Asymptotic Fringe Distributions for General Families of Random Trees

David Aldous

The Annals of Applied Probability, Vol. 1, No. 2 (May, 1991), 228-266.

Stable URL:
http://links.jstor.org/sici?sici=1050-5164%28199105%291%3A2%3C228%3AAFDFGF%3E2.0.CO%3B2-8

The Annals of Applied Probability is currently published by Institute of Mathematical Statistics.

The Annals of Applied Probability
1991, Vol. 1, No. 2, 228–266

ASYMPTOTIC FRINGE DISTRIBUTIONS FOR GENERAL FAMILIES OF RANDOM TREES[1]

BY DAVID ALDOUS

University of California, Berkeley

Consider some model of random finite trees of increasing size. It often happens that the subtree at a uniform random vertex converges in distribution to a limit random tree. We introduce some structure theory for such *asymptotic fringe distributions* and illustrate with many examples.

1. Introduction. Random finite trees occur in a wide variety of contexts. One large area of use is in theoretical computer science, where trees are used, for example, as data structures for sorting and searching. To study the typical performance of a tree-based algorithm, one starts with a probability model for the underlying data, and then the mathematical problem becomes to study the distribution of a certain functional (measuring some aspect of algorithm performance) on a certain random tree (constructed by the algorithm from the data). The standard way that such problems are treated by computer scientists is to set up exact expressions or recurrences involving $a(n)$ = expected performance on data-set of size n and (if no simple exact formula emerges) to obtain asymptotics by analytic methods. Examples of this style of work can be found in Knuth [24, 25], Flajolet [13], Vitter and Flajolet [36].

To modern probabilists, a somewhat different viewpoint is natural. The fact that random walk processes can be rescaled to converge to the Brownian motion process is the prototype example of weak convergence; the point of weak convergence is that it implies convergence of many different functionals of the processes. Similarly, instead of studying convergence of particular functionals of a sequence of random trees, it is natural to seek to formalize the notion that the random trees themselves converge in distribution to a limiting random tree. This will imply convergence of various functionals, and to obtain an explicit expression for the limit of a particular functional, one seeks to analyze directly the distribution of this functional of the limit tree. We call this the *process viewpoint*. Compared with the traditional analytic methods, its advantages are: (i) it separates the issue of proving convergence to some limit from the issue of evaluating the limit explicitly; (ii) it eliminates some technicalities such as Tauberian theorems needed in analytic treatments; (iii) it is the natural setting for discussion of robustness of limits under variations of model.

Received January 1990; revised June 1990.
[1]Research supported by NSF Grant MCS-87-01426.
AMS 1980 *subject classifications*. 60C05, 05C80.
Key words and phrases. Random tree, random graph, subtree, branching process, recursive tree, binary search tree, random trie, stable type.

On the other hand, its disadvantages are: (i) it only handles asymptotics, not exact results, and is hard to use to get rates of convergence; (ii) conceptual overhead in setting up notions of convergence of trees; (iii) its own technicalities, such as tightness (Lemma 4) and uniform integrability.

In this paper we are concerned with local limit results which describe the shape of the tree near a typical vertex. Natural questions in any random tree model are:

1. What proportion of vertices are leaves?
2. What is the distribution of degrees of vertices?
3. What is the distribution of the size of the subtree rooted at a uniform random vertex?
4. What is the distribution of the number of vertices within distance k of a uniform random vertex?

These are the kind of functionals whose asymptotic behavior can be handled in our framework. To be honest, the most interesting (both theoretically and for applications) problems about random trees concern global functionals such as heights of trees or average distances between random vertices. To handle such global functionals, one apparently needs several different formulations of convergence of rescaled trees to handle the qualitatively different behavior of different families (see [3] for discussion). In contrast, the formulation of local limit theory presented in this paper applies to almost all families of random trees which have occurred in applied models.

Given a random rooted tree $\mathscr{T}$, pick a vertex V uniformly at random and let $\mathscr{F}$ be the subtree rooted at V. For many families of random trees $(\mathscr{T}_n;\, n \geq 1)$, it turns out that as the size n of the underlying trees $\mathscr{T}_n$ tends to infinity, the size of these *fringe* subtrees $\mathscr{F}_n$ remains bounded and their distributions either converge or have a simple asymptotic cycling behavior. At first sight, this is just one of many possible types of limit results about random trees. The purpose of this paper is to show that, from this starting point, a lot of interesting "structure" follows.

1. It implies an a priori stronger result on local convergence to a limit random infinite tree (of a kind we term *sin*-tree; see Section 4).
2. These limit random sin-trees satisfy an invariance property; there is a fruitful analogy (Section 7.4) between invariant sin-trees and classical stationary processes.
3. We can introduce a concept (*coherence*; see Section 2.6) for families of random trees, which is one formalization of the idea of the same random tree model for trees of different sizes.

There is a close connection between our work and the well-studied notion of *stable type structure* in supercritical branching process models of single-sex biological populations, surveyed in Jagers and Nerman [22]. In rather general models where offspring behave independently of parents after birth (note: by allowing individuals to be of many types and allowing dependence between types of parents and types of offspring, we can fit in many models which at first sight involve dependence) the following fundamental result holds. Pick an individual uniformly from those born before time t; then the distribution of type of that individual converges as $t \to \infty$ to a limit stable type. Such results apply to random family trees for populations evolving with time and possessing certain assumed independence structure. Only two of the computer science models we discuss (Sections 3.2 and 3.3) can be fitted into this setting. A (perhaps surprising) point of this paper is that similar results often hold in models of combinatorial random trees where there is no prior notion of time or evolving populations or independence of evolution of offspring.

Much of the computer science work in this area deals with models which are sufficiently simple that the form of the asymptotic fringe distribution can be seen immediately. For harder models, it seems necessary to face up to the issue of determining the asymptotic fringe distribution even if one is interested only in simpler questions such as asymptotic proportion of leaves. This is one reason why it seems worthwhile to write out our theory carefully. We must admit that the structure theory developed in this paper is a little removed from the issues in analyzing particular cases, where one is concerned with proving convergence to an explicit asymptotic fringe distribution and evaluating distributions of functionals of that limit distribution. Nevertheless we have one result of immediate practical application. Let X_n be an empirical average associated with a realization $\mathcal{T}_n$ of a random tree, for example, the (random) proportion of vertices of $\mathcal{T}_n$ which are leaves. Existence of an asymptotic fringe distribution implies that $p = \lim EX_n$ exists. One might anticipate the stronger result $X_n \to p$ in probability and seek to prove this by estimating variance(X_n) and applying Chebyshev's inequality. But there is a much more powerful abstract method. If the asymptotic fringe distribution is extreme, then (Proposition 7, a kind of weak ergodic theorem) we immediately have convergence in probability for empirical averages of *all* local functionals. Then Theorem 13 gives checkable conditions for extremality, which apply to almost all the standard models.

In Section 2, we record theoretical results about asymptotic fringe distributions. Section 3 discusses several examples, well-studied from other viewpoints, in which the form of the asymptotic fringe is intuitively obvious. In Section 4, we give theoretical results about limit sin-trees, deferring the proof of the main result to Section 5. Section 6 describes some mathematically interesting (and not all well-understood) examples where invariant sin-trees seem to be the correct and necessary approach to understanding asymptotics. Further technical examples and remarks are in Section 7.

The length of the paper reflects in part our attempt to make it accessible both to mainstream applied probabilists not familiar with the study of random trees (who we hope to interest by providing numerous examples) and also to workers from computer science accustomed only to the standard methods of that field. For the latter, we have sought to be careful and precise in setting out the theory (Sections 2, 4). But the examples are treated sketchily, emphasizing the description of the limits rather than details of the proofs of convergence. Some of the novel examples will be treated in detail elsewhere.

A method of analysis of certain special random trees in computer science has been called fringe analysis: the connection with our work is mentioned in Section 6.4.

2. Definitions and first results. In the phrase *random tree* the word "random" has the probabilist's meaning (having some distribution) rather than the combinatorialists's meaning of "uniform random". We are thinking of trees generated as part of some specified stochastic process. Such trees typically have extra structure: labels on vertices, weights on edges, birth-times of vertices representing individuals, ordering of births, and so on. All such extra structure can be combined by introducing a mark-space S and considering trees whose vertices are marked with (not necessarily distinct) elements of S. Such marked trees are the conceptually natural subject of our study. However, the mark space S plays no essential role in the theory, so we avoid distraction by considering unmarked trees, and briefly discuss the effect of marks in Section 7.3. In particular we consider unordered trees, since the orders can be included in the mark-space, but the reader who prefers may regard trees as ordered without serious danger. It is important that our trees be *rooted*, to define fringe subtrees (but for unrooted trees one can simply introduce a uniform random root; see Section 2.3). The *degree* of a vertex v is the number of edges at v. For $v \neq$ root, out-degree(v) = degree(v) − 1; that is, the number of edges at v leading in a direction away from the root. *Root-degree*(t) is a convenient abbreviation for the degree of the root of the tree t.

2.1. *Fringe distributions.* Let T be the set of finite rooted trees. Define a matrix $Q = (Q(s,t)\colon s, t \in T)$ as follows. In a tree s, let the root have children $v_1, \dots, v_d$, $d \geq 0$. Let $f(s, v_1), \dots, f(s, v_d)$ be the subtrees rooted at $v_1, \dots, v_d$. That is, $f(s, v_i)$ contains the vertices w such that the path from the root to w goes first to v_i. Making v_i the root, we may regard $f(s, v_i)$ as an element of T. Note it may happen that the trees $f(s, v_i)$ and $f(s, v_j)$ are the same (i.e., isomorphic) for some $i \neq j$. Define

$$Q(s,t) = \sum_i 1_{(f(s,v_i)=t)}. \tag{1}$$

Call a probability distribution π on T a *fringe distribution* if

$$\sum_s \pi(s) Q(s,t) = \pi(t) \quad \text{for all } t \in T. \tag{2}$$

In matrix notation, this is: $\pi Q = \pi$. The row-sum $\sum_t Q(s,t)$ gives root-degree(s): note this equals 0 for the trivial tree *triv* consisting only of the root. The matrix powers Q^i have an obvious significance: $Q^i(s,t)$ is the number of subtrees of s rooted at distance i from the root of s which are isomorphic to t. In particular, writing $|t|$ for the number of vertices of t,

$$|s| = \sum_{i=0}^{\infty} \sum_t Q^i(s,t). \tag{3}$$

We record a simple lemma.

LEMMA 1. *Let π be a fringe distribution. Then*

(a) $0 < \pi(triv) < 1$.
(b) $\sum_t \pi(t)$ *root-degree*$(t) = 1$.
(c) $\sum_t \pi(t) \times$ (*number of leaves of* t) $= \infty$, *and so* $\sum_t \pi(t)|t| = \infty$.

PROOF. (a) This is clear from (2), using the interpretation of Q^i.

(b) The sum is equal to $\sum_t \sum_s \pi(t) Q(t,s)$: Sum first over t and use the definition of fringe distribution.

(c) Writing $[t]$ for the number of leaves of t, we have

$$[t] = \sum_{i=0}^{\infty} Q^i(t, \text{triv})$$

and so

$$\sum_t \pi(t)[t] = \sum_t \pi(t) \sum_{i=0}^{\infty} Q^i(t, \text{triv}) = \sum_{i=0}^{\infty} \pi(\text{triv}) = \infty,$$

the last step using (a). □

REMARK. Part (c) says the π-average tree size (number of vertices) is infinite. Examples will show that the π-average *height* may be finite or infinite. Recall that the height ht(t) is the maximum, over vertices v, of the distance from the root to v.

Given a fringe distribution π, we can define a Markov transition matrix P_π on $\{t: \pi(t) > 0\}$ by

(4) $$P_\pi(s,t) = \pi(t) Q(t,s) / \pi(s).$$

Then

$$\pi(s) P_\pi^n(s,t) \equiv \pi(t) Q^n(t,s), \qquad n \geq 1.$$

This is used in the proof of Proposition 2 below, and extensively in Section 4.

2.2. *Convergence of random fringe subtrees.* We now connect fringe distributions with the notion of fringe subtrees described in the Introduction. If s is a tree and v a vertex of s, then as above $f(s,v)$ is the fringe subtree rooted at v. Writing V for a uniform random vertex of s, we can consider the random fringe subtree $f(s,V)$ rooted at V as a random element of T. Define

$$F(s,t) = P(f(s,V) = t).$$

Given a random tree $\mathcal{T}$ we can consider the random fringe subtree $f(\mathcal{T}, V)$, where V is uniform on $\mathcal{T}$. There is a map taking the distribution μ of $\mathcal{T}$ to the distribution of $f(\mathcal{T}, V)$, and in vector-matrix notation this map is just

$$\mu \to \mu F.$$

We can write F in terms of Q via

$$F(s,t) = \frac{1}{|s|} \sum_{i=0}^{\infty} Q^i(s,t).$$

The key idea, formalized in Propositions 2 and 3 is: Fringe distributions are exactly the possible limit distributions for fringe subtrees $f(\mathcal{T}, V)$ as the size of trees $|\mathcal{T}| \to \infty$. To say this sharply, consider random trees $\mathcal{T}_k$ such that

$$|\mathcal{T}_k| \to_d \infty \quad \text{as } k \to \infty \tag{5}$$

but with otherwise arbitrary distributions. Write $\mathcal{F}_k \equiv f(\mathcal{T}_k, V_k)$, where V_k is uniform on $\mathcal{T}_k$, and call $\mathcal{F}_k$ the random fringe subtree associated with $\mathcal{T}_k$.

PROPOSITION 2. *Let π be a fringe distribution. Then there exist random trees $\mathcal{T}_k$ satisfying* (5) *such that*

$$P(\mathcal{F}_k = t) \to \pi(t), \quad \textit{all } t \in T.$$

In this setting, say π is the asymptotic fringe distribution associated with $\mathcal{T}_k$.

