A diffusion limit for a class of randomly-growing binary trees — **Source link** ↗

David Aldous, Paul C. Shields

Institutions: University of California, Berkeley, University of Toledo

Related papers:

- The Art of Computer Programming

- Evolution of random search trees

- Digital search trees revisited

- The Art of Computer Programming: Volume 3: Sorting and Searching

- Asymptotic behavior of the Lempel-Ziv parsing scheme and digital search trees

A Diffusion Limit for a Class
of Randomly-Growing Binary Trees

David Aldous[1] ★ and Paul Shields[2] ★★

[1] Department of Statistics, University of California, Berkeley CA 94720
[2] Department of Mathematics, University of Toled, 2801 W. Bancroft Street, Toledo OH 43606, USA

Summary. Binary trees are grown by adding one node at a time, an available node at height i being added with probability proportional to c^{-i}, $c > 1$. We establish both a "strong law of large numbers" and a "central limit theorem" for the vector $X(t) = (X_i(t))$, where $X_i(t)$ is the proportion of nodes of height i that are available at time t. We show, in fact, that there is a deterministic process $x_i(t)$ such that

$$\sum |X_i(t) - x_i(t)| \text{ converges to } 0 \text{ a.s.,}$$

and such that if $c > 2^{\frac{1}{2}}$,

$$Z_i^n(t) = 2^{n/2} \{ X_{n+1}(t\,c^n) - x_{n+i}(t\,c^n) \},$$

and $Z^n(t) = (Z_i^n(t))$, then $Z^n(t)$ converges weakly to a Gaussian diffusion $Z(t)$. The results are applied to establish asymptotic normality in the unbiased coin-tossing case for an entropy estimation procedure due to J. Ziv, and to obtain results on the growth of the maximum height of the tree.

1. Introduction

Nodes of the infinite, rooted binary tree T_∞ are labelled as binary strings, defined by labelling the root with the empty string ϕ, then proceeding by induction, labelling the two immediate successors of the node labelled with $b_1 b_2 \ldots b_h$ by $b_1 b_2 \ldots b_h 0$ and $b_1 b_2 \ldots b_h 1$. By a *tree* we shall mean a finite subtree of T_∞. We define and study limiting properties of a simple class of trees that grow randomly, one node at a time. Such trees arise naturally in Ziv's entropy estimation algo-

---

rithm, discussed below, which was the impetus for this paper, as well as in various models from physics and computer science.

For our purposes, a tree T is described by giving the set $A = A(T)$ of *available* nodes α, that is, nodes that are not in T but whose predecessors are in T. Trees evolve by selecting one available node and adding it to the tree to obtain a new tree; the chance that a given available node α is added being

$$c^{-h}(\alpha) / \sum_{\beta \in A} c^{-h}(\beta);$$

where $h = h(\alpha)$, the height of $\alpha = b_1 b_2 \ldots b_h$, and c is a parameter larger than 1. Initially the tree is empty. The first node added is the root ϕ. We write $\hat{T}(n)$ for the random tree with n nodes.

A standard trick in probability theory enables us to relate our tree growing process to another, more easily studied process. Think of time running continuously, starting at 0. Write $T(t)$ for the tree and $A(t)$ for the set of available nodes at time t. Given that α is available at time t, let the chance that α is added in the small time interval $[t, t + dt]$ be $c^{-h(\alpha)} dt$, independently for different α. This specifies a continuous-time tree process $T(t)$. Writing S_n for the time at which the n-th node is added, the process $T(S_n)$ is exactly the same as the process $\hat{T}(n)$ defined originally. But $T(t)$ has some independence properties which make it easier to analyze, and results for $T(t)$ lead to results for $\hat{T}(n)$ via the representation $\hat{T}(n) = T(S_n)$. We later make use, for example, of the fact that subtrees that grow from different available nodes are independent of each other. To be more precise, let $T_\alpha(s)$ denote the subset of $T(s)$ that contains only α and its successors. The continuous-time model assures that if α is available at time t the distribution on $T_\alpha(s)$, for $s > t$, is independent of the joint distributions on $T_\beta(s)$, $\beta \in A(t)$, $\beta \neq \alpha$. This independence property is an easy generalization of the standard "marking" argument for Poisson processes, (see, for example, [15, p. 38].)

Our results will be stated in terms of the vector of available node proportions $X(t) = (X_0(t), X_1(t), X_2(t), \ldots)$, where $X_h(t)$ is the proportion of nodes of height h that are available at time t. The vector $X(t)$ has nonnegative coordinates that sum to 1 and the process $(X(t))$ is a continuous-time Markov process on the countable state space of possible vectors of available node proportions with the transitions

$$(x_0, \ldots, x_i, x_{i+1}, \ldots) \rightarrow (x_0, \ldots, x_i - 2^{-i}, x_{i+1} + 2^{-i}, \ldots)$$

$$\text{at rate } 2^i x_i c^{-i}.$$

Although the vector $X(t)$ does not completely describe the tree $T(t)$, many interesting functionals of $T(t)$ are functionals of the vector $X(t)$. For example,

(a) $\sum 2^i X_i(t) - 1 = $ Number of nodes occupied at time t.

(b) $\sum_{i \geq 0} (i - 2) 2^i X_i(t) + 2 = $ Sum of heights of occupied nodes.   (1.1)

(c) $\text{Max}\{i: X_i(t) > 0\} - 1 = $ Height of tree at time t.

We write $\hat{X}(n) = (\hat{X}_0(n), \hat{X}_1(n), \hat{X}_2(n), \dots)$ for the vector of available node proportions for the discrete-time tree $\hat{T}(n)$ with n nodes, and note that $\hat{X}(n)$ is a discrete-time Markov chain. For convenience in stating and proving some of our results we shall extend the vectors $X(t)$ and $\hat{X}(n)$ to negative indices by defining $X_i(t) = \hat{X}_i(n) = 0$, for $i < 0$, for all n and t.

Our first fundamental result for $X(t)$, which says that $X(t)$ is almost deterministic for large t, is an analogue of the strong law of large numbers.

Theorem 1. *There is a deterministic vector process $x(t)$, which is "shift-periodic" in the sense that $x_i(t) = x_{i+1}(ct)$, for all i and t, such that*

$$\sum_{i=-\infty}^{i=\infty} |X_i(t) - x_i(t)|| \to 0 \text{ a.s.} \quad \text{as} \quad t \to \infty.$$

For the discrete-time process $\hat{X}(n)$ our result takes the following form.

