The Annals of Probability
1993, Vol. 21, No. 1, 248–289

THE CONTINUUM RANDOM TREE III

BY DAVID ALDOUS[1]

University of California, Berkeley

Let $(\mathscr{R}(k), k \geq 1)$ be random trees with k leaves, satisfying a consistency condition: Removing a random leaf from $\mathscr{R}(k)$ gives $\mathscr{R}(k-1)$. Then under an extra condition, this family determines a random continuum tree $\mathscr{S}$, which it is convenient to represent as a random subset of l_1. This leads to an abstract notion of convergence in distribution, as $n \to \infty$, of (rescaled) random trees $\mathscr{T}_n$ on n vertices to a limit continuum random tree $\mathscr{S}$. The notion is based upon the assumption that, for fixed k, the subtrees of $\mathscr{T}_n$ determined by k randomly chosen vertices converge to $\mathscr{R}(k)$.

As our main example, under mild conditions on the offspring distribution, the family tree of a Galton–Watson branching process, conditioned on total population size equal to n, can be rescaled to converge to a limit continuum random tree which can be constructed from Brownian excursion.

1. Introduction. Asymptotics for a particular model of random trees (the uniform random unordered labelled tree on n vertices) were discussed in Aldous (1991a) using an explicit algorithm for generating that random tree. After rescaling, those trees converge to a certain limit *continuum random tree*, which it is convenient to represent as a random subset, or more informatively as a random measure, in l_1. The first purpose of this paper is to discuss "convergence to a continuous limit" for general models of random finite trees. The notion of convergence of trees we develop is analogous to the classical treatment of weak convergence of processes. (Suppose finite-dimensional distributions converge, then the limit finite-dimensional distributions must be consistent, and under mild conditions define a "nice" process; then to show the original processes converge to this limit process, we need some extra "tightness" condition.) For convergence of rescaled random trees, we propose to use "subtree spanned by k randomly-chosen vertices" as the analog of "finite-dimensional distribution." If these subtrees converge (in the natural sense, for each fixed k), then the limit random trees $(\mathscr{R}(k))$ must satisfy a certain consistency condition (Definition 1). In Section 2 we study in detail the consequences of this consistency condition. Under a natural extra condition (Condition 2), the family $(\mathscr{R}(k))$ specifies a random continuum tree, formalized in general as a subset of l_1 (Section 2.3). Under further conditions, one can specify continuum trees via codings from continuous real functions (Sec-

Received May 1991; revised November 1991.
[1]Research supported by NSF Grants MCS-87-01426 and MCS-90-01710.
AMS 1991 *subject classifications*. 60C05, 60B10, 60J80.
Key words and phrases. Random tree, Galton–Watson branching process, Brownian excursion, weak convergence.

tion 2.7). The desired general notions of convergence of rescaled discrete trees to limit continuum trees follow easily (Section 3).

Examples of consistent families are given in Section 4. In particular, in Section 4.3 we discuss from the present viewpoint the particular continuum random tree of Aldous (1991a), which we rename the *Brownian CRT*. The special model of that paper can be viewed as the Galton–Watson branching process with Poisson(1) offspring distribution, conditioned on total population size equal to n. The second purpose of this paper is to show (Theorem 23 in Section 5) that the conditioned Galton–Watson branching process with more general (finite variance) offspring distribution converges to the same limit Brownian CRT, up to a scale constant.

This work provides the new mathematics needed for the "big picture" set out in detail in a companion survey paper Aldous (1991b). Here is a one-paragraph summary. Various combinatorial models of random trees correspond to conditioned critical Galton–Watson processes with specific offspring distributions. In particular, the uniform random *ordered* tree on n vertices corresponds to the shifted geometric $(1/2)$ offspring distribution. It has long been known that this particular model can be constructed from simple symmetric random walk on the integers, conditioned on first return to 0 being at time $2n$. This construction makes it simple to show, in this particular model, that asymptotics of functionals of the random tree can be expressed in terms of functionals of Brownian excursion. [Though obvious in retrospect, this connection was overlooked for many years: Louchard (1986) was perhaps the first to note the connection.] It seems intuitively clear that, in some sense, the random tree itself should converge to a limit tree constructible from Brownian excursion, and our results formalize this convergence. The "big picture" of Aldous (1991b) is that there are four distinct ways to look at the Brownian CRT:

1. Via the global construction of Aldous (1991a);
2. Via the general coding from continuous functions (Section 2.7), applied to Brownian excursion;
3. As the particular continuum random tree whose "finite-dimensional distributions" are specified by (33);
4. As the limit of conditioned Galton–Watson trees.

As previously mentioned, several different models of random labelled trees studied in the combinatorial literature are different special cases of conditioned Galton–Watson trees. A consequence of our Theorem 23 is that many functionals of these random trees have limiting distributions expressible in terms of Brownian excursion. This is discussed in detail in Aldous (1991b), Section 3.

2. Theory for general random tree models.

2.1. *Graph-theoretic trees.* To start with, let "tree" have its graph-theoretic meaning, involving a finite set of vertices and edges. Regard trees as *rooted*, that is, one vertex is distinguished and called the root. Picturing a "family tree," we can use the language of families. A vertex is an individual, the root being the progenitor. Each individual has some number of offspring (maybe 0), and each individual (except the root) has one parent. Initially we regard trees as unordered (= nonplanar), which means that we do not distinguish birth orders (first, second, ...) of offspring of an individual. (The ordered case is discussed in Section 2.6.) With these conventions, there is a finite set T_n of different (nonisomorphic, to be pedantic) trees with n vertices.

In this paper we shall use the word "tree" or "graph-theoretic tree" to mean a tree with one further piece of structure: Each edge e has a strictly positive "weight" or "length" $w(e)$. The *distance* $d(v, x)$ between two vertices v, x is the sum of the edge lengths along the path from v to x. If edge lengths are not specified, we use the natural default convention: Set each edge length equal to 1. Rescaling a tree by c (for constant $c > 0$) means multiplying each edge length by c.

Formally, a tree with n vertices (and hence $n - 1$ edges) could be represented as a point

$$t = (\hat{t}; x_1, \dots, x_{n-1}) \in T_n \times R^{n-1}, \tag{1}$$

where $\hat{t}$ is the corresponding tree without edge lengths, and the x_i are the edge lengths. Representation (1) enables us to formalize the notion of convergence of trees to a limit tree, when the numbers of vertices remain bounded, and to talk about "distributions" and convergence in distribution of random trees to a limit random finite tree.

Let t be a graph-theoretic tree, and let B be a subset of the vertices of t. Then we can define the *reduced subtree* $r(t, B)$ associated with t and B as follows. For vertices $v, v^* \in B$ the paths $(\text{root}, v_1, v_2, \dots, v)$ and $(\text{root}, v_1^*, v_2^*, \dots, v^*)$ share some maximal subpath $(\text{root}, v_1, v_2, \dots, b)$ where $b = b(v, v^*)$ is the *branchpoint* or *last common ancestor* of v and v^*. Let the vertices of $r(t, B)$ be the root of t, the elements of B and all the branchpoints. And let the length $w(u_1, u_2)$ of an edge (u_1, u_2) of $r(t, B)$ be just the distance $d(u_1, u_2)$ in t. In other words, an edge (u_1, u_2) of $r(t, B)$ corresponds to some path $(u_1 = v_1, v_2, \dots, v_l = u_2)$ in t, for which none of the intervening vertices are branchpoints or elements of B, and its length is the sum $\sum_i w(v_i, v_{i+1})$. Define a *proper k-tree* to be a graph-theoretic tree with exactly k leaves labelled $1, \dots, k$ and such that all internal nodes (branchpoints) have exactly two children. The root may have either one or two children. Such a tree has exactly $k - 1$ branchpoints, and has either $2k - 1$ or $2k - 2$ edges (depending on the degree of the root). A reduced subtree $r(t, B)$ with $|B| = k$ will be a proper k tree if the branchpoints are distinct and if B contains no ancestor-descendant pair. Figure 1 illustrates a reduced subtree which is a proper 5-tree.

Consider the following setup. For each n we have a random tree $\mathscr{T}_n$ on n vertices. (Imagine the edge lengths getting smaller as n increases.) Let $(V_{n,1}, \dots, V_{n,n})$ be a uniform random ordering of the vertices of $\mathscr{T}_n$. For $k < n$

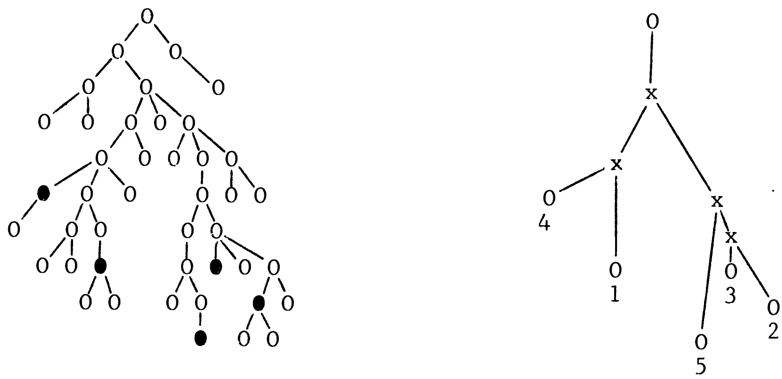

FIG. 1.

consider the reduced subtree

$$r(\mathscr{T}_n, \{V_{n,1}, \dots, V_{n,k}\})$$

in which the vertex $V_{n,i}$ is labelled i. Suppose that (for each k) as $n \to \infty$, the random trees $r(\mathscr{T}_n, \{V_{n,1}, \dots, V_{n,k}\})$ converge in distribution to a limit $\mathscr{R}(k)$ which is a random proper k-tree. [Here convergence in distribution has its natural meaning, formalized using (1).] Then the limit family $(\mathscr{R}(k);\ k \geq 1)$ must satisfy the following consistency conditions. First, for each k the labels $1, \dots, k$ of the leaves are exchangeable. Second, if we pick a leaf of $\mathscr{R}(k)$ at random, and remove the edge and branchpoint connecting that leaf to the remaining tree, then the remaining tree is distributed as $\mathscr{R}(k-1)$. These conditions are combined in the definition below.

DEFINITION 1. Let $(\mathscr{R}(k);\ k \geq 1)$ be a family of random proper k-trees. For $j \leq k$ let $(L_1^k, \dots, L_j^k)$ be uniform random choice of j distinct leaves of $\mathscr{R}(k)$. The family is *consistent* if, for each $1 \leq j \leq k < \infty$,

$$r\left(\mathscr{R}(k), \left\{L_1^k, \dots, L_j^k\right\}\right) =_d \mathscr{R}(j).$$

It is convenient to state here another condition.

DEFINITION 2. With the preceding notation, call the family $(\mathscr{R}(k),\ k \geq 1)$ *leaf tight* if

$$\min_{2 \leq j \leq k} d\left(L_1^k, L_j^k\right) \to_p 0 \quad \text{as } k \to \infty.$$

In words, "leaf tight" means the distance from a random leaf to the nearest of k random-chosen leaves tends to 0 as $k \to \infty$.

For the remainder of Section 2 we forget about the family $(\mathscr{T}_n)$ and just consider a consistent family $(\mathscr{R}(k))$ of random proper k-trees. We want to say that the family determines some "continuum random tree." One could talk abstractly about a projective limit $\mathscr{R}(\infty)$, but our goal is to give a concrete representation of a limit.

2.2. *Trees as subsets of* l_1. To discuss convergence of rescaled trees as the number of vertices $\to \infty$, we use the idea of l_1 representations introduced in Aldous (1991a). Consider a tree t with vertices (v_i). It is easy to see (an explicit construction is given below) that there exist points (v_i^*) in l_1 such that $\text{root}^* = 0$ and

$$\|v_i^* - v_j^*\| = d(v_i, v_j) \quad \text{for all } i, j, \tag{2}$$

where $\|\cdot\|$ is l_1 distance. For each edge (v_i, v_j) of t we can create a path in l_1 from v_i^* to v_j^* of length $\|v_j^* - v_i^*\|$ (in general there are many such possible paths). Let $S(t)$ be the closed subset of l_1 consisting of the vertices (v_i^*) and the connecting paths. Call $S(t)$ a *set representation* of t. Let $\mu(t)$ be the empirical probability distribution on (v_i^*). Call $\mu(t)$ a *measure representation* of t. We can now talk about convergence of set representations by using the Hausdorff topology on closed sets, and about convergence of measure representations by using weak convergence of probability measures.

REMARKS. In Aldous (1991a) we defined the set representation to include only the vertex set, not the edges. The present definition seems more convenient in general, and the arguments of Aldous (1991a) are unaffected by the change. One difficulty is that there seems no clean way to make representations include the order structure of the trees, in the finite case. But we shall see that in asymptotic results, the order structure can be included in the representations.

The sequential construction. This is the construction used in Aldous (1991a) for the special model considered there. Loosely, the construction is "add successive branches orthogonally." Let t be a graph-theoretic tree, and let $(w_1, \dots, w_k)$ be an enumeration of all the leaves of t. We embed the vertices (v_i) of t as points (v_i^*) of l_1 as follows. Let (z_i) be the unit vector basis of l_1. For the vertices $v_{1,i}$ in the path $(\text{root}, v_{1,1}, v_{1,2}, \dots, v_{1,m} = w_1)$ from the root to w_1, let $v_{1,i}^* = d(\text{root}, v_{1,i})z_1$. Inductively on j, regard $r(t, \{w_1, \dots, w_j\})$ as constructed from $r(t, \{w_1, \dots, w_{j-1}\})$ by adding a branch $(v_{j,0}, v_{j,1}, \dots, v_{j,m} = w_j)$ for which only $v_{j,0}$ is in $r(t, \{w_1, \dots, w_{j-1}\})$. Now $v_{j,0}^*$ has already been defined, so we may define $v_{j,i}^* = v_{j,0}^* + d(v_{j,0}, v_{j,i})z_j$. Eventually each vertex has been embedded. It is easy to check the "isometry" property (2). By including the straight-line edges $(v_{j,i}^*, v_{j,i+1}^*)$ we get a set representation $S(t)$ and a measure representation $\mu(t)$.

Given $S(t)$, write $[[x, y]]$ for the path in $S(t)$ from vertex x to vertex y. Note that we can define the notion of *reduced subtree* in terms of set representations $S(t)$: If B is a subset of vertices of $S(t)$, then

$$\bar{r}(S(t), B) = \bigcup_{x \in B} [[0, x]]. \tag{3}$$

This construction has some special properties, which are useful in later proofs (but not relevant to applications of our results). For $x \in l_1$, define the *special path* $[[0, x]]_{\mathrm{sp}}$ from 0 to x to be the piecewise linear path through

$$\begin{gathered} 0 \\ (x_1, 0, 0, 0, \dots) \\ (x_1, x_2, 0, 0, \dots) \\ (x_1, x_2, x_3, 0, \dots) \\ \dots . \end{gathered}$$

In the sequential construction, the path $[[0, x]]$ in $S(t)$ from 0 to a vertex x is always the special path $[[0, x]]_{\mathrm{sp}}$. So we can write

$$S(t) = \bigcup_{1 \le i \le k} [[0, w_i^*]]_{\mathrm{sp}}, \tag{4}$$

where (w_i^*) are the leaves of the embedded tree. Another special property is that the branchpoint $b = b(x, y)$ of vertices x, y of $S(t)$ is given by the coordinatewise minimum

$$b_i = x_i \wedge y_i. \tag{5}$$

2.3. *Continuum trees.* The purpose of embedding into l_1 is to be able to formalize the notion of trees with infinitesimally short edges.

DEFINITION. Let $S \subset l_1$ be a closed subset containing 0 and such that, for $x, y \in S$, there exists a unique path $[[x, y]]$ in S connecting x to y, and that path has length $\|x - y\|$. (Here a *path* is non-self-intersecting.) Such a set can be regarded as a topological tree with root 0. Note that the definition (3) of a reduced subtree $\bar{r}(S, B)$ still makes sense for a continuum tree, as does the notion of the branchpoint $b = b(x, y)$ of two points $x, y \in S$. We want the tree to be *binary*, and one way to state this is as follows.

(a) If $x_1, x_2, x_3 \in S$ are such that $b(x_1, x_2) = b(x_1, x_3) = b(x_2, x_3) = b$, say, then at least one of $\{x_1, x_2, x_3\}$ equals b.