PROOF. Write $L_k = \{t: |t| < k\}$. Define a measure θ_k on L_k by

$$\theta_k(w) = |w| \sum_{v \in T \setminus L_k} \pi(v) Q(v, w).$$

The associated fringe measure $\theta_k F$ is supported by L_k, and for $t \in L_k$,

$$\begin{aligned}
\theta_k F(t) &= \sum_{w \in L_k} \frac{1}{|w|} \theta_k(w) \sum_{i \ge 0} Q^i(w, t) \\
&= \sum_{v \in T \setminus L_k} \sum_{w \in L_k} \pi(v) Q(v, w) \sum_{i \ge 0} Q^i(w, t) \\
&= \pi(t) \sum_{v \in T \setminus L_k} \sum_{w \in L_k} \sum_{i \ge 0} P_\pi^i(t, w) P_\pi(w, v) \\
&= \pi(t).
\end{aligned}$$

Here the third equality follows from (4), and then the sum in that line is the expected number of times that the P_π-chain, started at t, crosses out of L_k. But this number is identically 1, because in a P_π-chain (X_i) the sizes $|X_i|$ are strictly increasing.

Since the map $\theta \to \theta F$ preserves mass, θ_k has total mass $\pi(L_k)$. Let $\mathcal{T}_k$ have distribution $\theta_k / \pi(L_k)$. Then the identity above says $P(\mathcal{F}_k = \cdot) = \pi(\cdot | L_k)$, and the result follows. □

To study a converse, let $\mathcal{T}_k$ be random trees and suppose

$$\pi(t) \equiv \lim_k P(\mathcal{F}_k = t) \quad \text{exists for each } t \in T. \tag{6}$$

Then $\sum_t \pi(t) \le 1$, by Fatou's lemma. By considering the case where $\mathcal{T}_k$ is the deterministic path $(1, 2, \ldots, k)$ rooted at 1, we may have $\pi \equiv 0$. So let us impose the condition

$$\sum_t \pi(t) = 1. \tag{7}$$

Another easy use of Fatou's lemma shows $\sum_t \pi(t)$root-degree$(t) \le 1$. By considering the case of a star graph rooted at its center, we may have $\pi(\text{triv}) = 1$, and so this sum may be 0. So in view of Lemma 1(b), let us impose the condition

$$(8) \qquad \sum_t \pi(t)\text{root-degree}(t) = 1,$$

which turns out to imply (7).

PROPOSITION 3. *Let $\mathcal{T}_k$ be random trees satisfying* (5). *Suppose the limit π in* (6) *exists and satisfies* (8). *Then π is a fringe distribution.*

PROOF. By considering s, t with $Q(s,t) > 0$, we see

$$(9) \qquad P(\mathcal{F}_k = t \mid \mathcal{T}_k) = \sum_s P(\mathcal{F}_k = s \mid \mathcal{T}_k) Q(s,t) + |\mathcal{T}_k|^{-1} 1_{(\mathcal{T}_k = t)}.$$

Taking expectations and limits, the left side converges to $\pi(t)$, and so it suffices to show

$$\sum_s P(\mathcal{F}_k = s) Q(s,t) \to \sum_s \pi(s) Q(s,t)$$

for fixed t. For each t, the right side is at most the lim inf of the left side, so it suffices to show that the double sums over s and t converge. Now the left double sum tends to 1, using (9), and the right double sum equals 1 by (8). □

We state without proof a routine tightness lemma.

DEFINITION. A sequence of random trees satisfying (5) is *fringe-tight* if every subsequence has a further subsequence which possesses an asymptotic fringe distribution.

LEMMA 4. *A sequence $\mathcal{T}_k$ of random trees satisfying* (5) *is fringe-tight iff*:

(i) $\lim_{a\to\infty} \limsup_{k\to\infty} P(|\mathcal{F}_k| > a) = 0$; *and*
(ii) $\lim_{d\to\infty} \limsup_{k\to\infty} E\, degree(V_k) 1_{(degree(V_k) > d)} = 0$.

In particular, if (5) holds and if the limits (6) exist, then π is the fringe distribution associated with $\mathcal{T}_k$ iff (i) and (ii) hold.

Informally, condition (i) says that the trees $(\mathcal{T}_k)$ should not be like the *path* tree and (ii) says they should not be like the *star* tree. In (i), the size $|f|$ could be replaced by the height ht(f).

Here is another easy fact:

LEMMA 5. *Let $(\mathcal{T}_k)$ be fringe-tight, and let $\mathcal{F}_k$ be the random fringe subtrees. Then*:

(i) $E|\mathcal{T}_k| \to \infty$.
(ii) $E|\mathcal{F}_k|/|\mathcal{T}_k| \to 0$.
(iii) $E\, \text{ht}(\mathcal{T}_k)/|\mathcal{T}_k| \to 0$.
(iv) $\text{ht}(\mathcal{T}_k) \to \infty$ *in probability*.

PROOF. Assertion (i) is immediate from Lemma 1(c) and assertion (ii) is a consequence of Lemma 4, since for fixed a, we have

$$E\frac{|\mathscr{F}_k|}{|\mathscr{T}_k|} \le E\frac{a}{|\mathscr{T}_k|} + P(|\mathscr{F}_k| > a).$$

For (iii), note that there are at least $\frac{1}{2}\,\mathrm{ht}(\mathscr{T}_k)$ vertices v for which the fringe subtree rooted at v has size at least $\frac{1}{2}\,\mathrm{ht}(\mathscr{T}_k)$. Thus

$$E(|\mathscr{F}_k|\,|\,\mathscr{T}_k) \ge \frac{1}{2}\frac{\mathrm{ht}(\mathscr{T}_k)}{|\mathscr{T}_k|} \times \frac{1}{2}\,\mathrm{ht}(\mathscr{T}_k)$$

and so

$$E\frac{|\mathscr{F}_k|}{|\mathscr{T}_k|} \ge \frac{1}{4}E\left(\frac{\mathrm{ht}(\mathscr{T}_k)}{|\mathscr{T}_k|}\right)^2$$

and then (iii) follows from (ii). Assertion (iv) is immediate from Proposition 11. □

2.3. *Unrooted trees.* Our treatment so far has involved rooted trees. Given an unrooted tree, we can make it a rooted tree by choosing a uniform random vertex as root: In this way we can talk about asymptotic fringe distributions for unrooted trees. We want to say it makes no difference whether the original trees are rooted or unrooted. This raises a question. Start with rooted trees, then consider them as unrooted by deleting the root, then randomly reroot: Does this affect asymptotic fringe distributions? Fortunately, the answer is no.

LEMMA 6. *Let $\mathscr{T}_k$ be random trees satisfying* (5) *with roots r_k. Let $\mathscr{S}_k$ be the same trees, rooted at uniform random vertices U_k. If either of the sequences $(\mathscr{T}_k)$, $(\mathscr{S}_k)$ has asymptotic fringe distribution π, then so does the other.*

PROOF. A branch of a rooted tree is a fringe subtree rooted at a child of the original root. Consider a vertex $v_k \in \mathscr{T}_k$. In each of $\mathscr{T}_k$ and $\mathscr{S}_k$, there is a fringe subtree rooted at v_k, and these two subtrees are identical provided r_k and U_k are in the same branch of the tree, considered as being rooted at v_k. Thus it is enough to prove

(10) $\quad P(U_k \text{ and } r_k \text{ in different branches of } \mathscr{T}_k \text{ rooted at } V_k) \to 0,$

where V_k is a uniform random vertex, independent of U_k.

First suppose an asymptotic fringe distribution exists for $\mathscr{T}_k$, considered as rooted at r_k. The event in (10) can be rewritten as:

$$U_k \text{ is in the fringe subtree rooted at } V_k.$$

This event has conditional probability (given $\mathscr{T}_k$ and V_k) equal to $|f(\mathscr{T}_k, V_k)|/|\mathscr{T}_k|$, and so has unconditional probability $E|f(\mathscr{T}_k, V_k)|/|\mathscr{T}_k|$. Thus (10) follows from Lemma 5(ii).

Conversely, suppose $\mathscr{S}_k$ has an asymptotic fringe distribution. Writing $\mathscr{F}_k$ for the random fringe subtree $f(\mathscr{S}_k, V_k)$ of $\mathscr{S}_k$, the event in (10) is the event

$\{\mathscr{T}_k$ contains $r_k\}$. This event has conditional probability (given $\mathscr{S}_k$ and r_k) equal to $h(r_k)/|\mathscr{S}_k|$, where $h(r_k)$ is the height of r_k in $\mathscr{S}_k$, and so the unconditional probability (10) equals $Eh(r_k)/|\mathscr{S}_k|$. The result now follows from Lemma 5(iii). □

REMARK. An alternative definition of *fringe subtree* for an unrooted tree t is as follows. For each vertex v, consider t as rooted at v, delete a branch of maximal size (choosing arbitrarily between equals), and let $f(t, v)$ be the remaining tree. Lemma 6 remains true with this definition.

2.4. *Empirical proportions and extreme fringe distributions.* The set of all fringe distributions is convex, that is, closed under mixtures. Call a fringe distribution π *extreme* if it has no representation as a mixture

$$\pi = a\pi_1 + (1-a)\pi_2, \qquad 0 < a < 1, \qquad \pi_i \text{ fringe distributions}$$

except for the trivial representation with $\pi_1 = \pi_2 = \pi$. Theorem 13 will give checkable conditions for extremality. Choquet theory (e.g., Dynkin [11]) says every fringe distribution is an integral mixture of extreme fringe distributions, though we shall not use that fact. Our interest comes from Proposition 7.

Recall $F(s,t)$ defined in Section 2.2. For a random tree $\mathscr{T}_k$, $F(\mathscr{T}_k, t)$ is the (random) *empirical* proportion of fringe subtrees of $\mathscr{T}_k$ which are isomorphic to t. If there is an associated asymptotic fringe distribution π, then by definition $EF(\mathscr{T}_k, t) \to \pi(t)$. If π is extreme, then we can automatically strengthen this to convergence in probability.

PROPOSITION 7. *Let π be the asymptotic fringe distribution associated with $\mathscr{T}_k$. Suppose π is extreme. Then for fixed $t \in T$, $\varepsilon > 0$,*

$$P(|F(\mathscr{T}_k, t) - \pi(t)| > \varepsilon) \to 0 \quad \text{as } k \to \infty.$$

PROOF. Suppose not. By passing to a subsequence, there exist $t \in T$ and $0 < b < \pi(t)$ such that $\lim_{k\to\infty} P(F(\mathscr{T}_k, t) \le b) = a$ exists and $0 < a < 1$. Define $B_1 = \{s\colon F(s,t) \le b\}$ and let B_2 be its complement. Write $\mu_k = \operatorname{dist}(\mathscr{T}_k)$ and $\mu_{k,i} = \operatorname{dist}(\mathscr{T}_k | \mathscr{T}_k \in B_i)$, $i = 1, 2$. Then μ_k is a mixture

$$\mu_k(\cdot) = \sum_{i=1}^{2} \mu_k(B_i)\mu_{k,i}(\cdot),$$

where the weights converge to $(a, 1-a)$. Using Lemma 4 we can pass to a further subsequence in which $(\mu_{k,i};\ k \ge 1)$ has an asymptotic fringe distribution, π_i say. Then $\pi = a\pi_1 + (1-a)\pi_2$. But then extremality implies $\pi_1 = \pi_2 = \pi$. But from the definition of B_1 we have $\pi_1(t) \le b < \pi(t)$, a contradiction.

In our general theory there is no explicit connection between the random trees $\mathscr{T}_k$ for different k. Suppose for a moment that we do have a tree-process

$(\mathcal{T}_u: 0 \le u < \infty)$ with continuous parameter u. In this setting, say the tree-process has an *a.s. fringe* if

(11) $$F(\mathcal{T}_u, t) \to \pi(t) \quad \text{a.s. as } u \to \infty.$$

When this holds, any family $(\mathcal{T}_k^*; k \ge 1)$ of the form

(12) $$\mathcal{T}_k^* = \mathcal{T}_{U_k}, \qquad U_k \to_p \infty$$

will have asymptotic fringe distribution π. Thus one may seek to show that a discrete family of random trees has an asymptotic fringe by embedding it via a random time-change into a continuous-time tree-process with an a.s. fringe. We shall see in Section 3 that this is a useful technique. □

2.5. *A generic counterexample.* A noteworthy feature of Proposition 7 is that the result requires only a hypothesis on the asymptotic fringe, rather than hypotheses on the underlying trees $(\mathcal{T}_k)$. The following construction shows that such a feature cannot be expected in more refined results. Given random trees $\mathcal{T}_k$ with $n(k)$ vertices and $\mathcal{S}_k$ with $m(k)$ vertices, form a combined tree $\mathcal{T}_k^*$ by identifying the root of $\mathcal{T}_k$ with some leaf of $\mathcal{S}_k$. If $(\mathcal{T}_k)$ has asymptotic fringe distribution π and $m(k) = o(n(k))$, then $(\mathcal{T}_k^*)$ also has asymptotic fringe distribution π. But it is clear that, if one is interested in (say) heights of trees, or central limit theorems for numbers of leaves, then the results may be very different for $(\mathcal{T}_k^*)$ than for $(\mathcal{T}_k)$. In other words, the theory developed in this paper can handle only laws of large numbers, and not more refined convergence assertions.

2.6. *Coherence.* Let $\psi: T \to \{0, 1, 2, \dots\}$ be a functional which we regard as measuring the size of a tree t. Usually we will have $\psi(t) = |t|$, the number of vertices of t, but it will sometimes be more natural to have ψ denote the number of leaves, or the height, of t. Write $T_n = \{t: \psi(t) = n\}$.

DEFINITION. Let μ_n be probability distributions on T_n, $n \ge 1$. Suppose (μ_n) is fringe-tight; and

(13) $$\text{dist}(\mathcal{F}_n | T_m) \to \mu_m \quad \text{as } n \to \infty, \text{each } m \ge 1,$$

where $\mathcal{F}_n$ is the random fringe subtree derived from μ_n. Then call (μ_n) *ψ-coherent.*

The definition is easier to understand in the context of the following (mathematically obvious) lemma.

LEMMA 8. *Let (π_n) be probability distributions on T_n, $n \ge 1$. Suppose (π_n) has asymptotic fringe distribution π. Then (π_n) is ψ-coherent iff*

(14) $$\pi_n(\cdot) = \pi(\cdot | T_n) \quad \textit{for each } n.$$

In other words, π is a mixture

$$\pi = \sum_n a_n \pi_n$$

of the (π_n), for some weights (a_n). When this holds, we also say that π is ψ-coherent. The intuitive idea of coherence is (14), but the general definition is designed to cover also cases where, instead of an asymptotic fringe distribution, there is asymptotic cycling behavior caused by discreteness effects.