Theorem 1 D. *There is a deterministic vector process $\hat{x}(n)$, which satisfies $\hat{x}_i(n) = \hat{x}_{i+1}(2n)$ for all i and n, such that*

$$\sum_{i=-\infty}^{i=\infty} |\hat{X}_i(n) - \hat{x}_i(n)| \to 0 \text{ a.s.} \quad \text{as} \quad n \to \infty.$$

Our second fundamental result is concerned with second-order properties, analogous to the central limit theorem. Here we found it necessary to assume that $c > 2^{\frac{1}{2}}$. A similar result holds for the discrete-time process $\hat{X}(n)$.

Theorem 2. *Let $Z_i^n(t) = 2^{n/2}(X_{n+i}(tc^n) - x_{n+1}(tc^n))$ and let $Z^n(t)$ denote the vector process $(Z_i^n(t): -\infty < i < \infty)$. For $c > 2^{\frac{1}{2}}$ the process $Z^n(t)$ converges in distribution (i.e., weakly) as $n \to \infty$ to a vector diffusion process $Z(t)$.*

The deterministic process $x(t)$ represents a scaling of the "drift" and could be replaced in the convergence statements of the theorems by the expected value, $p(t) = E(X(t))$. The "shift-periodic" property of $x(t)$ simplifies many of our arguments, however, and can be obtained from $p(t)$ by shifting time into the distant future and rescaling. The definition of $x(t)$ and elementary properties of $x(t)$ and $p(t)$ are given in Sect. 2. Bounds on the variance of $X_i(t)$ and proofs of Theorems 1 and 1 D are given in Sect. 3. The limiting diffusion process $Z(t)$ is defined and Theorem 2 is proved in Sect. 4. Limit properties of the various functionals defined in (1.1) are established in Sects. 5 and 6.

Various models for random trees have been studied in different contexts; two of these are described below and others are mentioned in our final section of remarks (sect. 7). Generally people study one model (i.e., without a free parameter c); our model is mathematically natural, and interesting because it exhibits at least 4 qualitative changes of behavior, namely,

$$c < 1, \quad c = 1, \quad 1 < c < 2^{\frac{1}{2}}, \quad 2^{\frac{1}{2}} < c.$$

Our results are also more precise than those previously obtained.

We close this section by describing the model that initiated our work and another model of interest.

Remark 1. The initial motivation for our work was the desire to obtain asymptotic normality properties for an entropy estimation algorithm due to Ziv, [17], which is based on an elegant data compression algorithm introduced by Ziv and Lempel, [18]. A binary sequence $\{b_n\}$ is parsed into blocks called words, where each new word is, by definition, the shortest consecutive sequence of symbols not seen in the past as a word. (This is a slight modification of the definition used in [17, 18].) Thus

$$11001010001000100\ldots \text{ parses into } 1, 10, 0, 101, 00, 01, 000, 100, \ldots$$

Ziv showed that if the sequence $\{b_n\}$ is a sample path of a stationary, ergodic process, and $W(n)$ is the number of words in the first n places, then $W(n) \cdot n^{-1} \log n$ converges almost surely, as $n \to \infty$, to the entropy of the process. Similar results for related algorithms are implicit in Pittel, [13]. Computer studies by Michael Barall and Shields indicated that if the process $\{b_n\}$ is independent and identically distributed then the number of words $W(n)$ is approximately normally distributed for large n. Attempts to prove this led to the ideas in this paper for, as Barall noted, the Ziv-Lempel parsing process grows a tree, where each time a new word α occurs the node α is added to the tree. The Ziv-Lempel parsing algorithm, when applied to unbiased coin tossing corresponds to the case when $c = 2$. This asymptotic normality suggests that $W(n)$ might be a useful test statistic for independence; power comparisons with various likelihood ratio tests are contained in [16].

Our diffusion limit result, Theorem 2, does lead to a proof of asymptotic normality of $W(n)$ in the case of unbiased coin-tossing (see Theorem 3 in Sect. 5). It is natural to conjecture that asymptotic normality holds for a large class of stationary processes satisfying some suitable mixing condition. But in view of the difficulty of even the simplest case, (i.e., the fair coin-tossing case we treat here), we are not optimistic about finding a general result. We believe the difficulty of our normality result is intrinsic (i.e., not just due to the limitations of our method); the estimator involves the $X_i(t)$ and as Theorem 2 says, their limiting behavior does involve an infinite-dimensional diffusion; we cannot imagine any way of proving asymptotic normality of the estimator without using the essence of Theorem 2 and thus having to face the issues we face.

Remark 2. We mention here an intuitively helpful "percolation" description for the continuous-time process that guided some of our thinking. With the edge connecting node α with its predecessor, associate a random variable ξ_α with exponential distribution, mean $c^{h(\alpha)}$, independent for different edges. Imagine water introduced at the base node, and percolating through the tree in such a way that node α is wetted a time ξ_α after its predecessor node is wetted. Then $T(t)$, the set of wet nodes at time t, is precisely our continuous-time process. Note, however, that standard percolation results, such as in Durrett, [6], are not helpful, due to nonhomogeneity.

2. Basic Properties of the Deterministic Approximation $x(t)$

Follow the progress of the tree along the left-most branch $0, 00, 000, \ldots$, and let $H(t)$ denote the height of the available node at time t. Then $H(t)$ is the continuous-time Markov chain with $H(0)=0$ and transitions $i \to i+1$ at rate c^{-i}. Let R_i be the first time t that $H(t)=i$, and let $p_i(t)=P(H(t)=i)$, so that

$$p_i(t) = E(X_i(t)) = P(R_i < t) - P(R_{i+1} < t).$$

Note that R_i has the same distribution as the sum $\sum_{j=0}^{i-1} c^j \xi_j$, where the ξ_j are independent and exponentially distributed random variables with mean 1.

Our deterministic approximation $x_i(t)$ to $X_i(t)$ is obtained by letting $H^*(t)$ denote the Markov chain on the positive *and* negative integers, started at $-\infty$ at time 0, with the transitions $i \to i+1$ at rate c^{-i} and defining $x_i(t)$ to be $P(H^*(t)=i)$. A precise definition of $H^*(t)$ is obtained by defining $R_i^* = \sum_{j=-\infty}^{i-1} c^j \xi_j$ and then defining $H^*(t)$ to be i iff $R_i^* < t \leq R_{i+1}^*$, so that

$$x_i(t) = P(H^*(t)=i) = P(R_i^* < t) - P(R_{i+1}^* < t).$$

Note that the vector $x(t)$ is nonnegative, sums to 1, and has the desired shift-periodic property $x_i(t) = x_{i+1}(ct)$.