Say $x \in S$ is in the *skeleton* of S if

$$x \in [[0, y[[\quad \text{for some } y \in S,$$

where $[[0, y[[$ denotes the path from 0 to y, excluding the endpoint y. If not, we mix metaphors and say x is a *leaf* of S.

Now let μ_0 be a nonatomic probability measure on l_1, related to S by the following:

(b) $\mu_0\{x\colon x \text{ is a leaf of } S\} = 1$.
(c) $\mu_0\{y\colon x \in [[0, y]]\} > 0$ for each x in the skeleton of S.

Then we say that the pair (S, μ_0) is a *continuum tree*.

Technical Remarks.

1. Such a tree must have uncountably many leaves, by (b) and nonatomicity.
2. If we just have a set S satisfying the conditions in the first part of the definition, up to and including (a), say S is a continuum tree set.
3. If the requirements except (c) are met, then we can make (c) true by pruning away sets $\{y: x \in [[0, y[[\}$ of zero measure.
4. As stated in Section 2.2, it is often convenient to work with *special* continuum trees, that is, those with the property:

(d) For each $x \in S$, the path $[[0, x]]$ in S from 0 to x is the special path $[[0, x]]_{\text{sp}}$.

So far, S and μ_0 have been deterministic. By a continuum random tree (a phrase more euphonic than the logical "random continuum tree") we mean a random pair $(\mathscr{S}, \mu)$ such that each realization $(\mathscr{S}(\omega), \mu(\omega))$ is a continuum tree.

If μ is a random probability measure on l_1 and $(Z_i;\ i \geq 1)$ are l_1 valued r.v.'s, the phrase "(Z_i) is an exchangeable sequence directed by μ" means that, conditional on $\mu = \mu_0$, the sequence (Z_i) is i.i.d. with distribution μ_0.

We now state the main result of Section 2.

THEOREM 3. (i) *Let $(\mathscr{R}(k);\ k \geq 1)$ be a consistent family of proper k-trees. Suppose the leaf-tight property (Definition 2) holds. Then there exists a special continuum random tree $(\mathscr{S}, \mu)$ with the following property. Let (Z_i) be an exchangeable sequence in l_1, directed by the random measure μ. Then for each k,*

$$\bar{r}(\mathscr{S}, \{Z_1, \dots, Z_k\}) \text{ is a set representation of } \mathscr{R}(k). \tag{6}$$

(ii) *Conversely, let $(\mathscr{S}, \mu)$ be a continuum random tree. Then* (6) *defines a family of graph-theoretic random trees $(\mathscr{R}(k))$, and this family is consistent and leaf-tight.*

REMARKS. When (6) holds, say $(\mathscr{S}, \mu)$ is a continuum random tree representing $(\mathscr{R}(k))$. Part (ii) is immediate from the definition of continuum tree: Conditions (a) and (b) and nonatomicity imply that each $\mathscr{R}(k)$ is a.s. proper, and the leaf-tight property follows from the result that an i.i.d. sequence is a.s. dense in its support. We included part (ii) merely for reassurance that the technical definition of "continuum tree" is exactly the right definition. The content of the theorem is part (i), and this will be proved in the next section. The proofs rest upon the exchangeability structure implicit in the consistency assumption. The simple exchangeability results we use are collected in Section 2.5. Cautionary remarks on the extent of applicability of Theorem 3 are given in Section 3.3.

From Theorem 3, it is easy to see relations between properties of $(\mathscr{R}(k))$ and properties of $(\mathscr{S}, \mu)$. We state some of these now.

In general, the conditions of Theorem 3 do not imply that $\mathscr{S}$ is a.s. a bounded or locally compact set. (See the technical example at the end of this section.) In fact it is clear that, in the terminology of Definition 1, the condition

$$\max_{1 \le i \le k} d(\text{root}, L_i^k) \text{ is tight} \quad \text{as } k \to \infty$$

is necessary and sufficient for support(μ), and hence $\mathscr{S}$, to be a.s. a bounded set. Similarly, define

$$D(n,k) = \max_{1 \le j \le n} \min_{1 \le i \le k} d(L_j^n, L_i^n).$$

Then $D(n,k)$ is stochastically increasing in n, so $D(n,k) \to_d D(\infty, k)$, say. The condition

$$(7) \qquad D(\infty, k) \to_d 0 \quad \text{as } k \to \infty$$

is necessary and sufficient for support(μ), and hence for $\mathscr{S}$, to be a.s. compact. To see why, consider an i.i.d. sequence (X_i) with distribution μ_0. Since an i.i.d. sequence is a.s. dense in its support, it is easy to see that the property "μ_0 has compact support" is equivalent to

$$(8) \qquad \sup\left\{\varepsilon: \mu_0\left\{x: \min_{1 \le i \le k} d(x, X_i) > \varepsilon\right\} > 0\right\} \to_d 0 \quad \text{as } k \to \infty.$$

By Theorem 3, $D(\infty, k)$ is distributed as

$$\sup\left\{\varepsilon: \mu\left\{x: \min_{1 \le i \le k} d(x, Z_i) > \varepsilon\right\} > 0\right\}.$$

So from (8) and de Finetti's theorem, $D(\infty, k) \to_d 0$ if and only if support(μ) is a.s. compact.

Return to the setup of a deterministic continuum tree (S, μ_0) at the beginning of the section. An obvious but fundamental fact about continuum tree-sets S is the following. Fix $x, y \in S$. The paths $[[0, x]]$ and $[[0, y]]$ coincide on $[[0, b]]$, where $b = b(x, y)$ is the branchpoint. The path $[[x, y]]$ is the union of the paths $[[0, x]] \setminus [[0, b]]$ and $[[0, y]] \setminus [[0, b]]$, and so has length

$$\|x - y\| = \|x - b(x,y)\| + \|y - b(x,y)\|.$$

So if x and y are close, then the paths from the root to x and to y coincide, except where near to x, and

$$(9) \qquad \|x - b(x,y)\| \le \|x - y\|.$$

More generally, it is easy to see:

LEMMA 4. *Let S be a continuum tree set. For all $x, x', y, y' \in S$,*

$$\|b(x,y) - b(x',y')\| \le \max(\|x - y\|, \|x' - y'\|).$$

Here are two typical technical uses of these facts.

LEMMA 5. *Let S be a continuum tree set. Let (z_i) in S be such that S is the closure of $\bar{r}(S,\{z_i\})$. Then the skeleton of S is* $\bigcup_{1\le i<\infty}[[0,z_i[[$.

PROOF. Fix $y \in S$. There is a subsequence z_{j_i} and $y_{j_i} \in [[0, z_{j_i}]]$ such that $y_{j_i} \to y$. The branchpoints $b(y, y_{j_i})$ lie in $[[0, y]]$ and (9) implies $b(y, y_{j_i}) \to y$. So

$$[[0,y[[\subseteq \bigcup_i [[0,y_{j_i}[[\subseteq \bigcup_i [[0,z_i]]. \qquad \square$$

LEMMA 6. *Let S be a compact continuum tree set. Then condition* (c) *of the definition of continuum tree is equivalent to*

$$\{x\colon x \text{ is a leaf of } S\} \subseteq \text{support}(\mu_0).$$

PROOF. Suppose (c) holds. Let z be a leaf of S. For small $\varepsilon > 0$ let $z(\varepsilon)$ be the point of $[[0, z]]$ with $\|z(\varepsilon) - z\| = \varepsilon$. By (c) there exists $y(\varepsilon) \in \text{support}(\mu_0)$ such that $z(\varepsilon) \in [[0, y(\varepsilon)]]$. By compactness there is a subsequence $y(\varepsilon_n) \to y$, say. Now

$$\|z - b(z, y(\varepsilon_n))\| \le \varepsilon,$$

$$b(z, y(\varepsilon_n)) \to b(z, y) \quad \text{by Lemma 4}$$

and so $z = b(z, y)$. That is, $z \in [[0, y]]$. But z is a leaf, so $y = z$ and so $z = \lim_n y(\varepsilon_n) \in \text{support}(\mu_0)$.

Conversely, suppose (c) fails at some point x in the skeleton of S. By compactness, we can choose y which attains the maximum of $\|y\|$ subject to $x \in [[0, y]]$. Then y is a leaf of S, but y is not in the support of μ_0. $\square$

REMARK. Lemma 6 and condition (b) imply that for a continuum tree with S compact,

$$\text{support}(\mu_0) = \overline{\{\text{leaves of } S\}}.$$

In general, condition (b) implies $\text{support}(\mu_0) \subseteq S$. Botanical trees rarely have their trunks completely covered by leaves: Mathematically, this is the property

$$(10) \qquad S = \text{support}(\mu_0).$$

Call this property "leaf-dense." It plays a role in Section 2.7.

Technical example. Informally, let (S_n, μ_n), $n \ge 1$ be copies of some continuum tree, with μ_n reweighted to have total mass 2^{-n}. Then make a continuum tree rooted at 0 containing an infinite path $[0, \infty)$ with S_n attached at point n on the infinite path. This gives an unbounded tree. If instead of the infinite path we take a path $[0, 1]$ and attach S_n at $1 - 1/n$, then the resulting tree is not locally compact because of orthogonality of branches.

2.4. *Proof of Theorem* 3. Fix a consistent family $(\mathscr{R}(k))$. Let $(L_1^k, \dots, L_k^k)$ be a uniform random ordering of the k leaves of $\mathscr{R}(k)$. Apply the sequential construction of Section 2.2 to build a set-representation $S(k)$ of $\mathscr{R}(k)$. By consistency (and the Kolmogorov extension theorem) we can do this simultaneously for all k. Precisely, there exist random elements (L_i) of l_1 such that for each k,

$$S(k) = \bigcup_{1 \le i \le k} [[0, L_i]]_{\mathrm{sp}} \tag{11}$$

is a set representation of a random graph-theoretic tree distributed as $\mathscr{R}(k)$. Here we are using the special property (4) of the sequential construction.

LEMMA 7. *Suppose the family $(\mathscr{R}(k))$ is leaf tight. Let $\nu^k(\omega, \cdot)$ be the empirical distribution of $(L_1, \dots, L_k)$. Then*

$$\nu^k(\omega, \cdot) \to \mu(\omega, \cdot) \quad \text{a.s. as } k \to \infty, \text{ for some random measure } \mu. \tag{12}$$

Here ν^k is a random probability measure on l_1, so convergence is weak convergence of probability measures.

PROOF. Let π_j: $l_1 \to l_1^j$ be the projection map $x \to (x_1, \dots, x_j)$ onto the first j coordinates. We first quote a standard weak convergence fact [c.f. Billingsley (1968), Theorem 4.2]. Let (β_k) be probability measures on l_1. Suppose we know $\lim_k \pi_j \beta_k$ exists, for each j. In order that $\lim_k \beta_k$ exists, it is necessary and sufficient that

$$\lim_{j \to \infty} \limsup_k \beta_k\{x: \|x - \pi_j x\| > \varepsilon\} \to 0 \quad \text{for each } \varepsilon > 0. \tag{13}$$

For $j < k$, $\pi_j L_k$ is the point at which the path in $\mathscr{R}(k)$ from 0 to L_k branches away from the subtree $\mathscr{R}(j)$. Now fix j. Then using consistency we see that $(\pi_j L_k;\ k > j)$ is exchangeable, and so by de Finetti's theorem (see Section 2.5) has a directing random measure $\theta_j(\omega, \cdot)$, say, for which

$$\pi_j \nu^k(\omega, \cdot) \to \theta_j(\omega, \cdot) \quad \text{a.s. as } k \to \infty.$$

So by (13), to prove (12) it suffices to prove

$$\lim_{j \to \infty} \limsup_k \frac{1}{k} \sum_{i=j+1}^{k} 1_{(\|L_i - \pi_j L_i\| > \varepsilon)} = 0 \quad \text{a.s.} \tag{14}$$

By exchangeability and de Finetti's theorem, the lim sup equals $P(\|L_{j+1} - \pi_j L_{j+1}\| > \varepsilon | \theta_j)$. The lim sup in (14) is decreasing in j, so to prove a.s. convergence to 0 as $j \to \infty$ it suffices to prove the expectation converges to 0, that is, it suffices to prove

$$P\big(\|L_{j+1} - \pi_j L_{j+1}\| > \varepsilon\big) \to 0 \quad \text{as } j \to \infty.$$

But

$$\begin{aligned} \|L_{j+1} - \pi_j L_{j+1}\| &\le \min_{1 \le i \le j} \|L_{j+1} - L_i\| \\ &=_d \min_{1 \le i \le j} d\big(L_{j+1}^{j+1}, L_i^{j+1}\big) \\ &=_d \min_{2 \le i \le j+1} d\big(L_1^{j+1}, L_i^{j+1}\big), \end{aligned}$$

using exchangeability. Thus the leaf-tight property (Definition 2) implies Lemma 7. □

The sequence (L_i) in (11), obtained from the sequential construction, is certainly not exchangeable in l_1 (e.g., because L_1, but no other L_i, is a multiple of z_1). But we can improve matters with a rerandomization argument.

LEMMA 8. *In the setting of Lemma 7, let (Z_i) be an exchangeable sequence in l_1, directed by the random measure μ. Then for each k,*

$$S_0(k) = \bigcup_{1 \le i \le k} [[0, Z_i]]_{\mathrm{sp}}$$

is a set representation of $\mathscr{R}(k)$.

PROOF. Let $S(m)$ be the special set representation of $\mathscr{R}(m)$ given by (11). For $k < m$ let

$$S(k, m) = \bar{r}\big(S(m), \{L_{\pi_m(1)}, \dots, L_{\pi_m(k)}\}\big),$$

where $(\pi_m(\cdot))$ is a uniform random permutation of $(1, \dots, m)$. By consistency, $S(k, m)$ is a set representation of $\mathscr{R}(k)$. Letting $m \to \infty$, Lemma 7 and Lemma 9 (stated in the next section) imply that the position of the k leaves of $S(k, m)$ converges in distribution to the position of the k leaves of $S_0(k)$. But by (5) the positions of leaves determine the positions of branchpoints, and so the positions of all the vertices of $S(k, m)$ converge to the positions of all the vertices of $S_0(k)$. Hence $S_0(k)$ is indeed a set representation of $\mathscr{R}(k)$. □

PROOF OF THEOREM 3. Now define $\mathscr{S}$ to be the closure of $\bigcup_k S_0(k)$, that is, the closure of $\bigcup_{1 \le i < \infty} [[0, Z_i]]_{\mathrm{sp}}$. The issue is to show that $(\mathscr{S}, \mu)$ is a random continuum tree, and then the required property (6) is immediate from Lemma 8.

First go back to the construction (11) and fix a realization of (L_i). In what follows we omit "a.s.," where no subtleties are involved. Write $S(\infty) = \bigcup_{1 \le i < \infty} [[0, L_i]]_{\mathrm{sp}}$ and let $\mathscr{S}^*$ be the closure of $S(\infty)$. In the sequential construction, the jth branch (i.e., the edge connecting L_j to the preexisting tree) attaches to some point in the kth branch, for some $k < j$. Write $k = \phi(j)$ to describe this relationship. It is intuitively clear that each point in $\mathscr{S}^*$ can be associated with the sequence of branches used (partially) by the path from the root to that point. It is not hard to verify the following precise formulation. Each $x \in \mathscr{S}^* \setminus S(\infty)$ corresponds to a sequence $(1 = j_1, j_2, j_3, \dots)$ with $\phi(j_i) = j_{i-1}$ and with $\|b(L_{j_i}, L_{j_{i-1}})\|$ bounded. The "correspondence" is that the special path $[[0, x]]_{\mathrm{sp}}$ is the piecewise linear path through the branchpoints $b(L_{j_i}, L_{j_{i-1}})$. From this correspondence, it is easy to see that $\mathscr{S}^*$ has the required "unique paths" property of a continuum tree. And the "binary" property (a) follows from the fact that each $\mathscr{R}(k)$, and hence $S(\infty)$, is a binary tree.

