 - t)\right).$$

In other words, for the time-rescaled process new branch-points occur at rate approximately F_t/k.

Now suppose $U_i < i - 1$ for $i = tn/k$, say. Then a new branch is grown, from a uniform random point on the existing tree, with length

$$L = \min\{j > i: U_j < j - 1\} - i.$$

This L has approximately geometric distribution with mean n/i, that is approximately exponential distribution with rate t/k. Moreover, the (rescaled) time taken to grow this branch is Lk/n, which is of order $k^2/n \to 0$.

Rescaling space by the factor $1/k$, these dynamics become those of $\mathscr{R}_t$. □

PROOF OF THEOREM 12. Part (ii) of Theorem 11 implies that, for $k = k(n)$ as before, the rescaled entire n-tree converges to Process $\hat{2}$, that is, to $\mathscr{R}$, say. Restricting to the portion attached to the first unit segment of the baseline,

$$\sigma_{1/k}(\mathscr{S}_n(n) \cap D) \rightsquigarrow_d \mathscr{R} \cap D$$

and the limit is a.s. compact. So Lemma 13 shows that each $\mathscr{R}_t$ is equivalent in distribution to some subset of $\mathscr{R} \cap D$, and it follows that the union $\bigcup_t \mathscr{R}_t$ is also equivalent in distribution to some subset of $\mathscr{R} \cap D$. Thus $\mathscr{R}^*$ is compact, which is assertion (i). To prove (ii), recall that τ_m is the time at which the mth branch is created. Recall that π_m is the projection $l_1 \to l_1^m$. We shall show

(57) $$\left(\pi_m(t^{-1}\nu_t): t \ge \tau_{m-1}\right) \text{ is a measure-valued martingale}$$

and then the argument following (22) applies unchanged to establish (ii).

To argue (57), fix t_0 and condition on the tree $\mathscr{R}_{t_0}$ constructed up to time t_0. Let J be a subset of that tree of positive length, and color J black. As the tree grows, color new branches black if they are attached to a black point on the existing tree. For $t \ge t_0$, let X_t be the length of the black subset of $\mathscr{R}_t$.

Then, given $X_t = x$,

$$X_{t+dt} = \begin{cases} x & \text{with probability } 1 - x\,dt, \\ x + t^{-1}\xi & \text{with probability } x\,dt, \end{cases} \tag{58}$$

where ξ has exponential(1) distribution. It follows that $(t^{-1}X_t;\ t \ge t_0)$ is a martingale, and this leads to (57).

To prove (iii), the key fact [cf. (28)] is that, for X_t defined previously,

$$\overline{X}_\infty \equiv \lim_t t^{-1}X_t > 0 \quad \text{a.s.};$$

the rest of the argument is similar to that in Theorem 3. The proof of (28) used an abstract urn model result, but here we can do exact calculations. Set $\psi(t,\theta) = E\exp(-\theta X_t)$. From (58) we get the equation,

$$\frac{d\psi}{dt} = E\left(X_t \exp(-\theta X_t)\left(1 - E\exp\left(\frac{\theta\xi}{t}\right)\right)\right)$$
$$= \frac{\theta}{t+\theta}\frac{d\psi}{d\theta}.$$

Given $X_{t_0} = x_0$, the solution for $t \ge t_0$ is

$$\psi(t,\theta) = \exp\left\{-x_0\left(\sqrt{t_0^2 + \theta^2 + 2t\theta} - t_0\right)\right\}.$$

So $\overline{X}_\infty$ has transform

$$E\exp(-\theta\overline{X}_\infty) = \exp\left\{-x_0\left(\sqrt{t_0^2 + 2\theta} - t_0\right)\right\} \tag{59}$$

and, putting $\theta = \infty$, we see that $P(\overline{X}_\infty = 0) = 0$ as required.