The forward differential equations for the process $H(t)$ (and for $H^*(t)$) are

$$dz_i(t) = c^{-(i-1)} z_{i-1}(t) \, dt - c^{-i} z_i(t) \, dt. \tag{2.1}$$

A solution $z(t) = (z_i(t))$ to this system with

$$z_i \equiv 0, \quad i < 0, \qquad z_0(0) = 1, \qquad z_i(0) = 0, \quad i > 0,$$

is given by $z_i(t) = p_i(t)$. Similarly, $x(t) = (x_i(t))$ is a solution vector, satisfying

$$\sum_i x_i(t) \equiv 1, \qquad x_i(0+) = 0, \qquad \text{for all } i.$$

Informally, $x(t)$ is the solution that starts with unit mass located at "position" $-\infty$. Similarly, if we define the conditional probabilities

$$p_{j,i}(t) = P(H(u+t) = i \mid H(u) = j)$$

then $z^j(t) = (p_{j,i}(t))$ is a solution to the forward equations satisfying

$$p_{j,i}(t) \equiv 0, \quad i < j, \qquad p_{j,j}(0) = 1, \qquad p_{j,i}(0) = 0, \quad i > j.$$

The forward equations are linear, so that given a bounded set of values at time s, $\{z_i(s)\}$, there is a solution valid for $t > s$ of the form

$$z_j(t) = \sum_{i=-\infty}^{j} p_{i,j}(t-s) \, z_i(s). \tag{2.2}$$

It is clear that $p_i(t) \approx x_i(t)$, for large i and t, because R_i and R_i^* almost surely grow without bound as i grows and differ by the almost surely finite quantity R_0^*. A precise result is contained in the following lemma.

Lemma 1. *For $j \geq 0$; $t, u \geq 0$,*

    (a) $|p_j(t+u) - p_j(t)| \leq u c^{1-j}$,

    (b) $|p_j(t) - x_j(t)| \leq (1 - 1/c)^{-1} c^{1-j}$,

    (c) $\displaystyle\sum_{i=0}^{j} p_i(t) \leq \sum_{i=-\infty}^{j} x_i(t)$,

    (d) $\displaystyle\sum_{j=0}^{\infty} |p_j(t) - x_j(t)| \to 0$, as $t \to \infty$.

Proof. Part (a) follows from the mean value theorem and the forward equations. The probabilistic interpretation for $x(t)$ yields $x_j(t) = E(p_j(t - R_0^*))$, so that (a) implies

$$|p_j(t) - x_j(t)| \leq c^{1-j} E R_0^*,$$

giving (b). The proof of (c) is straightforward. Part (d) follows from part (b) and the fact that both $p_j(t)$ and $x_j(t) \to 0$ as $t \to \infty$, for fixed j.

Next we present the tail estimates we shall need. Keep in mind the intuitive idea that, at time t near c^i, the height $H(t)$, which is the height of a typical available node, is around i. First we estimate the size of $x_i(t)$ for t large relative to i.

Lemma 2. *There is a constant K, which depends only on c, such that for all $i \geq 0$,*

$$x_i(t) \leq P(R_{i+1}^* \geq t) \leq K \exp(-t c^{-(i+1)}),$$
$$p_i(t) \leq P(R_{i+1} \geq t) \leq K \exp(-t c^{-(i+1)}).$$

Proof. The first inequality is clear in each case, and $R_{i+1}^* \geq R_{i+1}$, so it is enough to bound $P(R_{i+1}^* > t)$. For any $\theta > 0$,

$$P(R_{i+1}^* > t) \leq e^{-\theta t} E \exp(\theta R_{i+1}^*),$$

so that if we choose $\theta = c^{-(i+1)}$ and use the representation $R_{i+1}^* = \sum_{j=-\infty}^{i} c^j \xi_j$, we obtain

$$E \exp(\theta R_{i+1}^*) = E \exp\left(c^{-1} \sum_{j=-\infty}^{0} c^j \xi_j\right) = K < \infty.$$

This proves the lemma.

Next is our basic upper tail estimate.

Lemma 3. $p_{i+k}(t c^i) \leq t^k c^{-k(k-1)/2}$, $i, k \geq 0$, $t \geq 0$. *The same result holds for $x_{i+k}(t c^i)$.*

Proof. From the relation $p_j(t) = P(R_j < t) - P(R_{j+1} \le t)$ we get

$$p_{i+k}(tc^i) \le P(R_{i+k} \le tc^i) = P\left(\sum_{j=0}^{i+k-1} c^j \xi_j \le tc^i\right)$$

$$\le P\left(\sum_{j=i}^{i+k-1} c^j \xi_j \le tc^i\right) \le P\left(\sum_{j=0}^{k-1} c^j \xi_j \le t\right)$$

$$\le P(\xi_j \le tc^{-j}; \ 0 \le j \le k-1)$$

which gives

$$p_{i+k}(tc^i) \le \prod_{0 \le j \le k-1} P(\xi_j \le tc^{-j}) \le \prod_{0 \le j \le k-1} tc^{-j} = t^k c^{-k(k-1)/2},$$

proving the lemma.

The above tail estimate lemmas guarantee that many formal manipulations with various series in the x_i and p_i can be rigorously justified. For example, the forward equations (2.1) show that the formal derivative of

$$h(t) = \sum_{i=-\infty}^{\infty} 2^i x_i(t) \quad \text{is} \quad h'(t) = \sum_{i=-\infty}^{\infty} 2^i c^{-i} x_i(t).$$

The tail estimate lemmas guarantee uniform convergence of both series on any compact subinterval of $(0, \infty)$, so that such termwise differentiation is indeed valid. We generally omit details of such convergence proofs in our subsequent discussion.

3. Variance and first order results for $X(t)$

We first establish our basic estimate of the variance of $X_j(t)$, the fraction of nodes of depth j that are available at time t. We make use of our earlier definition

$$p_{k,j}(t) = P(H(u+t) = j \mid H(u) = k),$$

noting that we can rescale to obtain $p_{i,j}(u) = p_{j-i}(uc^{-i})$.

