Recursive Self-Similarity for Random Trees, Random Triangulations and Brownian Excursion

David Aldous

Stable URL:

http://links.jstor.org/sici?sici=0091-1798%28199404%2922%3A2%3C527%3ARSFRTR%3E2.0.CO%3B2-A

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

1992 WALD MEMORIAL LECTURE

RECURSIVE SELF-SIMILARITY FOR RANDOM TREES, RANDOM TRIANGULATIONS AND BROWNIAN EXCURSION[1]

BY DAVID ALDOUS

University of California, Berkeley

Recursive self-similarity for a random object is the property of being de-composable into independent rescaled copies of the original object. Certain random combinatorial objects—trees and triangulations—possess approximate versions of recursive self-similarity, and then their continuous limits possess exact recursive self-similarity. In particular, since the limit continuum random tree can be identified with Brownian excursion, we get a nonobvious recursive self-similarity property for Brownian excursion.

1. Introduction. There is a Big Picture of a world of discrete and continuous random trees, surveyed several years ago in [2]. This paper treats a new aspect of this world, which hopefully will be interesting and comprehensible to the reader who has not seen this world before. In Section 1 we jump right into the specific issue of this paper and lay out some concrete calculations in the discrete world of random trees and triangulations. In Section 2 we use the Big Picture to see consequences in the continuous world. In Section 3 we do some explicit distributional calculations, which can be viewed as exact calculations for continuum random trees or Brownian excursion, or as asymptotics for discrete random trees and triangulations.

I do want to make an ideological point, illustrated by the following joke. A tourist was enjoying a beautiful panoramic vista of mountains interspersed with lakes when she heard a neighbor comment, "This country would be so much tidier if all the mountains were moved over *there* and all the lakes were moved over *there*!" To me, the beauty of this topic is the interaction between the discrete and continuous worlds. It is possible to be tidy-minded and treat asymptotics of discrete random objects via discrete methods which pay no attention to the existence of continuous limit objects, and to treat continuous random objects by continuous methods without reference to discrete approximations, but why?

1.1. Recursive self-similarity. The idea (I shall not attempt a precise definition) of recursive self-similarity of a random set is that the set can be decomposed as a union of (perhaps randomly) rescaled independent copies
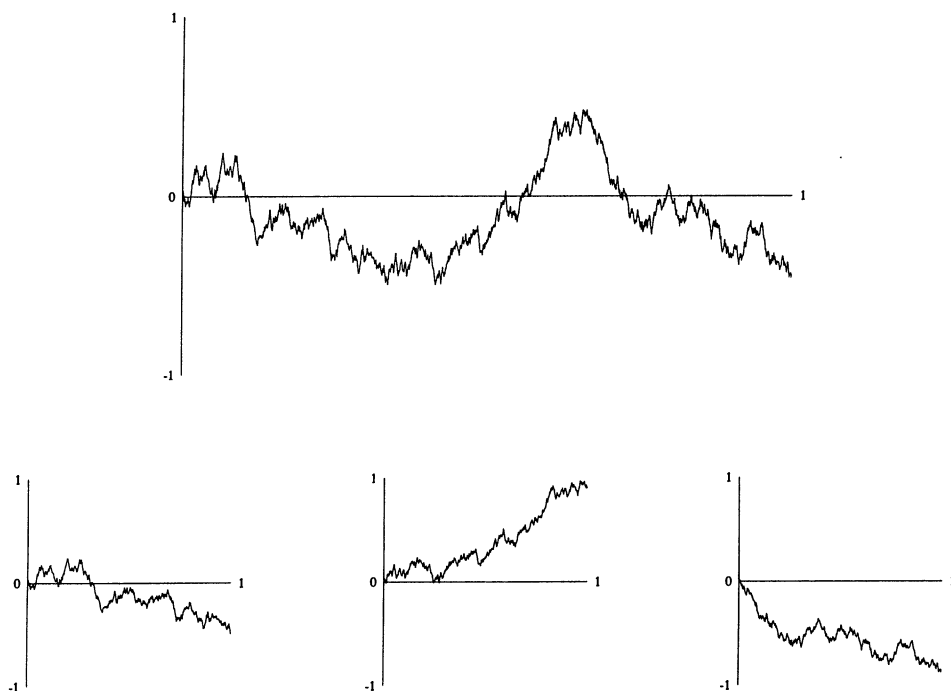
FIG. 1.

of itself. The most familiar example is the Brownian motion sample path $\mathbf{B} = \{B(t): 0 \leq t \leq 1\}$. The prescription (see Figure 1)

$$B(t) = \begin{cases} 3^{-1/2}B_1(3t), & 0 \leq t \leq \frac{1}{3}, \\ 3^{-1/2}\big(B_1(1) + B_2(3t-1)\big), & \frac{1}{3} \leq t \leq \frac{2}{3}, \\ 3^{-1/2}\big(B_1(1) + B_2(1) + B_3(3t-2)\big), & \frac{2}{3} \leq t \leq 1, \end{cases}$$

decomposes $\mathbf{B}$ into three independent copies $\mathbf{B}_1, \mathbf{B}_2, \mathbf{B}_3$ of $\mathbf{B}$. (Yes, we could have used just two copies, but three is more in keeping with the theme of the paper.) A simple application of Donsker's theorem for triangular arrays shows that Brownian motion is (up to scaling) the unique continuous random process with this particular decomposition property.

A somewhat different, large class of examples are obtained by suitably randomizing the type of recursive constructions of deterministic fractals—such as the Cantor set and the Sierpinski gasket—with which the reader is surely familiar. For definiteness, let us specify an example, essentially what Mandelbrot [15] calls canonical curdling.

Split the unit square into four equal subsquares. Choose one subsquare at random and delete it, retaining the other three. Repeat recursively within each retained subsquare. This yields a decreasing sequence of random sets, whose intersection S is a random set with dimension $\log_2 3$. Also, S has an obvious
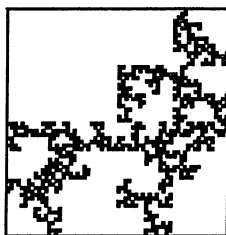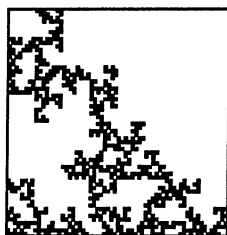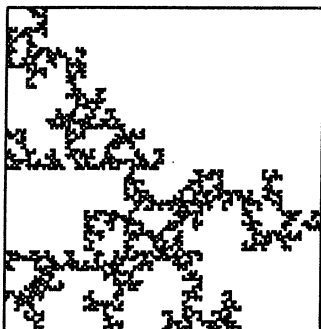
FIG. 2.

self-similarity property which can be written as

$$(1) \qquad S =_d \left(\tfrac{1}{2}S_1 + a_1\right) \cup \left(\tfrac{1}{2}S_2 + a_2\right) \cup \left(\tfrac{1}{2}S_3 + a_3\right),$$

where the S_i are independent copies of S and the a_i are sampled without re-placement from $\left\{(0,0), \left(0, \tfrac{1}{2}\right), \left(\tfrac{1}{2}, 0\right), \left(\tfrac{1}{2}, \tfrac{1}{2}\right)\right\}$. See Figure 2. Taking the discarded subsquares to be open makes S closed, and it is easy to check that S is the unique nonempty closed random set satisfying (1).