When $\psi(t) = |t|$, the number of vertices, we just say *size-coherent*. The definitions have obvious modifications in the case where $\pi(T_n)$ is nonzero only for some subsequence of integers n.

An ultimate hope is that coherence may be useful as a general hypothesis for proving deeper results about families of random trees. See Section 7.5 for one conjecture. But this is pure speculation. In this paper we merely verify that the standard examples in Section 3 are coherent. In particular, the three basic random tree models in computer science discussed in Flajolet [14] are coherent.

3. Standard examples.

3.1. *The general continuous-time branching process*. This process is parametrized by a random collection $(C; \beta_i, 1 \le i \le C)$, where $0 \le \beta_i < \infty$, the β_i are nondecreasing in i and where for our purposes we may take $C \ge 1$, $EC > 1$.

Each individual has C children, at times β_i after the individual's birth; these times and numbers of children being independent for different individuals. At time 0, a single individual is born. For $0 \le t < \infty$, let $\mathscr{T}_t$ be the family tree of this process at time t. Under minor technical assumptions on (β_i) (nonexplosiveness and nonlattice assumptions), the following fundamental results hold. See Jagers [21], Jagers and Nerman [22] and Nerman and Jagers [32] for details.

1. There is a *Malthusian parameter* $\theta \in (0, \infty)$ such that

$$E|\mathscr{T}_t| \sim ce^{\theta t} \quad \text{as } t \to \infty \quad \text{for some } 0 < c < \infty. \tag{15}$$

2. $(\mathscr{T}_t)$ has an asymptotic fringe distribution $\pi = \text{dist}(\mathscr{F})$ [in fact, it has an a.s. fringe in the sense of (11)] given by

$$\mathscr{F} = \mathscr{T}_L, \quad \text{where } L \text{ has exponential}(\theta) \text{ distribution and is independent of } (\mathscr{T}_t). \tag{16}$$

We can then write down expressions for probabilities involving the asymptotic fringe distribution. For example,

$$P(\text{root-degree}(\mathscr{F}) \ge d) = \int_0^\infty \theta e^{-\theta s} P(\beta_d < s)\, ds = Ee^{-\theta\beta_d}, \tag{17}$$

where β_d is the dth birth time. Also,

$$P(|\mathscr{F}| \ge 3)$$
$$= \int_0^\infty \left(P(\beta_2 \le s) + \int_0^s P(\beta_1 \in du, \beta_2 > s) P(\beta_1 \le s - u)\, du \right) \theta e^{-\theta s}\, ds.$$

Consider now the embedded discrete trees

$$\mathscr{T}_n^* = \mathscr{T}_{U_n}, \qquad U_n = \inf\{u : |\mathscr{T}_u| \geq n\}.$$

Then $(\mathscr{T}_n^*)$ also has asymptotic fringe $\mathscr{F}$, by the a.s. fringe property. In general these embedded discrete random trees are not simple to describe, and are not coherent, but there are interesting special cases. Consider the very special property: For each t, n,

(18) $\qquad \mathscr{T}_n^*$ has the conditional distribution of $\mathscr{T}_t$ given $|\mathscr{T}_t| = n$.

This implies that, for L as in (16),

$\mathscr{T}_n^*$ has the conditional distribution of $\mathscr{T}_L$ given $|\mathscr{T}_L| = n$.

Then (16) implies

(19)
$$P(\mathscr{F} = \cdot) = \sum_n a_n P(\mathscr{T}_n^* = \cdot),$$

where

(20)
$$a_n = P(|\mathscr{T}_L| = n) = \int_0^\infty \theta e^{-\theta s} P(|\mathscr{T}_s| = n)\, ds.$$

So $(\mathscr{T}_n^*)$ is a *coherent* family.

In fact condition (18) is so special that it occurs essentially only in the following two examples. But it does enable these (at first sight rather different) two examples to be treated in almost exactly the same way.

3.2. *Yule trees/recursive trees.* Consider the simple continuous-time pure birth process: Initially there is one individual and each individual gives birth throughout all time at exponential rate 1. This process (or rather, the total population at time t) is often called the *Yule process*. Let $\mathscr{T}_t$ denote the family tree of individuals born before time t. It is hard to resist calling $(\mathscr{T}_t;\, 0 \leq t < \infty)$ the *Yule tree* process.

The embedded discrete-time random trees $(\mathscr{T}_n^*)$ can be described in a different way. Construct a tree on labeled vertices $(1, 2, \ldots)$ by making 1 the root and using the recursion

(21) $\qquad$ vertex j is connected to vertex U_j,

where U_j is uniform on $(1, 2, \ldots, j-1)$ and the (U_j) are independent. Removing labels gives a process identical to $(\mathscr{T}_n^*)$. Of course, this is a very natural model for random trees, which has been discussed by many authors, and is sometimes called the *random recursive tree*—Dondajewski, Kirschenhofer and Szymanski [10].

The Yule tree process is the case of the general continuous-time branching process for which $C = \infty$ and (β_i) are the times of a Poisson(1) process. It is clear that $N(t) \equiv |\mathscr{T}_t|$ satisfies

$$E(dN(t) | N(t) = n) = n\, dt$$

and so $E|\mathscr{T}_t| = e^t$, giving Malthusian parameter $\theta = 1$.

Fix t and condition on the times of births of $(\mathcal{T}_u: 0 \le u \le t)$ being $(\gamma_1, \gamma_2, \dots, \gamma_n)$. Then the embedded discrete-time tree still grows according to the recursion (21), by elementary properties of the exponential distribution. Hence the very special property (18) holds, so the family of recursive trees $(\mathcal{T}_n^*)$ is coherent and we can apply (19, 20). Here it is elementary that $a_n = \int_0^\infty e^{-\theta s} P(N(t) = s)\, ds$ satisfies $a_1 = \frac{1}{2}$, $a_n = na_n - (n-1)a_{n-1}$, which has solution $a_n = 1/(n(n+1))$. Thus the asymptotic fringe $\mathcal{F}$ satisfies

$$(22) \qquad P(\mathcal{F} = \cdot\,) = \sum_n \frac{1}{n(n+1)} P(\mathcal{T}_n^* = \cdot\,)$$

and in particular,

$$P(|\mathcal{F}| = n) = \frac{1}{n(n+1)}, \qquad n \ge 1.$$

So this is the limit distribution of size of random subtrees in the random recursive tree.

A calculation direct from (17) gives

$$P(\text{root-degree}(\mathcal{F}) = d) = \left(\tfrac{1}{2}\right)^d, \qquad d \ge 0.$$

So this geometric distribution is the limit distribution of out-degrees in the random recursive tree. Further properties of $\mathcal{F}$ may be studied using standard branching process methods.

3.3. *Binary search tree.* Let $\mathcal{T}_t$ be the special case of the general continuous-time branching process, where $C = 2$ and $(\beta_1, \beta_2) = (\min(\xi_1, \xi_2), \max(\xi_1, \xi_2))$ for independent ξ_1, ξ_2 with exponential(1) distributions. Note that the initial (time 0) birth rate is 2 and each time a birth occurs, the overall birth rate increases by 1. Thus $N(t) = |\mathcal{T}_t|$ satisfies

$$E(dN(t) | N(t) = n) = (n+1)\, dt$$

giving $EN(t) = 2e^t - 1$. So the Malthusian parameter $\theta = 1$.

We now describe a discrete-time random tree process $(\mathcal{T}_n^*)$. By binary tree we mean here a rooted tree with out-degrees at most 2, regarded as an *ordered* tree (i.e., distinguishing between left and right offspring). Given a binary tree t, there are exactly $|t| + 1$ different (ordered) binary trees obtainable by adding a new vertex and edge. Thus we can define $\mathcal{T}_n^*$ recursively as

(23) $\mathcal{T}_{n-1}^*$ plus a new vertex chosen uniformly randomly from the set of n vertices it is possible to add.

Call this the *binary search tree* process. These are the trees constructed by a standard algorithm from i.i.d. continuous r.v.'s.; see Knuth [[25], Section 6.2.2].

As in the preceding example, fix t and condition on the times of births of $(\mathcal{T}_u: 0 \le u \le t)$ being $(\gamma_1, \gamma_2, \dots, \gamma_n)$. Then the embedded discrete-time tree still grows according to the recursion (23), by elementary properties of the exponential distribution. Hence the very special property (18) holds, so the family of binary search trees ($\mathscr{T}_n^*$) is coherent, and we can apply (19, 20). As in the above section, an elementary argument gives the form of the asymptotic fringe distribution $\mathscr{F}$:

$$P(\mathscr{F} = \cdot) = \sum_n \frac{2}{(n+1)(n+2)} P(\mathscr{T}_n^* = \cdot). \tag{24}$$

In particular,

$$P(|\mathscr{F}| = n) = \frac{2}{(n+1)(n+2)}, \qquad n \geq 1.$$

Using (17) we have

$$P(\text{root-degree}(\mathscr{F}) \geq 1) = Ee^{-\beta_1} = \tfrac{2}{3},$$
$$P(\text{root-degree}(\mathscr{F}) \geq 2) = Ee^{-\beta_2} = \tfrac{1}{3}.$$

In other words,

$$\text{root-degree}(\mathscr{F}) \text{ is uniform on } \{0, 1, 2\}. \tag{25}$$

3.4. *Conditioned critical branching processes.* Consider a simple discrete-time (Galton–Watson) branching process with offspring distribution ξ, and suppose we are in the *critical* setting $E\xi = 1$. To avoid technicalities, suppose $P(\xi = 0) > 0$, $E\xi^2 < \infty$ and that ξ is non(sub)lattice. Write $\mathscr{T}$ for the family tree of the entire process started with one progenitor. By criticality, $|\mathscr{T}| < \infty$ a.s. For $n \geq 1$, write $\mathscr{T}_n$ for $\mathscr{T}$ conditioned on $|\mathscr{T}| = n$. The next lemma implies that ($\mathscr{T}_n$) is a (size-) coherent family with asymptotic fringe $\mathscr{T}$.

LEMMA 9. *Let $\mathscr{F}_n$ be the fringe subtree associated with $\mathscr{T}_n$. Then*

$$P(\mathscr{F}_n = t) \to P(\mathscr{T} = t) \quad \text{as } n \to \infty, t \in \mathbf{T}.$$

Although I do not know an explicit reference, this is closely related to work of the Russian school on aspects of conditioned branching processes motivated by random mappings and (combinatorial) random trees. Their work is summarized in the book of Kolchin [26]. To make an explicit proof, note that the special case of the lemma with t = triv can be written as

$$E(\text{proportion of leaves in } \mathscr{T}_n) \to P(\xi = 0). \tag{26}$$

This result (in fact the deeper CLT) is from [26] (Theorem 2.3.1, page 113). We also need the tail estimate ([26], Lemma 2.1.4, page 105)

$$P(|\mathscr{T}| = n) \sim n^{-3/2} (2\pi \operatorname{var}(\xi))^{-1/2}. \tag{27}$$

PROOF OF LEMMA 9. Fix n and trees s, t with $|s| + |t| = n + 1$. Let V be a uniform random vertex of $\mathscr{T}$. Let $\mathscr{F}^0$ be the fringe subtree at V and let $\mathscr{S}^0$ be the remainder of $\mathscr{T}$ after $\mathscr{F}^0$ (but not vertex V) is removed. It is easy to see

$$P(\mathscr{S}^0 = s, \mathscr{F}^0 = t) = \frac{[s]}{n} \frac{P(\mathscr{T} = s)}{P(\xi = 0)} P(\mathscr{T} = t),$$

where $[s]$ is the number of leaves of s. Now

$$P(\mathscr{F}_n = t) = \sum_{s:|s|=n-|t|+1} P(\mathscr{S}^0 = s, \mathscr{F}^0 = t)/P(|\mathscr{T}| = n)$$

and this rearranges to

$$\frac{P(\mathscr{F}_n = t)}{P(\mathscr{F} = t)} = \frac{P(|\mathscr{T}| = n - |t| + 1)}{P(|\mathscr{T}| = n)} \frac{n - |t| + 1}{n} \frac{E(\text{proportion of leaves in } \mathscr{T}_{n-|t|+1})}{P(\xi = 0)}.$$

As $n \to \infty$, each ratio $\to 1$, by (26, 27). □

Remark. With special distributions for ξ, the conditioned Galton–Watson tree becomes the uniform distribution on certain combinatorial sets of rooted labeled trees. This model has been studied by combinatorialists (e.g., Meir and Moon [27]) under the name *simply generated* trees. In particular, the case where ξ has Poisson(1) distribution gives the uniform random unordered labeled tree. The case of the geometric distribution $P(\xi = i) = (\frac{1}{2})^i$, $i \geq 0$, gives the uniform random ordered labeled tree.

3.5. *Trees and strings*. There is a general construction of trees whose vertices are marked with strings. Fix $d \geq 2$ and let $\mathbf{b}$ denote a string $\mathbf{b} = (b_1, b_2, \dots, b_i)$; $i \geq 1$, $b_i \in \{1, 2, \dots, d\}$. Call $(b_1, \dots, b_{j-1})$ the *parent* of the string $(b_1, \dots, b_j)$ and for $i \leq j$, call the string $(b_1, \dots, b_i)$ a *prefix* of the string $(b_1, \dots, b_j)$. Now let $\mathbf{B}$ be a finite family of strings satisfying:

If $\mathbf{b} \in \mathbf{B}$, then the parent $\mathbf{b}'$ of $\mathbf{b}$ is in $\mathbf{B}$.

Then we can regard $\mathbf{B}$ as a *tree* in a natural way: the vertices are the strings $\mathbf{b}$ and the edges are $(\mathbf{b}, \mathbf{b}')$, where $\mathbf{b}'$ is the parent of $\mathbf{b}$. The root of the tree is the null string.

The trees in this and the following two sections use the construction above. We first consider two simple deterministic examples. Let $d = 2$ and let $t(n)$ be the tree on the n strings

root, 0, 1, 00, 01, 10, 11, 000, 001, 010, 011, 100, 101, … .

Note that $t(2^{h+1} - 1)$ is the complete binary tree of height h. Let $\mathscr{F}_n$ be the random fringe subtree of $t(n)$. It is easy to see that

(28) if $|\mathscr{F}_n| = m$, then, considered as rooted unlabeled trees, $\mathscr{F}_n = t(m)$.