Lemma 4. $\text{var } X_j(t) \le 2^{-j} \left\{ p_j(t) + \sum_{k=1}^{j} 2^{j-k} \text{var } p_{k,j}(t - R_k) \right\}.$

Proof. Let $A_\alpha(t)$ be the event that a node α is available at time t. Fix two nodes, α and β, at height j and suppose their highest common ancestor, γ, is at height $k-1$. Let τ be the time at which γ is filled. Then

$$P(A_\alpha(t) \mid \tau = s) = p_{k,j}(t - s),$$

$$P(A_\alpha(t) \text{ and } A_\beta(t) \mid \tau = s) = p_{k,j}^2(t - s).$$

Note our use of the continuous-time model to obtain these results, for, conditioned on the value of $\tau < t$, $A_\alpha(t)$ and $A_\beta(t)$ are independent. The conditional variance formula then gives

$$\text{cov}(A_\alpha(t), A_\beta(t)) = \text{var } p_{k,j}(t-\tau).$$

Now $X_j(t) = 2^{-j} \sum_{h(\alpha)=j} 1_{A_\alpha(t)}$, where 1_A denotes the indicator function of a set A. There are at most $2^j \cdot 2^{j-k}$ pairs of nodes of height j whose highest common ancestor is at height $k-1$. A direct calculation then establishes the lemma, since $\text{var } A_\alpha(t) \leq p_j(t)$ and τ has the distribution of R_k.

Lemma 4 gives us the following variance bounds.

Lemma 5. *There is a constant B that depends only on c such that*

(a) *If $c = 2^{\frac{1}{2}}$ then $\text{var } X_j(t) \leq Bj2^{-j}$.*

(b) *If $c < 2^{\frac{1}{2}}$ then $\text{var } X_j(t) \leq Bc^{-2j}$.*

(c) *If $c > 2^{\frac{1}{2}}$ then $\text{var } X_j(tc^j) \leq B \cdot 2^{-j} \exp(-t/2c)$.*

Proof. For any function f, $0 \leq f \leq 1$, any $t_0 > 0$, and any random variable R, it is not hard to show that

$$\text{var } f(R) \leq \max_{u \geq t_0} [f'(u)]^2 \text{ var } R + P(R \leq t_0). \tag{3.1}$$

We first apply this with

$$t_0 = \infty, \quad f(R) = p_{j-k}(R), \quad R = (t - R_k)c^{-k},$$

along with Lemma 1(a) to obtain

$$\text{var } p_{k,j}(t - R_k) = \text{var } p_{j-k}((t - R_k)c^{-k}) \leq c^{2-2j} \text{ var } R_k.$$

Lemma 4 then yields the bound

$$\text{var } X_j(t) \leq 2^{-j} \left\{ 1 + \sum_{k=1}^{j} 2^{j-k} c^{2-2j} \text{ var } R_K \right\}.$$

Now $\text{var } R_K = \sum_{i=0}^{k-i} c^{2i}$, and simple algebra leads to parts (a) and (b) and the following weak version of (c).

(c') *If $c > 2^{\frac{1}{2}}$ then $\text{var } X_j(t) \leq B 2^{-j}$, $t \geq 0$.*

To improve (c') to (c) we apply (3.1) with $f(R) = \text{var } p_{j-k}(R)$, $t_0 = \frac{1}{2} tc^{j-k}$, and $R = tc^{j-k} - c^{-k} R_k$ to obtain

$$\text{var } p_{k,j}(tc^j - R_k) = \text{var } p_{j-k}(R)$$
$$\leq [\max_{u \geq t_0} f'(u)]^2 c^{-2k} \text{ var } R_k + P(R_k \geq \tfrac{1}{2} tc^j).$$

We then upper bound var R_k by c^{2k} and use the bounds

$$p_i'(t) \leq c^{1-i} P(R_i > t) = O(c^{-i} \exp(-tc^{-(i+1)})),$$
$$P(R_k \geq t) = O(\exp(-tc^{-(k+1)})),$$

which come from the proof of Lemma 2(a), to obtain

$$\operatorname{var} p_{k,j}(tc^j - R_k) = O(c^{-2(j-k)} \exp(-tc^{-1})) + O(\exp(-\tfrac{1}{2}c^{-1}tc^{j-k})).$$

In view of (c'), we may assume that $t > 2c$ and hence bound both terms by $O(c^{-2(j-k)} \exp(-t/2c))$. This bound together with the Lemma 2 bound $p_j(tc^j) = O(\exp(-t/c))$, can then be used in the inequality of Lemma 4 to obtain the desired result (c).

We shall later need two extensions of Lemma 5, which we state as the next two lemmas.

Lemma 5A. *Let $V_j(t)$ be the fraction of nodes of height j that have become available before time t. Then the bounds of Lemma 5 hold for V_j in place of X_j.*

Proof. Just replace the phrase "available at time t" in the proof of Lemma 4 by "available before time t" and repeat the arguments of Lemma 4 and Lemma 5.

Lemma 5B. *If $c > 2^{\frac{1}{2}}$ then there is a bound B that depends only on c such that*

$$\sum_{k=-\infty}^{j} 2^{j-k} \operatorname{var} p_{k,j}(t - R_k^*) \leq B.$$

Proof. This is proved by considering the process $X^*(t)$ obtained by waiting the random time R_0^* until the root node becomes available, that is, $X^*(t) = X(t - R_0^*)$ if $t \geq R_0^*$, $X^*(t) = 0$ otherwise.

We shall also need another consequence of Lemma 5, which we state as follows.

Lemma 5C. *There are positive constants B and α, which depend only on c, such that*

$$E \sum_j |X_j(t) - p_j(t)| \leq B \cdot t^{-\alpha} \log t, \qquad t \geq e.$$

Proof. Since $E|X_j(t) - p_j(t)|$ is bounded by $2p_j(t)$ and by $[\operatorname{var} X_j(t)]^{\frac{1}{2}}$, we can write

$$E \sum_j |X_j(t) - p_j(t)| \leq \sum_{j \leq j_0} 2p_j(t) + \sum_{j > j_0} [\operatorname{var} X_j(t)]^{\frac{1}{2}}$$

for any $j_0 = j_0(t)$. Suppose for the moment that $c \geq 2^{\frac{1}{2}}$ and let $\gamma = \log c$. Choose α so that $\frac{1}{4} < \alpha\gamma < \frac{1}{2}$. Given t, choose j_0 so that $t \leq c^{2j_0} \leq c^2 t$. Lemmas 2 and 5 then give

$$\sum_{j \leq j_0} 2p_j(t) \leq 2j_0 K \cdot \exp(-tc^{-(j_0+1)}) = O(t^{-\alpha} \log t),$$

$$\sum_{j > j_0} [\operatorname{var} X_j(t)]^{\frac{1}{2}} \leq 4B^{\frac{1}{2}} 2^{-j_0/2} = O(t^{-\alpha}),$$

which establishes Lemma 5C in the case $c \geq 2^{\frac{1}{2}}$. The proof for $c < 2^{\frac{1}{2}}$ is similar and even easier for one can take $\alpha = \frac{1}{2}$.