By Lemma 7, our realization of (L_i) has limit empirical distribution μ_∞, say, a realization of μ. In Lemma 8 the (Z_i) are i.i.d. with distribution μ_∞. Because each point of support(μ_∞) is a limit point of the (L_i), we see that each Z_i is in $\mathscr{S}^*$. In other words, the $\mathscr{S}$ defined above is a reduced subtree of $\mathscr{S}^*$, and so inherits the properties we established above for $\mathscr{S}^*$. (In fact $\mathscr{S} = \mathscr{S}^*$, but we can avoid having to prove that fact.) It remains to prove properties (b) and (c) of *continuum tree*. By the definition of "proper k-tree," μ must be nonatomic and

$$P(Z_1 \in [[0, Z_i[[) = 0.$$

Then by Lemma 5,

$$P(Z_1 \text{ is in the skeleton of } \mathscr{S}) = 0.$$

This implies (b). To prove (c), the fact that Z_i is a.s. in support(μ) and that μ is nonatomic implies that for $\varepsilon > 0$,

$$\mu\{x: \|x - Z_i\| < \varepsilon\} > 0 \quad \text{a.s.}$$

Write $a(Z_i, \varepsilon)$ for the point $a \in [[0, Z_i]]$ such that $\|a - Z_i\| = \min(\varepsilon, \|Z_i\|)$. Then by (9), for each x with $\|x - Z_i\| < \varepsilon$ we have $a(Z_i, \varepsilon) \in [[0, x]]$. Thus

$$\mu\{x: a(Z_i, \varepsilon) \in [[0, x]]\} > 0 \quad \text{a.s.}$$

Letting $\varepsilon \to 0$ and applying Lemma 5 gives (c). □

2.5. *Exchangeability lemmas.* Let μ be a random probability measure. Let (Z_i) be random variables whose joint distribution, conditioned on $\mu = \mu_0$, is i.i.d. (μ_0). In the terminology of Aldous (1985), (Z_i) is an exchangeable sequence directed by μ. One statement of de Finetti's theorem [e.g., Aldous (1985), Section 3] asserts that every infinite exchangeable sequence is of this form, for some directing random measure.

Here is a standard weak convergence result for exchangeable processes [see, e.g., Kallenberg (1973) or Aldous (1985), Proposition 7.20].

LEMMA 9. *Let $(Z_i,\ i \ge 1)$ be an exchangeable sequence directed by a random measure μ. For each n let $(\xi_1^n, \dots, \xi_n^n)$ be exchangeable with empirical distribution μ_n. Then the following are equivalent*:

(i) $$(\xi_1^n, \dots, \xi_k^n) \to_d (Z_1, \dots, Z_k) \quad \textit{as } n \to \infty, \textit{ for each } k.$$

(ii) $$\mu_n \to_d \mu.$$

The next result is well known in various guises.

LEMMA 10. *Let $\prec$ be a random linear ordering of $\{1, 2, 3, \dots\}$ such that, for each k, the $k!$ orderings of $\{1, \dots, k\}$ are equally likely. Then there exist independent random variables (U_i) distributed uniformly on $(0, 1)$ such that*

$$i \prec j \quad \textit{if and only if } U_i < U_j.$$

OUTLINE OF PROOF. Let $U_{n,i}$ be the increasing rank of i in the ordering $\prec$ on $\{1, 2, \dots, n\}$. Then $U_i = \lim_n n^{-1} U_{n,i}$ exists a.s. and has the asserted properties.

2.6. *Ordered trees.* Often it is natural to regard graph-theoretic trees as ordered (= planar). That is, we do distinguish birth orders (first, second, ...) of offspring of an individual. Of course, in such a case we could ignore the order structure and apply the previous theory of unordered continuum trees $(\mathscr{S}, \mu)$. But it would be preferable to include the order structure in $(\mathscr{S}, \mu)$.

In a graph-theoretic tree, the order defines a linear order $\prec$ on the vertices of t, as follows. Say $v_1 \prec v_2$ if either: (i) v_1 is on the path from the root to v_2; or (ii) at the branchpoint $b(v_1, v_2)$, the edge leading toward v_1 is earlier in the ordering than the edge leading toward v_2.

Conversely, one could specify an *order* on the tree by specifying a linear order on the vertices, though the linear order must satisfy certain compatibility conditions (which we leave to the reader to formulate).

Now reconsider Lemma 8, regarding the $(\mathscr{R}(k))$ as ordered trees. The linear ordering $\prec$ on the vertices of $\mathscr{R}(k)$ induces a linear ordering (which we also call $\prec$) on $(Z_1, \dots, Z_k)$, and hence on $(Z_i, 1 \le i < \infty)$. It is easy to show this extends to a linear order on $\mathscr{S}$.

Thus we can define a (deterministic) ordered continuum tree to be a triple $(\mathscr{S}, \mu_0, \prec)$, where $\prec$ is a linear order on $\mathscr{S}$ which satisfies certain compatibility conditions. But there is an alternative way, well-known in the combinatorial literature, of looking at ordered trees, which we now discuss.

Coding ordered trees as walks. Let t be an ordered graph-theoretic tree on n vertices. Then we can define $\tilde{f}$: $\{1, 2, \dots, 2n-1\} \to \{\text{vertices of } t\}$, which we regard as a walk ("depth-first search") around t, as follows.

$\tilde{f}(1) = \text{root}$.

Given $\tilde{f}(i) = v$, choose, if possible, the first (in the ordering) child w of v which has not already been visited, and let $\tilde{f}(i+1) = w$. If not possible, let $\tilde{f}(i+1)$ be the parent of v.

This walk traverses each edge once in each direction. The walk gives the order structure on the vertices: $v_1 \prec v_2$ if and only if $\tilde{f}^{-1}(v_1) < \tilde{f}^{-1}(v_2)$, where $\tilde{f}^{-1}(v) = \min\{t\colon \tilde{f}(t) = v\}$.

Next we define the *search-depth* function f: $\{1, 2, \dots, 2n-1\} \to [0, \infty)$ by:

$$(15) \qquad f(i) = d\big(\text{root}, \tilde{f}(i)\big), \qquad 1 \le i \le 2n-1.$$

It is easy to check that the function f determines the ordered tree t. This construction suggests two possibilities for continuum trees.

(a) That an ordered continuum tree $(\mathscr{S}, \mu_0, \prec)$ might more simply be described via a function $\tilde{f}$: $[0,1] \to l_1$, where $\mathscr{S}$ is the range of $\tilde{f}$, μ_0 is the measure induced by $\tilde{f}$ from Lebesgue measure and $\prec$ is the induced ordering.

(b) In terms of the search-depth function f, the distance between leaves of t is given by

$$d\big(\tilde{f}(t_1), \tilde{f}(t_2)\big) = \Big(f(t_1) - \min_{t_1 \le t \le t_2} f(t)\Big) + \Big(f(t_2) - \min_{t_1 \le t \le t_2} f(t)\Big).$$

So instead of using l_1 to describe the idea of a continuum tree, one could perhaps define such a tree via a real-valued function f, thinking of the vertices being labelled by $t \in (0,1)$ and the distances between vertices being given by the formula above.

The next section gives the mathematics connecting this approach to our previous theory; further discussion is in Section 3.3. The idea of constructing continuous trees from continuous functions was treated more briefly in Le Gall (1991).

For later use, we record two simple lemmas for ordered graph-theoretic trees. The first is straightforward.

Lemma 11. *For vertices $v_1 \prec v_2 \prec v_3$ in an ordered graph-theoretic tree,*

$$d(v_1, b(v_1, v_2)) \le d(v_1, b(v_1, v_3)).$$

The same holds in an ordered continuum tree, with $d(\cdot, \cdot)$ replaced by $\| \; \|$.

Next, we need a little care in talking about "random vertices" and "random times on the walk." If U_n is uniform on $\{1, 2, \dots, 2n-1\}$ and $\tilde{f}$ is the walk around t, then $\tilde{f}(U_n)$ is a random vertex of t, but is not a *uniform* random vertex. Fortunately it can be almost matched with a uniform random vertex, as follows.

Lemma 12. *Let $\tilde{f}$ be the walk around an ordered graph-theoretic tree t on n vertices. Then we can construct (U, V) such that the following hold:*

(i) *U is uniform on $\{1, 2, \dots, 2n-1\}$.*
(ii) *V is uniform on the vertices of t.*
(iii) *$P(d(\tilde{f}(U), V) > d^*) \le 1/(2n-1)$, where d^* is the maximum edge length in t.*

Proof. Define $\theta(0) = \theta(2n-1) = \text{root}$, and for $1 \le i \le 2n-2$ define

$$\begin{aligned}\theta(i) &= \tilde{f}(i+1) \quad \text{if } \tilde{f}(i+1) \text{ is child of } \tilde{f}(i),\\ &= \tilde{f}(i) \quad \text{if } \tilde{f}(i+1) \text{ is parent of } \tilde{f}(i).\end{aligned}$$

Then each vertex v occurs as $\theta(i)$ for exactly two i's. So if U^* is uniform on $\{0, 1, 2, \dots, 2n-1\}$, then $\theta(U^*)$ is uniform on the vertices of t. And

$$d\big(\theta(U^*), \tilde{f}(U^*)\big) \le d^* \quad \text{provided } U^* \ne 0.$$

Specifying U so that $U = U^*$ on $\{U^* \ge 1\}$, the result follows. □

2.7. *Trees and excursions.* Let $f\colon [0,1] \to [0,\infty)$ satisfy:

CONDITION 1. (i) $f(0) = f(1) = 0$, and $f(t) = 0$ for at most one value $0 < t < 1$.

(ii) f is continuous.

(iii) The set of times of strict local minima is dense.

(iv) If $t_1 < t_2$ are strict local minima with $f(t_2) = f(t_1)$, then

$$\inf_{t_1 < t < t_2} f(t) < f(t_1).$$

(v) The set of times of one-sided local minima has Lebesgue measure 0.

Saying t is a one-sided local minimum means there exists $\varepsilon > 0$ such that $f(t) = \inf\{f(s)\colon t \le s \le t + \varepsilon\}$ or $f(t) = \inf\{f(s)\colon t - \varepsilon \le s \le t\}$.

For future reference note that if $t_2 \in (0,1)$ is the time of a strict local minimum then there exists an increasing triple (t_1, t_2, t_3) satisfying:

(vi) $f(t_1) = f(t_2) = f(t_3)$, $f(t) > f(t_1)$ on $(t_1, t_2) \cup (t_2, t_3)$.

Note also that for any t_2 and any $0 \le h < f(t_2)$, the increasing pair (t_1, t_3) defined by

$$t_1 = \max\{t < t_2\colon f(t) = h\},$$
$$t_3 = \min\{t > t_2\colon f(t) = h\}$$

satisfies:

(vii) $f(t_1) = f(t_3)$, $f(t) > f(t_1)$ on (t_1, t_3).

THEOREM 13. *Let f satisfy Condition 1. Then there exists a continuous function $\tilde{f}\colon [0,1] \to l_1$ such that*

$$\mathscr{S}_f = \{\tilde{f}(t)\colon 0 \le t \le 1\};\ \mu_f(\cdot) = Leb\{t\colon \tilde{f}(t) \in \cdot\}$$

defines a special continuum tree $(\mathscr{S}_f, \mu_f)$ satisfying

$$\|\tilde{f}(t_2) - \tilde{f}(t_2)\| = \Big(f(t_1) - \min_{t_1 \le t \le t_2} f(t)\Big) + \Big(f(t_2) - \min_{t_1 \le t \le t_2} f(t)\Big), \quad t_1 < t_2. \tag{16}$$

REMARKS. Regard points of $\mathscr{S}_f$ as labelled by $t \in [0,1]$. Note that for a pair (t_1, t_3) satisfying (vii) we have $\tilde{f}(t_3) = \tilde{f}(t_1)$ by (16). In other words, t_1 and t_3 will label the same point of the continuum tree. We may regard $\mathscr{S}_f$ as ordered by setting $s_1 \prec s_2$ if and only if $\tilde{f}^{-1}(s_1) < \tilde{f}^{-1}(s_2)$, where $\tilde{f}^{-1}(s) = \min\{t\colon \tilde{f}(t) = s\}$.

PROOF. For each $\mathbf{t} = (t_1, \dots, t_k)$ with $0 < t_1 < \cdots < t_k < 1$, we shall define an ordered graph-theoretic tree $\mathscr{R}_f(\mathbf{t})$ containing vertices labelled as $(t_1, \dots, t_k)$: A vertex may be multiply labelled, and there may be unlabelled branchpoints. Set $b_i = \min_{t_i \le t \le t_{i+1}} f(t)$. Draw an edge of length $f(t_1)$ and label one end "root" and the other end "t_1." Inductively, from t_i move back a distance $f(t_i) - b_i$ toward the root, draw a new edge of length $f(t_{i+1}) - b_i$,

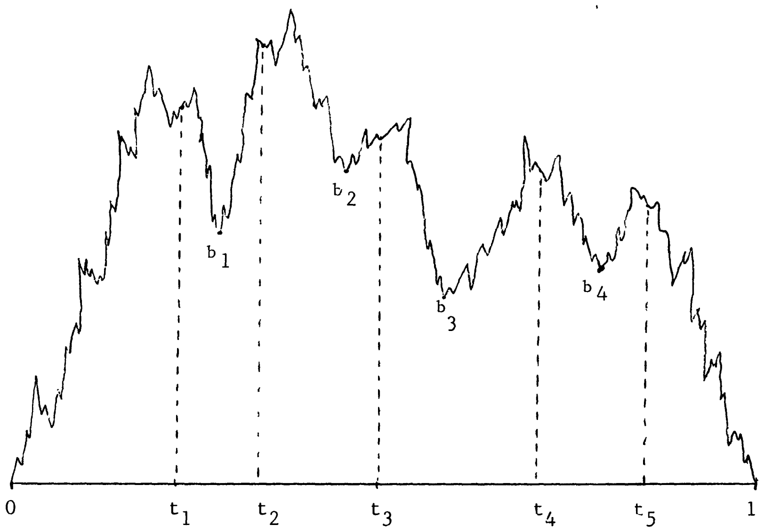

FIG. 2.

label the new endpoint "t_{i+1}" and at the new branchpoint regard the new edge as "after" the preexisting edge in the ordering. This procedure is illustrated in Figures 2 and 3. It is not hard to verify that as $\mathbf{t}$ varies these trees are consistent (under taking reduced subtrees). One can also check by induction that the distances between vertices t_i, t_j of $\mathscr{R}_f(\mathbf{t})$ are given by (16).

LEMMA 14. *Let f satisfy Condition 1. Then there exists a continuous map $\tilde{f}\colon [0,1] \to l_1$ such that $\|\tilde{f}(t)\| = f(t)$ and such that, for each $\mathbf{t} = (t_1, \dots, t_k)$, defining $\mathscr{S}_f(\mathbf{t})$ to be $\bigcup_{t \in \mathbf{t}}[[0, \tilde{f}(t)]]_{\mathrm{sp}}$,*

$$\mathscr{S}_f(\mathbf{t}) \text{ is a set representation of } \mathscr{R}_f(\mathbf{t}).$$

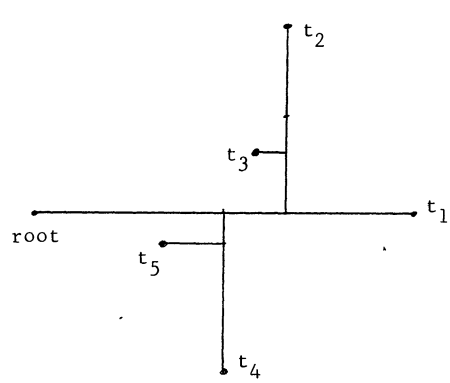

FIG. 3.