Falconer [7], Graf [10] and Mauldin and Williams [16] abstracted this type of example to define a notion of "random recursive construction" which allows, for instance, the recursive step to consist of replacing a square by randomly po-sitioned subsquares of differing sizes. See [11, Section 6] for examples. See also Patzschke and Zahle [22] for how these constructions fit into a general "self-similarity" framework, and see Falconer and Grimmett [8] for further results. These authors were mainly concerned with fractal dimension, instead of the distributional questions which underlie this paper. Also, I view these construc-tions rather differently from how I view Brownian motion. Brownian motion is a "natural" process which happens to have a recursive self-similarity property, whereas the constructions above are "artificial" and are specifically designed to have a recursive self-similarity property. Are there other "natural" examples? The next two sections treat two different-looking discrete-world examples.
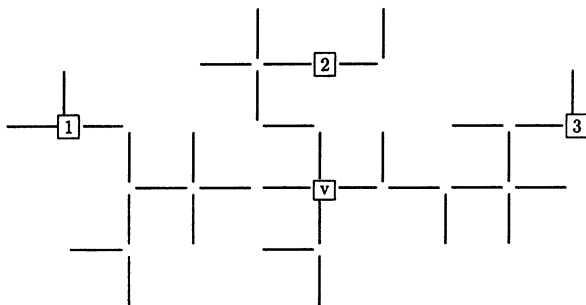
FIG. 3.

1.2. *Trees and branches.* Figure 3 illustrates the idea that, given three typical vertices $\{1, 2, 3\}$ in a typical large tree, these should lie in different branches from a branchpoint v, and these three branches should contain almost all the vertices of the tree. This suggests an approximate recursive decomposition of random trees into three subtrees. To say this carefully we recall some terminology. A finite *tree* is a connected graph without cycles. A *path* $v_0, v_1, \ldots, v_k$ is a sequence of distinct vertices for which each (v_i, v_{i+1}) is an edge. Saying a tree is *rooted* just means that one of the vertices is distinguished and called the root. In a rooted tree we may use the language of families. That is, in a path (root $= v_0, v_1, \ldots, v_k$) we call v_{i+1} a *child* of v_i, and we call v_i the *parent* of v_{i+1}. If the root has d children $w_1, \ldots, w_d$, then the tree has d *branches*, each branch being the subtree consisting of some w_i and its descendants. (Our convention is to exclude the root from the branches). The *size* of a tree or subtree is its number of vertices.

The precise version of our initial idea is as follows. Consider an n-vertex tree with vertices labeled $\{1, 2, \ldots, n\}$. Consider three distinct vertices, say, 1, 2, 3. There are two alternatives: (a) One of these vertices is on the path joining the other two. (b) There exists a vertex $v > 3$ such that, considering the tree rooted at v, vertices 1, 2 and 3 are in different branches, and these branches have sizes m_1, m_2 and m_3 satisfying

$$m_1 \geq 1, \qquad m_2 \geq 1, \qquad m_3 \geq 1, \qquad m_1 + m_2 + m_3 \leq n - 1.$$

We want to argue that (b) is the "likely" alternative. Cayley's formula says there are n^{n-2} different unrooted trees on n labeled vertices. So consider a "uniform random n-tree", that is, a uniform random choice from these n^{n-2} trees. Using the fact (e.g., [19], Theorem 7.8) that

$$P(\text{distance between vertices } 1 \text{ and } 2 = k) = \frac{k+1}{n-1} \prod_{i=1}^{k} \frac{n-i}{n},$$

it easily follows that the probability of (a) equals

$$\sum_{k=1}^{n-1} \frac{3(k+1)(k-1)}{(n-1)(n-2)} \prod_{i=1}^{k} \frac{n-i}{n} \sim 3\sqrt{\frac{\pi}{2}}\, n^{-1/2}.$$

So far large n, (b) is indeed the more likely possibility. Write $q_n(m_1, m_2, m_3)$ for the probability of (b). An elementary counting exercise shows

$$q_n(m_1, m_2, m_3) = \frac{1}{n^{n-2}} \times (n-3)$$

(2)
$$\times \frac{(n-4)!}{(m_1-1)!(m_2-1)!(m_3-1)!(n-1-m_1-m_2-m_3)!}$$
$$\times m_1^{m_1-1} m_2^{m_2-1} m_3^{m_3-1} (n-m_1-m_2-m_3)^{n-m_1-m_2-m_3-2},$$

for there are $(n-3)$ choices for the root v; the next term is the multinomial for the number of ways to split the remaining $n-4$ vertices into subsets of size $m_i - 1$ to comprise the other vertices in the branches containing i, $i = 1, 2, 3$; the terms $m_i^{m_i-1}$ give the number of ways to put tree-structure on these subsets, with one vertex distinguished as being attached to v; the final term is the number of ways to put tree-structure on the remaining $n - m_1 - m_2 - m_3$ vertices, with vertex v as root.

It is straightforward to use Stirling's approximation to get asymptotics. Write

$$\rho(m) = \frac{m^{m-1}}{m!} e^{-m}, \qquad m \geq 1.$$

For t_1, t_2, t_3 satisfying

(3) $$t_1 > 0, \qquad t_2 > 0, \qquad t_3 > 0, \qquad t_1 + t_2 + t_3 = 1$$

we find

(4) $$q_n(m_1, m_2, m_3) \sim n^{-2}(2\pi)^{-1}(t_1 t_2 t_3)^{-1/2} \rho(m_4)$$

as $n \to \infty$, $m_i/n \to t_i$, $i = 1, 2, 3$, where $m_4 = n - m_1 - m_2 - m_3$.