Proof of Theorem 1. As noted in Lemma 1 we have

$$\lim_{t \to \infty} \sum_i |p_i(t) - x_i(t)| = 0,$$

so it suffices to show that

$$\sum_j |X_j(t) - p_j(t)| \to 0 \text{ a.s.} \quad \text{as} \quad t \to \infty.$$

Let $\theta_u \downarrow 1$. Lemma 5C and the Borel-Cantelli principle imply that a.s.

$$\sum_j |X_j(\theta_u^n) - p_j(\theta_u^n)| \to 0, \quad \text{as} \quad n \to \infty, \quad \text{for each } u \geq 1.$$

Fix ω where this holds and define

$$A(t) = \max_j |\sum_{i \leq j} X_i(t) - \sum_{i \leq j} p_i(t)|.$$

We will show that $A(t) \to 0$ a.s. as $t \to \infty$. Towards this end note first that $\sum_{i \leq j} X_i(t)$ and $\sum_{i \leq j} p_j(t)$ are increasing in j. Fix u, and define $m = m(t)$ by $\theta_u^m \leq t \leq \theta_u^{m+1}$. Then $A(t)$ is bounded by the sum of the three terms

$$\max_j |\sum_{i \leq j} X_i(\theta_u^{m+1}) - \sum_{i \leq j} X_i(\theta_u^m)|,$$

$$\max_j |\sum_{i \leq j} X_i(\theta_u^m) - \sum_{i \leq j} p_i(\theta_u^m)|,$$

$$\max_j |\sum_{i \leq j} p_i(\theta_u^m) - \sum_{i \leq j} p_i(\theta_u^{m+1})|.$$

Letting $t \to \infty$ and using the fact that $\sum_j |X_i(\theta_u^m) - p_i(\theta_u^m)|$ approaches 0 for each $u \geq 1$, we obtain

$$\limsup_{t \to \infty} A(t) \leq 2 \cdot \limsup_{m \to \infty} \max_j |\sum_{i \leq j} p_i(\theta_u^{m+1}) - \sum_{i \leq j} p_i(\theta_u^m)|.$$

As $\theta_u \downarrow 1$, the right side tends to 0, from the limiting shift-periodicity and continuity of the $p_i(t)$.

The fact that $A(t)$ goes to 0, together with the limiting shift-periodic property of $p_i(t)$ implies that $\sum_j |X_j(t) - p_j(t)|$ goes to 0. (This follows by shifting and using the analysis principle that $\lim_n \sum_{i=0}^{\infty} |p(n, i) - \hat{p}(n, i)| = 0$, if $p_n = (p(n. i))$ and $\hat{p}_n = (\hat{p}(n, i))$ are sequences of probability vectors such that

$$\lim_n \max_j |\sum_{i \leq j} [p(n, i) - \hat{p}(n, i)]| = 0,$$

and if there is a probability vector $q = (q(i))$ such that

$$\lim_n \sum_{i \leq j} [\hat{p}(n, i) - q(i)] = 0,$$

for each j.) This completes the proof of Theorem 1.

Proof of Theorem 1 D. Define

$$h(t) = \sum_{i = -\infty}^{\infty} 2^i x_i(t).$$

The forward equations, (2.1), yield the derivative $h'(t) = \sum 2^i c^{-i} x_i(t)$, hence $h(t)$ is an increasing function. Let $N(t) = \sum_{i=0}^{\infty} 2^i X_i(t)$, so that $N(t) - 1$ is the number of nodes filled at time t. We first prove

$$N(t)/h(t) \to 1, \quad \text{a.s.} \quad \text{as } t \to \infty. \tag{3.2}$$

This is proved as follows. First note that the shift-periodic property $x_i(t) = x_{i+1}(ct)$ gives the corresponding property for h, $h(ct) = 2h(t)$. Choose $i = i(t)$ such that $c^i \leq t < c^{i+1}$. We first show that there is a $\gamma > 0$ such that

$$2^{-i} E |N(t) - EN(t)| = O(t^{-\gamma}), \quad \text{as } t \to \infty.$$

Since $EN(t) = \sum 2^i p_i(t)$ this reduces to showing

$$\sum_{j=\infty}^{\infty} 2^j E |X_{i+j}(t) - p_{i+j}(t)| = O(t^{-\gamma}).$$

Since $E |X_j(t) - p_j(t)| \leq 2 p_j(t)$, we can write

$$\sum_{j=-\infty}^{\infty} 2^j E |X_{i+j}(t) - p_{i+j}(t)|$$

$$\leq 2^{j_0} \sum_j E |X_{i+j}(t) - p_{i+j}(t)| + \sum_{j > j_0} 2^j p_{i+j}(t),$$

for any j_0. If $j_0 = [di]$, where $[u]$ denotes the greatest integer in u, then for a given c we can choose $d > 0$ and use Lemma 5C to bound the first term by $O(t^{-\gamma})$. Lemma 3 then gives the same bound for the second term. A similar (but simpler) argument shows that $2^{-i} |EN(t) - h(t)| \to 0$. The two facts,

$$2^{-i} E |N(t) - EN(t)| = O(t^{-\gamma}), \quad \text{and} \quad \lim_t 2^{-i} |EN(t) - h(t)| = 0,$$

imply that for each $\theta > 1$,

$$N(\theta^n)/h(\theta^n) \to 1, \quad \text{a.s.,} \quad \text{as } n \to \infty.$$

The monotonicity of $N(t)$ and $h(t)$, together with the shift-periodic property of h then imply (3.2).