Proof of Lemma 14. Take a dense sequence of points (t_i) in $(0,1)$ which are not one-sided local minima of f. Let $\mathbf{t}_k$ be the increasing reordering of $(t_1,\dots,t_k)$. Construct $\mathscr{R}_f(\mathbf{t}_k)$ as before. Then copy the sequential construction in Section 2.2, using the trees $\mathscr{R}_f(\mathbf{t}_k)$ in place of $\mathscr{R}(k)$. This construction gives a sequence of points (v_i) in l_1 such that a set representation of $\mathscr{R}_f(\mathbf{t}_k)$ is given by $\bigcup_{1\le i\le k}[[0,v_i]]_{\mathrm{sp}}$. Consider the set of points $\{(t_i,v_i): 1\le i<\infty\}$ in $[0,1]\times l_1$. Define $\tilde f(t_i)=v_i$. Then $\tilde f$ extends to a continuous function on $[0,1]$, because $\|\tilde f(t_i)-\tilde f(t_2)\|$ is given by (16) and because f is continuous. By construction, the assertion of the Lemma holds for k-tuples $\mathbf{t}$ with entries from (t_i), and by taking limits it holds for all k-tuples. □

Given Lemma 14, we can define $\mathscr{S}_f$ to be $\bigcup_{t\in[0,1]}[[0,\tilde f(t)]]_{\mathrm{sp}}$, and define μ_f to be the probability measure on l_1 induced by $\tilde f$ from Lebesgue measure on $[0,1]$. We must show that the pair $(\mathscr{S}_f,\mu_f)$ is a continuum tree, in the sense of Section 2.3. One path in $\mathscr{S}_f$ from the root to $\tilde f(t_0)$ is the path

$$b\to\tilde f(s_-(b)): 0\le b\le f(t_0),$$

where

$$s_-(b)=\max\{t\le t_0: f(t)=b\}.$$

This is the same as the path

$$b\to\tilde f(s_+(b)): 0\le b\le f(t_0),$$

where

$$s_+(b)=\min\{t\ge t_0: f(t)=b\},$$

because $\tilde f(s_+(b))=\tilde f(s_-(b))$, as remarked after the statement of the theorem. This path has length $f(t_0)$. The "unique paths" property of *continuum tree* now follows easily from the construction: The branchpoint of $\tilde f(t_0)$ and $\tilde f(t_1)$ is $\tilde f(t)$, where t is the time at which the minimum of f over (t_0,t_1) is attained. Moreover the fact that $b\to\tilde f(s_-(b))$ is the path $[[0,\tilde f(t_0)]]_{\mathrm{sp}}$ shows that the definition above of $\mathscr{S}_f$ is equivalent to the definition in the statement of the theorem. Next, note that for a triple (t_1,t_2,t_3) as in (vi), the point $\tilde f(t_0)=\tilde f(t_1)=\tilde f(t_2)$ is a branchpoint of $\mathscr{S}_f$, and each branchpoint arises this way. But the analog of (vi) cannot happen with 4-tuples, by (iv), and this establishes the *binary* property of $\mathscr{S}_f$. Finally, the skeleton of $\mathscr{S}_f$ is the set of points $x=\tilde f(t_1)=\tilde f(t_2)$, for pairs (t_1,t_2) as in (vii): That is to say, the set labelled by the times of one-sided local minima of f. So condition (b) of *continuum tree* follows from (v), and condition (c) is satisfied because the measure in question is at least t_2-t_1. □

Remark. Let us briefly discuss uniqueness in Theorem 13. By analogy with isomorphism of graph-theoretic trees, call continuum trees $(\mathscr{S}_i,\mu_i)$, $i=1,2$ *equivalent* if there is a (nonlinear) isometry from $\mathscr{S}_1$ onto $\mathscr{S}_2$ which maps μ_1 to μ_2 (and preserves order, in the ordered case). Given a deterministic ordered continuum tree which is equivalent to both $(\mathscr{S}_f,\mu_f)$ and to $(\mathscr{S}_g,\mu_g)$, there is an induced isometry between $\mathcal{S}_f$ and $\mathcal{S}_g$. By the measure- and order-preserving property of the isometry, it must be the map $\tilde{f}(t) \to \tilde{g}(t)$, and it follows that $f = g$. This is the uniqueness property for deterministic trees. In the setting of Theorem 15, it follows that the distribution of $(\mathcal{R}(k),\ k \geq 1)$ determines the distribution of the random function f.

It is easy to see that $(\mathcal{S}_f, \mu_f)$ has the "leaf-dense" property (10). We actually need a "one-sided" version of this property for ordered trees. Let

$$\mu_x^+(\cdot) = \mu\{y \in \mathcal{S}: y \in \cdot\,,\ x \prec y\}$$

and similarly for μ_x^-. Say $(\mathcal{S}, \mu_0)$ is *order-leaf-dense* if for each leaf x of $\mathcal{S}$

$$[[0, x]] \subseteq \operatorname{support}(\mu_x^+) \cap \operatorname{support}(\mu_x^-). \tag{17}$$

The next result connects this construction via functions with the general theory. The "intrinsic" condition (iii) corresponding to the existence of representing functions is complicated (see Section 3.3 for further discussion).

THEOREM 15. *Let $(\mathcal{R}(k),\ k \geq 1)$ be a consistent family of random ordered proper k trees. The following are equivalent*:

(i) *There exists a random function f satisfying Condition* 1, *and such that the continuum random tree $(\mathcal{S}_f, \mu_f)$ represents $(\mathcal{R}(k))$ in the sense* (6).

(ii) *The family $(\mathcal{R}(k),\ k \geq 1)$ has some representing ordered continuum random tree $(\mathcal{S}, \mu, \prec)$ which is "order-leaf-dense"* (17) *and for which $\mathcal{S}$ is compact.*

(iii) *The family $(\mathcal{R}(k),\ k \geq 1)$ satisfies the compactness condition* (7) *and also the following conditions*:

(a) *For each $0 < a < 1$ and each $\delta > 0$, the probability of the following event tends to* 1 *as $k \to \infty$*:

$$\exists\, 1 < j \leq k: |d(root, B_j^k) - a d(root, L_1^k)| \leq \delta,$$

$$d(L_j^k, B_j^k) \leq \delta,$$

the branch at B_j^k toward L_j^k is before *that toward L_1^k.*

Here B_j^k is the branchpoint in $\mathcal{R}(k)$ of L_1^k and L_j^k, and before *refers to the order structure of $\mathcal{R}(k)$.*

(b) *The same as* (a), *with* before *replaced by* after.

PROOF. (i) *implies* (iii). It is enough to consider the case where f is nonrandom. Clearly $\mathcal{S}_f$ is compact, so the issue is to check (iii)(a), because (b) is similar. Fix $0 < a < 1$ and $\delta > 0$. Define

$$v(t_0) = \sup\{t < t_0: f(t) = af(t_0)\}.$$

By uniform continuity, there exists $\varepsilon > 0$ such that: for each $0 < t_0 < 1$ and each $t \in (v(t_0) - \varepsilon, v(t_0))$,

$$|f(t) - af(t_0)| \le \delta,$$

$$\left| f(t) - \min_{t \le u \le v(t_0)} f(u) \right| \le \delta.$$

Let (U_i) be i.i.d. uniform on $(0,1)$. Saying that $(\mathscr{S}_f, \mu_f)$ represents $(\mathscr{R}(k))$ is saying that $\mathscr{R}(k)$ is distributed as $\mathscr{R}_f(U_1^k, \dots, U_k^k)$, where $(U_1^k, \dots, U_k^k)$ is the increasing reordering of $(U_1, \dots, U_k)$. The event in (a) holds provided $U_j \in (v(U_1) - \varepsilon, v(U_1))$ for some $1 < j \le k$, and this plainly has chance $\to 1$ as $k \to \infty$.

(iii) *implies* (ii). The compactness condition (7) is stronger than the leaf-tight condition (2). So by Theorem 3 the family $(\mathscr{R}(k))$ has some representing continuum random tree $(\mathscr{S}, \mu)$, and by (7) $\mathscr{S}$ is a.s. compact. As stated in Section 2.6, $\mathscr{S}$ inherits an ordering $\prec$ from the ordering on $(\mathscr{R}(k))$. So the issue is to verify (17). If (z_i) and x are leaves of $\mathscr{S}$, if $z_i \to x$ and if (17) holds for each z_i, then it is easy to check that (17) holds for x. Appealing to Lemma 6, it suffices to verify (17) for a sequence (z_i) of leaves dense in support(μ). Now the exchangeable sequence (Z_i) appearing in Theorem 3 has these properties a.s., so it suffices to verify that (17) holds a.s. for $x = Z_1$. But this is what condition (iii) is set up to do. Indeed, by considering points $Z_2(\omega), \dots, Z_k(\omega)$, which are a.s. in support$(\mu(\omega, \cdot))$, we see that the probability of the following event is at least the probability of the event in (iii)(a):

(18) support(μ_x^-) contains a point y such that
$$-\delta \le \|b(y,x)\| - a\|x\| \le \delta, \qquad \|b(y,x) - y\| \le \delta.$$

So by assumption (iii), this event has probability 1. So a.s. (18) holds for all a and δ, implying $[[0,x]] \subseteq \text{support}(\mu_x^-)$.

(ii) *implies* (i). By (ii), we have a continuum random tree $(\mathscr{S}, \mu)$ such that (6) holds:

$$\bar{r}(\mathscr{S}, \{Z_1, \dots, Z_k\}) \text{ is a set representation of } \mathscr{R}(k).$$

The order structure on $\mathscr{R}(k)$ determines an ordering $\prec$ on $(Z_1, \dots, Z_k)$, and hence on $(1, \dots, k)$. By consistency (Definition 1) all such orderings on $(1, \dots, k)$ are equally likely. Appealing to Lemma 10, there exist i.i.d. uniform $(0,1)$ r.v.'s (U_i) such that

$$Z_i \prec Z_j \quad \text{if and only if } U_i < U_j.$$

Fix a realization of $\mathscr{S}$, μ and (Z_i). Define $\tilde{f}(U_i) = Z_i$. Our main goal is to show that $\tilde{f}$ can be extended to a continuous l_1-valued function.

Suppose that for some t_0,

$$\lim_{U_i \downarrow t_0} Z_i \text{ does not exist}.$$

Then by compactness we can choose $i(m)$, $j(m)$ such that

$$i(m) < j(m) < i(m+1) < \cdots$$

$$U_{i(m)} \downarrow t_0, \qquad U_{j(m)} \downarrow t_0, \qquad Z_{i(m)} \to z_1, \qquad Z_{j(m)} \to z_2 \neq z_1.$$

Applying Lemma 11 to $Z_{i(m+1)} \prec Z_{j(m)} \prec Z_{i(m)}$,

$$\begin{aligned}\|Z_{i(m+1)} - b(Z_{i(m+1)}, Z_{j(m)})\| &\leq \|Z_{i(m+1)} - b(Z_{i(m+1)}, Z_{i(m)})\| \\ &\leq \|Z_{i(m+1)} - Z_{i(m)}\| \\ &\to 0.\end{aligned}$$

So $b(Z_{i(m+1)}, Z_{j(m)}) \to z_1$ and thus by Lemma 4, $b(z_1, z_2) = z_1$. But the same argument applied to $Z_{j(m+1)} \prec Z_{i(m+1)} \not\prec Z_{j(m)}$ gives $b(z_2, z_1) = z_2$.

This argument, and its symmetric version, show that we may assume that $\tilde{f}$ has right-hand and left-hand limits everywhere.

Now fix t_0 and consider

$$z_1 = \lim_{t \uparrow t_0} \tilde{f}(t), \qquad z_2 = \lim_{t \downarrow t_0} \tilde{f}(t).$$

Now $\|b(z_1, z_2)\| \leq \min(\|z_1\|, \|z_2\|)$, and if $\|b(z_1, z_2)\| = \|z_1\| = \|z_2\|$, then $b(z_1, z_2) = z_1 = z_2$, and so $\tilde{f}$ is continuous at t_0. To argue by contradiction, suppose $\|b(z_1, z_2)\| < \|z_2\|$, say. Choose $a < 1$ such that

$$\|b(z_1, z_2)\| < a\|z_2\|. \tag{19}$$

By (17) and the density of (Z_j) in support(μ), for each Z_i with $U_i > t_0$ there exists some $Z_{j(i)}$ with $U_{j(i)} < U_i$ and with

$$\|Z_{j(i)} - b(Z_{j(i)}, Z_i)\| \leq 2^{-i},$$

$$\|b(Z_{j(i)}, Z_i)\| \in \left(a\|Z_i\| - 2^{-i}, a\|Z_i\| + 2^{-i}\right).$$

Now consider a sequence $U_i \downarrow t_0$. The inequalities above imply $\|Z_{j(i)}\| \to a\|z_2\|$. We cannot have $U_{j(i)} \downarrow t_0$, else $\|Z_{j(i)}\| \to \|z_2\|$; and we cannot have $U_{j(i)} \uparrow t_0$, else (using Lemma 4) $\|Z_{j(i)}\| \to \|b(z_1, z_2)\|$. The only other possibility is $\limsup U_{j(i)} < t_0$. But then we can find $k(i)$ with $U_{j(i)} < U_{k(i)} \uparrow t_0$ ultimately. Then by Lemma 11,

$$\|Z_i - b(Z_i, Z_{k(i)})\| \leq \|Z_i - b(Z_i, Z_{j(i)})\|,$$

in other words,

$$\|b(Z_i, Z_{k(i)})\| \geq \|b(Z_i, Z_{j(i)})\|.$$

But the left side converges to $\|b(z_1, z_2)\|$ and the right side to $a\|z_2\|$, contradicting (19).

Thus $\tilde{f}$ is continuous. Now define $f(t) = \|\tilde{f}(t)\|$. We want to show that f gives distances correctly; that is,

$$\begin{aligned}\|\tilde{f}(t_1) - \tilde{f}(t_2)\| = &\left(\|\tilde{f}(t_1)\| - \inf_{t_1 \leq t \leq t_2} \|\tilde{f}(t_2)\|\right) \\ &+ \left(\|\tilde{f}(t_1)\| - \inf_{t_1 \leq t \leq t_2} \|\tilde{f}(t_2)\|\right).\end{aligned} \tag{20}$$

Now for $U_i < U_k < U_j$ we have $\|Z_k\| \geq \|b(Z_i, Z_j)\|$ and so

$$\|Z_i - Z_j\| \geq \big(\|Z_i\| - \|Z_k\|\big) + \big(\|Z_j\| - \|Z_k\|\big).$$

Then by the density of (U_i) and the continuity of $\tilde{f}$, we get a $\geq$ in (20). To get equality, we shall show that there exists $t \in [t_1, t_2]$ such that $\tilde{f}(t) = b(\tilde{f}(t_1), \tilde{f}(t_2))$. But by continuity, as t increases from t_1 to t_2, the path $\tilde{f}(t)$ moves continuously in $\mathscr{S}$ from $\tilde{f}(t_1)$ to $\tilde{f}(t_2)$, and hence must cover the unique non-self-intersecting path from $\tilde{f}(t_1)$ to $\tilde{f}(t_2)$, which passes through the branchpoint.

We need to check that f represents the continuum random tree $(\mathscr{S}, \mu)$ that we started with. Note that for any j we could have started by setting $\tilde{f}(U_i) = Z_i$, $i > j$, and we would have gotten the same $\tilde{f}$ by continuous extension. Thus $\tilde{f}$ is $\sigma(\mathscr{S}, \mu, Z_i, i > j)$ measurable for each j, and hence $\tilde{f}$ is $\sigma(\mathscr{S}, \mu)$ measurable. So given a realization of $(\mathscr{S}, \mu)$ we can define the realization of $\tilde{f}$. Now define $(\mathscr{S}_f, \mu_f)$ in terms of $\tilde{f}$ as in Theorem 13. Then $\mu_f = \mu$ because μ is the limit empirical distribution of (Z_i) and μ_f is the limit empirical distribution of $(\tilde{f}(U_i))$. And $\mathscr{S}_f = \mathscr{S}$ follows from order-leaf-tightness.

Finally, the remaining requirements on f in Condition 1 correspond to parts of the definition of *continuum tree*, as in the proof of Theorem 13.

2.8. *Perfectly balanced trees and ultrametrics.* Theorem 15 dealt with one specialization of the notion of continuum random tree, to the case where the tree could be represented by a continuous random function. A much simpler specialization involves trees in which all leaves are at the same distance from the root. Here the limit continuum random tree can be represented by a point process. The results are stated below: The proofs are straightforward.

In Section 2.1 we defined the reduced subtree $r(\mathscr{R}(k), \{L_1, \dots, L_j\})$ as the subtree spanned by $L_1, \dots, L_j$ and the root. In this section it is convenient to redefine it as the subtree spanned by $L_1, \dots, L_j$ only, and rooted at the last common ancestor of (L_i) in $\mathscr{R}(k)$. This redefinition implies a redefinition of *consistency* at Definition 1.

Call a random proper k-tree *perfectly balanced* if there is a random variable D_k such that each leaf is at distance D_k from the root. Regard trees as ordered (by introducing a random order, if necessary). Theorem 3 has the following corollary.