Now ρ is the distribution of a random variable Z [which can be interpreted as the total population of the Galton–Watson branching process with Poisson(1) offspring], and $(2\pi)^{-1}(t_1 t_2 t_3)^{-1/2}$ is the Dirichlet$(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$ density of random variables $(\Delta_1, \Delta_2, \Delta_3)$, say, on the simplex (3). Thus (4) is the local limit theorem corresponding to the following (formally weaker) convergence in distribution assertion.

PROPOSITION 1.   *In the uniform random n-tree, in case (b) let $\Delta_1^{(n)}, \Delta_2^{(n)}$ and $\Delta_3^{(n)}$ be the sizes of the branches containing vertices 1, 2 and 3, and in case (a) let $\Delta_1^{(n)}, \Delta_2^{(n)}$ and $\Delta_3^{(n)}$ be arbitrary. Then, as $n \to \infty$,*

$$(n^{-1}\Delta_1^{(n)}, n^{-1}\Delta_2^{(n)}, n^{-1}\Delta_3^{(n)}, n - \Delta_1^{(n)} - \Delta_2^{(n)} - \Delta_3^{(n)}) \to_d (\Delta_1, \Delta_2, \Delta_3, Z),$$

where Z and $(\Delta_1, \Delta_2, \Delta_3)$ are independent with the distributions specified above.

Implicit in the argument for (2) is the fact that, conditional on $\{\Delta_i^{(n)} = m_i; i = 1, 2, 3\}$, the three branches are independent uniform random m_i-trees, and the three vertices in these branches which attach to the root are independent uniform choices. With these observations, Proposition 1 provides an approximate recursive decomposition of random n-trees as a joining-together of three smaller random trees. This is only *approximate* for two reasons: There is some chance that the three vertices do not determine a branchpoint, and there are $O(1)$ vertices left out of the three branches. However, in the $n \to \infty$ limit these difficulties vanish; see Theorem 2 later.

1.3. *Triangulations of the n-gon.*   An undergraduate-level exposition of some of the ideas in this paper relating to triangulations was given in [4], but the presentation there has a different emphasis.

A *triangulation* of a finite set Q in the plane is a collection of nonintersecting line segments with endpoints from Q such that the convex hull of Q is partitioned into triangular regions. We shall be concerned only with the cases Q_n consisting of the vertices of a convex n-gon, which are labeled as $1, 2, \ldots, n$ in clockwise order from an arbitrary starting vertex. In such a triangulation each point is linked to its neighbor on either side and may or may not be linked to other points. It is well known ([23] and [9], Chapter 20) that the number of different triangulations of Q_n is given by the $(n-1)$st Catalan number c_{n-1}, where

$$(5) \qquad\qquad c_m = \frac{(2m-2)!}{(m-1)!m!}, \qquad m \geq 1.$$

One way to check this is from the recursion (6). The edge $(1, n)$ is part of a triangle whose other vertex is some j with $2 \leq j \leq n-1$, and then the whole triangulation decomposes into a triangulation of vertices $1, 2, \ldots, j$ and a triangulation of vertices $j, j+1, \ldots, n$. Thus the number $\tilde{c}_{n-1}$ of triangulations of Q_n satisfies

$$(6) \qquad\qquad \tilde{c}_{n-1} = \sum_{j=2}^{n-1} \tilde{c}_{j-1}\tilde{c}_{n-j}, \qquad n \geq 3,$$

where $\tilde{c}_1 = 1$, and this is a standard recursion for the Catalan numbers.

Consider now the uniform random triangulation of Q_n. What is the chance it contains the triangle with vertices (i_1, i_2, i_3), where $1 \leq i_1 < i_2 < i_3 \leq n$? If it does contain this triangle, then the triangulation decomposes into three triangulations of the vertices $\{i_1, \ldots, i_2\}$, $\{i_2, \ldots, i_3\}$ and $\{i_3, \ldots, n, 1, \ldots, i_1\}$, respectively. So

P(random triangulation of Q_n contains triangle (i_1, i_2, i_3))

$$(7) \qquad\qquad = \frac{c_{i_2-i_1}c_{i_3-i_2}c_{n+i_1-i_3}}{c_{n-1}}.$$

Noting that (5) and Stirling's formula give

$$c_m \sim \pi^{-1/2} m^{-3/2} 2^{2m-2},$$

we can take asymptotics in (7) to get

$$P\big(\text{random triangulation of } Q_n \text{ contains triangle } (i_1, i_2, i_3)\big)$$

$$\sim n^{-3} \phi(t_1, t_2, t_3) \quad \text{as } \left(\frac{i_1}{n}, \frac{i_2}{n}, \frac{i_3}{n}\right) \to (t_1, t_2, t_3),$$

where

(8)
$$\phi(t_1, t_2, t_3) = \frac{1}{4\pi}(t_2 - t_1)^{-3/2}(t_3 - t_2)^{-3/2}(1 + t_1 - t_3)^{-3/2},$$

$$0 \le t_1 < t_2 < t_3 \le 1.$$

To get a recursive decomposition, consider the case where n is odd. A triangulation of Q_n contains a unique triangle with vertices $1 \le i_1 < i_2 < i_3 \le n$ such that

(9)
$$\max(i_2 - i_1, i_3 - i_2, i_1 + n - i_3) \le n/2$$

because, when Q_n is the *regular* n-gon inscribed in a circle, this is the triangle containing the center of the circle. Arguing as in (6), the uniform random triangulation can be decomposed into three conditionally independent uniform random triangulations of vertices $\{I_1, \ldots, I_2\}$, $\{I_2, \ldots, I_3\}$ and $\{I_3, \ldots, n, 1, \ldots, I_1\}$, where (I_1, I_2, I_3) is the triangle in the original random triangulation containing the center of the circle. Figure 4 is an illustration, where the triangulations of $\{I_1, \ldots, I_2\}$ and so on are redrawn as triangulations of *regular* polygons.

This recursion is not exact for even n, because there may be an edge through the center of the circle. However, we shall see in subsection 2.3 that there is a notion of "random triangulation of the circle" which does have exact recursive self-similarity.