Now we let $s(u)$ be the inverse function of the increasing function $h(t)$, so that, in particular, $h(s(n)) = n$. Next we define

$$S_n = \min\{t: N(t) = n+1\},$$

that is, S_n is the time until n nodes are filled. The shift property, $h(ct) = 2h(t)$, together with (3.2), then leads to the following inverse form of (3.2).

$$S_n/s(n) \to 1, \quad \text{a.s.,} \quad \text{as } n \to \infty. \tag{3.3}$$

To prove this we define integers $a(n)$, $b(n)$ and numbers u_n, v_n such that

$$S_n = c^{a(n)} u_n, \quad s(n) = c^{b(n)} v_n, \quad 1 \leq u_n, v_n \leq c.$$

Since $N(t)/h(t) \to 1$, a.s., and u_n and v_n are bounded, we can drop to a subsequence and relabel to obtain

$$\frac{N(S_n)}{h(S_n)} \to 1, \quad u_n \to u, \quad \text{and} \quad v_n \to v \quad \text{as } n \to \infty.$$

The shift-periodic property $h(ct) = 2h(t)$, together with $N(S_n) = n+1$ and $h(s(n)) = n$, then gives

$$\frac{h(S_n)}{h(s(n))} = 2^{a(n)-b(n)} \frac{h(u_n)}{h(v_n)} \to 1, \quad \text{as } n \to \infty.$$

However, $h(u_n)/h(v_n) \to h(u)/h(v)$, so that $a(n) - b(n)$ is eventually a constant, say k, and $2^k = h(v)/h(u)$. Since h is strictly increasing we must have $v = c^k u$ and hence

$$\frac{S_n}{s(n)} = c^{a(n)-b(n)} \frac{u_n}{v_n} \to \frac{c^k u}{v} = 1.$$

This proves (3.3). Note that since $1 \leq u, v \leq c$, the integer k can only have the values -1, 0, 1. Thus, if we require only that $c^{-1} \leq u_n, v_n \leq c$, we can if necessary replace u_n by $c^{-1} u_n$ or v_n by $c^{-1} v_n$ and thereby force k to be 0 and $a(n)$ and $b(n)$ to be eventually equal. We shall use this observation in the next paragraph.

We now set $\hat{x}(n) = x(s(n))$ and $\hat{X}(n) = X(S_n)$, so that $\hat{X}(n)$ is the discrete-time process. The shift properties, $x_i(t) = x_{i+1}(ct)$ and $h(ct) = 2h(t)$, imply the shift-periodicity of $\hat{x}$, i.e., $\hat{x}_i(n) = \hat{x}_{i+1}(2n)$. To complete the proof of Theorem 1D we must show that

$$\lim_{n \to \infty} \sum |\hat{X}_i(n) - \hat{x}_i(n)| = \lim_{n \to \infty} \sum |X_i(S_n) - x_i(s(n))| = 0, \quad \text{a.s.}$$

By Theorem 1 it is enough to show that

$$\lim_{n \to \infty} \sum |x_i(S_n) - x_i(s(n))| = 0, \quad \text{a.s.}$$

Using the notation of the proof of (3.3) and the observation at the end of that proof we can assume that u_n and v_n both go to the same value u and

that $a(n) = b(n)$, for n sufficiently large. If n is large enough we use the shift property of x to obtain

$$\sum |x_i(S_n) - x_i(s(n))| = \sum |x_{i-a(n)}(u_n) - x_{i-a(n)}(v_n)|,$$

which approaches 0 as $n \to \infty$ since u_n and v_n both approach the same value u. This completes the proof of Theorem 1 D.

For use in Sect. 6, note that the shift-periodic property $h(ct) = 2h(t)$ implies that $\log_2 h(t) - \log_c t$ is bounded as $t \to \infty$, and hence the inverse function $s(n)$ satisfies

$$\log_2 n - \log_c s(n) \quad \text{is bounded as } n \to \infty. \tag{3.4}$$

4. Second-order Structure

We start by recalling some standard theory. For each $t \geq 0$ let $A(t)$ and $\sigma(t)$ be nonrandom $d \times d$ matrices, varying continuously with t. Let $W(t) = (W_i(t): 1 \leq i \leq d)$ be d-dimensional Brownian motion. the stochastic differential equation

$$dZ(t) = A(t) Z(t) dt + \sigma(t) dW(t)$$

together with an initial distribution $Z(0)$, defines a d-dimensional diffusion, (i.e., a continuous-path, Markov process) which is also a Gaussian process, provided $Z(0)$ is Gaussian. Writing $\Delta X = X(t + \Delta t) - X(t)$, we have

$$\begin{aligned} E(\Delta Z | Z(t) = z) &\sim A(t) z \Delta t \\ E(\Delta Z_i \Delta Z_j | Z(t) = z) &\sim (\sigma(t) \sigma^T(t))_{ij} \Delta t, \end{aligned} \tag{4.1}$$

where we use the symbol $\sim$ to indicate equality up to terms of order $o(\Delta t)$ (in probability) as $\Delta t \to 0$ for fixed t.

We quote a standard result.

Proposition 1. *Let $A(t)$, $\sigma(t)$, and $Z(t)$ be as above. For each n, let $Z^n(t)$ be an R^d-valued Markov process. Suppose*

   (i) $E(\Delta Z^n | Z^n(t) = z) \sim A(t) z \Delta t$,

   (ii) $E(\Delta Z_i^n \Delta Z_j^n | Z^n(t) = z) \sim B_{ij}^n(t) \Delta t$,

where $B^n(t)$ is a random matrix process such that for t_0 fixed

$$\sup_{t \leq t_0} |B_{ij}^n(t) - (\sigma(t) \sigma^T(t))_{ij}| \underset{p}{\to} 0 \quad \text{as } n \to \infty.$$

   (iii) *There exist constants $\delta_n \downarrow 0$ such that the jumps $|Z^n(t) - Z^n(t-)|$ are bounded by δ_n.*

   (iv) $Z^n(0) \overset{\mathscr{D}}{\to} Z(0)$ *as $n \to \infty$.*

Then $Z^n \overset{\mathscr{D}}{\to} Z$ as $n \to \infty$, in the usual sense of weak convergence of processes.

The underlying idea is that $Z(t)$ is determined by the properties (4.1) and by having continuous paths; the hypotheses of Proposition 1 ensure that any possible weak limit process must have these properties and therefore must be Z; so it is only necessary to prove tightness of (Z^n). A treatment of these ideas can be found in [8, Sect. II.3].

Our main result, Theorem 2, uses an infinite-dimensional version of Proposition 1. Let $X(t)$ be the vector of available node proportions for our random continuous-time tree $T(t)$ and let $x(t)$ be its deterministic approximation defined in Sect. 2. A straightforward calculation using the infinitesimal transition rates for $X(t)$ yields the following conditional means and covariances.