COROLLARY 16. *Let $(\mathscr{R}(k),\ k \geq 1)$ be perfectly balanced proper k-trees, consistent in the sense above. Suppose the leaf-tight property (Definition 2) holds, and suppose $(D_k,\ k \geq 1)$ is tight. Then there exists a continuum random tree $(\mathscr{S}, \mu)$ representing $(\mathscr{R}(k))$ in the sense of (6), and a r.v. D_∞ such that*

$$\mu\{x \in \mathscr{S}: \|x\| = D_\infty\} = 1 \quad a.s.,$$

$$D_k \to_d D_\infty.$$

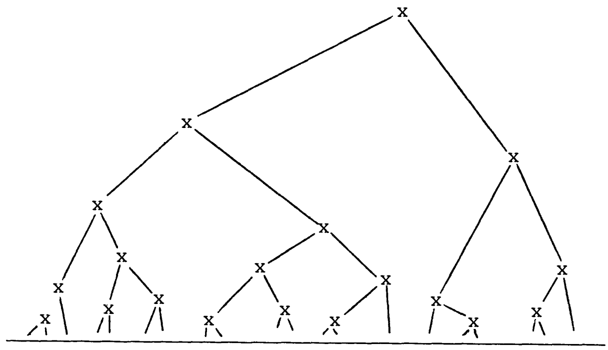

FIG. 4.

Define a (deterministic) nice point process on $[0,1] \times (0,\infty)$ to be a countably infinite set Ξ of points $\xi = (\xi_1, \xi_2)$ such that:

(i) $|\{\xi: \xi_2 > \delta\}| < \infty$ for each $\delta > 0$;
(ii) The closure of Ξ contains $[0,1] \times \{0\}$;
(iii) If ξ, ξ' have $\xi_1 < \xi_1'$ and $\xi_2 = \xi_2'$, then there exists ξ^* such that $\xi_1 < \xi_1^* < \xi_1'$ and $\xi_2^* > \xi_2$.

Given a nice point process, define a metric d on $I^* = [0,1] \setminus \bigcup\{\xi_1: \xi \in \Xi\}$ by

$$d(u, u') = 2\max\{\xi_2: \xi = (\xi_1, \xi_2) \in \Xi \text{ for some } u < \xi_1 < u'\}. \tag{21}$$

This d is an *ultrametric*. That is, for each triple u_1, u_2, u_3 there exists a reordering (in our case, the increasing ordering) such that $d(u_{(1)}, u_{(3)}) = \max(d(u_{(1)}, u_{(2)}), d(u_{(2)}, u_{(3)}))$.

Given a nice point process and $u_1, \dots, u_k \in I^*$ we can define a perfectly balanced proper k-tree with leaves (u_i) for which distances are given by (21) and branchpoints correspond to the ξ appearing in (21). See Figure 4. Assumptions (ii) and (iii) imply that such a ξ is unique and that ξ_2 is strictly positive.

COROLLARY 17. *Given a random nice point process, let $\mathscr{R}(k)$ be the tree constructed as above from i.i.d. uniform $U_1, \dots, U_k$. Then $(\mathscr{R}(k), k \geq 1)$ is a consistent leaf-tight family of perfectly balanced proper random k-trees. Conversely, each such family arises in this way from some random nice point process.*

3. Convergence of rescaled trees. The payoff from our detailed study of continuum random trees in Section 2 is that we can easily formulate abstract conditions for weak convergence of rescaled finite trees to a continuum limit tree. Throughout Section 3, "convergence" is as $n \to \infty$.

3.1. *Unordered trees.* We now return to the set-up of Section 2.1. For each n we have a random tree $\mathscr{T}_n$ on n vertices. Let $(V_{n,1}, \dots, V_{n,n})$ be a uniform random ordering of the vertices of $\mathscr{T}_n$, and for $k < n$ consider the reduced subtree $r(\mathscr{T}_n, \{V_{n,1}, \dots, V_{n,k}\})$.

THEOREM 18. *Suppose that, for each k, $r(\mathscr{T}_n, \{V_{n,1}, \dots, V_{n,k}\}) \to_d \mathscr{R}(k)$, where $\mathscr{R}(k)$ is a random proper k-tree. The family $(\mathscr{R}(k);\ k \geq 1)$ is automatically consistent: Suppose it is also leaf-tight. Let $(\mathscr{S}, \mu)$ be the continuum random tree representing $(\mathscr{R}(k))$, given by Theorem 3. Then there exist measure-representations μ_n of $(\mathscr{T}_n)$ such that $\mu_n \to_d \mu$.*

PROOF. For each n, embed $(V_{n,i})$ into l_1 as in the sequential construction of Section 2.2. By hypothesis,

$$(V_{n,1}, \dots, V_{n,k}) \to_d (L_1, \dots, L_k) \quad \text{for all } k, \tag{22}$$

where (L_i) are as in Lemma 7, with limit empirical distribution μ. Now fix j. We have

$$(\pi_j V_{n,j+1}, \dots, \pi_j V_{n,j+k}) \to_d (\pi_j L_{j+1}, \dots, \pi_j L_{j+k}) \quad \text{for all } k.$$

These random k-tuples are exchangeable, and the infinite exchangeable sequence $(\pi_j L_{j+1},\ i \geq 1)$ has directing random measure $\pi_j \mu$. So by Lemma 9,

$$\pi_j \mu_{n,j} \to_d \pi_j \mu, \tag{23}$$

where

$$\mu_{n,j} = \text{empirical distribution } (V_{n,j+1}, \dots, V_{n,n}).$$

Also, (22) implies $V_{n,j+1} - \pi_j V_{n,j+1} \to_d L_{j+1} - \pi_j L_{j+1}$ and hence

$$E\|V_{n,j+1} - \pi_j V_{n,j+1}\| \wedge 1 \to_d E\|L_{j+1} - \pi_j L_{j+1}\| \wedge 1.$$

Now the families $(\|V_{n,j+i} - \pi_j V_{n,j+i}\|;\ 1 \leq i \leq n - j)$ and $(\|L_{j+i} - \pi_j L_{j+i}\|;\ i \geq 1)$ are exchangeable, so by computing the expectations above in terms of empirical distributions,

$$E\int \|x - \pi_j x\| \wedge 1\, \mu_{n,j}(dx) \to E\int \|x - \pi_j x\| \wedge 1\, \mu(dx). \tag{24}$$

Now (23) and (24) remain true after replacing $\mu_{n,j}$ by $\mu_{n,0}$. Appealing to the random measure version of (13), we see these imply $\mu_{n,0} \to_d \mu$, as required. □

The next corollary states the natural extra condition needed to go from weak convergence of measures to Hausdorff convergence of their supports.

COROLLARY 19. *In the setting of Theorem 18, let $\mathscr{S}_n(\omega) = \text{support}(\mu_n(\omega, \cdot\,))$ and suppose $\mathscr{S}$ is a.s. compact. Write*

$$\Delta(n,k) = \max_{v \in \mathscr{T}_n} \min_{w \in r(\mathscr{T}_n, \{V_{n,1}, \dots, V_{n,k}\})} d(v, w).$$

If

$$\lim_k \limsup_n P(\Delta(n,k) > \varepsilon) = 0, \quad \textit{each } \varepsilon > 0, \tag{25}$$

then $\mathcal{S}_n \to_d \mathcal{S}$.

3.2. *Ordered trees.* The next result is the analog of Corollary 19 where trees are ordered and where we represent the limit continuum random tree via a random function.

THEOREM 20. *Let* $(\mathcal{R}(k);\ k \geq 1)$ *be ordered proper* k*-trees satisfying the conditions of Theorem* 15, *and let* $f\colon [0,1] \to [0,\infty)$ *be a random function satisfying part* (i) *of that theorem. Let* $(\mathcal{T}_n;\ n \geq 1)$ *be random ordered graph-theoretic trees on* n *vertices, and let* $f_n\colon \{1,2,\dots,2n-1\} \to [0,\infty)$ *be the search-depth process for* $\mathcal{T}_n$ *defined at* (15). *Define* $\bar{f}_n\colon [0,1] \to [0,\infty)$ *by*

$$\bar{f}_n(i/2n) = f_n(i), \qquad 1 \leq i \leq 2n-1; \qquad \bar{f}_n(0) = \bar{f}_n(1) = 0,$$

with linear interpolation between these values. Then the following are equivalent:

(i) $\bar{f}_n \to_d f$, *in the sense of weak convergence on* $C[0,1]$.
(ii) *The family* $(\mathcal{T}_n)$ *satisfies* (25), *and for each* k,

$$r(\mathcal{T}_n, \{V_{n,1}, \dots, V_{n,k}\}) \to_d \mathcal{R}(k), \tag{26}$$

where we regard these graph-theoretic trees as ordered.

PROOF. It is easy to see that each of (i) and (ii) implies that the maximal edge length tends to 0. Using Lemma 12, we can replace the random vertices $\{V_{n,1}, \dots, V_{n,k}\}$ in (26) and in the definition of $\Delta(n,k)$ by $\{\tilde{f}_n(U_{n,1}), \dots, \tilde{f}_n(U_{n,k})\}$, where $\tilde{f}_n$ is the walk around $\mathcal{T}_n$ described in Section 2.6, and $(U_{n,i})$ are independent uniform on $\{1,2,\dots,2n-1\}$. With this replacement, (26) becomes, for each k,

$$\begin{aligned}
&\left(\bar{f}_n\left(\frac{U_{n,(1)}}{2n}\right), \inf_{U_{n,(1)}/2n < t < U_{n,(2)}/2n} \bar{f}_n(t), \bar{f}_n\left(\frac{U_{n,(2)}}{2n}\right),\right.\\
&\qquad \left.\inf_{U_{n,(2)}/2n < t < U_{n,(3)}/2n} \bar{f}_n(t), \dots, \bar{f}_n\left(\frac{U_{n,(k)}}{2n}\right)\right)\\
&\qquad \to_d \left(f(U_{(1)}), \inf_{U_{(1)} < t < U_{(2)}} f(t), f(U_{(2)}), \inf_{U_{(2)} < t < U_{(3)}} f(t), \dots, f(U_{(k)})\right),
\end{aligned} \tag{26a}$$

where $U_{n,(i)}$ are the order statistics of $U_{n,1}, \dots, U_{n,k}$ and $U_{(i)}$ are the order statistics of k i.i.d. uniform $(0,1)$ r.v.'s. And with this replacement, (25) becomes

$$\lim_k \limsup_n P(\Delta^*(n,k) > \varepsilon) = 0, \quad \text{each } \varepsilon > 0, \tag{25a}$$

where

$$\Delta^*(n,k) = \max_{0\le a\le 2n} \min_{1\le i\le k} d\left(\bar{f}\left(\frac{a}{2n}\right), \bar{f}\left(\frac{U_{n,(i)}}{2n}\right)\right).$$

In what follows we shall write $U_{n(0)} = 1$, $U_{n,(k+1)} = 2n - 1$, $U_{(0)} = 0$, $U_{(k+1)} = 1$.

If (i) holds, then (26a) follows immediately. To check that (25a) is also a consequence of (i), consider the distance from $v \in \mathscr{T}_n$ to the next (in the order on vertices) vertex of the form $\tilde{f}_n(U_{n,i})$: We see that

$$\Delta^*(n,k) \le \max_{0\le a\le 2n}\left(\bar{f}_n\left(\frac{a}{2n}\right) + \bar{f}_n\left(\frac{U_{n,i(a)}}{2n}\right) - 2\min_{a/2n\le j\le U_{n,i(a)}/2n} \bar{f}_n\left(\frac{j}{2n}\right)\right),$$

where $U_{n,i(a)}$ is the smallest element of $\{U_{n,1},\dots,U_{n,k},2n\}$ greater than a. As $n\to\infty$ this bound converges in distribution to

$$\sup_{0\le t\le 1}\left(f(t) + f(U_{i(t)}) - 2\inf_{t\le s\le U_{i(t)}} f(s)\right), \tag{27}$$

where $U_{i(t)}$ is the smallest element of $\{U_1,\dots,U_k,1\}$ greater than t. By continuity of f and the elementary fact

$$\max_{0\le i\le k}\left(U_{(i+1)} - U_{(i)}\right) \to_d 0 \quad \text{as } k\to\infty, \tag{28}$$

the bound (27) converges to 0 in distribution as $k\to\infty$, establishing (25a).

To prove (ii) implies (i), the issue is to show that (25a) and (26a) imply tightness of $(\bar{f}_n;\ 1\le n<\infty)$: It is then easy to use (26a) to show that any subsequential weak limit is distributed as $\bar{f}$. We first want to use $\Delta^*(n,k)$ to upper bound certain types of oscillations of $\bar{f}_n$. Consider the walk $(\tilde{f}_n(j):\ 1\le j\le 2n-1)$. The construction of the walk leads to the first equality below. For $j_1<j_2<j_3$,

$$\begin{aligned}
&\min_{j\notin(j_1,j_3)} d\left(\tilde{f}_n(j_2),\tilde{f}_n(j)\right)\\
&\quad= \min\left(\tilde{f}_n(j_2) - \min_{j_1\le j\le j_2}\tilde{f}_n(j),\ \tilde{f}_n(j_2) - \min_{j_2\le j\le j_3}\tilde{f}_n(j)\right)\\
&\quad\ge \min\left(\left(\tilde{f}_n(j_2)-\tilde{f}_n(j_1)\right)^+,\left(\tilde{f}_n(j_2)-\tilde{f}_n(j_3)\right)^+\right).
\end{aligned}$$

Applying this to $j_1 = U_{n,(i)}$ and $j_3 = U_{n,(i+1)}$ gives

$$\begin{aligned}
\Delta^*(n,k) \ge \max_{0\le i\le k}\ \sup_{U_{n,(i)}/2n\le t\le U_{n,(i+1)}/2n} \min&\left(\bar{f}_n(t) - \bar{f}_n\left(\frac{U_{n,(i)}}{2n}\right),\right.\\
&\left.\ \bar{f}_n(t) - \bar{f}_n\left(\frac{U_{n,(i+1)}}{2n}\right)\right).
\end{aligned} \tag{29}$$

Next, it is simple to check that, for any $t_1 < t < t_2$ and any function g,

$$|g(t) - g(t_1)| \le |g(t_2) - g(t_1)| + \big(g(t_1) - g(t)\big)^+ + \min\Big(\big(g(t) - g(t_1)\big)^+, \big(g(t) - g(t_2)\big)^+\Big).$$

Applying this to $g = \bar{f}_n$ and $t_1 = U_{n,(i)}/2n$ and $t_2 = U_{n,(i+1)}/2n$,

$$\max_{0\le i\le k} \sup_{U_{n,(i)}/2n \le t \le U_{n,(i+1)}/2n} \left| \bar{f}_n(t) - \bar{f}_n\left(\frac{U_{n,(i)}}{2n}\right)\right| \qquad \left(\equiv w_k^0(\bar{f}_n), \text{ say}\right)$$

$$\le \max_{0\le i\le k} \left| \bar{f}_n\left(\frac{U_{n,(i+1)}}{2n}\right) - \bar{f}_n\left(\frac{U_{n,(i)}}{2n}\right)\right| \qquad \left(\equiv w_k^1(\bar{f}_n), \text{ say}\right)$$

$$+ \max_{0\le i\le k} \left(\bar{f}_n\left(\frac{U_{n,(i)}}{2n}\right) - \inf_{U_{n,(i)}/2n \le t \le U_{n,(i+1)}/2n} \bar{f}_n(t)\right) \qquad \left(\equiv w_k^2(\bar{f}_n), \text{ say}\right)$$

$$+ \max_{0\le i\le k} \sup_{U_{n,(i)}/2n \le t \le U_{n,(i+1)}/2n} \min\left(\bar{f}_n(t) - \bar{f}_n\left(\frac{U_{n,(i)}}{2n}\right), \bar{f}_n(t) - \bar{f}_n\left(\frac{U_{n,(i+1)}}{2n}\right)\right) \qquad \left(\equiv w_k^3(\bar{f}_n), \text{ say}\right).$$

We now claim that for $\alpha = 1, 2, 3$,

$$\lim_{k\to\infty} \limsup_{n\to\infty} P\left(w_k^\alpha(\bar{f}_n) > \varepsilon\right) = 0, \qquad \varepsilon > 0. \tag{30}$$

For $\alpha = 3$ this holds by (29) and (25a). For $\alpha = 1, 2$ it follows easily from (26a), continuity of $\bar{f}$ and (28). Thus (30) holds for $\alpha = 0$. In terms of the usual modulus of continuity $w(\bar{f}_n, \delta)$, the condition for tightness is

$$\lim_{\delta\to\infty} \limsup_{n\to\infty} P\left(w(\bar{f}_n, \delta) > 3\varepsilon\right) = 0, \qquad \varepsilon > 0. \tag{31}$$

Now an easy "triangle inequality" argument shows

$$w(\bar{f}_n, \delta) \le 3 w_k^0(\bar{f}_n) \quad \text{on} \left\{ \min_{0\le i\le k} \left(\frac{U_{n,(i+1)}}{2n} - \frac{U_{n,(i)}}{2n}\right) > \delta \right\}.$$

It follows that the double limit in (31) is bounded by the $\limsup_{n\to\infty}$ in (30) for fixed k. Thus assertion (30) for $\alpha = 0$ implies the tightness condition (31).