So the family $(t(n))$ is coherent; indeed, we have equality in (13) rather than merely convergence. It is also easy to see that there is an asymptotic fringe distribution $\mathscr{F}$ given by

$$P(\mathscr{F} = t(2^{h+1} - 1)) = \left(\tfrac{1}{2}\right)^h, \qquad h \geq 0. \tag{29}$$

Now consider the trees $u(n)$ defined similarly, except that the rows are now filled out in reversed significance order, for example,

000, 100, 010, 110, 001, 101, 011, 111.

Property (28) still holds, so the family is coherent. If we consider a subsequence $n_j \sim (1+\alpha)2^j$, $0 < \alpha < 1$, then it is not hard to see (draw a picture) that this subsequence has asymptotic fringe distribution $\mathcal{F}_\alpha$ specified as follows.

Let α have binary expansion $\sum_{i \geq 1} b_i 2^{-i}$ and define

$$[\alpha]_h = \sum_{i=1}^{h} b_i 2^i,$$

$$\langle \alpha \rangle_h = \sum_{i \geq 1} b_{h+i} 2^{-i}.$$

Then

$$P(\mathcal{F}_\alpha = u(1)) = \frac{1}{2(1+\alpha)}, \qquad \alpha \leq \frac{1}{2}$$

$$= \frac{\alpha}{(1+\alpha)}, \qquad \alpha \geq \frac{1}{2},$$

$$P(\mathcal{F}_\alpha = u(2^h + [\alpha]_h - 1)) = \frac{1 - \langle \alpha \rangle_h}{(1+\alpha)2^h}, \qquad h \geq 2, \text{ or } h \geq 1 \text{ and } [\alpha]_h > 0,$$

$$P(\mathcal{F}_\alpha = u(2^h + [\alpha]_h)) = \frac{\langle \alpha \rangle_h}{(1+\alpha)2^h}, \qquad h \geq 1.$$

Since this depends on α, the entire sequence $(u(n))$ cannot have an asymptotic fringe distribution; instead, the fringes asymptotically cycle continuously through $(\mathcal{F}_\alpha, 0 \leq \alpha \leq 1)$.

This example is somewhat artificial, but the same cyclic behavior occurs naturally in the following two examples. Indeed, in natural examples of nondeterministic trees, we see cyclic behavior iff $(H_n - \text{median}(H_n))$ is stochastically bounded, where H_n is the height of a uniform random vertex. But it does not seem easy to formalize this idea.

3.6. *Tries.* Fix $d \geq 2$. Consider a collection of k distinct strings $(\mathbf{b}_1, \ldots, \mathbf{b}_k)$, where each string is of the form $\mathbf{b} = (b_1, b_2, \ldots)$, $b_i \in \{1, 2, \ldots, d\}$. Then replace the collection $(\mathbf{b}_j)$ with the collection $(\mathbf{b}'_j)$, where $\mathbf{b}'_j$ is the shortest prefix of $\mathbf{b}_j$ which is not a prefix of any other $\mathbf{b}$ in the collection. Then set

$$\mathbf{B} = \{\mathbf{a}\colon \mathbf{a} \text{ is a prefix of some } \mathbf{b}'_j\}.$$

As described in the previous section, we can regard $\mathbf{B}$ as a tree. The leaves of the tree are exactly the $\mathbf{b}'$s.

This construction gives the *trie* on $(\mathbf{b}_1, \ldots, \mathbf{b}_k)$; see, for example, Knuth ([25], Section 6.3). Note that the trie has exactly k leaves, but a variable total number of vertices. By taking each string $\mathbf{b}$ to have independent uniform random entries and different strings independent, we get a simple *random* trie $\mathcal{T}_k$.

Given such a random trie, for each string **a**, let

$\phi_k(\mathbf{a})$ = number of **b**'s in the collection for which **a** is a prefix of **b**′.

For fixed **a**, if any random **b** has **a** as a prefix, then the remaining entries are still i.i.d. uniform. So, for fixed **a**, if $\phi_k(\mathbf{a}) = m \geq 2$, then the fringe subtree rooted at **a**, considered as a rooted unlabeled tree, has the same distribution as $\mathcal{T}_m$. Thus the family $(\mathcal{T}_k;\ k \geq 1)$ is *leaf-number* coherent.

Now fix $0 < u < 1$ and consider a subsequence $k_j = d^{j+u_j}$, $u_j \to u$. A straightforward argument, sketched below, shows this subsequence has a certain asymptotic fringe distribution $\mathcal{F}_u$.

Take j large and consider a string **a** of length $j - i$. Then $\phi_{k_j}(\mathbf{a})$ has approximately Poisson distribution with mean $k_j d^{-(j-i)} \sim d^{u+i}$. So the number of internal nodes **a** with $\phi_{k_j}(\mathbf{a}) = m \geq 2$ works out to be asymptotic to $d^j c_u(m)$, where

$$c_u(m) = \sum_{i=-\infty}^{\infty} d^{-i} \exp(-d^{u+i}) d^{(u+i)m}/m!, \qquad m \geq 2. \tag{30}$$

The number of leaves is $k_j \sim d^j c_u(1)$ for

$$c_u(1) = d^u.$$

Thus by coherence, the asymptotic fringe distribution for this subsequence is

$$P(\mathcal{F}_u = \mathcal{T}_m) = c_u(m) / \sum_{\nu \geq 1} c_u(\nu); \qquad m \geq 1. \tag{31}$$

Again the asymptotic fringe distribution $(\mathcal{F}_u)$ cycles as $0 \leq u \leq 1$. Note that the asymptotic size of the random trie on k strings is k/(leaf-probability), and so is $k \sum_m c_u(m)$, where u is the fractional part of $\log_d k$; see Jacquet and Regnier [19, 20] for a more extensive analytic treatment.

3.7. *Digital storage trees*. The trie on $(\mathbf{b}_1, \ldots, \mathbf{b}_k)$ is a data structure in which the strings $\mathbf{b}_i$ are stored in the leaves of the trie and where inclusion of new strings may cause existing strings to be moved. There is a related data structure where strings are stored at all vertices of the tree and their location is not changed by inclusion of further strings. In this *digital storage tree* we simply store $\mathbf{b}_j$ at the vertex $\mathbf{b}'_j$ which is the shortest prefix of $\mathbf{b}_j$ which is not the prefix of any of $\mathbf{b}'_1, \ldots, \mathbf{b}'_{j-1}$. As before, suppose we have k strings with i.i.d. uniform entries, and then we get random trees $\mathcal{T}_k^*$. For simplicity consider the binary ($d = 2$) case.

The coherence property is clear: Given that the fringe subtree at $\mathbf{b}'_j$ has size m, there are m strings amongst $\mathbf{b}_{j+1}, \ldots, \mathbf{b}_k$ which have prefix $\mathbf{b}'_j$, and the subsequent entries on those strings are still i.i.d. uniform and hence the fringe subtree is distributed as $\mathcal{T}_m^*$.

As with the trees of Sections 3.2 and 3.3, these embed naturally into continuous-time Markov random trees $(\mathcal{T}_t;\ t \geq 0)$ described as follows. Say $(a_1, \ldots, a_i)$ is available at time t if that vertex is not in the tree $\mathcal{T}_t$ but its parent is in $\mathcal{T}_t$. Then the tree grows according to the rule: Each available vertex $(a_1, \dots, a_i)$ is added at rate 2^{-i}. This nonhomogeneity means we cannot directly apply the standard theory of Section 3.1. Note however that the rate of growth of $|\mathcal{T}_t|$ is always 1.

The asymptotics can be described simply in terms of the distribution function

$$(32) \quad G(t) = P\left(\sum_{i=0}^{\infty} 2^{-i}\xi_i \le t\right), \qquad (\xi_i) \text{ independent, exponential } (1).$$

Indeed, the chance that a vertex $(a_1, \dots, a_h)$ of height h is in $\mathcal{T}_t$ is exactly $P(\sum_{i=1}^{h} 2^i \xi_i \le t)$. And this is asymptotically $G(t2^{-h})$.

Fix $0 < u < 1$ and consider a subsequence $n_j = 2^{j+u_j}$, $u_j \to u$. Then the families $(\mathcal{T}_{n_j}^*)$ and $(\mathcal{T}_{n_j})$ have asymptotic fringe distribution $\mathcal{F}_u$ described as follows.

By coherence, $\mathcal{F}_u$ must be of the form

$$P(\mathcal{F}_u \in \cdot) = \sum_{m \ge 1} c_u(m) P(\mathcal{T}_m^* \in \cdot)$$

for constants $c_u(m)$. By considering the chance that a vertex $\mathbf{b}'$ at height $j - i$ is added at time $2^{j-i}s$, and then the conditional probability of exactly $m - 1$ more strings with prefix $\mathbf{b}'$ being seen before time n_j, a brief calculation gives

$$(33) \qquad c_u(m) = \sum_{i=-\infty}^{\infty} 2^{-i-u} \int_0^{2^{u+i}} \Lambda(2^{u+i} - s, m - 1) G(ds),$$

where $\Lambda(\lambda, \cdot)$ is the Poisson(λ) probability function.

Related questions are studied analytically by Flajolet and Sedgewick [15], who report (Theorem 2) that a cycle-average of the leaf-probabilities $c_u(1)$ equals approximately 0.372 and (pages 763–764) that a cycle-average of $P(\text{root-degree}(\mathcal{F}_u) = 2)$ equals $1/(4 \log 2)$. Probabilistic analysis related to central limit theorems is in Aldous and Shields [4].

4. Sin-trees. In this section we show that the results of Section 2.2 imply apparently stronger results about convergence to infinite discrete trees, and explore structural properties of the limit trees.

4.1. *Definition and convergence.* For a rooted tree $\mathbf{t}$ with an infinite number of vertices, but where each vertex has finite degree, the following are equivalent. (i) There exists a unique infinite path (root $= v_0, v_1, v_2, \dots$) from the root; (ii) for each vertex w, there exists a unique infinite path ($w = w_0, w_1, w_2, \dots$) from w; (iii) there exists no doubly-infinite path $(\dots, w_{-1}, w_0, w_1, \dots)$.

(A *path* has distinct vertices). Call such a tree a *sin*-tree (for single infinite path): an alternative name is tree with one end. For each vertex v in a sin-tree $\mathbf{t}$, let $f(\mathbf{t}, v)$ contain the vertices w such that the unique infinite path from w hits v; consider $f(\mathbf{t}, v)$ as a finite tree rooted at v. Given a sin-tree $\mathbf{t}$ and the path (v_i) as in (i), let $b_0 = f(\mathbf{t}, \text{root})$ and for $i \ge 1$, let the ith branch b_i be the set of vertices x such that the unique infinite path from x hits v_i but does not hit v_{i-1}. Consider b_i as a rooted (at v_i) unlabeled tree. The map

$$\mathbf{t} \leftrightarrow (b_0, b_1, b_2, \dots)$$

defines an isomorphism between the set of sin-trees and the product set $T \times T \times \cdots$. Call this the branch representation of $\mathbf{t}$. With this isomorphism in mind, we may write T^∞ for the set of sin-trees and can define convergence of sin-trees (or convergence in distribution of random sin-trees) via convergence in the product topology on T^∞ (where the countable set T is given the discrete topology).

The matrix Q at (1) extends in a natural way to a kernel Q^∞ on T^∞ as follows. If $Q(t, s) \geq 1$, write $t \setminus s$ for the tree obtained from t by deleting a first-generation subtree isomorphic to s. Then set

$$Q^\infty((t, b_1, b_2, \dots), (s, t \setminus s, b_1, b_2, \dots)) = Q(t, s),$$

where we are using the branch representation of sin-trees. Call a probability distribution μ on T^∞ invariant if it is invariant under this kernel Q^∞. Call a random sin-tree $\mathbf{F}$ invariant if its distribution is invariant.

There is a slightly different representation of sin-trees. Let $(t_0, t_1, t_2, \dots)$ be an infinite sequence in T such that

$$Q(t_i, t_{i-1}) \geq 1, \qquad i \geq 1.$$

Then there is a unique sin-tree $\mathbf{t}$ such that $t_i = f(\mathbf{t}, v_i)$ for each $i \geq 0$, where (root $= v_0, v_1, v_2, \dots$) is the unique infinite path in $\mathbf{t}$ from the root. Call this $(t_0, t_1, \dots)$ the *monotone representation* of $\mathbf{t}$. In terms of monotone representations, the kernel Q^∞ is defined by

$$Q^\infty((t_0, t_1, \dots), (t_{-1}, t_0, t_1, \dots)) = Q(t_0, t_{-1}).$$

PROPOSITION 10. *Let μ be a probability distribution on T^∞, let π be a probability distribution on T and consider the relation*

(34) $\qquad$ *π is the image of μ under the map $\mathbf{t} \to f(\mathbf{t}, root)$.*

(a) *If μ is invariant, then the π defined by* (34) *is a fringe distribution.*

(b) *If π is a fringe distribution, then there exists a unique invariant measure μ satisfying* (34).

As the proof below shows, μ is constructed from π using the Markov chain P_π of (4).

Now let v be a vertex at height h in a finite rooted tree t. Let $b_0(t, v)$ be the fringe subtree at v and for $1 \leq i \leq h$, define branches $b_i(t, v)$ as above, with "path from x to root" in place of "infinite path from x". We may regard $\mathbf{f}(t, v) = (b_i(t, v), 0 \leq i < \infty)$ as an element of T^∞ by setting

$$b_i(t, v) = \varnothing, \qquad i > h,$$

where $\varnothing$ is a symbol appended to T. Note that $b_0(t, v)$ is the fringe subtree $f(t, v)$ of Section 2.2. Call $\mathbf{f}(t, v)$ the *extended fringe* rooted at v.

As in Section 2.2, consider a sequence of random finite trees $\mathscr{T}_k$ and let V_k be a uniform random vertex of $\mathscr{T}_k$. Then we can consider the associated extended fringe $\mathbf{F}_k = \mathbf{f}(\mathscr{T}_k, V_k)$. Note that the first entry $F_{k,0}$ of $\mathbf{F}_k = (F_{k,0}, F_{k,1}, F_{k,2}, \dots)$ is just the random fringe subtree $\mathscr{F}_k$ of Section 2.2.

PROPOSITION 11. *Let $(\mathscr{T}_k)$ be a sequence of random trees with asymptotic fringe distribution π. Then the extended fringes $(\mathbf{F}_k)$ converge in distribution to μ, the unique invariant measure satisfying* (34).