(a) $E(\Delta X_i | X(t)) \sim c^{-(i-1)} X_{i-1}(t) \Delta t - c^{-i} X_i(t) \Delta t,$

(b) $E([\Delta X_i]^2 | X(t)) \sim (2c)^{-i} X_i(t) \Delta t + (2c)^{-(i-1)} X_{i-1}(t) \Delta t,$ (4.2)

(c) $E(\Delta X_i \Delta X_{i-1} | X(t)) \sim (2c)^{-(i-1)} X_{i-1}(t) \Delta t,$

(d) $E(\Delta X_i \Delta X_j | X(t)) \sim 0,$ if $|j-i| > 1.$

Now define the infinite-dimensional process $Z^n(t) = (Z_i^n(t))$ by

$$ Z_i^n(t) = 2^{n/2} \{ X_{n+i}(t\,c^n) - x_{n+i}(t\,c^n) \}, \qquad -\infty < i < \infty. $$

Note that we center by $x_{n+i}(t\,c^n)$, instead of $p_{n+i}(t\,c^n) = E X_{n+i}(t\,c^n)$. However, from Lemma 1(b),

$$ 2^{n/2} |x_{n+i}(t\,c^n) - p_{n+i}(t\,c^n)| \leq (c-1)\,c^{-i}(2^{\frac{1}{2}}\,c^{-1})^n, $$

which goes to 0 as $n \to \infty$, so the difference is asymptotically negligible.

Of course, Z^n inherits the Markov property of X. The conditional mean and covariance results, (4.2), and the fact that the x_i satisfy the forward equations, (2.1), combine to give the following.

(a) $E(\Delta Z_i^n | Z^n(t) = z) \sim \{ c^{-(i-1)} z_{i-1} - c^{-i} z_i \} \Delta t.$ (4.3)

(b) $E([\Delta Z_i^n]^2 | Z^n(t) = z) \sim (2c)^{-i} X_{n+i}(t\,c^n) \Delta t + (2c)^{-(i-1)} X_{n+i-1}(t\,c^n) \Delta t.$

(c) $E(\Delta Z_i^n \Delta Z_{i-1}^n | Z^n(t) = z) \sim -(2c)^{-(i-1)} X_{n+i-1}(t\,c^n) \Delta t.$

(d) $E(\Delta Z_i^n \Delta Z_j^n | Z^n(t) = z) \sim 0$ for $|j-i| > 1.$

Theorem 1 tells us that for large n, $X_{n+i}(t\,c^n) \approx x_{n+i}(t\,c^n) = x_i(t)$. These calculations, and the finite-dimensional result, Proposition 1, strongly suggest the following result, which is the explicit form of Theorem 2 that we shall prove.

Theorem 2. *Assume that $c > 2^{\frac{1}{2}}$. Then $Z^n(t)$ converges weakly to $Z(t)$ as $n \to \infty$, where $Z(t) = (Z_i(t); -\infty < i < \infty)$ is an infinite-dimensional Gaussian diffusion specified by $Z_i(0) = 0$ and the following four conditions.*

(a) $E(\Delta Z_i | Z(t) = z) \sim \{ c^{-(i-1)} z_{i-1} - c^{-i} z_i \} \Delta t,$

(b) $E([\Delta Z_i]^2 | Z(t) = z) \sim \{ (2c)^{-i} x_i(t) + (2c)^{-(i-1)} x_{i-1}(t) \} \Delta t,$

(c) $E(\Delta Z_i \Delta Z_{i-1} | Z(t) = z) \sim -(2c)^{-(i-1)} x_{i-1}(t) \Delta t,$

(d) $E(\Delta Z_i \Delta Z_j | Z(t) = z) \sim 0$ *for* $|j-i| > 1.$

To outline the structure of the proof, recall that to prove $Z^n \overset{\mathscr{D}}{\to} Z$ on a metric space (S, d) it suffices to construct $Z^{n,\,i_0}$, $Z^{\infty,\,i_0}$ such that

(a) $Z^{\infty,\,i_0} \overset{\mathscr{D}}{\to} Z$, as $i_0 \to \infty$,

(b) $Z^{n,\,i_0} \overset{\mathscr{D}}{\to} Z^{\infty,\,i_0}$, as $n \to \infty$ for each i_0,

(c) $\displaystyle\lim_{i_0 \to \infty} \limsup_n P(d(Z^{n,\,i_0}, Z^n) > \varepsilon) = 0$, for each $\varepsilon > 0$. $\qquad$ (4.4)

We shall use this technique by defining $Z^{\infty,\,i_0}$ to be the process satisfying the four conditions (a)–(d) of Theorem 2, but with $Z_i(t) = 0$, for $i \leq i_0$. Later we shall define $Z^{n,\,i_0}$ as the normalised version of the tree process which has been "controlled" by making the nodes at height $n - i_0$ become available at deterministic times.

We start by considering the question of the existence of the process Z. Translating (a)–(d) into the language of stochastic differential equations gives

$$dZ_i(t) = c^{-(i-1)} Z_{i-1}(t)\,dt - c^{-i} Z_i(t)\,dt$$
$$+ (2c)^{-\frac{1}{2}(i-1)} x_{i-1}^{\frac{1}{2}}(t)\,dW_{i-1}(t) - (2c)^{-\frac{1}{2}i} x_i^{\frac{1}{2}}(t)\,dW_i(t) \qquad (4.5)$$

where the $W_i(t)$ are independent Brownian motions. Using formula (2.2) we can just write down a solution of the stochastic equation, namely,

$$Z_j(t) = \sum_{i=-\infty}^{j} p_{i,j}(t)\,Z_i(0)$$

$$+ \sum_{i=-\infty}^{j} \int_0^t p_{i,j}(t-s)\,\{(2c)^{-\frac{1}{2}(i-1)} x_{i-1}^{\frac{1}{2}}(s)\,dW_{i-1}(s)$$

$$- (2c)^{-i/2} x_i^{\frac{1}{2}}(s)\,dW_i(s)\}. \qquad (4.6)$$

We can make a precise statement by truncation. Fix $i_0 < \infty$. Interpret the stochastic differential equations (4.5) as equations for $Z_i(t)$, $i \geq -i_0$, by putting $W_i(t) \equiv 0$ and $Z_i(t) \equiv 0$, for $i < -i_0$, and $Z_i(0) = 0$, for all i. Then the infinite series (4.6) does indeed form a solution; later we will refer to this solution as $Z_j^{\infty,\,i_0}(t)$.