1.4. *Combinatorial recursions.* What we have done looks strange from a purely combinatorial viewpoint. Instead of the simple recursion (5) for the number of triangulations, we are using in (6) the more complicated and incomplete recursion

$$c_{n-1} = \sum_{\substack{1 \le i_1 < i_2 < i_3 \le n, \\ \max(i_2 - i_1, i_3 - i_2, n + i_1 - i_3) \le n/2}} c_{i_2 - i_1} c_{i_3 - i_2} c_{n + i_1 - i_3}, \quad n \text{ odd.}$$

Similarly, the natural recursion for the number of rooted trees of size n is obtained by considering branch sizes at the root, and this is simpler than the approximate recursion implicit in subsection 1.2. However, these sorts of recursions (see [18] for the case of rooted trees), from the probability viewpoint,
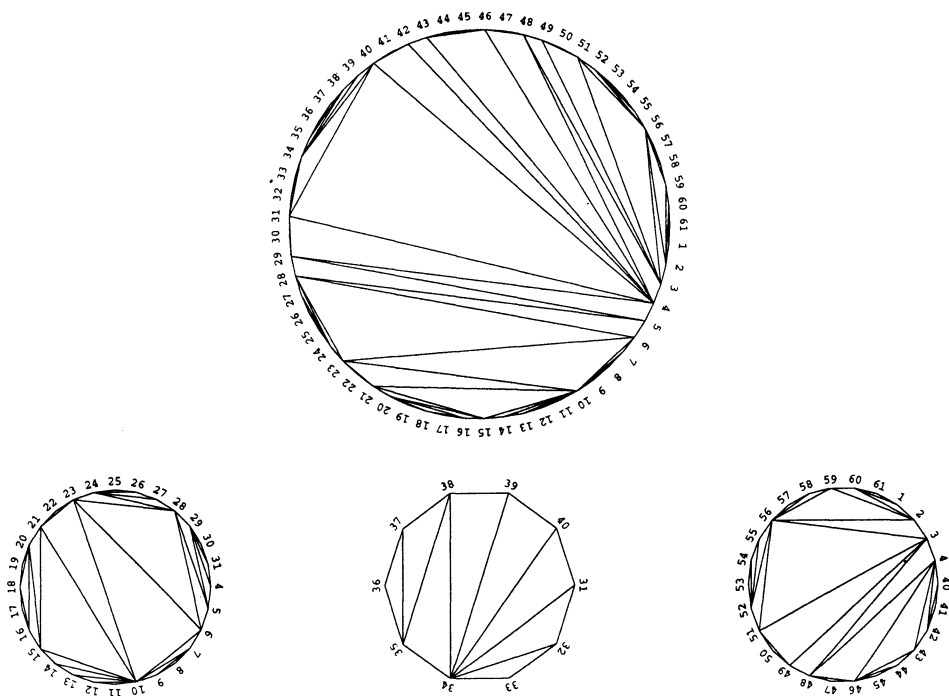
FIG. 4.

split random size-n objects into smaller random objects of typical sizes $O(1)$ and $n - O(1)$, whereas we want to split into smaller objects whose sizes are order n rather than smaller order. More vividly, we are "splitting logs" instead of "chipping away at a glacier."

2. The continuous world.

2.1. *The continuum random tree.* The main idea of this paper is that the approximate recursive decompositions of discrete objects in subsections 1.2 and 1.3 become exact recursive decompositions in the continuous limit (Theorems 2 and 4). Here we need to start giving some of the Big Picture from [2]. The classical invariance principle formalizes the idea that if partial sums grow as order $n^{1/2}$, then the rescaled partial sum process should converge to Brownian motion. There is an analogous invariance principle for trees. If random n-vertex trees are such that the distance between typical vertices grows as order $n^{1/2}$, then the rescaled trees (make each edge have length $n^{-1/2}$) should converge to a particular limit, the Brownian continuum random tree (CRT). This is proved in [3] for a particular family of random tree models, which include those appearing in this paper. Abstractly, a continuum tree is a pair $(\mathcal{S}, \mu)$, where $\mathcal{S}$ is a set representing the vertices and with certain natural properties (in particular, that there is a unique path in $\mathcal{S}$ between any two vertices) and μ is a probability

measure on S representing picking a vertex uniformly at random.

In this paper we will not give details of weak convergence arguments—[3] is a kind of "proof intimidation" that it is indeed possible to write out details. But granted a weak convergence theorem [3, Theorem 23] saying that uniform random n-trees can be rescaled to converge to the Brownian CRT, Proposition 1 and the remarks following it imply an exact recursive self-similarity property for the limit Brownian continuum random tree. This property is stated as Theorem 2. "Brownian-scaling by c" means multiplying distances by $c^{1/2}$ and measures by c.

THEOREM 2. *Given the Brownian CRT (S, μ), pick three vertices Z_1, Z_2, Z_3 independently according to μ, let V be the branchpoint of these three vertices, let B_1, B_2, B_3 be the branches and let $\Delta_i = \mu(B_i)$. Then the following hold:*

(a) *$\Delta = (\Delta_1, \Delta_2, \Delta_3)$ has the* Dirichlet $\left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$ *density $(2\pi)^{-1} (t_1 t_2 t_3)^{-1/2}$.*

(b) *For $i = 1, 2, 3$, let (S_i, μ_i) be the branch (B_i, μ) Brownian-scaled by Δ_i^{-1}, and let Z_i^* be the position of Z_i after scaling. Then the (S_i, μ_i) are distributed as the Brownian CRT, independently of each other and of Δ, and each Z_i^* is distributed as a μ_i-random vertex in S_i.*

Let us say this property the other way round. Take three independent copies of the Brownian CRT. Pick a vertex uniformly at random in each. Take an independent Dirichlet-distributed $(\Delta_1, \Delta_2, \Delta_3)$. Brownian-scale the ith tree by Δ_i, and join the trees at the randomly picked vertices. Then the resulting tree is also distributed as the Brownian CRT. We can envisage the decomposition as in Figure 3, but adding more and more smaller and smaller branches (cf. [1], Figure 2).

2.2. *Coding from Brownian excursion.* Another aspect of the Big Picture is that the Brownian CRT can be constructed from Brownian excursion (of unit length). This story really goes back forty years to Harris [12, Theorem 5], who noted a one-to-one correspondence between excursions of walks with ± 1 steps and a certain kind of tree (rooted, ordered trees). With the advantage of hindsight, it is obvious that Harris's mapping {excursions of walks} $\rightarrow$ {discrete trees} has a continuous analog, a mapping

{excursions of continuous functions} $\rightarrow$ {continuum trees}.

In fact, mappings of this type were discovered independently in the 1980's in the context of Brownian motion [21, 20, 14], and the general case is discussed in [3] and [13].