PROOF OF PROPOSITIONS 10 AND 11. Assertion (a) is immediate from the definitions. Let $X_0, X_1, \dots$ be the monotone representation of some random sin-tree, and let $\mu_i = \text{dist}(X_0, \dots, X_i)$. Then the random sin-tree is invariant iff for each i,

$$\mu_i(t_0, \dots, t_i) = \mu_{i-1}(t_1, \dots, t_i) Q(t_1, t_0).$$

By induction, this is equivalent to: For each i,

$$\mu_i(t_0, \dots, t_i) = \pi(t_i) \prod_{j=1}^{i} Q(t_j, t_{j-1}), \tag{35}$$

where $\pi = \mu_0$. This implies uniqueness in assertion (b). Existence is established by taking $(X_i;\ i \geq 0)$ to be the P_π-chain of Section 2.1 started with distribution π, for which

$$\begin{aligned}\mu_i(t_0, \dots, t_i) &= \pi(t_0) \prod_{j=0}^{i-1} P_\pi(t_j, t_{j+1}) \\ &= \pi(t_i) \prod_{j=1}^{i} Q(t_j, t_{j-1}) \quad \text{by (4)}.\end{aligned}$$

To prove convergence, let $(X_0^k, X_1^k, \dots)$ be the monotone representation of the extended fringe of $\mathscr{T}_k$. Then for trees $t_i \in T \setminus \varnothing$,

$$P\big((X_0^k, \dots, X_i^k) = (t_0, \dots, t_i)\big) = P(\mathscr{F}_k = t_i) \prod_{j=1}^{i} Q(t_j, t_{j-1})$$

using the interpretation of the product term as the number of paths (root $= v_0, v_1, \dots, v_i$) in the tree t_i such that for all j the fringe subtree $f(t_i, v_j)$ equals t_j.

Proposition 11 now follows from (35). □

4.2. *Convergence of functionals.* We have taken for granted that the reader understands the connection between convergence of trees and convergence of the kind of explicit statistics (e.g., degrees of vertices) mentioned in the introduction. Here we spell out the connection for the reader unaccustomed to thinking about convergence of random processes.

As in Section 2.2, let $f(\mathscr{T}, V)$ be the random fringe subtree rooted at a uniform random vertex of a random tree $\mathscr{T}$. Let $g\colon T \to R$ be an arbitrary function. Then we can define

$$\Gamma_g(\mathscr{T}) \equiv g(f(\mathscr{T}, V)). \tag{36}$$

Think of Γ_g as a functional on random trees (pedantically, it is a functional of the trees and the external randomization defining V). If $(\mathscr{T}_k)$ has asymptotic fringe $\mathscr{F}$, then

(37) $$\Gamma_g(\mathscr{T}_k) \to_d g(\mathscr{F})$$

for all functions $g\colon T \to R$. And convergence of expectations

$$E\Gamma_g(\mathscr{T}_k) \to Eg(\mathscr{F})$$

will follow upon verification of the usual uniform integrability condition.

For example, the functionals

$\Gamma(\mathscr{T})$ = proportion of vertices of $\mathscr{T}$ which are leaves,

$\Gamma(\mathscr{T})$ = out-degree of uniform random vertex,

$\Gamma(\mathscr{T})$ = size of subtree of uniform random vertex,

$\Gamma(\mathscr{T})$ = height of subtree of uniform random vertex,

arise from the functions

$$g(t) = 1_{(t=\mathrm{triv})},$$
$$g(t) = \text{root-degree}(t),$$
$$g(t) = |t|,$$
$$g(t) = \mathrm{ht}(t).$$

Aside from the results about out-degrees mentioned in Section 3, few special cases appear in the literature. References to calculations with the general continuous-time branching process are in Jagers and Nerman [22]. Moon and Meir ([28], Corollary 2.2) give the result for heights for conditioned Galton–Watson trees.

Convergence of extended fringes enables us to extend (37) to other functionals. Let $g\colon T^\infty \to R$ be a function on sin-trees. Let $\mathbf{f}(\mathscr{T}, V)$ be the extended fringe of $\mathscr{T}$, as in the previous section. Let

(38) $$\Gamma_g(\mathscr{T}) \equiv g(\mathbf{f}(\mathscr{T}, V)).$$

Convergence of fringe distributions implies convergence of extended fringe distributions to an invariant sin-tree $\mathbf{F}$ (Proposition 11) and hence

(39) $$\Gamma_g(\mathscr{T}_k) \to_d g(\mathbf{F})$$

for each continuous $g\colon T^\infty \to R$. Such functionals Γ_g are the local functionals to which our theory applies. Note that continuity is a strong requirement for g, since T^∞ has the product topology. Roughly, g must be almost unaffected by any modification of a tree which affects only relations of the root who have no recent common ancestor with the root.

Simple examples are

$\Gamma(\mathscr{T})$ = distance from uniform random vertex to nearest leaf,

$\Gamma(\mathscr{T})$ = number of vertices at distance j from uniform random vertex, which arise from

$$g(\mathbf{t}) = \text{distance from root to nearest leaf},$$

$$g_j(\mathbf{t}) = \text{number of vertices at distance } j \text{ from root}.$$

Grimmett [18] studies the former functional of the uniform random labeled tree model.

Returning to convergence of fringe subtrees, here is a more elaborate example. Say s is a root-subtree of t if s can be obtained from t by repeating the operation "delete some leaf of current tree". Fix a tree t. Writing v for a generic vertex of t, define

$$N(t,s) = \sum_v 1_{(s \text{ is a root subtree of } f(t,v))}.$$

Then $N(t, s)$ counts the number of times s appears as a pattern in t, in the terminology of Steyaert and Flajolet [35] who study its properties in the context of the conditioned critical Galton–Watson model. Obviously, knowing that a family $(\mathscr{T}_k)$ has asymptotic fringe distribution π tells us the asymptotic proportion of such appearances:

$$E\frac{N(\mathscr{T}_k, s)}{|\mathscr{T}_k|} \to \rho(s) \equiv \sum_t \pi(t) 1_{(s \text{ is a root-subtree of } t)}.$$

Finally, convergence of random fringe subtrees $f(\mathscr{T}_k, V_k)$ to $\mathscr{F}$ implies convergence of $h \circ f(\mathscr{T}_k, V_k)$ to $h(\mathscr{F})$ for any function $h\colon T \to T$. For instance, consider the mapping $h\colon T \to T$ which eliminates vertices of out-degree 1 by taking each path segment $v_0, v_1, \ldots, v_j$ for which the interior vertices (but not v_0 and v_j) have out-degree 1 and collapsing the segment into a single edge (v_0, v_j). This is a standard compaction idea. If $\mathscr{T}_n$ is the digital storage tree, then the PATRICIA tree of [25] (Section 6.3) is defined to be $h(\mathscr{T}_n)$. Thus the form of the asymptotic fringe for random PATRICIA trees can be deduced from the form for digital storage trees.

4.3. *Reduced Markov descriptions*. There is a natural construction of random *marked* trees, which we now describe (omitting measure-theoretic details). Take a mark-space S. For each $x \in S$, let there be given a distribution $(C_x; \gamma_i(x), 1 \le i \le C_x)$, where $C_x \ge 0$ is integer-valued and each $\gamma_i(x)$ is S-valued. Write

$$R(x, \cdot) = \sum_i P(\gamma_i(x) \in \cdot),$$

so that R is a nonnegative kernel on S. Suppose ν is a probability measure on S which is invariant under R. For each x we can construct a random marked tree $\mathscr{F}_x$ as follows. Mark the root x. Let the root have C_x offspring, marked $\gamma_1, \ldots, \gamma_{C_x}$. Conditional on these values, the children reproduce independently, a child marked z having C_z offspring marked $\gamma_i(z)$, $1 \le i \le C_z$ and so on. Suppose these trees $\mathscr{F}_x$ are a.s. finite. Then let $\mathscr{F} = \mathscr{F}_X$, where X has distribution ν. Considered as an unmarked tree, $\mathscr{F}$ has distribution π, say. It is easy to see (from the assumption that ν is invariant under R) that π is a fringe distribution.

Conversely, given a fringe distribution π, a *Markov description* of π is a collection of a mark space S, offspring mark distributions $(C_x; \gamma_i(x), 1 \le i \le C_x)$ and an R-invariant measure ν such that the tree $\mathscr{F}$ constructed above has distribution π.

Such a description always exists, for a trivial reason: We can take $S = T$ and mark each vertex v of the random tree by the subtree rooted at v. This is the unreduced Markov description, where R is just the matrix Q of (1). What is interesting is the case where we may take S to be a simpler set than T, in which case we say we have a *reduced* Markov description. Examples are given in the next two sections.

The corresponding invariant sin-tree, considered as a marked tree, can be constructed as follows. Let the vertices on the infinite path from the root be marked $(X_0, X_1, X_2, \dots)$, where X_0 has distribution ν and $(X_i, i \ge 0)$ is the S-valued Markov chain with transition kernel

$$P^*(x, dy) = \nu(dy) R(y, dx)/\nu(dx). \tag{40}$$

Conditional on this sequence, let the branches $(B_i, i \ge 0)$ be independent, with B_0 distributed as $\mathscr{F}_{X_0}$ and B_i distributed as $\mathscr{F}_{X_i, X_{i-1}}$, $i \ge 1$, for $\mathscr{F}_{x,y}$ defined as follows.

$\mathscr{F}_{x,y}$ is constructed as $\mathscr{F}_x$, except that the distribution of the number and marks of first-generation offspring is changed to the following distribution:

$$P(\hat{C}_x = c; \hat{\gamma}_1(x) \in dg_1, \dots, \hat{\gamma}_c(x) \in dg_c)$$
$$= \frac{\sum_i P(C_x = c + 1; \gamma_1(x) \in dg_1, \dots, \gamma_i(x) \in dy, \dots, \gamma_{c+1}(x) \in dg_c)}{\Theta(c + 1, x, dy)},$$

where

$$\Theta(c + 1, x, dy) = \sum_i P(C_x = c + 1, \gamma_i(x) \in dy).$$

The argument that this construction does give the invariant sin-tree is very similar to the argument in the special case [22] for the time-homogeneous general branching process—we omit the details.

Call the chain $(X_i; i \ge 0)$ the *ancestor chain*.

Note that, in the unreduced description, (40) is just the definition (4) of P_π:

$$P_\pi(s, t) = \pi(t) Q(t, s)/\pi(s).$$

In other words, P_π is the transition matrix for the ancestor chain in the unreduced description of the invariant sin-tree. That was the essence of the proof of Proposition 10.

4.4. *The one-dimensional Markov case.* Suppose we may take the mark-space S to be $[0, \infty)$ or a subset thereof and suppose the kernel R has the monotonicity property

$$R(x, (x, \infty)) = 0 \quad \text{for each } x. \tag{41}$$

In this case we have a one-dimensional Markov description of the fringe distribution and the associated invariant sin-tree. Here are the examples.

The fundamental example is the general continuous-time branching process of Section 3.1. Here $(\gamma_i(x), i \geq 1)$ are the nonnegative values of $(x - \beta_i, i \geq 1)$. That is, we also have time-homogeneity for the offspring distributions. In this example, the ancestor chain (X_i) is a renewal process on $[0, \infty)$. Proposition 11 and (16) imply convergence of the extended fringe to the invariant sin-tree and this latter type of result is the main issue of [32].

(*Remark*: In the context of population processes, one can also consider the process of descendants into the indefinite future, which does not make sense for the general random trees considered in this paper.)

This example suggests the intuitive interpretation of the marks as time of birth, with time measured backward from the present time 0. This interpretation is the motivation for assumption (41), which says that children are not born before their parents.

In the special cases of the Yule tree (Section 3.2) and the binary search tree (Section 3.3), the ancestor chain (X_i) is just a Poisson(1) counting process.

For the conditioned critical Galton–Watson process with ξ offspring (Section 3.4), the asymptotic fringe distribution is the unconditioned Galton–Watson family tree $\mathscr{T}$. This is one-dimensional Markov in a degenerate sense: We can take a single mark 0 and set the offspring number C equal to ξ. Here the branches $(B_i, i \geq 0)$ are independent, $B_0 =_d \mathscr{T}$, and for $i \geq 1$, each branch B_i has the distribution of the family tree of the modified Galton–Watson process described as follows. The second and subsequent generation offspring distribution is ξ; the first-generation offspring distribution is $\hat{\xi}$, where

$$P(\hat{\xi} = c) = (c + 1)P(\xi = c + 1)/E\xi, \qquad c \geq 0.$$

Almost the same sin-tree occurs in a slightly different context; see Section 7.5(d).

For random tries (Section 3.6) think of the mark of a vertex v as an integer m indicating the number of strings with prefix v. Here $m \geq 1$ is an integer. The distribution $(C_m; \gamma_i(m), 1 \leq i \leq C_m)$ is as follows. $C_1 = 0$. For $m \geq 2$, throw m balls at random into d boxes. Let C_m be the number of occupied boxes and $\gamma_1(m), \ldots$ be the numbers of balls in the occupied boxes. Note that this description of the "Markovian evolution" does not involve the parameter u appearing in the asymptotic fringe distribution for subsequences. The parameter u does appear in the description (31) of the invariant distribution ν. In other words, what happens here is that the kernel R at (41) has a (cyclic) one-parameter family of invariant distributions.

(Actually, for random tries and digital storage trees, it turns out to be more natural to use two-dimensional Markov descriptions, specified in the next section.)

4.5. *Extremality*. Aside from permitting concise descriptions of invariant sin-trees, reduced Markov descriptions are technically useful for verifying the extremality condition used in Proposition 7 on convergence of empirical proportions. Note that extremality of a fringe distribution is equivalent to extremality of the associated invariant sin-tree, by Proposition 10.

We first quote a Markov chain result, a variation of the standard result about coupling. As noted by Hermann Thorisson (personal communication), it can be deduced from more elaborate results in Greven [17].