This completes the proof of Theorem 12. □

REMARK. Putting $t_0 = 0$, $x_0 = 1$ in (59), we see that the ν^*-measure of $\mathscr{R}^*$ has Laplace transform $\exp(-\sqrt{2\theta})$, and so has the positive stable distribution of exponent $\frac{1}{2}$. Returning to Process $\hat{2}$, $(\mathscr{R}, \nu)$, by construction the process $u \to \nu\{x: \pi_1(x) \le u\}$ has stationary independent increments, and by self-similarity (51) it must be the positive stable $(\frac{1}{2})$ process, up to some scale constant c. A calculation from (50) specifies c and verifies that $\nu^*(\mathscr{R}^*) =_d \nu(\mathscr{R} \cap D)$.

It remains to show that Processes 2 and $\hat{2}$ are equivalent in distribution, a fact made plausible by the preceding remark. To outline a proof, let $\nu_n(t)$ be counting measure on the restricted partial trees $\mathscr{S}_n(t) \cap D$ considered in Lemma 13. Let $k \to \infty$, $k = o(n^{1/2})$. Lemma 13 asserted convergence of set-representations, but here the trees have only finitely many branches and the measures involved are essentially uniform on the trees, so it is clear that the corresponding measure-representations converge:

$$\sigma_{1/k}\left(k^{-1}\nu_n(tn/k)\right) \to_d \nu_t \quad \text{as } n \to \infty,\ t \text{ fixed}.$$

Part (ii) of Theorem 11 implies convergence of the complete finite trees

$$\sigma_{1/k}\left(k^{-2}\nu_n(n)\right) \rightsquigarrow_d \nu(\cdot \cap D) \quad \text{as } n \to \infty.$$

And part (ii) of Theorem 12 says

$$t^{-1}\nu_t \to_d \nu^* \quad \text{as } t \to \infty.$$

So, to show $\nu^* \approx_d \nu$, we need to prove the analog of (35):

$$(60) \quad \lim_{t\to\infty} \limsup_{n\to\infty} P\Big(d_W\Big(\sigma_{1/k}\big((kt)^{-1}\nu_n(tn/k)\big), \sigma_{1/k}\big(k^{-2}\nu_n(n)\big)\Big) > \varepsilon\Big) = 0.$$

The proof uses the martingale property (57) applied to the discrete trees in much the same way as the proof of (35). We omit the details. □

REFERENCES

[1] ALDOUS, D. J. (1986). Some interesting processes arising as heavy traffic limits in a M/M/∞ storage process. *Stochastic Process. Appl.* **22** 291–313.

[2] ALDOUS, D. J. (1988). The random walk construction of uniform spanning trees and uniform labelled trees. *SIAM J. Discrete Math.* To appear.

[3] BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

[4] DUDLEY, R. M. (1989). *Real Analysis and Probability*. Wadsworth, Monterey, Calif.

[5] FLAJOLET, P. and ODLYZKO, A. (1982). The average height of binary trees and other simple trees. *J. Comput. System Sci.* **25** 171–213.

[6] GRIMMETT, G. R. (1980). Random labelled trees and their branching networks. *J. Austral. Math. Soc. Ser. A* **30** 229–237.

[7] KOLCHIN, V. F. (1986). *Random Mappings*. Optimization Software, New York.

[8] LE GALL, J.-F. (1989). Brownian excursions, trees and measure-valued branching processes. Technical Report, Université P. et M. Curie.

[9] MEIR, A. and MOON, J. W. (1978). On the altitude of nodes in random trees. *Canad. J. Math.* **30** 997–1015.

[10] MOON, J. W. (1970). *Counting Labelled Trees*. Canadian Mathematical Congress, Montreal.

[11] PEMANTLE, R. (1989). A time-dependent version of Polya's urn. *J. Theoret. Probab.* To appear.

[12] RÉNYI, A. and SZEKERES, G. (1967). On the height of trees. *J. Austral. Math. Soc.* **7** 497–507.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720