The mapping is illustrated in Figure 5, which shows part of the tree derived from an *excursion function f*. [The essential requirements of an excursion function f are that $f(0) = f(1) = 0$, $f(\cdot) > 0$ on $(0, 1)$ and that f be continuous. There are other technical requirements, adapted to the technical definition of *continuum tree*. See [3], subsection 2.7.] Vertices of the tree are labeled by numbers $0 \leq t \leq 1$, with the root labeled 0, and vertex t is at distance $f(t)$ from the root. A
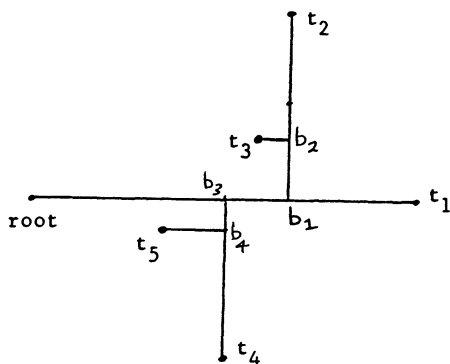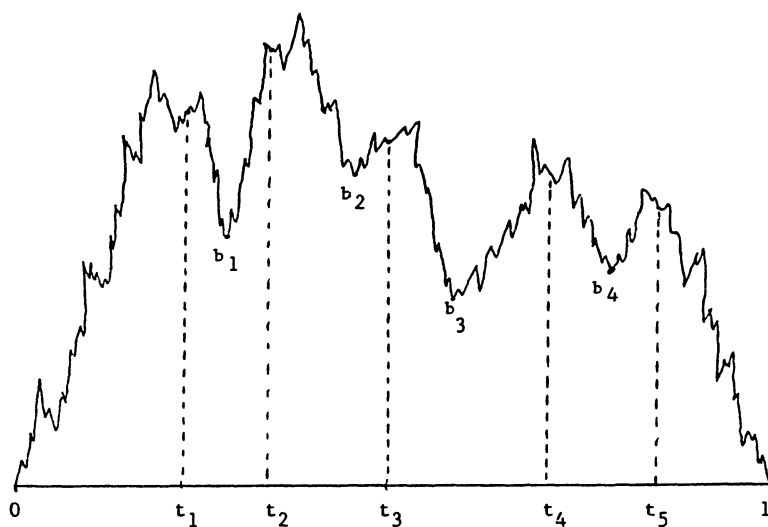
FIG. 5.

key fact is that branchpoints in the tree correspond to (times of) local minima of f. Thus, given $t_1 < t_2$, consider $b_1 = \arg\min_{[t_1, t_2]} f$. Then b_1 is the branchpoint for the triple (root, t_1, t_2) and the distances from b_1 to the root, to t_1 and to t_2 are $f(b), f(t_1) - f(b)$ and $f(t_2) - f(b)$, respectively. Another fact is that Lebesgue measure on [0, 1] is carried by the mapping to a probability measure μ on the continuum tree, and we interpret μ as picking a vertex uniformly at random from the continuum tree.

By [3], Corollary 22, applying this mapping to Brownian excursion gives the Brownian CRT. [With the standardization convention of [3] we really use $2B(t)$, where B is standard Brownian excursion, but the factor 2 makes no difference
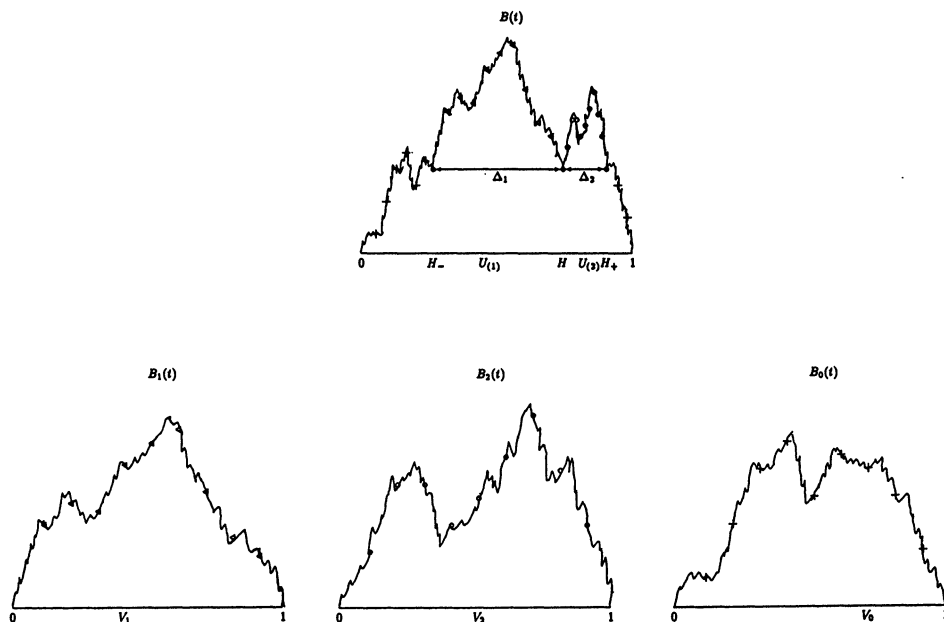
FIG. 6.

to the issues of the present paper.] Theorem 2 can therefore be rewritten as a recursive self-similarity property of Brownian excursion. To state this property, it is simplest to regard one of the three random vertices in Theorem 2 as the root, so we only pick two more random vertices, say, U_1 and U_2. Below is the precise definition (Figure 6 is more understandable) of the rescaled Brownian excursions (B_1, B_2, B_3) coding the branches in the decomposition of Theorem 2.

Start with an independent triple (B, U_1, U_2), where B is Brownian excursion and U_1 and U_2 are $U(0, 1)$. Define

$$U_{(1)} = \min(U_1, U_2), \qquad U_{(2)} = \max(U_1, U_2);$$

$$H = \arg \min_{U_{(1)} \le t \le U_{(2)}} B(t),$$

$$H_- = \sup \{t < U_{(1)} : B(t) = B(H)\},$$

$$H_+ = \inf \{t > U_{(2)} : B(t) = B(H)\};$$

$$\Delta_1 = H - H_-, \qquad \Delta_2 = H_+ - H, \qquad \Delta_0 = 1 + H_- - H_+;$$

$$V_1 = \frac{U_{(1)} - H_-}{\Delta_1}, \qquad V_2 = \frac{U_{(2)} - H}{\Delta_2}, \qquad V_0 = \frac{H_-}{\Delta_0};$$

$$B_1(t) = \Delta_1^{-1/2}\big(B(H_- + t\Delta_1) - B(H)\big), \qquad 0 \le t \le 1,$$

$$B_2(t) = \Delta_2^{-1/2}\big(B(H + t\Delta_2) - B(H)\big), \qquad 0 \le t \le 1,$$