PROPOSITION 12. *Let P^* be a Markov transition matrix on a mark space S. Call a Markov chain with transition matrix P^* and initial state s a (P^*, s)-chain. The following are equivalent*:

(a) *There are no nonconstant bounded P^*-harmonic functions.*

(b) *For each pair $s_1, s_2 \in S$, we can construct a (P^*, s_1)-chain $(X_n^1\colon n \geq 0)$, a (P^*, s_2)-chain $(X_n^2\colon n \geq 0)$, a random time $\tau < \infty$ a.s. and a random shift $-\infty < \sigma < \infty$ such that for each n,*

$$X_n^1 = X_{n+\sigma}^2 \quad a.s.\ on\ \{n \geq \tau\}.$$

We call the construction in Proposition 12(b) a shift-coupling. The next result, proved in Section 5, is the only mathematically serious result of the paper.

THEOREM 13. *Let $(X_i;\ i \geq 0)$ be the ancestor chain in some reduced Markov description of an invariant sin-tree μ. Suppose the conditions of Proposition 12 are satisfied by (X_i). Then μ is extreme.*

We now use Theorem 13 to prove extremality of the limits in the random tree models of Section 3. For a conditioned critical Galton–Watson process (Section 3.4), we noted in the previous section that the ancestor process has a single state, so all is obvious. For the Yule tree (Section 3.2) and the binary search tree (Section 3.3), we noted that the ancestor process was the Poisson(1) counting process, which certainly admits shift-coupling. Indeed, for the general continuous-time branching process, the one-dimensional ancestor process is a renewal process and (provided it is nonlattice), the classical Choquet–Deny theorem says that condition (a) holds, so we have extremality.

Now consider tries (Section 3.6) and the subsequential asymptotic fringe distribution $\mathscr{F}_u$ defined there. As stated in the previous section, there is a one-dimensional Markov description in which a vertex v of the fringe distribution is marked by the number x of leaves in its subtree. But it is more natural to use a two-dimensional reduced Markov description with mark space $S = \{(i, m)\colon -\infty < i < \infty,\ m \geq 1\}$. Here a mark (i, m) indicates that the vertex (in the trie with $\approx d^{j+u}$ leaves) is at height $j - i$ and its fringe subtree has m leaves. The offspring of (i, m) have labels $(i - 1, \gamma_1(m)), (i - 1, \gamma_2(m), \dots)$, for $\gamma(m)$ as in the above section. This description gives a simpler form to the ancestor process: It has transition matrix specified via

$$(i, m) \to (i + 1, m + \beta_{i+1}), \tag{42}$$

where β_i has Poisson distribution with mean $d^{u+i}(1 - 1/d)$.

It is easy to see that, in any Markov chain of form (42), a sufficient condition for shift-coupling is

$$\lim_{i \to \infty} \sum_b |P(\beta_i = b) - P(\beta_i = b+1)| = 0. \tag{43}$$

This certainly holds in the present case, so $\mathscr{F}_u$ is an extreme distribution.

For the sin-tree associated with the fringe $\mathscr{F}_u$ of the digital storage tree of Section 3.7, there is a natural two-dimensional mark space $S = \{(i,t): -\infty < i < \infty,\ 0 \le t \le 2^u\}$. Here a mark (i,t) indicates a vertex in the tree at time $\approx 2^{j+u}$ which is at height $j - i$ and which was inserted at time $t2^j$. In the reduced Markov description, the individual (i,t) has at most two offspring labeled $(i-1, t + 2^{1-i}\xi_1)$ and $(i-1, t+2^{1-i}\xi_2)$, where ξ_1, ξ_2 are independent exponential (1) and the offspring exist if $t + 2^{1-i}\xi \le 2^u$. The ancestor chain has transitions of the form

$$(i,t) \to (i+1, s), \qquad 0 < s < t,$$

occurring at rate

$$2^i \frac{f_{i+1}(s)}{f_i(t)} \exp(-(t-s)2^i),$$

where f_i is the density function of the r.v. $\sum_{m=-\infty}^{-i} 2^m \xi_m$ for independent (ξ_m) with exponential (1) distribution. Now write

$$(0,x), (1, X_1), (2, X_2), \ldots$$

for the ancestor chain started at $(0,x)$. It is enough to show that chains $(X_n;\ n \ge 0)$ with different starting points x can be coupled. But the process $Y_n \equiv 2^n X_n$ is a homogeneous Markov chain with transition kernel

$$\hat{P}(y, dz) = \frac{f_0(z)}{f_0(y)} \exp\left(-\left(y - \frac{z}{2}\right)\right), \qquad 0 \le z \le 2y.$$

It is enough to show that $\hat{P}$ defines a Harris recurrent chain, for then (Asmussen [7], Proposition 6.3.13) (Y_n), and hence (X_n), admits coupling. Now $\hat{P}$ has stationary density f_0, and proof of Harris recurrence is just a simple variation on [7] (Example 6.3.1).

5. Proof of Theorem 13. Fix an invariant sin-tree distribution $\mu = \text{dist}(\mathbf{F})$ and its associated fringe distribution π. Let $(B_0, B_1, B_2, \ldots)$ be the branch representation of $\mathbf{F}$ and let $(X_0, X_1, X_2, \ldots)$ be the monotone representation. Recall that $(X_n;\ n \ge 0)$ evolves as the P_π-chain, for P_π defined at (4). Say $h: T \to R$ is bounded P_π-harmonic if h is bounded and

$$h(s) = \sum_t P_\pi(s,t)h(t) \quad \text{for all } s.$$

Note that $(h(X_n);\ n \ge 0)$ is a bounded martingale, when h is bounded harmonic and (X_n) is a P_π-chain.

LEMMA 14. *The following are equivalent*:

(a) π *is extreme.*
(b) μ *is extreme.*
(c) *Each bounded* P_π*-harmonic function is constant on* $\{t: \pi(t) > 0\}$.

PROOF. The equivalence of (a) and (c) is routine: Given harmonic h, consider the distribution π_h defined by $d\pi_h/d\pi = ch$, $c^{-1} = \int h\,d\pi$. As previously noted, the equivalence of (a) and (b) follows from Proposition 10. □

Now fix $b^1, b^2 \in T$ and for $j = 1, 2$, let $\mathbf{F}^j$ have the conditional distribution of $\mathbf{F}$ given $B_0 = X_0 = b^j$. Write $(b^j, B_1^j, B_2^j, \dots)$ and $(b^j, X_1^j, X_2^j, \dots)$ for the branch and monotone representations of $\mathbf{F}^j$.

LEMMA 15. *Under the hypothesis of Theorem* 13, *we can construct a joint distribution* $(\mathbf{F}^1, \mathbf{F}^2)$, *a random time* τ *and a random shift* σ *such that*

$$B_n^1 = B_{n+\sigma}^2 \quad on\ \{n \geq \tau\}.$$

PROOF. We may suppose the trees possess the marks given by the reduced Markov construction. Fix arbitrary marks s^1, s^2 and for $j = 1, 2$, let $(Z_n^j; n \geq 0)$ be the ancestor process with $Z_0^j = s^j$. By Proposition 12, we can construct a shift-coupling of these ancestor processes using times τ', σ. Then by considering the construction [below (40)] of the invariant sin-tree from the ancestor process, we see that if $(Z_{n-1}^1, Z_n^1) = (Z_{n+\sigma-1}^2, Z_{n+\sigma}^2)$, then we may take $B_n^1 = B_{n+\sigma}^2$. Thus the lemma holds for $\tau = \tau' + 1$. □

From now on, let $(\mathbf{F}^1, \mathbf{F}^2)$ have the joint distribution given by Lemma 15. Consider the monotone representations $(X_n^j; n \geq 0)$ of the conditioned sin-trees $\mathbf{F}^j$; these are P_π-chains with initial states b^j. Informally, X_n^j has the same information as does $(b^j, B_1^j, \dots, B_n^j)$ except that the location of the original root is forgotten. If the shift-coupling of Lemma 15 could be extended to a shift-coupling of the processes $(X_n^j; n \geq 0)$, then we would be finished (by Proposition 12 applied to P_π). But we cannot carry through such an argument, so instead use a more devious method. Fix a harmonic function h with $0 \leq h \leq 1$. By the martingale convergence theorem, the limits

$$L^j = \lim_n h(X_n^j) \tag{44}$$

exist a.s. and $h(b^j) = EL^j$. We shall show

$$L^1 = L^2 \quad \text{a.s.} \tag{45}$$

and then $h(b^1) = h(b^2)$, so by Lemma 14, the theorem is proved.

For $\mathbf{t} \in T^\infty$ and v a vertex of $\mathbf{t}$, write $\mathbf{t}[v]$ for the tree rerooted at v. Let Λ be an operator which, to each $\mathbf{t} \in T^\infty$, associates a probability measure $\Lambda(\mathbf{t}, \cdot)$ on the vertices of $\mathbf{t}$. Let $\xi_{\mathbf{t}}$ denote a random vertex of $\mathbf{t}$ with distribution $\Lambda(\mathbf{t}, \cdot)$. Say Λ is $\mathbf{F}$*-symmetric* if the joint distribution $(\mathbf{F}, \mathbf{F}[\xi_{\mathbf{F}}])$ on $T^\infty \times T^\infty$ is symmetric. This implies, in particular, that the randomly rerooted tree $\mathbf{F}[\xi_{\mathbf{F}}]$ has the same distribution as $\mathbf{F}$.

Here is a technical lemma whose proof is deferred.

LEMMA 16. *There exist operators Λ_N, $N \geq 1$ with the following properties.*

(a) $\rho(\mathbf{t}, \xi_{\mathbf{t}}^N) \to_p \infty$ *as* $N \to \infty$ *for each* $\mathbf{t}$ (ρ *is defined below*).

(b) *For any two sin-trees* $\mathbf{t}_1$, $\mathbf{t}_2$ *which differ by only finitely many vertices, we can construct a joint distribution for* $\xi_{\mathbf{t}_1}^N$ *and* $\xi_{\mathbf{t}_2}^N$ *such that* $P(\xi_{\mathbf{t}_1}^N \neq \xi_{\mathbf{t}_2}^N) \to 0$ *as* $N \to \infty$.

(c) *Each* Λ_N *is symmetric for each invariant sin-tree.*

(d) *For each invariant sin-tree* $\mathbf{F}$,

$$P\big(\xi_{\mathbf{F}}^N \text{ is on the infinite path from the root}\big) \to 0 \quad \text{as } N \to \infty.$$

We need yet more definitions. Fix $\mathbf{t} \in T^\infty$. Write $S(\mathbf{t})$ for the set of vertices of $\mathbf{t}$ which are neither in the infinite path from the root, nor in the fringe subtree at the root. For $v \in S(\mathbf{t})$, the infinite path $(v = v_0, v_1, v_2, \dots)$ meets the infinite path $(\text{root} = w_0, w_1, w_2, \dots)$ at some vertex $v_a = w_b$, say, where $a, b \geq 1$. Define

$$\rho(\mathbf{t}, v) = b,$$
$$v_+ = v_a,$$
$$v_- = v_{a-1}.$$

For v outside $S(\mathbf{t})$, set $\rho(\mathbf{t}, v) = 0$. Using the fixed bounded harmonic function h, define

$$\Delta(\mathbf{t}, v) = \begin{cases} |h(f(\mathbf{t}, v_+)) - h(f(\mathbf{t}, v_-))|, & v \in S(\mathbf{t}), \\ 1 & \text{else}, \end{cases}$$

where $f(\mathbf{t}, v)$ is the fringe subtree rooted at v.

LEMMA 17. *For operators Λ_N as in Lemma 16 and for any invariant sin-tree* $\mathbf{F}$,

$$E\Delta\big(\mathbf{F}, \xi_{\mathbf{F}}^N\big) \to 0 \quad \text{as } N \to \infty.$$

PROOF. For any $\mathbf{F}$-symmetric operator and any $L \geq 1$,

$$E\Delta(\mathbf{F}, \xi_{\mathbf{F}}) \leq s(L) + P\big(\rho(\mathbf{F}, \xi_{\mathbf{F}}) = 0 \quad \text{or} \quad |(\xi_{\mathbf{F}})_+| < L\big), \tag{46}$$

where

$$s(L) = E \sup_n |h(X_n) - h(X_{n-1})| 1_{(|X_n| \geq L)}.$$

Consider the sequence Λ_N of operators given by Lemma 16. By properties (a) and (d), as $N \to \infty$ (L fixed), the final $P(\cdot)$ term of (46) tends to 0. Thus it is enough to show that $s(L) \to 0$ as $L \to \infty$. But this is an easy consequence of the upcrossing inequality for the martingale $(h(X_n), n \geq 0)$.

Now consider a sin-tree $\mathbf{t}$, and let $\mathbf{t}^1$, $\mathbf{t}^2$ be sin-trees obtained by deleting a finite number of vertices from $\mathbf{t}$ and rerooting. Then $(\mathbf{t}^1, \mathbf{t}^2)$ admits coupling of side-branches, in the sense of Lemma 15. Conversely, any pair $(\mathbf{t}^1, \mathbf{t}^2)$ which admits coupling of side-branches is of this form, for some $\mathbf{t}$. Let Λ_N be operators as in Lemma 16 and let $\xi_j^N \equiv \xi_{\mathbf{t}^j}^N$ be the corresponding random vertices of $\mathbf{t}^j$. By Lemma 16(a, b), we may suppose

$$P(\xi_1^N \neq \xi_2^N) \to 0 \quad \text{as } N \to \infty,$$
$$\rho(\mathbf{t}^1, \xi_1^N) \to_p \infty \quad \text{as } N \to \infty.$$

Therefore, also using Lemma 16(d),

$$P\big(\xi_j^N \in S(\mathbf{t}^j),\ j = 1,2 \text{ and } f\big(\mathbf{t}^1, (\xi_1^N)_-\big) = f\big(\mathbf{t}^2, (\xi_2^N)_-\big)\big) \to_p 1 \quad \text{as } N \to \infty,$$

where $(\xi_j^N)_-$ is defined as $(v)_-$ above. But for each N, on the event above, we have

$$|h\big(f\big(\mathbf{t}^1, (\xi_1^N)_+\big)\big) - h\big(f\big(\mathbf{t}^2, (\xi_2^N)_+\big)\big)| \le \sum_{j=1}^{2} \Delta\big(\mathbf{t}^j, \xi_1^N\big)$$

for our fixed harmonic function $0 \le h \le 1$. Suppose now that the limits

$$l^j = \lim_n h\big(x_n^j\big), \qquad j = 1, 2,$$

exist, where $(x_n^j;\ n \ge 0)$ is the monotone representation of $\mathbf{t}^j$. Then the foregoing implies

$$|l^1 - l^2| \le \liminf_N E \sum_{j=1}^{2} \Delta\big(\mathbf{t}^j, \xi_j^N\big).$$

Applying this to the shift-coupled realizations of $(\mathbf{F}^1, \mathbf{F}^2)$ given in Lemma 15,

$$E|L^1 - L^2| \le \liminf_N E \sum_{j=1}^{2} \Delta\big(\mathbf{F}^j, \xi_{\mathbf{F}^j}^N\big).$$

But the limit is 0 by Lemma 17, so $L^1 = L^2$ a.s., giving (45) and establishing the Theorem. □

PROOF OF LEMMA 16. Fix a sin-tree $\mathbf{t}$. Let $(W_u;\ u \ge 0)$ be continuous-time random walk on the vertices of $\mathbf{t}$: more exactly, the continuous-time Markov chain with transition rates

$$\text{rate}(v \to w) = 1, \quad \text{each edge } (v, w).$$

This chain is null-recurrent, by considering the induced walk on the infinite path from the root. And the uniform distribution is invariant for this chain. Write $\Lambda_N(\mathbf{t}, \cdot\,)$ for the distribution of W_N, given $W_0 = \text{root}$. Null-recurrence implies that, for each fixed vertex v, $P(W_N = v) \to 0$ as $N \to \infty$, and this establishes part (a) of the lemma.