3.3. *Remarks.*

REMARK 1. Heuristically, our results apply to trees in which the number of vertices within distance d of a typical vertex grows as d^α for some $1 < \alpha < \infty$. They do not apply to most "Markovian growth" models of random trees, whose behavior is fundamentally different from that described in this paper. Consider for instance a supercritical branching process conditioned on non-extinction. Let $\mathcal{T}_n$ be its family tree, at the first time that the total number of births reaches n. Then the rescaled limit $\mathcal{R}(k)$ in Theorem 18 would be the tree consisting of k leaves, each attached to the root by a unit edge. This limit is not "proper" and, much worse, completely fails the leaf-tight property (Definition 2). The same limit can be expected in most trees for which the number of vertices within distance d of a typical vertex grows exponentially with d.

REMARK 2. We set up the theory around binary limit trees, because that is what occurs in the natural examples. One could handle nonbinary limit trees with various minor changes in the definitions and statements of results.

REMARK 3. The avoidance of any tightness (as $n \to \infty$) requirement in Theorem 18 is perhaps analogous to the following result of Kallenberg (1973). For $k = 1, 2, \dots, \infty$ let $(X^k(t)\colon 0 \le t \le 1)$ be a process with paths in $D[0,1]$ and with exchangeable increments. Then to prove weak convergence in $D[0,1]$, it is sufficient to prove convergence of finite-dimensional distributions.

REMARK 4. In Theorem 18, the fact that we are picking a *uniform* random subset of k vertices is unimportant. More generally, suppose that with each realization t of $\mathscr{T}_n(\omega)$ is associated a probability distribution $p_n(t, \cdot)$ on the vertices of t. Then for each k we could choose $V_{n,1}, \dots, V_{n,k}$ via i.i.d. choices from p_n, conditioned on distinctness:

$$P(V_{n,i} = v_i, 1 \le i \le k \mid \mathscr{T}_n = t) = c_{k,n}(t) \prod_{i=1}^{k} p_n(t, v_i) \quad \text{if } (v_1, \dots, v_k) \text{ distinct.}$$

Additionally, we redefine the measure-representation μ_n to have probability distribution $p_n(t, \cdot)$ (instead of the uniform distribution) on the vertices of t. Under the natural condition

$$\sup_{v \in \mathscr{T}_n} p_n(\mathscr{T}_n, v) \to_p 0 \quad \text{as } n \to \infty,$$

the theorem goes through unchanged.

In most applications, all vertices have the same standing—some happen to be leaves, and others happen not to be leaves. In some applications, there is a qualitative difference between leaves and internal vertices. For instance, one might seek to describe some abstract notion of "strength of relationship" between m objects by setting up a tree in which each object is a leaf of the tree, and where the tree is used to define "distance" between objects. Here the internal vertices have a different meaning from the leaves. In such a case, it would be natural to consider limit theorems based upon picking k *leaves* at random, and this is a setting where the modification of Theorem 18 described above would be appropriate.

REMARK 5. Since our main example is the continuum random tree of Section 4.3 which can be constructed from Brownian excursion, the reader may ask "why not set up the entire theory in terms of codings of continuous functions, and forget about the l_1 story?" One answer is provided by comparing Theorems 15 and 3: The intrinsic conditions on the tree which permit representation by a continuous function are rather messy. Another answer comes from Section 2.8: Perfectly balanced trees are natural objects which cannot be represented by continuous functions. It is presumably true that general ordered continuum trees can be represented by some subset of upper semicontinuous functions, in which case one could discuss weak convergence of random trees in terms of weak convergence of upper semicontinuous random processes. The latter has recently been studied in the context of extremal theory [see O'Brien, Torfs and Vervaat (1990) for references].

4. Examples of consistent families.

4.1. *Kingman's coalescent.* This process, introduced by Kingman (1982), has become a basic tool in mathematical population genetics. Fix k. At time $t = 0$ there are k individuals $(1, \dots, k)$. At each time $t > 0$ there is a partition of the individuals into j "clusters," $1 \le j \le k$. The process evolves according to the rule: In time $[t, t + dt]$, each of the $\binom{j}{2}$ pairs of clusters has chance dt to coalesce into a single cluster. This family of processes (as k varies) is called *the coalescent*. Let us see how the coalescent can be regarded as a continuum random tree (without claiming that it is useful to do so). There is no natural ordering, so regard trees as unordered.

For fixed k the process specifies a random proper k-tree $\mathscr{R}^*(k)$ in the obvious way: The k leaves are the individuals, the branchpoints indicate coalescent events, the edge lengths are the times that a particular cluster remains unchanged and the depth D_k of the root is the time taken for all the individuals to coalesce into a single cluster. The family $(\mathscr{R}^*(k);\ k \ge 2)$ is perfectly balanced and consistent in the sense of Section 2.8. To apply Corollary 16, we need to check the leaf-tight property (Definition 2). Since the pairs involved in successive coalescences are chosen randomly, we can calculate

$$P(\{1\} \text{ is a cluster at time } t | N_k(s),\ s \le t) = \frac{N_k(t)(N_k(t) - 1)}{k(k-1)},$$

where $N_k(t)$ is the number of clusters at time t. So in the notation of Definition 2,

$$P\Big(\min_{2 \le j \le k} d\big(L_1^k, L_j^k\big) > 2t\Big) = \frac{EN_k(t)(N_k(t) - 1)}{k(k-1)}.$$

Now $N_k(t)$ is a certain continuous-time pure death process started at k but with transition rates independent of k. It is easy to show that as $k \to \infty$, $EN_k(t)(N_k(t) - 1) \uparrow EN_\infty(t))(N_\infty(t) - 1) < \infty$, and so the leaf-tight property holds. Similarly, $ED_k \to ED_\infty < \infty$. Corollary 16 says there exists a representing continuum random tree $(\mathscr{S}, \mu)$.

4.2. *Coalescing Brownian motions.* Here is a simple way to combine spatial motion with the coalescing mechanism in the previous example. Take k

points in the unit interval, let a particle start at each point and perform independent Brownian motions, coalescing when they meet. The $k \to \infty$ limit process, "coalescing Brownian motions started from every point" is well known [but apparently written down only in the Ph.D. thesis of Arratia (1979)] as a process defined for times $t > 0$. Our notion of continuum trees permits rigorous extension to $t = 0$. More generally, in studying systems of diffusing particles in R^d in which two particles annihilate each other when they (nearly) collide, it is useful [e.g., Sznitman (1989)] to first forget the annihilations and consider a "collision tree" as follows. Fix a particle i at time t, then consider the particles j which, at some time t_j during $[0, t]$ have (nearly) collided with particle i, then for each j consider the particles k which, at some time t_k during $[0, t_j]$ have (nearly) collided with particle j, and so on. As the number of particles $\to \infty$ and "nearly" is varied to keep the mean free path constant, one gets a limit finite tree. Presumably, if instead the mean free path tends to 0, then we will get a continuum limit tree qualitatively similar to that of the coalescent.

Here is a further modification of these ideas, motivated by a particular lattice model discussed later. Let $(X_t^+, X_t^-: 0 \le t \le 1)$ satisfy

$$X_0^+ = X_0^- = 0; \qquad X_t^+ > X_t^-, 0 < t < 1; \qquad X_1^+ = X_1^-. \tag{32}$$

Fix k. Choose k points (t_i, x_i) uniformly from product Lebesgue measure on $\{(t, x)\colon 0 < t < 1,\ X_t^- < x < X_t^+\}$. We now specify a "birth and coalescence" process involving k particles $1, \dots, k$ and two extra particles $\oplus, \ominus$. The particles $\oplus$ and $\ominus$ are born at time 0 and predestined to follow the paths X^+ and X^-. Particle i is born at time t_i and position x_i. Particle i evolves as a standard Brownian motion, independent of other particles, until it meets and coalesces with some other particle. If it coalesces with some particle $j \ne i$, the combined particle continues to evolve as a Brownian motion; whereas if it coalesces with $\oplus$ or $\ominus$, the combined particle follows the predestined path.

We now define a tree, illustrated in Figure 5. There are k leaves, corresponding to the births of the k particles. Branchpoints correspond to coales-

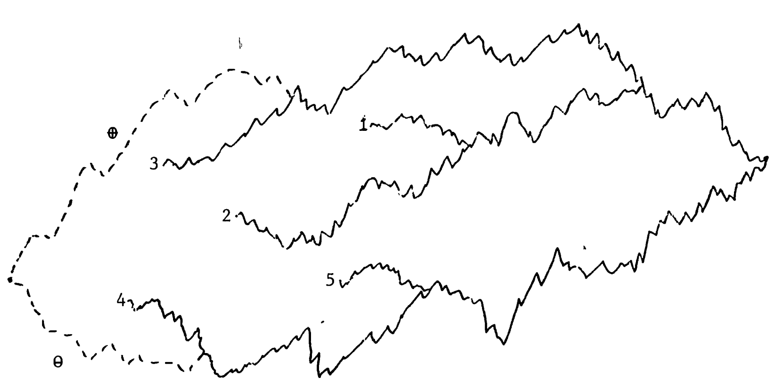

FIG. 5.

cences, and an edge corresponds to the period of time between a birth or coalescence and the subsequent coalescence, and the edge length is the duration of that period of time. This specifies a random graph-theoretic tree $\mathscr{R}(k)$. We clearly have the consistency of Definition 1, and (from elementary properties of Brownian motion) the leaf-tight property holds. So Theorem 3 establishes the existence of a representing continuum random tree. One can also give $\mathscr{R}(k)$ the order induced from the order on the one-dimensional state space, and then verify the conditions of Theorem 15 to show that this continuum random tree can be coded from some continuous random function.

We briefly mention the underlying discrete model [see Nguyen (1990) for more details]. On the vertex set $\{(m,n)\colon m \geq 0,\ m+n \text{ even}\} \subset Z^2$ draw an edge from vertex (m,n) to either $(m-1,n-1)$ or $(m+1,n-1)$, choosing uniformly and independently for each vertex. The component of this random graph containing $(0,0)$ can be regarded as a random tree $\mathscr{T}$ rooted at $(0,0)$, with height $H = \max\{m\colon (m,n) \in \mathscr{T}\}$. Write $\mathscr{T}_h$ for $\mathscr{T}$ conditioned on $H = h$.

It is intuitively clear that, rescaling the edges of $\mathscr{T}_h$ by $1/h$, these random trees converge to the continuum random tree defined above, in which X^+ and X^- are independent Brownian motions conditioned on event (32). This could be proved using the ideas of Nguyen (1990) and Section 3.

4.3. *The Brownian CRT.* A proper k-tree in which the root has degree 1 has exactly $2k-1$ edges. As at (1), such an (unordered) graph-theoretic tree can be specified as

$$t = \left(\hat{t}; x_1, \dots, x_{2k-1}\right) \in T_{2k}^* \times R^{2k-1},$$

where $\hat{t}$ describes the shape of the tree and $(x_1, \dots, x_{2k-1})$ describes the lengths of the edges. Here the set T_{2k}^* of "shapes of trees" refers to binary trees with exactly k leaves, the leaves being distinguished. In constructing such a tree by adding the leaves sequentially, the edge to leaf k could be attached to any of the $(2k-3)$ preexisting edges, so the number of possible shapes is

$$|T_{2k}^*| = \prod_{i=1}^{k-1} (2i-1).$$

LEMMA 21. *There exists a consistent family* $(\mathscr{R}(k);\ k \geq 1)$ *of random proper k-trees such that* $\mathscr{R}(k)$ *has density*

$$f\left(\hat{t}; x_1, \dots, x_{2k-1}\right) = s \exp\left(-s^2/2\right), \quad s = \sum_{i=1}^{2k-1} x_i. \tag{33}$$

This is one of the main components of the big picture discussed in Aldous (1991b), and could be proved in several ways. We give a "constructive" argument below. Note that (33) implies that the $2k-1$ edge lengths have exchangeable joint distribution, and this joint distribution is independent of the "shape" $\hat{t}$ of the tree. (This is much stronger than the property of "exchangeability of the k leaf labels" which holds in any consistent family.) Note also that the assertion that (33) defines a *density* is the calculus assertion

$$\int_{x_1>0} \cdots \int_{x_{2k-1}>0} s \exp(-s^2/2)\, dx_1 \cdots dx_{2k-1} = \prod_{i=1}^{k-1} \frac{1}{2i-1}.$$

PROOF. The following construction [Aldous (1991a), Process 3] gives random proper k-trees $\mathscr{R}(k)$. We shall show by induction on k that these have density (33).

Let $(C_1, C_2, \dots)$ be the times of a nonhomogeneous Poisson process on $(0, \infty)$ with rate $r(t) = t$. Let $\mathscr{R}(1)$ consist of an edge of length C_1 from a root to leaf 1. Inductively, obtain $\mathscr{R}(k+1)$ from $\mathscr{R}(k)$ by attaching an edge of length $C_{k+1} - C_k$ to a uniform random point of $\mathscr{R}(k)$ (i.e., a point chosen from normalized Lebesgue measure on the edges), labelling the new leaf $k+1$.

To analyze the densities, let $(t^*; x_1^*, x_2^*, \dots, x_{2k+1}^*)$ be a tree with leaves $1, \dots, k, k+1$, and let $(\hat{t}; x_1, x_2, \dots, x_{2k-1})$ be the reduced subtree associated with t^* and leaves $1, \dots, k$. Then t^* is obtained from $\hat{t}$ by creating a new internal vertex (splitting an edge of length x_j into two edges of lengths $x_{j_1}^*$ and $x_{j_2}^*$, say, with $x_j = x_{j_1}^* + x_{j_2}^*$, and joining leaf $k+1$ to that new internal vertex by an edge of length $x_{j_3}^*$, say. From the construction,

$$f(t^*; x_1^*, \dots, x_{2k+1}^*) = \frac{s^*}{s} \exp\left(-\frac{1}{2}(s^{*2} - s^2)\right) f(\hat{t}, x_1, \dots, x_{2k-1}),$$

$$s^* = \sum x_i^*,$$

where the term $1/s$ reflects the chance that the $k+1$ segment is attached at a particular place on the existing tree, and the term $s^* \exp(-(1/2)(s^{*2} - s^2))$ reflects the chance that the $k+1$ segment has length $x_{j_3}^* = s^* - s$. Plainly (33) follows by induction.

Finally, if we deterministically specify $k-1$ distinct leaves $(l_1, \dots, l_{k-1})$ of $\mathscr{R}(k)$ and consider the reduced subtree $r(\mathscr{R}(k), \{l_1, \dots, l_{k-1}\})$, then the distribution of the reduced subtree does not depend on the choice of $(l_1, \dots, l_{k-1})$, by exchangeability of edge lengths in (33). So $\mathscr{R}(k-1)$, which by the construction above is the reduced subtree on leaves $(1, \dots, k-1)$, has the same distribution as the reduced subtree on $k-1$ randomly-picked distinct leaves. This verifies consistency, and establishes the lemma. □

It is easy to use the construction above to show that $(\mathscr{R}(k);\ k \geq 1)$ is leaf tight. Briefly, consider the initial edge in $\mathscr{R}(1)$. The lengths of added edges tend to zero, and an infinite number of edges are attached to this initial edge at uniform positions; it follows that leaf 1 is indeed a limit of some sequence of other leaves. Theorem 3 now allows us to make the following definition:

DEFINITION. The *Brownian CRT* is the continuum random tree which represents the family $(\mathscr{R}(k);\ k \geq 1)$ in Lemma 21.

So far we have regarded these trees as unordered. A proper k-tree where the root has degree 1 has 2^{k-1} possible orders. So if we choose an order uniformly at random, (33) becomes

$$(34) \qquad f(\hat{t}; x_1, \dots, x_{2k-1}) = 2^{-(k-1)} s \exp(-s^2/2), \qquad s = \sum_{i=1}^{2k-1} x_i,$$

where the order is included in $\hat{t}$. If in the proof of Lemma 21 we flip fair coins to decide on the order at the branchpoint each time we add a new branch, then we get a consistent family of ordered k-trees $(\mathscr{R}(k))$ with density (34).