$$B_0(t) = \Delta_0^{-1/2} B(t\Delta_0), \qquad 0 \le t \le V_0$$

$$= \Delta_0^{-1/2} B\big(H_+ + (t - V_0)\Delta_0\big), \qquad V_0 \le t \le 1.$$
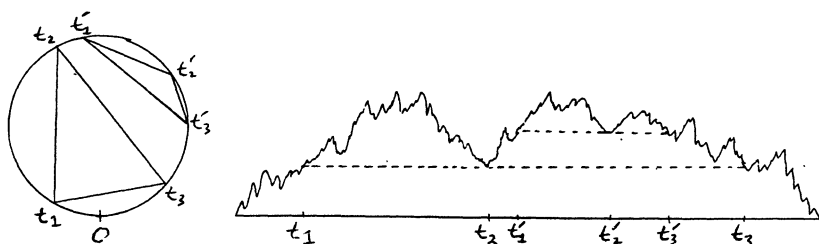
1. There exists a unique centroid.
2. n is even and there exist two centroids v_1 and v_2 such that (v_1, v_2) is an edge which partitions the vertices into two subsets of size $n/2$ each.

Turning to continuum trees, there is an elegant way of looking at the centroid. The key idea is that the two-way correspondence (Figure 5) between continuum trees and excursion functions can be extended to a three-way correspondence involving also *triangulations of the circle*.

A local minimum of an excursion function f can be regarded as a triple (t_1, t_2, t_3) with

$$f(t_1) = f(t_2) = f(t_3), \qquad f(\cdot) > f(t_2) \text{ on } (t_1, t_2) \cup (t_2, t_3).$$

Take a circle of unit circumference, with points on the circumference labeled by $0 \le t < 1$. For each local minimum (t_1, t_2, t_3) of f, draw a triangle with vertices (t_1, t_2, t_3). Figure 7 illustrates that, for a fixed function f, the triangles associated with different local minima are typically disjoint.

One could take several pages to discuss carefully, in the deterministic context, the correspondence {excursion functions} $\leftrightarrow$ {triangulations of the circle}, in the style of the discussion in [3, Section 2.7] of the correspondence {excursion functions} $\leftrightarrow$ (a certain subclass of){continuum trees}. However, this does not seem worthwhile. The essential feature of the three-way correspondence is that each triangle (t_1, t_2, t_3) of the triangulation corresponds to a branchpoint of the continuum tree, where the branches have size (i.e., μ-measures) $t_2 - t_1$, $t_3 - t_2$ and $1 + t_1 - t_3$, as well as to a local minimum of the excursion function. In particular, the centroid of the continuum tree corresponds to the triangle in the triangulation which contains the center of the circle. The point is that we can "see" the centroid in the triangulation representation much more easily than in a drawing of the tree itself or of an excursion function.

Now consider the random setting. Applying the coding of Figure 7 to Brownian excursion gives a certain *random triangulation of the circle*. A key fact is that this is the limit of the uniform random triangulations of the regular n-gon discussed in subsection 1.3. I do not want to give details of the weak convergence argument, but here is an outline. There is a classical correspondence [9, 4] between triangulations of the n-gon and a certain type of binary tree. Under this correspondence, a branchpoint in the tree corresponds to a triangle in the tri-

angulation, and a leaf in the tree corresponds to an edge of the convex hull of the n-gon. So the relative sizes of the three branches from a branchpoint v in the binary tree are essentially just the relative numbers of points of the n-gon cut off by the three sides of the triangle corresponding to v in the triangulation. The invariance principle shows that uniform random binary trees rescale to the Brownian CRT. So the asymptotic relative sizes of triangles in uniform random triangulations of the n-gon are just the sizes of branches from branchpoints in the Brownian CRT. Using the coding from Brownian excursion gives the construction of Figure 7.

What calculations can we do with this limit continuous "random triangulation of the circle"? There is a mean intensity function $\phi(t_1, t_2, t_3)$ for the positions of triangles in the random triangulation of the circle, or, equivalently, for the local minima of Brownian excursion. To spell this out, representing the vertices of a triangle by their distances t' around the circumference,

$$\phi(t_1, t_2, t_3)\, dt_1\, dt_2\, dt_3$$
$$= \text{mean number of triangles } (t'_1, t'_2, t'_3) \text{ with } t'_i \in [t_i, t_i + dt_i], \qquad i = 1, 2, 3.$$

And representing local minima of Brownian excursion as triples (t'_1, t'_2, t'_3), where t'_2 is the time of the local minimum,

$$\phi(t_1, t_2, t_3) dt_1\, dt_2\, dt_3 = \text{ mean number of local minima with } t'_i \in [t_i, t_i + dt_i],$$
$$i = 1, 2, 3.$$

The simplest way to get the explicit formula for ϕ is via the elementary computation with discrete triangulations (8) and the weak convergence argument outlined above, which gives

$$\phi(t_1, t_2, t_3) = \frac{1}{4\pi}(t_2 - t_1)^{-3/2}(t_3 - t_2)^{-3/2}(1 + t_1 - t_3)^{-3/2},$$

(10)
$$0 \le t_1 < t_2 < t_3 \le 1.$$

We invite the reader to try to derive (10) directly from the Brownian excursion description—the $t^{-3/2}$ power law is rather clear, being the law of excursion lengths under the Itô σ-finite measure on excursions of variable lengths, but getting the numerical constant correct is a little tricky.

It is clear that (with probability 1) the center of the circle will be in the interior of some triangle of the random triangulation. In other words, in the Brownian CRT there will be (with probability 1) a unique centroid, that is to say a branchpoint for which all three branches have size less than $\frac{1}{2}$. Let S_1, S_2, S_3 be the sizes of the branches from the centroid, arranged in random order. From (10) we may derive the density function

$$f_{S_1, S_2, S_3}(x_1, x_2, x_3) = \frac{1}{12\pi}(x_1 x_2 x_3)^{-3/2}$$

(11)
$$\text{on } \left\{ x_i > 0,\ x_1 + x_2 + x_3 = 1,\ \max_i x_i < \tfrac{1}{2} \right\}.$$

In deriving (11), the issue is that branchpoints with branch sizes $(x_1, x_2, 1 - x_1 - x_2)$ arise with probability 1/3! from triangles $(t, t + x_1 \bmod 1, t + x_1 + x_2 \bmod 1)$ or $(t, \ t + x_2 \bmod 1, \ t + x_1 + x_2 \bmod 1)$ for arbitrary $0 < t < 1$, so expression (11) on $\{x_i > 0, \ x_1 + x_2 + x_3 = 1\}$ is the mean intensity of the point process of branch sizes at *all* branchpoints, and restricting to $\{\max_i x_i < \frac{1}{2}\}$ picks out the centroid.