Here is a sketch of the proof of (b). The general case reduces to the case where $\mathbf{t}_1$, $\mathbf{t}_2$ differ by only one leaf; this further reduces to the case of walks on

$(0, 1, 2, 3, \dots)$ and on $(1, 2, 3, \dots)$; here the simplest independent coupling idea works.

For (d), the event under consideration can be rewritten as

$$\mathrm{root}(\mathbf{F}) \quad \text{is in the fringe of } \mathbf{F}[\xi_{\mathbf{F}}^N],$$

where fringe means fringe at the root. By the symmetry property, this has the same probability as the event

$$\mathrm{root}(\mathbf{F}[\xi_{\mathbf{F}}^N])\ (\text{which} = \xi_{\mathbf{F}}^N) \quad \text{is in the fringe of } \mathbf{F}.$$

But this probability tends to 0 by (a).

It remains to prove the symmetry property. Note that we can define the random walk (W_u) on any finite tree t. Then the stationary distribution is uniform: give W_0 this uniform distribution. Then $(W_0, W_N) =_d (W_N, W_0)$, by time-reversibility, and applying this construction to a random finite tree $\mathscr{T}$ gives

$$(\mathscr{T}, W_0, W_N) =_d (\mathscr{T}, W_N, W_0). \tag{47}$$

Now let $\mathbf{F}$ be an invariant sin-tree. Appealing to Propositions 2 and 11, there exist random trees $\mathscr{T}^k$ such that

$$\mathbf{f}(\mathscr{T}^k, U^k) \to_d \mathbf{F}$$

for uniform random vertices U^k. Applying this with W_0^k and W_N^k (the positions at times 0 and N of stationary random walk on $\mathscr{T}^k$) in place of U^k, it is easy to see the existence of a limit joint distribution

$$\left(\mathbf{f}(\mathscr{T}^k, W_0^k), \mathbf{f}(\mathscr{T}^k, W_N^k)\right) \to_d (\mathbf{F}_1, \mathbf{F}_2), \quad \text{say}.$$

This limit joint distribution is that of $(\mathbf{F}, \mathbf{F}[\xi_F^N])$, and the joint distribution is symmetric by (47). □

6. More examples. In the examples of Section 3, the nature of the asymptotic fringe distributions was rather obvious. Here are some examples where the behavior is less clear.

6.1. *Uniform random spanning trees of graphs.* Let G be a finite connected graph on n vertices. A *spanning tree* of G is a subgraph—with the same vertices but in general fewer edges—which is a tree. Let $(X_j;\ j \geq 0)$ be random walk on G, that is the discrete-time Markov chain with transition matrix P of the form

$$P(v, w) = \begin{cases} 1/r_v & \text{if } (v, w) \text{ is an edge}, \\ 0 & \text{if not}, \end{cases}$$

where r_v is the degree of v. Let X_0 be arbitrary. For each vertex v let T_v be the first hitting time:

$$T_v = \min\{j \geq 0 \colon X_j = v\}.$$

Define a subgraph $\mathscr{T}(G)$ of G to consist of the $n - 1$ edges

$$(X_{T_v - 1}, X_{T_v}), \qquad v \neq X_0.$$

It is clear that $\mathscr{T}(G)$ is a random spanning tree. It is not obvious, but is nevertheless true (see Aldous [2], Broder [8]) that $\mathscr{T}(G)$ is a *uniform* random spanning tree of G.

So for a sequence of graphs G_n we may consider asymptotic fringe distributions for $\mathscr{T}(G_n)$. A simple example is where G_n is the complete graph on n vertices. Then $\mathscr{T}(G_n)$ is the uniform random tree on n labeled vertices, which, as remarked earlier, is the same as the conditioned critical Galton–Watson branching process whose offspring distribution is Poisson(1). Thus the asymptotic fringe is $\mathscr{T}_1$, the corresponding unconditioned branching process.

It is proved in [2] (Proposition 10) that, for the cube graphs $G_d = \{0, 1\}^d$ and related highly-symmetric graphs, the asymptotic distribution of out-degree in $\mathscr{T}(G_d)$ is Poisson(1). Modifications of the arguments there yield the stronger conclusion that the asymptotic fringe in this example is the same $\mathscr{T}_1$ as occurs with the complete graphs.

Finally, fix $d \geq 2$ and consider the discrete torus $G_k = \{0, 1, \dots, k - 1\}^d$, $k \geq 1$. As $k \to \infty$, it is intuitively apparent that the asymptotic fringe distribution $\mathscr{F}_d$ exists and may be regarded as a random subtree of Z^d. But it is not obvious whether $\mathscr{F}_d$ contains all the vertices of Z^d. Pemantle [33] has shown this happens iff $d \leq 4$, using ingenious arguments with loop-erased random walk. Further study of these limit random trees, which (like those in the next section) presumably have interesting fractal properties in low dimension, remains a hard open project.

6.2. *Euclidean minimal spanning trees.* Let $\xi_1, \dots, \xi_n$ be independent random points in R^d with some common density function. Let $\mathscr{T}_n$ be the Euclidean minimum spanning tree on these n random points. As $n \to \infty$, the distribution of points around a typical point ξ, suitably rescaled, converges to a point process $\mathscr{N}^0$ which is a Poisson (rate 1) point process with an extra point at 0. It is therefore plausible that the extended fringe associated with $(\mathscr{T}_n)$ should converge to an infinite sin-tree $\mathscr{T}$ which can be interpreted as the Euclidean minimum spanning tree on the infinite vertex-set $\mathscr{N}^0$. This is discussed in Aldous and Steele [5], where a type of convergence is proved to a limiting *forest* on vertex-set $\mathscr{N}^0$. Proving that this limit forest is a.s. a single sin-tree seems a difficult problem related to continuum percolation.

6.3. *Greedy undirected tree.* The random recursive tree of Section 3.2 could be regarded as a spanning tree of the complete graph, where we put i.i.d. random weights on the edges and grow the tree by attaching vertices in prespecified order to the existing tree using the minimum-weight edge. In this setting, if we do not specify the order of attaching vertices, but instead use the usual greedy algorithm for constructing minimum-weight spanning trees, then we get the following tree. Fix n. Define forests $(\mathscr{R}_i;\ 0 \leq i \leq n - 1)$ on vertices $\{1, 2, \dots, n\}$ as follows. $\mathscr{R}_0$ has no edges. $\mathscr{R}_i$ has the edges of $\mathscr{R}_{i-1}$ and an extra edge (v, w) chosen uniformly from the set of edges with end-vertices in different components of $\mathscr{R}_{i-1}$. Then $\mathscr{R}_{n-1}$ is a random tree: call it $\mathscr{T}_n$.

It is shown in Aldous [1] that $(\mathscr{T}_n;\ n \geq 1)$ has asymptotic fringe distribution π described there as follows.

Write $\mathscr{G}_s$ for the family tree of a Galton–Watson branching process with one progenitor and Poisson(s) offspring distribution, considered as a rooted unlabeled tree. Construct a random tree process $(\mathscr{H}_s\colon 0 \leq s < \infty)$ as follows. $\mathscr{H}_0$ is a root only. During time $[s, s + ds]$, each vertex v of $\mathscr{H}_s$ has chance ds to have attached to it an independent copy of $\mathscr{G}_s$, if the realization of $\mathscr{G}_s$ is finite: If infinite, nothing is attached, but at the first time τ^* when some infinite tree is attempted, the vertex v at which this occurs is distinguished. This construction yields a finite tree $\mathscr{H}_\infty$ with one distinguished vertex. Delete the branch containing the distinguished vertex (if it is not the root): Call the remaining tree $\mathscr{H}_\infty^0$. Then π is the distribution of $\mathscr{H}_\infty^0$.

Without giving details, let us describe the corresponding invariant sin-tree. Use the construction above to define a marked tree, where the marks are defined as follows. Let $s^*(v)$ be the time at which v is added to the tree-process. Then give v the mark $s(v) = \max(s^*(v), \tau^*) > 1$. In the corresponding invariant sin-tree, these marks *decrease* along the infinite path of the sin-tree. To be consistent with our setup in Section 4.4, we want to reverse time, and it is convenient to replace marks $s \in (1, \infty)$ by marks $u(s) \in (0, 1)$ using the decreasing function

$$u(s) = sF(s), \qquad F(s) = P(|\mathscr{G}_s| < \infty).$$

It turns out that the invariant sin-tree has a one-dimensional Markov description. An individual marked u has a Poisson($\lambda(u)$) number of offspring each marked u and also a Poisson process with rate $\rho(u')$ of offspring with marks $u' < u$, where

$$\lambda(u) = \frac{u \log u}{u - 1},$$

$$\rho(u) = \frac{1 - u + \log u}{(1 - u)^2} = \frac{d\lambda(u)}{du}.$$

The invariant measure ν of Section 4.3 works out to be

$$\nu[0, u] = \lambda^{-1}(u), \qquad 0 \leq u \leq 1,$$

where λ^{-1} denotes the inverse function. The transition kernel P^* for the ancestor chain is

$$P^*(u, \{u\}) = \lambda(u),$$

$$P^*(u, dx) = \rho(u)\frac{\nu(dx)}{\nu(du)}\,du, \qquad u < x < 1.$$

As $u \uparrow 1$, $\lambda(u) \sim 1 - (1 - u)/2$ and $\rho(u)\nu(dx)/\nu(du) \sim \frac{1}{2}$ on $\{u < x < 1\}$. It is now not difficult to use Theorem 13 to prove extremality.

Remarks. Informally, the invariant sin-tree for this model is asymptotically like the invariant sin-tree for the conditioned Galton–Watson process of

Section 3.4, with Poisson(1) offspring distribution. This observation strongly suggests that global statistics (diameter, mean distances between random vertices, etc.) in this model are asymptotically exactly the same as for the Galton–Watson model, which in this case is the uniform random labeled tree. Proving such assertions remains an open question.

It is not hard to show by direct calculation that this model is *not* coherent.

A well-studied problem (e.g., Frieze [16]) concerns the weight W_n of the minimum-weight spanning tree of the complete graph on n vertices, where the edges are given i.i.d. nonnegative weights. This minimum-weight spanning tree is in fact the same random tree as $\mathscr{T}_n$ above. By considering the edges of $\mathscr{T}_n$ as being marked by weights, the ideas of this paper are used in [1] to recover and slightly improve known results on the asymptotics of W_n.

6.4. 2–3 *trees and fringe analysis.* An interesting technique, sometimes called *fringe analysis*, applicable to certain families of random trees was introduced by Yao [37]. Having nothing new to say, I shall not go into details, but shall briefly indicate the connection with the present work.

Write T_h^* for the set of all rooted trees for which: (i) all leaves are at height h; (ii) all internal nodes have out-degree 2 or 3. Let $T^* = \bigcup_h T_h^*$. Given $t \in T^*$ and $h \geq 1$, there is an h-fringe of t, the multiset of all subtrees of t which have height h. Such subtrees are in T_h^*. Write $f_h(t, s)$ for the proportion of the subtrees in the h-fringe which are isomorphic to s.

Yao [37] studied a certain family $(\mathscr{T}_n;\ n \geq 1)$ of T^*-valued random trees and observed the following special property. For each h, the process of mean proportions of fringe subtrees

$$(48) \qquad n \to \left(Ef_h(\mathscr{T}_n, s);\ s \in T_h^*\right)$$

evolves in the same way as the process

$$(49) \qquad n \to \left(P\left(Z_n^h = s\right);\ s \in T_h^*\right)$$

for a certain T_h^*-valued Markov chain $(Z_n^h;\ n \geq 1)$. This chain has some stationary distribution $\pi_h(\cdot)$ on T_h^*, which can be calculated numerically for small h. And global functionals of the tree can be bounded by functionals on T_h^*, so numerical bounds for asymptotics of such global functionals can be computed.

In the terminology of the present paper, the family $(\mathscr{T}_n;\ n \geq 1)$ has an asymptotic fringe distribution of the form

$$\pi = \sum_h c_h \pi_h$$

for some constants (c_h); the limits (π_h) must fit together via

$$c_h \pi_h Q = c_{h-1} \pi_{h-1}$$

for Q as at (1).

An abstract treatment of this idea of models where fringes evolve with n as Markov chains was given by Eisenbarth, Ziviani, Gonnet, Mehlhorn and Wood [12]; more examples are given there and in subsequent papers, for example, [34, 9]. Aldous, Flannery and Palacios [6] observed that, in the abstract setting proposed in [12], the Markov chain arises from an urn model and that one could deduce a.s. convergence of fringe proportions. That is, there is an a.s. fringe π in the sense of (11).

These models are the simplest examples where one can show that asymptotic fringe distributions π exist without being able to describe π explicitly. An open question about these models concerns the asymptotic behavior of heights, that is to say the evaluation of the limit constant

$$\gamma = \lim_n \frac{\mathrm{ht}(\mathcal{T}_n)}{\log n}.$$

Here γ can be expressed in terms of the limiting sin-tree. There are some heuristic arguments using sin-trees to obtain explicit values of γ in certain models; it is a challenging open problem to formalize these arguments.