In Aldous (1991a), Theorem 3, a complicated direct proof of compactness was given, and [from our Theorem 15(iii)] this is essentially enough to show that CRT under discussion, regarded as ordered, can be represented by some random continuous function. But the remarkable fact (motivating the name) is:

COROLLARY 22. *The Brownian CRT is the continuum random tree represented in the sense of Theorem* 13 *by the random function* $f(t) = 2B_t$, $0 \leq t \leq 1$, *where B is standard Brownian excursion of duration* 1.

In other words, the k-tree constructed from $2B_t$ and i.i.d. $U(0,1)$ r.v.'s $(U_1, \dots, U_k)$ as in the proof of Theorem 13 has density (33). I do not know a direct proof of this from known properties of Brownian excursion, though such a proof must be possible. But the result can be obtained as a corollary of our Theorem 23, and that proof is given after the statement of Theorem 23. Neveu and Pitman (1989) construct trees from Brownian excursion conditioned on height exceeding h, and Le Gall (1991) from the σ-finite measure on Brownian excursions of varying duration.

Finally, in Aldous (1991a) the measure-representation μ of the Brownian CRT was obtained as $\mu = \lim \mu_{C_k}$ a.s., where μ_{C_k} is Lebesgue measure on the k edges of $\mathscr{R}(k)$ in the previous construction, renormalized to have total mass 1. In the preceding definition, Lemma 7 gives μ as the limit $\mu = \lim \nu^k$ a.s., where ν^k is the empirical distribution of the k leaves of $\mathscr{R}(k)$. It is not hard to verify that the two limits μ are identical, and hence that the supporting random sets $\mathscr{S}$ are identical.

4.4. *Subtrees of Z^d.* It is tempting to believe that several models of random subtrees of the integer lattice Z^d, for example, the "uniform random spanning trees" studied by Pemantle (1991), can be rescaled to converge to some continuum random tree. Proving such results seems beyond current technology, but we will report elsewhere on simulation studies in progress.

5. Convergence of conditioned Galton–Watson trees. Let $\mathscr{T}$ denote the critical Galton–Watson branching process started with a single progenitor and considered as an ordered rooted tree, where the offspring distribution has finite variance σ^2. Let $\mathscr{T}_n$ be $\mathscr{T}$ conditioned on $|\mathscr{T}|$, the total population size until extinction, being exactly n. Our main result, Theorem 23, is an invariance principle for such families: After rescaling, there is a limit random tree, which is the Brownian CRT of Section 4.3, and the offspring distribution only affects the limit via a scale factor $1/\sigma$.

There is of course a vast literature on branching process asymptotics. This particular conditioning has been studied most extensively by the Soviet school, and their results (through the early 1980's) are set forth in the monograph of Kolchin (1986). They use analytic methods to prove asymptotic convergence of various statistics of the trees: We shall combine some of their analytic estimates (stated in Section 5.1) with the "general abstract nonsense" of Section 3 to prove Theorem 23. This particular conditioning occurs only occasionally in the Western literature on branching processes [e.g., Kennedy (1975)] although Western combinatorialists, starting with Moon and Meir (1978), have developed independently many results similar to the Soviets in different (mathematical) language. Much of the motivation for studying this model is that, for various special cases of offspring distribution, the random tree $\mathscr{T}_n$ becomes the uniform random tree in some class of n-vertex trees of specified combinatorial type [see Aldous (1991b), Section 2.1 for a translation guide]. The consequences of Theorem 23 for combinatorial models of random trees—that limit distributions and interesting "global" functions can be expressed as functionals of Brownian excursion—are detailed in Aldous (1991b) Section 3.

THEOREM 23. *Let $\mathscr{T}_n$ be a conditioned Galton–Watson tree whose offspring distribution ξ satisfies*

$$E\xi = 1,$$

$$0 < \mathrm{var}(\xi) = \sigma^2 < \infty,$$

$$g.c.d.\{j: P(\xi = j) > 0\} = 1.$$

Rescale the edges of $\mathscr{T}_n$ to have length $\sigma n^{-1/2}$. Let f_n: $\{1, 2, \dots, 2n-1\} \to [0, \infty)$ be the search-depth process for rescaled $\mathscr{T}_n$ defined at (15). *Define $\bar{f}_n$: $[0,1] \to [0, \infty)$ by*

$$\bar{f}_n(i/2n) = f_n(i), \qquad 1 \le i \le 2n-1; \qquad \bar{f}_n(0) = \bar{f}_n(1) = 0,$$

with linear interpolation between these values.

Then $(\bar{f}_n(t),\ 0 \le t \le 1) \to_d (2B(t),\ 0 \le t \le 1)$ *on* $C[0,1]$, *where B is standard Brownian excursion.*

Here is the structure of the proof. In Section 5.2 we verify conditions (ii) in Theorem 20, for $(\mathscr{R}(k))$ defined in Lemma 21. The proof is then completed by the following argument [said more leisurely in Aldous (1991b), Section 2]. A special case of the Galton–Watson tree $\mathscr{T}_n$ is where the offspring distribution is $P(\xi = i) = 2^{-i}$, $i \geq 0$. In this case the search-depth process (before any rescaling) is exactly simple symmetric random walk conditioned on first return to 0 being at time $2n$. But it is well known that this conditioned random walk rescales to Brownian excursion, and a brief computation with our precise rescaling conventions here reveals that $f_n \to_d 2B$ in this special case. Now Theorem 20 applied to this special case gives Corollary 22, that is, shows that $(\mathscr{R}(k))$ is represented by $2B(t)$, and then we can apply Theorem 20 to the general case to establish Theorem 23.

In general the search-depth process of a random finite tree is not a tractable random process, even for Galton–Watson trees. So it is not profitable to seek to prove directly the convergence of search-depth processes, and this is the point of our abstract theory in Section 2. An obvious exception is the preceding special case. As another exception, in a different special case of binary Galton–Watson trees, the convergence of the search-depth process (precisely, the search-depth process restricted to leaves) to Brownian excursion has recently been proved directly by Gutjahr and Pflug (1992). They use exact combinatorial formulas for the finite-dimensional distributions of the search-depth process.

Technical remarks.

1. Whenever Theorem 20 holds, we can ignore the order structure and instead obtain the conclusions of Theorem 18 and Corollary 19. Thus in Theorem 23 we could ignore order structure on $\mathscr{T}_n$ and say: There exist measure- and set-representations $\mu^n, \mathscr{S}^n$ of rescaled $\mathscr{T}_n$ such that

$$\mu^n \to_d \mu, \qquad \mathscr{S}^n \to_d \mathscr{S},$$

where μ and $\mathscr{S}$ are measure and set representations of the Brownian CRT.

2. The "sublattice" case where the support of ξ has g.c.d. $= d \geq 2$ involves only minor modifications.

3. Analogously to random walks and Brownian motion, the case of infinite second moments is qualitatively different, and one does not get the Brownian CRT as a limit.

4. Because we condition on total population size, the distribution of $\mathscr{T}_n$ is unchanged by replacing ξ with another distribution χ in the same exponential family

$$P(\xi = i) = c\theta^i P(\chi = i), \qquad i \geq 0 \text{ for some } c, \theta.$$

Thus there is no essential loss of generality in considering only critical branching processes.

5.1. *Analysis estimates.* Here are the analysis facts we need. Let $\operatorname{ht}(\mathscr{T})$ be the height of $\mathscr{T}$, that is the number of generations until extinction.

PROPOSITION 24. *For the Gaton–Watson tree $\mathscr{T}$ with offspring distribution ξ satisfying the assumptions of Theorem* 23,

$$P(\mathrm{ht}(\mathscr{T}) > h) \sim \frac{2}{\sigma^2 h} \quad \text{as } h \to \infty, \tag{35}$$

$$P(|\mathscr{T}| = n) \sim (2\pi)^{-1/2}\sigma^{-1}n^{-3/2} \quad \text{as } n \to \infty, \tag{36}$$

$$n^{1/2}P(\mathrm{ht}(\mathscr{T}) > bn^{1/2}, |\mathscr{T}| < \delta n) \to \sigma^{-1}\delta^{-1/2}G(b\delta^{-1/2}\sigma) \quad \text{as } n \to \infty, \tag{37}$$

where $G(\cdot)$ *satisfies*

$$G(x) \le \kappa_1 \exp(-x/\kappa_2), \qquad 0 < x < \infty$$

for certain constants $\kappa_i < \infty$.

Let (X_i) *be i.i.d. and distributed as* $|\mathscr{T}|$. *Then*

$$\max_{j \ge 0} \left| m^2 P\left(\sum_{i=1}^{m} X_i = j \right) - \sigma^2 g\left(\frac{\sigma^2 j}{m^2} \right) \right| \to 0 \quad \text{as } m \to \infty, \tag{38}$$

where

$$g(x) = (2\pi)^{-1/2} x^{-3/2} \exp\left(-\frac{1}{2x} \right) \tag{39}$$

is the positive stable (1/2) *density.*

Result (35) is classical [see Kolchin (1986), Theorem 2.1.2]. Fact (36) is Kolchin [(1986), Lemma 2.1.4], and is a simple consequence of the local central limit theorem and a ballot-type identity due to Dwass [an elegant proof of the latter is in Waymire (1991)]. Next, the limit result [Kolchin (1986), Theorem 2.4.3] on height conditioned on size can be rewritten in terms of the maximum W^* of standard Brownian excursion [c.f. Aldous (1991b)] as

$$P(\mathrm{ht}(\mathscr{T}) > xn^{1/2} \,|\, |\mathscr{T}| = n) \to P(2\sigma^{-1}W^* > x).$$

Then by scaling, (37) holds with

$$G(x) = (2\pi)^{-1/2} \int_0^1 P(2u^{1/2}W^* > x) u^{-3/2}\, du,$$

from which it is easy to obtain the asserted exponential tail bound. Finally, in view of (36), assertion (38) is just the local limit theorem for convergence of i.i.d. sums to the positive stable (1/2) limit [see Gnedenko and Kolmogorov (1954), Section 50].

For later use, associate with ξ the distributions $\hat{\xi}$ and ξ^* defined by

$$P(\hat{\xi} = i) = (i+1)P(\xi = i+1), \qquad i \ge 0, \tag{40}$$

$$P(\xi^* = i) = \sigma^{-2}(i+1)(i+2)P(\xi = i+2), \qquad i \ge 0, \tag{41}$$

and note that

$$E\hat{\xi} = \sigma^2.$$

5.2. *Proof of Theorem* 23. The aim is to verify condition (ii) of Theorem 20. Although that involves rescaled and reduced subtrees, for most of the argument we do neither, working with "ordinary" trees and subtrees whose edges have length 1. Regard trees as ordered. A subtree u of a tree τ is a connected subset of vertices of τ containing the root. Subtrees u_1, u_2 are different if they are different as subsets of τ; they are isomorphic if they are isomorphic as rooted ordered trees in themselves. Recall that $|u|$ denotes the number of vertices of u.

Fix $k \geq 2$. Let a k-tree be a tree with exactly k leaves labelled $(1, \dots, k)$ and such that the reduced subtree $r(t, \{\text{leaves}\})$ is a proper k-tree. For k-trees t_1, t_2 to be isomorphic, the isomorphism must preserve the leaf labelling. Consider distinct vertices $(v_1, \dots, v_k)$ of a tree τ. Call the subtree t spanned by the (v_i) and the root, and with v_i labelled as i, a *k-subtree* of τ. So there are $|\tau|!/(|\tau| - k)!$ different k-subtrees of τ. Here is the fundamental identity.

LEMMA 25. *Let $\mathscr{T}$ be the unconditioned Galton–Watson random tree. Let t be a k-tree. Then*

$$E(\textit{number of different } k\textit{-subtrees of } \mathscr{T} \textit{ isomorphic } (\textit{as } k \textit{ trees}) \textit{ to } t)1_{(|\mathscr{T}|=n)}$$

$$= \left(\frac{\sigma^2}{2}\right)^{k-1} P\big(S_{L(t)+k} = n - |t| - L(t)\big),$$

where

$$L(t) = \sum_{i=1}^{|t|-2k+1} \hat{\xi}_i + \sum_{i=1}^{k-1} \xi_i^*, \qquad S_m = \sum_{i=1}^{m} X_i,$$

the r.v.'s $(\hat{\xi}_i, \xi_i^*, X_i)$ are independent, with the distributions of $\hat{\xi}_i$ and ξ_i as specified at (40) *and* (41), *and with $X_i =_d |\mathscr{T}|$.*

PROOF. Call a subtree u^* of $\mathscr{T}$ *full* if, whenever u^* contains an individual, it also contains all the siblings of that individual. I assert that for any tree u,

$$E(\text{number of full subtrees of } \mathscr{T} \text{ isomorphic to } u)1_{(|\mathscr{T}|=n)}$$

$$= P\big(S_{n(u;0)} = n - |u|\big) \prod_{v:\, d(u;v) \geq 1} P(\xi = d(u;v)) \tag{42}$$

$$= f(u), \quad \text{say,}$$

where $n(u;i)$ is the number of vertices of u with exactly i children, and $d(u;v)$ is the number of children of individual v in u. For $\mathscr{T}$ can have at most one full subtree isomorphic to u, and this happens if and only if the individuals corresponding to internal vertices of u have exactly the right numbers of children; then the descendants of the leaves of u may be arbitrary, but the total number of such descendants must be $n - |u|$ if the total population size of $\mathscr{T}$ is to be n.

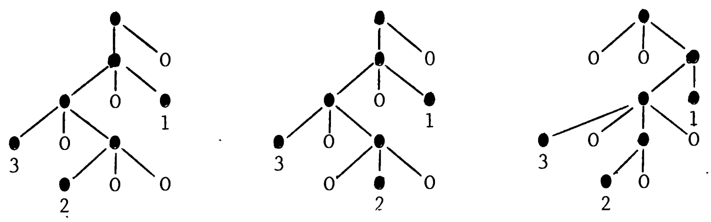

FIG. 6.

Now fix a k-tree t. Let t^* be a subtree of $\mathscr{T}$ with k labelled leaves which is isomorphic to t, and let u be the subtree consisting of the individuals in t^* together with all their siblings. Consider the pair (t^*, u) without reference to the rest of $\mathscr{T}$. This pair has the following properties:

(i) t^* is isomorphic to t.
(ii) t^* is a subtree of u.
(iii) Each individual of u is either in t^* or is a sibling of an individual in t^*.

So the expectation in Lemma 25 equals

$$\sum_{(t^*, u)} f(u), \tag{43}$$

where the sum is over all pairs (t^*, u) satisfying (i)–(iii).

Figure 6 illustrates three pairs (t^*, u) which contribute to the sum (43) for a particular t. Note that the left and center pairs are counted as distinct, even though the u's are isomorphic.

Given a pair (t^*, u) satisfying (i)–(iii), for each internal node v of t^* let $s(v)$ be the number of extra children added at v in making u, that is to say,

$$s(v) = d(u;v) - d(t^*;v) \quad \text{for vertices } v \text{ with } d(t^*;v) \geq 1.$$

Let s denote the vector $(s(v))$. Note that u has $k + |\mathbf{s}|$ leaves and $|t| + |\mathbf{s}|$ vertices, where $|\mathbf{s}| = \Sigma_v s(v)$. So we can write $f(u)$ in terms of t^* and $\mathbf{s}$ as

$$P\big(S_{k+|\mathbf{s}|} = n - |t^*| - |\mathbf{s}|\big) \prod_{v:\, d(t^*, v) \geq 1} P(\xi = d(t^*;v) + s(v)).$$

Given t^* and $\mathbf{s}$, the number of nonisomorphic pairs (t^*, u) satisfying (i)–(iii) and with $d(u;v) - d(t;v) \equiv s(v)$ equals

$$\left(\prod_{v:\, d(t;v)=1} (s(v) + 1)\right)\left(\prod_{v:\, d(t;v)=2} \frac{(s(v) + 1)(s(v) + 2)}{2}\right),$$

because for an individual with one child in t and s other children in u, we have $s + 1$ positions in which the child in t might be placed. Similarly, for an individual with two children in t in specified relative order and with s other children in u, there are $(s + 1)(s + 2)/2$ choices for the positions of the two children of t.