The approximate recursive decomposition of triangulations of the n-gon, illustrated in Figure 4, becomes in the limit an exact recursive decomposition of the random triangulation of the circle about the triangle containing the center of the circle. Translating into an assertion about the Brownian CRT gives a recursive self-similarity property involving decompositions about the centroid.

THEOREM 4. *Given the Brownian CRT* $(\mathcal{S}, \mu)$, *let* D_1, D_2, *and* D_3 *be the branches from the centroid, in random order, and let* $S_i = \mu(D_i)$. *Then the following hold:*

(a) $S = (S_1, S_2, S_3)$ *has the density* (11).

(b) *For* $i = 1, 2, 3$, *let* $(\mathcal{S}_i, \mu_i)$ *be the branch* (D_i, μ) *Brownian-scaled by* S_i^{-1}, *and let* Z_i^* *be the position of the original centroid in* $\mathcal{S}_i$. *Then the* $(\mathcal{S}_i, \mu_i)$ *are distributed as the Brownian CRT, independently of each other and of* S, *and each* Z_i^* *is distributed as a* μ_i-*random vertex in* $\mathcal{S}_i$.

We could of course rephrase this as a decomposition of Brownian excursion, analogous to Corollary 3, but we refrain from doing so, since the concept of the centroid is rather unnatural in the coding as Brownian excursion.

3. Calculations.

3.1. *Sizes of branches from the centroid.* As discussed in [2], if we explicitly know some distribution associated with Brownian excursion, then the invariance principle shows that the distribution is the limit distribution of some functional of combinatorial random trees. It turns out that, for most natural and interesting functionals, these limit distributions have been investigated in the combinatorial literature by generating function methods, independently of the stochastic process literature studying distributions of Brownian excursion. However, properties involving centroids of random trees have apparently not been studied in the combinatorial literature, so (11) implies a new result for random trees. To say this carefully, the invariance principle applies to what Meir and Moon [17] call 'simply generated families' of n-vertex random trees, which to a probabilist (see [3], subsection 2.1) are Galton–Watson branching processes with mean offspring equal to 1, conditioned on the total family size until extinction being equal to n. Let $S_1^{(n)}, S_2^{(n)}$ and $S_3^{(n)}$ be the sizes of the three largest branches (in random order) from the centroid of a random tree in any

simply generated family. Then the invariance principle implies

$$n^{-1}(S_1^{(n)}, S_2^{(n)}, S_3^{(n)}) \to_d (S_1, S_2, S_3),$$

where the limit has density (11).

We now set about doing calculations based upon (11). We present these as exact calculations for the Brownian CRT, but as above they all have obvious interpretations as asymptotic distributions for random trees from any simply generated family. Some of the calculations make use of the indefinite integral formula

$$(12) \qquad \int y^{-3/2}(a - y)^{-3/2} \, dy = 2a^{-2}y^{-1/2}(a - y)^{-1/2}(2y - a), \qquad 0 < y < a.$$

Unexplained integrations below were done using MATHEMATICA. I do not have a good "geometric" explanation of why surprisingly many of the integrals can be done explicitly.

First consider the size S_1 of a typical branch B_1 from the centroid, that is, of a branch chosen uniformly from the three branches. Its density is obtained by integrating out x_2 in (11):

$$
\begin{aligned}
(13) \qquad f_{S_1}(x) &= \int_{1/2-x}^{1/2} \frac{1}{12\pi}\big(xx_2(1 - x - x_2)\big)^{-3/2} \, dx_2 \\
&= \frac{2}{3\pi}x^{-1/2}(1 - 2x)^{-1/2}(1 - x)^{-2}, \qquad 0 < x < \frac{1}{2},
\end{aligned}
$$

using (12).

Now consider size-biasing. Let S_1^* be the size of the branch from the centroid containing a randomly chosen point. Then

$$f_{S_1^*}(x) = \frac{xf_{S_1}(x)}{ES_1}$$

and $ES_1 = \frac{1}{3}$ by symmetry. So the density of S_1^* is

$$(14) \qquad f_{S_1^*}(x) = \frac{2}{\pi}x^{1/2}(1 - x)^{-2}(1 - 2x)^{-1/2}.$$

In particular, the mean size ES_1^* of the branch containing a randomly chosen vertex is

$$ES_1^* = \int_0^{1/2} xf_{S_1^*}(x) \, dx = 2^{1/2} - 1 \approx 0.414.$$

Now suppose that we order the sizes of the branches as $L_1 > L_2 > L_3$. Then the density is

$$
\begin{aligned}
(15) \qquad f_{L_1, L_2, L_3}(x_1, x_2, x_3) &= \frac{1}{2\pi}(x_1 x_2 x_3)^{-3/2} \\
&\text{on } \Big\{ \tfrac{1}{2} > x_1 > x_2 > x_3 > 0; \ \textstyle\sum x_i = 1 \Big\},
\end{aligned}
$$

by (11), with an extra factor $3! = 6$ coming from the number of possible reorderings. We can now compute the marginal density of the largest branch size L_1 on $\{\frac{1}{3} < x < \frac{1}{2}\}$,

$$
\begin{aligned}
(16) \qquad f_{L_1}(x) &= \int_{1-2x}^{(1-x)/2} \frac{1}{2\pi}(xx_3(1-x-x_3))^{-3/2}\, dx_3 \\
&= \frac{3x-1}{\pi x^2(1-x)^2(1-2x)^{1/2}},
\end{aligned}
$$

using (12) again. As remarked in [4], this shows that the maximal branch size is strongly biased toward the upper end of the interval $\left[\frac{1}{3}, \frac{1}{2}\right]$. One can calculate

$$
P(L_1 \le x) = \frac{\arctan 3^{-1/2} - \arctan(1-2x)^{1/2}}{6\pi} - \frac{(3x-1)(1-2x)^{1/2}}{\pi x(1-x)},
$$

$$
EL_1 = \frac{2\operatorname{arctanh} 3^{-1/2} - 2\arctan 3^{-1/2} - 3^{1/2}}{\pi} \approx 0.465,
$$

$$
\text{median}(L_1) \approx 0.479.
$$

Similarly we can consider the smallest branch size L_3. Its density on $\{0 < x_3 < \frac{1}{3}\}$ is

$$
\begin{aligned}
(17) \qquad f_{L_3}(x_3) &= \int_{(1-x_3)/2}^{\min(1-2x_3,1/2)} \frac{1}{2\pi}\big(x_1 x_3(1-x_1-x_3)\big)^{-3/2}\, dx_1 \\
&= \pi^{-1} x_3^{-3/2}(1-x_3)^{-2}(1-2x_3)^{-1/2} q(x_3),
\end{aligned}
$$

where

$$
q(x_3) = \begin{cases} 2x_3, & x_3 < \frac{1}{4} \\ x_3^{-1/2}(1-3x_3), & x_3 > \frac{1}{4}. \end{cases}
$$

The distribution is biased toward the lower end of the interval $[0, \frac{1}{3}]$, and numerically the median is 0.118.