7. Technical examples and remarks.

7.1. *k-type critical Galton–Watson branching process.* Here is an example of a fringe distribution which can be described simply but which does *not* have a one-dimensional reduced Markov description.

An extension of Example 3.4 is to replace the offspring distribution ξ by a family $(\xi_{i,j};\ 1 \le i, j \le k)$. Suppose that the matrix of means

$$M_{i,j} = E\xi_{i,j}$$

is an irreducible *stochastic* matrix, which therefore has a unique stationary distribution ρ. Suppose that a type i individual has (for each j) $\xi_{i,j}$ offspring of type j, who themselves have offspring independently. This specifies a critical k-type Galton–Watson process, and under a minor nondeterminism assumption, the family tree $\mathcal{T}_i$ for the process started with a single type i individual is a.s. finite. It is straightforward to check that the mixture

$$\pi = \sum_i \rho(i)\mathrm{dist}(\mathcal{T}_i)$$

is a fringe distribution. The associated invariant sin-tree is given by the construction in Section 4.3, with (M, ρ) replacing (R, ν). However, we lose the monotonicity property (41) and so the limit is not one-dimensional Markov as defined in Section 4.4.

One's initial reaction to this example is to say: change the definition of one-dimensional Markov to allow multiple types. But this is not satisfactory, since if we were to allow an infinite number of types, then (as observed in Section 4.3) we could declare tree t to be "type t" and the notion of "Markov" becomes vacuous. Allowing only finitely many types does not resolve the issue, because the current example could be restated with infinitely many types.

REMARK. It is undoubtedly true (but may be technically messy to prove) that this π is the asymptotic fringe distribution (as $n \to \infty$) associated with the critical k-type branching process conditioned to have population size equal to n.

7.2. *Balanced deterministic trees.* Recall from Section 4.1 the branch representation $(B_i;\ i \ge 0)$ of an invariant sin-tree. In the critical Galton–Watson limit case, the branches are i.i.d. for $i \ge 1$. In all the other natural examples, the size of the branches $|B_i|$ tends to increase with i. One might conjecture that this is a general fact: Here is a counterexample to several formalizations of this idea.

Let $d_1, d_2, \dots \ge 1$ be deterministic and such that

$$D \equiv \sum_{i \ge 0} \delta_i < \infty, \quad \text{where } \delta_i = (d_1 d_2 \cdots d_i)^{-1}, \qquad \delta_0 = 1.$$

For each h, let t_h be the deterministic tree of height h in which the root has degree d_h, each first-generation offspring has out-degree d_{h-1}, each second-generation offspring has out-degree d_{h-2} and so on. Then $(t_h;\ h \ge 0)$ is a (height- or size-) coherent family with asymptotic fringe distribution

$$\pi(t_h) = \delta_h / D, \qquad h \ge 0.$$

Consider the case where

$$d_{j^2} = 2, \qquad j \ge 1, \qquad d_i = 1 \text{ otherwise}.$$

Then the branch size process $(|B_i|;\ i \ge 0)$ is of the form $(b_{i+\eta};\ i \ge 0)$, where η is a random shift and $b_i = 1$ for $i \ne j^2$. Clearly this process is not increasing in any strong sense.

7.3. *Marked trees.* We set up our general theory for rooted trees without any extra structure. Most types of structure can be handled by introducing some suitable mark space S and considering marked trees. For instance:

1. To handle ordered trees, we introduce a mark space $S_{\text{ord}} = \{1, 2, \dots\}$ and mark the ordered offspring $(v_1, v_2, \dots)$ of each vertex v with marks $1, 2, \dots$.
2. To handle a tree with edge-weights (e.g., in the random minimum-weight spanning tree problem mentioned in Section 6.3), we mark each vertex with the weight of the edge leading from that vertex to the root, using label-space $S = (0, \infty)$.
3. For trees built with vertices in R^d, we just mark each vertex by its position in R^d.
4. In studying continuous-time branching processes (Section 3.1), each vertex of $\mathscr{T}_t$ represents an individual born before time t. It is natural to mark the individual with its birth-time (most conveniently measured backwards from time t).

Several types of structure may be present simultaneously: Just use a product mark space.

To make a theory, let S be either countable (with the discrete topology) or uncountable and equipped with a Polish topology. Write T_S for the set of finite rooted trees with marks from S. Thus $t \in T_S$ has a vertex-set $\mathbf{v}(t)$, an edge-set $\mathbf{e}(t)$ and marks $\{s(v): v \in \mathbf{v}(t)\}$. Two marked trees are identical if there is an isomorphism between vertices which preserves edges and marks. To discuss convergence of trees or random trees, we require that the mark-space S be the same for all the trees under discussion. For marked trees t_n, t_∞, the assertion $t_n \to t_\infty$ means: For each sufficiently large n, there exists an isomorphism ϕ_n: $\mathbf{v}(t_\infty) \to \mathbf{v}(t_n)$ which preserves edges and is such that $s_n(\phi_n(v)) \to s(v)$ as $n \to \infty$ for each $v \in \mathbf{v}(t_\infty)$.

The general results of this paper apply to marked trees with essentially no change. In Lemma 4 (fringe-tightness) we need to add a condition

(iii) $$\{s(V_k): k = 1, 2, \dots\} \text{ is tight on } S.$$

Theorem 13 becomes nicer in this setting. Regarding an invariant sin-tree in a reduced Markov description as an invariant marked sin-tree, the shift-coupling of ancestor chains property becomes *necessary*, as well as sufficient, for extremality.

7.4. *Analogy with stationarity.* Let us list analogies between our results and standard elementary facts about stationary sequences.

Fix a finite set S. Suppose that for each n, we have a sequence of S-valued random variables $(X_{n,1}, \dots, X_{n,n})$. Let U_n be uniform on $\{1, \dots, n\}$, independent of the X's. For each n, define $Y_i^n = X_{n, U_n + i}$, with some (unimportant) convention about overflow. Two elementary facts are: If

(50) $$(Y_i^n; i \geq 0) \to_d (Y_i; i \geq 0),$$

then the limit process $(Y_i;\ i \geq 0)$ is stationary and every stationary sequence arises this way. Our Propositions 2 and 3 are clearly analogs of these facts. Proposition 7 is analogous to a weak (convergence in probability) ergodic theorem for stationary sequences (which can be proved in the same soft way, though is not often given in textbooks). Proposition 10 is analogous to the fact that the distribution of a one-sided stationary sequence $(Y_i;\ i \geq 0)$ determines uniquely the distribution of a two-sided stationary sequence $(Y_i;\ -\infty < i < \infty)$. Proposition 11 is analogous to the fact that convergence of one-sided sequences in (50) implies convergence of two-sided sequences.

Our reduced Markov descriptions in Section 4.3 are analogous to the issue of representing a stationary process (Y_n) as a function $Y_n = h(X_n)$ of a stationary Markov process (X_n). There is always a trivial such representation with $X_n = (\dots, Y_{n-1}, Y_n)$, but this gives no information. But if there is a representation with a simple Markov chain, then we can deduce useful results about (Y_n).

7.5. *Final remarks.*

1. There is a coding between trees and walks which enables invariant sin-trees to be identified with a class of stationary processes. This coding yields alternative proofs of some simple results (e.g., Proposition 7) but does not help with Theorem 13. It also provides a starting place for the study of random tree analogs of deeper facts about stationary processes. For example, we are studying how the entropy of a limit sin-tree relates to issues in data compression, that is, ability to code the finite random trees as smaller trees.

2. We used random walks on invariant sin-trees as a technical tool in the proof of Theorem 13, but they have some intrinsic interest which may be pursued elsewhere. Essentially, the rate of convergence of $P(W_t = W_0)$ to 0 reflects the growth rate of the random tree, and in particular provides useful definitions of trees of exponential/polynomial growth. The case of the conditioned critical Galton–Watson sin-tree has been studied in detail by Kesten [23].
3. In the examples of Section 3, central limit theorems for the number of leaves are known; see Najock and Heyde [31] for the Yule tree, Kolchin ([26], Theorem 2.3.1) for the conditioned critical Galton–Watson process, Aldous and Shields [4] for the digital storage tree, Jacquet and Regnier [20] for tries. It would be nice to find some abstract structure which implied such a CLT, and our best guess is coherence.
4. Our results concern asymptotics of the local structure of trees around a random vertex. One can also study asymptotics of the local structure around the root. This is comparatively easy to study directly in concrete models, but less amenable to structure theory, because obviously *any* limit infinite tree can occur. The particular case of conditioned critical Galton–Watson processes has been studied (e.g., Kesten [23], Lemmas 1.14 and 2.2), because here the limit tree has the interesting interpretation of the critical branching process, conditioned to live forever. In this example, it turns out that the limit root-based sin-tree is almost the same as the limit invariant sin-tree (based on a random vertex) we derived in Section 4.4; the only change is that B_0 has the distribution as B_i, $i \geq 1$. Some functionals of finite conditioned branching processes, which we can interpret as functionals of the root-based fringe, were studied analytically in Meir and Moon [29, 30].
5. Our work also suggests various structural questions, such as "describe constructively all coherent families with a one-dimensional reduced Markov description".

Acknowledgment. I thank a referee for careful reading and the concise proof of Lemma 1(c).

REFERENCES

[1] ALDOUS, D. J. (1990). A random tree model associated with random graphs. *Random Structures and Algorithms*. 1 383–402.

[2] ALDOUS, D. J. (1990). The random walk construction of uniform spanning trees and uniform labelled trees. *SIAM J. Discrete Math.* **3** 450–465.

[3] ALDOUS, D. J. (1991). The continuum random tree III. In preparation.

[4] ALDOUS, D. and SHIELDS, P. (1988). A diffusion limit for a class of randomly-growing binary trees. *Probab. Theory Related Fields* **79** 509–542.

[5] ALDOUS, D. J. and STEELE, J. M. (1990). Asymptotics for Euclidean minimal spanning trees on random points. In preparation.

[6] ALDOUS, D. J., FLANNERY, B. and PALACIOS, J. (1988). Two applications of urn processes. *Probab. Engrg. Inform. Sci.* **2** 293–307.

[7] ASMUSSEN, S. (1987). *Applied Probability and Queues*. Wiley, New York.

[8] BRODER, A. (1989). Generating random spanning trees. In *Symp. Foundations of Computer Sci.* 442–447. IEEE, New York.

[9] CUNTO, W. and GASCON, J. L. (1987). Improving time and space efficiency in generalized binary search trees. *Acta Inform.* **24** 583–594.

[10] DONDAJEWSKI, M., KIRSCHENHOFER, P. and SZYMANSKI, J. (1985). Vertex-degrees in strata of a random recursive tree. *Ann. Discrete Math.* **28** 99–105.

[11] DYNKIN, E. B. (1978). Sufficient statistics and extreme points. *Ann. Probab.* **6** 705–730.

[12] EISENBARTH, B., ZIVIANI, N., GONNET, G., MEHLHORN, K., and WOOD, D. (1982). The theory of fringe analysis and its application to 2–3 trees and *b*-trees. *Inform. and Control* **55** 125–174.

[13] FLAJOLET, P. (1988). Mathematical methods in the analysis of algorithms and data structures. In *Trends in Theoretical Computer Science* 225–304. Computer Science Press, Rockville, Md.

[14] FLAJOLET, P. (1988). Random tree models in the analysis of algorithms. In *Performance '87* (P. J. Courtois and G. Latouche, eds.) 171–187. Elsevier, New York.

[15] FLAJOLET, P. and SEDGEWICK, R. (1986). Digital search trees revisited. *SIAM J. Comput.* **15** 748–767.

[16] FRIEZE, A. M. (1985). On the value of a random minimum spanning tree problem. *Discrete Appl. Math.* **10** 47–56.

[17] GREVEN, A. (1987). Couplings of Markov chains and randomized stopping times. *Probab. Theory Related Fields* **75** 195–212.

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

[19] JACQUET, P. and REGNIER, M. (1988). Normal limiting distribution for the size and the external path length of tries. Technical report, INRIA.

[20] JACQUET, P. and REGNIER, M. (1989). New results on the size of tries. *IEEE Trans. Inform. Theory* **35** 203–205.

[21] JAGERS, P. (1975). *Branching Processes with Biological Applications*. Wiley, New York.

[22] JAGERS, P. and NERMAN, O. (1984). The growth and composition of branching processes. *Adv. in Appl. Probab.* **16** 221–259.

[23] KESTEN, H. (1987). Subdiffusive behavior of random walk on a random cluster. *Ann. Inst. H. Poincare Sect. B* **22** 425–487.

[24] KNUTH, D. E. (1968). *The Art of Computer Programming* **1**. Addison-Wesley, Reading, Mass.

[25] KNUTH, D. E. (1973). *The Art of Computer Programming* **3**. Addison-Wesley, Reading, Mass.

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

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

[28] MEIR, A. and MOON, J. W. (1981). The outer-distance of nodes in random trees. *Aequationes Math.* **23** 204–211.

[29] MEIR, A. and MOON, J. W. (1981). Terminal path numbers for certain families of trees. *J. Austral. Math. Soc. Ser. A* **31** 429–438.

[30] MEIR, A. and MOON, J. W. (1986). On major and minor branches of rooted trees. *Canad. J. Math.* **3** 673–693.

[31] NAJOCK, D. and HEYDE, C. C. (1982). On the number of terminal vertices in certain random trees with an application to stemma construction in philology. *J. Appl. Probab.* **19** 675–680.

[32] NERMAN, O. and JAGERS, P. (1984). The stable doubly infinite pedigree process of supercritical branching populations. *Z. Wahrsch. Verw. Gebiete* **65** 445–460.

[33] PEMANTLE, R. (1991). Choosing a Spanning Tree for the Integer Lattice Uniformly. *Ann. Probab.* To appear.

[34] POBLETE, P. and MUNRO, J. (1985). The analysis of a fringe heuristic for binary search trees. *J. Algorithms* **6** 336–350.

[35] STEYAERT, J.-M. and FLAJOLET, P. (1983). Patterns and pattern-matching in trees: An analysis. *Inform. and Control* **58** 19–57.

[36] VITTER, J. S. and FLAJOLET, P. (1990). Analysis of algorithms and data structures. In *Handbook of Theoretical Computer Science*. North–Holland, Amsterdam.

[37] YAO, A. (1978). On random 2–3 trees. *Acta Inform.* **9** 159–170.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720