So (43) becomes

$$\sum_{\mathbf{s}} P\big(S_{k+|\mathbf{s}|} = n - |t| - |\mathbf{s}|\big) \times \prod_{v:\, d(t;v)=1} (s(v)+1)P(\xi = s(v)+1)$$
$$\times \prod_{v:\, d(t;v)=2} \frac{(s(v)+1)(s(v)+2)P(\xi = s(v)+2)}{2}.$$

Using (40, 41), this becomes $(\sigma^2/2)^{k-1}$ times

$$\sum_{\mathbf{s}} P\big(S_{k+|\mathbf{s}|} = n - |t| - |\mathbf{s}|\big)\Bigg(\prod_{v:\, d(t;v)=1} P(\hat{\xi} = s(v))\Bigg)\Bigg(\prod_{v:\, d(t,v)=2} P(\xi^* = s(v))\Bigg).$$

But the sum can be rewritten as the probability term in Lemma 25. For because of the "proper" in the definition of k-tree, there are no individuals in t with more than two children, exactly $k-1$ individuals with two children, exactly k with no children, and hence the remaining $|t| - 2k + 1$ individuals have one child.

Now consider the conditioned Galton–Watson tree $\mathscr{T}_n$. Let $\mathscr{R}(n,k)$ be the k-subtree spanned by the root and k randomly chosen distinct vertices. The equality in the next lemma is immediate from Lemma 25.

LEMMA 26. *Let t be a k-tree. Then*

$$P(\mathscr{R}(n,k) \text{ is isomorphic to } t)$$
$$= \frac{(n-k)!}{n!2^{k-1}}\sigma^{2(k-1)}\frac{P\big(S_{L(t)+k} = n - |t| - L(t)\big)}{P(|\mathscr{T}| = n)}$$
$$\sim \left(\frac{1}{2}\right)^{k-1}\left(\frac{\sigma}{n^{1/2}}\right)^{2k}|t|\exp\left(-\frac{|t|^2\sigma^2}{2n}\right) \quad \textit{as } n \to \infty,\ |t| = \Theta(n^{1/2}).$$

($|t| = \Theta(n^{1/2})$ *means the ratio* $|t|/n^{1/2}$ *is bounded away from* 0 *and* ∞.)

To prove the asymptotics, in view of (36), what we have to prove is

$$(44)\quad P\big(S_{L(t)+k} = n - |t| - L(t)\big) \sim (2\pi)^{-1/2}\sigma|t|n^{-3/2}\exp\left(-\frac{\sigma^2|t|^2}{2n}\right)$$

as $n \to \infty$, $|t| = \Theta(n^{1/2})$. Note that the right side is $\Theta(n^{-1})$.

Since $E\hat{\xi} = \sigma^2$, the weak law of large numbers implies there exist $\varepsilon_n \downarrow 0$ such that

$$(45)\qquad P\left(\left|\frac{L(t)}{\sigma^2|t|} - 1\right| > \varepsilon_n\right) \to 0.$$

Now (38) implies $\sup_s P(S_m = s) = O(m^{-2})$, and so by conditioning on $L(t)$ and using (45),

$$(46)\quad P\left(S_{L(t)+k} = n - |t| - L(t),\ \frac{L(t)}{\sigma^2|t|} > 1 + \varepsilon_n\right) = o(|t|^{-2}) = o(n^{-1}).$$

Similarly, (38) implies

$$\sup_{(1/2)\sigma^2|t| \le m \le \sigma^2|t|+k} \sup_s P(S_m = s) = O(|t|^{-2}) = O(n^{-1})$$

and again by conditioning on $L(t)$ and using (45),

$$(47) \qquad P\left(S_{L(t)+k} = n - |t| - L(t),\ \frac{1}{2} \le \frac{L(t)}{\sigma^2|t|} < 1 - \varepsilon_n\right) = o(n^{-1}).$$

And also

$$(48) \qquad P\left(\frac{L(t)}{\sigma^2|t|} \le \frac{1}{2}\right) = o(n^{-1}),$$

because, by truncating $\hat{\xi}$ at some point making its mean greater than $1/2$, the large deviation theorem implies the probability in (48) decreases geometrically fast in $|t|$. Now by (46)–(48), in proving (44) we may replace $L(t)$ by deterministic $l(t) \sim \sigma^2|t|$. Now from (38) and a little rearrangement, if $m = \Theta(n^{1/2})$ and $q = \Theta(n^{1/2})$, then

$$P(S_m = n - q) \sim (2\pi)^{-1/2} m n^{-3/2} \sigma^{-1} \exp\left(-\frac{m^2}{2\sigma^2 n}\right).$$

Applying this with $m = l(t) + k \sim \sigma^2|t|$ and $q = |t| + l(t)$ establishes (44).

We are now halfway to our goal. Rescale the edges of $\mathscr{T}_n$ as in the statement of Theorem 23. Let $\mathscr{R}^*(n,k)$ be the reduced subtree associated with k random vertices of rescaled $\mathscr{T}_n$. Then Lemma 26 implies

$$(49) \qquad \mathscr{R}^*(n,k) \to_d \mathscr{R}(k) \quad \text{as } n \to \infty,$$

where $\mathscr{R}(k)$ has the density (34). This verifies (26), and so it remains to verify the "tightness" condition (25).

Let $\mathscr{F}(n,k)$ denote the full subtree of $\mathscr{T}_n$ associated with $\mathscr{R}(n,k)$. The argument for (42) gives the following distributional relationship. Conditional on $\mathscr{F}(n,k) = u$, the tree $\mathscr{T}_n$ can be regarded as built from u and subtrees $(\mathscr{T}_v^*)$ appended at the leaves of v of u, and this family $(\mathscr{T}_v^*; v \text{ a leaf of } u)$ consists of i.i.d. copies of $\mathscr{T}$, conditioned on the total number of vertices $\sum_v |\mathscr{T}_v^*|$ being exactly $n - |u|$. Let (X_i, Y_i) be independent (in i) copies of the joint distribution $(|\mathscr{T}|, \operatorname{ht}(\mathscr{T}))$ and set

$$S_m = \sum_{i=1}^{m} X_i, \qquad Y_m^* = \max_{1 \le i \le m} Y_i.$$

Let $D(n,k)$ be the maximum, over vertices v of $\mathscr{T}_n$, of the distance from v to $\mathscr{F}(n,k)$. Let $L(n,k)$ be the number of leaves of $\mathscr{F}(n,k)$. The previous distributional relationship implies

(∗) the conditional distribution of $D(n,k)$ given $\mathscr{F}(n,k)$ is the conditional distribution of $Y^*_{L(n,k)}$ given $S_{L(n,k)} = n - |\mathscr{F}(n,k)| + L(n,k)$.

Now (49) implies

$$\sigma n^{-1/2}|\mathscr{R}(n,k)| \to_d |\mathscr{R}(k)|, \tag{50}$$

where $|\mathscr{R}(k)|$ denotes the total length (i.e., the sum of edge lengths) of $\mathscr{R}(k)$. The argument for (45) shows $L(n,k)/|\mathscr{R}(n,k)| \to_p \sigma^2$ and hence, by (50),

$$n^{-1/2}L(n,k) \to_d \sigma|\mathscr{R}(k)|. \tag{51}$$

Now fix $b > 0$ and $0 < c_1 < c_2 < \infty$. Then

$$\begin{aligned} P(D(n,k) > bn^{1/2}) \le P(n^{-1/2}L(n,k) \notin [c_1, c_2]) \\ + P(|\mathscr{F}(n,k)| - L(n,k) > n^{2/3}) \\ + P(D(n,k) > bn^{1/2}, n^{-1/2}L(n,k) \in [c_1, c_2], \\ |\mathscr{F}(n,k)| - L(n,k) \le n^{2/3}). \end{aligned}$$

The first term in the bound tends to $P(\sigma|\mathscr{R}(k)| \notin [c_1, c_2])$ by (51). The second term tends to 0 by (50), because $|\mathscr{F}(n,k)| - L(n,k) \le |\mathscr{R}(n,k)|$. And by conditioning on $\mathscr{F}(n,k)$ and using the representation $(*)$, the final term is bounded by the conditional probability appearing in Lemma 27 below. So by the conclusion of Lemma 27,

$$\begin{aligned} &\limsup_{n\to\infty} P(n^{-1/2}D(n,k) > b) \\ &\quad \le a(b, c_1) + P(\sigma|\mathscr{R}(k)| \le c_1) + P(\sigma|\mathscr{R}(k)| \ge c_2). \end{aligned} \tag{52}$$

First let $c_2 \to \infty$, so the final term vanishes. Note $|\mathscr{R}(k)| \to_d \infty$ as $k \to \infty$. So letting $k \to \infty$, then $c_1 \to \infty$,

$$\lim_{k\to\infty} \limsup_{n\to\infty} P(n^{-1/2}D(n,k) > b) = 0.$$

Now the distance from a vertex to $\mathscr{R}(n,k)$ is at most 1 more than the distance to $\mathscr{F}(n,k)$. So after rescaling edge lengths, the $\Delta(n,k)$ in (25) satisfies

$$\Delta(n,k) \le \sigma n^{-1/2}(D(n,k) + 1).$$

This and the previous equation verify condition (25) and complete the proof of Theorem 23. □

LEMMA 27. *For* $0 < c_1 < c_2 < \infty$,

$$\limsup_{n\to\infty} \sup_{c_1 n^{1/2} \le m \le c_2 n^{1/2}} \sup_{n - n^{2/3} \le n^* \le n} P(Y_m^* \ge bn^{1/2} | S_m = n^*) \le a(b, c_1),$$

where the constants $a(b, c_1)$ *satisfy*

$$a(b, c_1) \to 0 \quad \text{as } c_1 \to \infty, \text{ all } b > 0.$$

PROOF. By taking subsequences, it is enough to prove: If $m = m(n) \sim cn^{1/2}$ and $n^* = n^*(n) \sim n$, then

$$\limsup_{n\to\infty} P(Y_m^* \ge bn^{1/2} | S_m = n^*) \le a(b, c). \tag{53}$$

To prove (53), note

$$P(Y_m^* \geq bn^{1/2}|S_m = n^*) \leq mP(Y_1 \geq bn^{1/2}|S_m = n^*)$$
$$= m\sum_x P(Y_1 \geq bn^{1/2},\, X_1 = x)\frac{P(S_{m-1} = n^* - x)}{P(S_m = n^*)}$$

and by (38) this is asymptotically bounded by

$$cn^{1/2}\sum_x P(Y_1 \geq bn^{1/2},\, X_1 = x)\frac{g((\sigma^2/c^2)((n-x)/n))}{g(\sigma^2/c^2)}. \tag{54}$$

We may suppose that c is sufficiently large that the density g at (39) is increasing on $(0, \sigma^2/c^2)$. Then the fraction in (54) is at most 1. For $1/2 > \delta > 0$, it is easy to verify

$$g(\sigma^2(1-\delta)^2/c^2)\big/g(\sigma^2/c^2) \leq 2^{3/2}\exp(-c^2\delta/(2\sigma^2)).$$

So for $x \geq \delta n$ the right side is an upper bound for the fraction in (54). By splitting the sum (54) at $x = \delta n$ we see that (54) is bounded by

$$cn^{1/2}P(Y_1 \geq bn^{1/2},\, X_1 \leq \delta n) + cn^{1/2}P(Y_1 \geq bn^{1/2})2^{3/2}\exp(-c^2\delta/(2\sigma^2)).$$

Appealing to (35) and (37), the inequality in (53) holds when we define

$$a(b,c) = c\sigma^{-1}\delta^{-1/2}\kappa_1\exp\left(-\frac{b\delta^{-1}\sigma}{\kappa_2}\right) + c2^{3/2}\exp\left(-\frac{c^2\delta}{2\sigma^2}\right)\frac{2}{\sigma^2 b}.$$

Choosing $\delta = 1/c$ completes the proof. □

5.3. *Final remarks.* Other models of random trees where one might expect the Brownian CRT (or some other continuum random tree) limit are mentioned in Aldous [(1991b), Section 4]. Here let me discuss, from the viewpoint of this paper, the recent work of Durrett, Kesten and Waymire (1991). They consider the conditioned Galton–Watson trees $\mathscr{T}_n$, but instead of having unit edge lengths they allow the edge lengths to be random, i.i.d. copies of W, say, where $W \geq 0$ and $EW = 1$. Their main result implies that, provided

$$EW^2 < \infty, \tag{55}$$

the height of $\mathscr{T}_n$ behaves asymptotically as if $W \equiv 1$. Here (55) is the natural condition, because if

$$P(W > w) \sim cw^{-\alpha}, \quad \text{where } 1 < \alpha < 2, \tag{56}$$

then the height is at least order $n^{1/\alpha}$ by considering the longest edge. The first half of our proof in Theorem 23 (ignoring order structure) gave convergence of measure representations when $W \equiv 1$, and by the weak law of large numbers this extends unchanged to the "random W" model, assuming only first moments. Thus the case (56) is a natural example where measure representations converge but set representations do not. By assuming existence of sufficiently high moments of W it is not hard to check the hypotheses of Corollary 19 to show that the set representations converge, but I do not know if second moments suffice.

NOTE ADDED IN PROOF. A direct proof of Corollary 22 in terms of Brownian excursion only has been given by Le Gall (1992).

Acknowledgment. I thank the referees for pointing out obscurities in the original draft, and for encouraging me to remove moment conditions from Theorem 23.

REFERENCES

ALDOUS, D. J. (1991a). The continuum random tree. I. *Ann. Probab.* **19** 1–28.

ALDOUS, D. J. (1991b). The continuum random tree II: an overview. In *Proc. Durham Symp. Stochastic Analysis 1990* (M. T. Barlow and N. H. Bingham, eds.) 23–70. Cambridge Univ. Press.

ALDOUS, D. J. (1985). Exchangeability and related topics. *Ecole d'Eté St Flour 1983. Lecture Notes in Math.* **1117**. 1–198. Springer, Berlin.

ARRATIA, R. (1979). Coalescing Brownian motions on the line. Ph.D. dissertation, Univ. Wisconsin, Madison.

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

DURRETT, R., KESTEN, H. and WAYMIRE, E. (1991). On weighted heights of random trees. *J. Theoret. Probab.* **4** 223–237.

GNEDENKO, B. V. and KOLMOGOROV, A. N. (1954). *Limit Distributions for Sums of Independent Random Variables*. Addison-Wesley, Reading, MA.

GUTJAHR, W. and PFLUG, G. CH. (1992). The asymptotic contour process of a binary tree is Brownian excursion. *Stochastic Proc. Appl.* **41** 69–90.

KALLENBERG, O. (1973). Canonical representations and convergence criteria for processes with interchangeable increments. *Z. Wahrsch. Verw. Gebiete* **27** 23–36.

KENNEDY, D. P. (1975). The Galton–Watson process conditioned on the total progeny. *J. Appl. Probab.* **12** 800–806.

KINGMAN, J. F. C. (1987). The coalescent. *Stochastic Process. Appl.* **13** 235–248.

KOLCHIN, V. F. (1986). *Random Mappings*. Optimization Software, New York. (Translation of Russian original.)

LE GALL, J. F. (1991). Brownian excursions, trees and measure-valued branching processes. *Ann. Probab.* **19** 1399–1439.

LE GALL, J. F. (1992). The uniform random tree in a Brownian excursion. Unpublished.

LOUCHARD, G. (1986). Brownian motion and algorithmic complexity. *BIT* **26** 17–34.

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

NEVEU, J. and PITMAN, J. (1989). Renewal property of the extrema and tree property of a one-dimensional Brownian motion. *Seminaire de Probabilities XXIII. Lecture Notes in Math.* **1372**. 239–247. Springer, Berlin.

NGUYEN, B. G. (1990). Percolation of coalescing random walks. *J. Appl. Probab.* **27** 269–277.

O'BRIEN, G. L., TORFS, P. J. J. F. and VERVAAT, W. (1990). Stationary self-similar extremal processes. *Probab. Theory Related Fields*. **87** 97–120.

PEMANTLE, R. (1991). Choosing a spanning tree for the integer lattice uniformly. *Ann. Probab.* **19** 1559–1574.

SZNITMAN, A. S. (1989). A limiting result for the structure of collisions between many independent diffusions. *Probab. Theory Related Fields* **81** 353–381.

WAYMIRE, E. C. (1991). On network structure function computations. In *Time Series* (D. Brillinger, M. Rosenblatt and M. Taqqu, eds.). Springer, Berlin.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720