The next calculations relate the centroid to the decomposition of Theorem 2. Recall that Z_1, Z_2 and Z_3 were three randomly chosen vertices of the Brownian CRT and that V was their branchpoint.

PROPOSITION 5.

(a) *The probability that V is the centroid is $3 \cdot 2^{-1/2} - 2 \approx 0.121$.*

(b) *The probability that the centroid lies on the spanning subtree joining Z_1, Z_2, Z_3 is $4 - 9 \cdot 2^{-3/2} \approx 0.818$.*

PROOF. (a) In the notation of Theorem 2, the condition for V to be the centroid is $\max(\Delta_1, \Delta_2, \Delta_3) \le \frac{1}{2}$. By exchangeability this has probability $1 - 3P\big(\Delta_1 > \frac{1}{2}\big)$. However, Δ_1 has density $x^{-1/2}/2$, so $P\big(\Delta_1 > \frac{1}{2}\big) = 1 - \sqrt{1/2}$.

(b) The only way that the centroid could not be in the spanning tree is if the event

(18)   all three of Z_1, Z_2, Z_3 lie in the same branch from the centroid

occurs. Conditional on the size S_1^* of the branch from the centroid which contains Z_1, the chance of event (18) equals $(S_1^*)^2$. Thus the probability in (b) equals

$$1 - E(S_1^*)^2 = 1 - \int_0^{1/2} x^2 f_{S_1^*}(x)\, dx,$$

for $f_{S_1^*}$ defined by (14). Evaluating the integral gives the result stated. $\square$

As a final calculation, recall that the median x of a random variable X can be interpreted as $\arg\min_x E|X - x|$. The analog holds for the centroid of a tree. Let's find the asymptotics of $\overline{D}_n$, the distance from the centroid to a uniform random vertex in an n-vertex tree from a simply generated family. Unlike previous results, here we need family-dependent scaling constants. It is perhaps easiest to work by comparison with the known result for the distance D_n from the *root* to a uniform random vertex. Here

(19) $$ED_n \sim \sqrt{\frac{\pi}{2}} \sigma^{-1} n^{1/2},$$

where σ depends on the family (in the Galton–Watson description, σ is the offspring standard deviation). This is [17, Theorem 4.6], and a derivation using the invariance principle is given in [3, equation (36)], where $\sqrt{\pi/2}$ is misprinted as $\sqrt{2/\pi}$. The recursive self-similarity result, Theorem 4, immediately gives

$$E\overline{D}_n \sim ED_n\, E\sqrt{S_1^*},$$

where S_1^* is the size-biased branch size (14). Evaluating the relevant integral gives $E\sqrt{S_1^*} = 2/\pi$ and hence

(20) $$E\overline{D}_n \sim \sqrt{\frac{2}{\pi}} \sigma^{-1} n^{1/2}.$$ $\square$

REFERENCES

[1] ALDOUS, D. J. (1991). The continuum random tree. I. *Ann. Probab.* **19** 1–28.
[2] ALDOUS, D. J. (1991). The continuum random tree II: An overview. In *Stochastic Analysis* (M. T. Barlow and N. H. Bingham, eds.) 23–70. Cambridge Univ. Press.
[3] ALDOUS, D. J. (1993). The continuum random tree III. *Ann. Probab.* **21** 248–289.
[4] ALDOUS, D. J. (1994). Triangulating the circle, at random. *Amer. Math. Monthly* **101** 223–233.

[5] BERTOIN, J. and PITMAN, J. (1994). Path transformations connecting Brownian bridge, excursion and meander. *Bull. Sci. Math.* To appear.

[6] BUCKLEY, F. and HARARY, F. (1990). *Distance in Graphs.* Addison-Wesley, New York.

[7] FALCONER, K. J. (1986). Random fractals. *Math. Proc. Cambridge Philos. Soc.* **100** 559–582.

[8] FALCONER, K. J. and GRIMMETT, G. R. (1992). On the geometry of random Cantor sets and fractal percolation. *J. Theoret. Probab.* **5** 465–486.

[9] GARDNER, M. (1987). *Time Travel and Other Mathematical Bewilderments.* Freeman, New York.

[10] GRAF, S. (1987). Statistically self-similar fractals. *Probab. Theory Related Fields* **74** 357–392.

[11] GRAF, S., MAULDIN, R. D. and WILLIAMS, S. C. (1988). The exact Hausdorff dimension in random recursive constructions. *Memo. Amer. Math. Soc.* **71** 1–121.

[12] HARRIS, T. E. (1956). First passage and recurrence distributions. *Trans. Amer. Math. Soc.* **73** 471–486.

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

[14] LE GALL, J.-F. (1989). Marches aléatoires, mouvements Brownien et processus de branchement. *Séminaire de Probabilités XXIII. Lecture Notes in Math.* **1372** 258–274. Springer, Berlin.

[15] MANDELBROT, B. (1983). *The Fractal Geometry of Nature.* Freeman, New York.

[16] MAULDIN, R. D. and WILLIAMS, S. C. (1986). Random recursive constructions: Asymptotic geometric and topological properties. *Trans. Amer. Math Soc.* **295** 325–346.

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

[18] MEIR, A. and MOON, J. W. (1987). On major and minor branches of rooted trees. *Canad. J. Math.* **3** 673–693.

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

[20] NEVEU, J. and PITMAN, J. (1989). The branching process in a Brownian excursion. *Séminaire de Probabilités XXIII. Lecture Notes in Math.* **1372** 248–257. Springer, Berlin.

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

[22] PATZSCHKE, N. and ZAHLE, U. (1990). Self-similar random measures IV—the recursive construction model of Falconer, Graf and Mauldin and Williams. *Math. Nachr.* **149** 285–302.

[23] POLYÁ, G. (1956). On picture-writing. *Amer. Math. Monthly* **51** 689–697.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720