Thus the only real issue in proving that the infinite series (4.6) makes sense and defines a solution of the full set of stochastic differential equations (4.5), with the initial condition $Z(0) = 0$, say, is to prove convergence of the sum (4.6). Recall that $\int f(s)\,dW(s)$ has variance $\int f^2(s)\,ds$. Thus we can group the two terms in $dW_i(s)$ and use independence of the W_i to obtain

$$EZ_j^2(t) = \sum_{i=-\infty}^{j} \int_0^t \{p_{i+1,j}(t-s) - p_{i,j}(t-s)\}^2 (2c)^{-i} x_i(s)\,ds. \qquad (4.7)$$

It is enough to prove the convergence of this sum. The task is not hard if we think probabilistically, for if Theorem 2 is correct, then $Z(t)$ is a limit of

rescaled $X_n(t')$ and therefore EZ_j^2 must be a limit of rescaled $\mathrm{var}(X_n(t'))$, so formula (4.7) must be a disguised and rescaled version of the bound on the variance of $X_n(t)$ which we derived in Lemma 4. Let us show how.

As in Sect. 2, let $H^*(t)$ be the Markov chain on the integers, started at $-\infty$ at time $t=0$, with transitions $i \to i+1$ at rate c^{-i}, and let R_i^* be the waiting time until the process reaches i. We showed in Lemma 5B that if $c > 2^{\frac{1}{2}}$ then there is a constant B, independent of j, such that

$$\sum_{k=-\infty}^{j} 2^{j-k} \mathrm{var}\, p_{k,j}(t-R_k^*) \leq B. \tag{4.8}$$

We shall use this to show that the sum (4.7) indeed converges.

The next lemma provides the link between (4.7) and (4.8).

Lemma 6. *Let σ be a state of a continuous-time Markov chain $V(t)$. Let T_0 be the first hitting time on σ and T_1 the time at which σ is exited (so that $T_1 - T_0$ is exponentially distributed). Let A be an event such that $P(A \mid T_1, V(u), u \leq T_1) = f(T_1)$, for some function $f(t)$. Let $g(t) = P(A \mid V(t) = \sigma)$. Suppose f and g are differentiable. Then*

$$\mathrm{var}\, f(T_1) \geq E\{f(T_1) - g(T_1)\}^2.$$

Proof. We shall prove the stronger identity

$$E(\mathrm{var}(f(T_1) \mid T_0)) = E\{f(T_1) - g(T_1)\}^2.$$

Without loss of generality we can take $T_1 - T_0$ to have mean 1 so that $g'(t) = g(t) - f(t)$.

We shall do some stochastic calculus with the martingale $M_t = P(A \mid V(t))$. (See [7] for a treatment of the martingale ideas used here.) We have $M_t = g(t)$ on $T_0 \leq t < T_1$, $M_{T_1} = f(T_1)$ and

$$d\langle M, M \rangle_t = (f(t) - g(t))^2\, dt, \tag{4.9}$$

$$d(f(t)\, M_t) = f'(t)\, M_t\, dt + f(t)\, dM_t. \tag{4.10}$$

Integration over (T_0, T_1) gives

$$Ef^2(T_1) - Eg^2(T_0) = E(M_{T_1}^2 - M_{T_0}^2) = E \int_{T_0}^{T_1} d\langle M, M \rangle_t$$

$$= E \int_{T_0}^{T_1} (g(t) - f(t))^2\, dt \quad \text{(by (4.9))}$$

$$= E \int_{T_0}^{T_1} g'(t)\, g(t)\, dt - E \int_{T_0}^{T_1} f(t)\, g'(t)\, dt \quad \text{(since } g' = g - f\text{)}$$

$$= \tfrac{1}{2} E\{g^2(T_1) - g^2(T_0)\} - Ef(T_1)\, g(T_1)$$

$$+ Ef(T_0)\, g(T_0) + E \int_{T_0}^{T_1} f'(t)\, g(t)\, dt \tag{4.11}$$

and

$$E(f^2(T_1) - f(T_0) g(T_0)) = E(f(T_1) M_{T_1} - f(T_0) M_{T_0})$$

$$= E \int_{T_0}^{T_1} f'(t) M_t \, dt \quad \text{(by (4.10))}$$

$$= E \int_{T_0}^{T_1} f'(t) g(t) \, dt. \tag{4.12}$$

Subtracting (4.12) from (4.11) and rearranging gives

$$Ef(T_1) g(T_1) = \tfrac{1}{2} \{ E g^2(T_1) + E g^2(T_0) \}$$

so that

$$E\{f(T_1) - g(T_1)\}^2 = Ef^2(T_1) - E g^2(T_0). \tag{4.13}$$

The proof is completed by first noting that $g(T_0) = M_{T_0} = E(M_{T_1} | T_0)$ $= E(f(T_1) | T_0)$, so that $\mathrm{var}(f(T_1) | T_0) = E(f^2(T_1) | T_0) - g^2(T_0)$, and hence

$$E \, \mathrm{var}(f(T_1) | T_0) = Ef^2(T_1) - E g^2(T_0).$$

This final equality combines with (4.13) to establish the desired result

$$E(\mathrm{var}(f(T_1) | T_0)) = E\{f(T_1) - g(T_1)\}^2.$$

This completes the proof of Lemma 6.

We apply the lemma with

$$V(t) = H^*(t); \qquad \sigma = i; \qquad A \text{ the event } ``H^*(t) = j"$$

so that

$$T_0 = R_i^*, \qquad T_1 = R_{i+1}^*, \qquad f(s) = p_{i+1, j}(t-s), \qquad g(s) = p_{i, j}(t-s).$$

The conclusion of the lemma is

$$\mathrm{var} \, p_{i+1, j}(t - R_{i+1}^*) \geqq E\{p_{i+1, j}(t - R_{i+1}^*) - p_{i, j}(t - R_{i+1}^*)\}^2$$

$$= \int_0^t \{p_{i+1, j}(t-s) - p_{i, j}(t-s)\}^2 c^{-i} x_i(s) \, ds,$$

where the second equality holds because R_{i+1}^* has density $c^{-i} x_i(s)$. This, combined with (4.7) and (4.8), shows that $EZ_j^2(t)$ is bounded by B, independent of t, which completes the proof of the existence of $Z(t)$.

This finishes part (a) of the outline (4.4). Before proceeding to the other parts, it is convenient to establish a lemma. Fix k and let $q_k(m)$ and $q_k^*(m)$ be the respective times t and t^* such that

$$P(R_k \leqq t) = \sum_{i \geqq k} p_i(t) = m \, 2^{-k}; \qquad P(R_k^* \leqq t^*) = \sum_{i \geqq k} x_i(t^*) = m \, 2^{-k}.$$

Let $Q_k(m)$ be the random time at which the m-th node at height k becomes available. (In fact, pairs of nodes become available simultaneously at the time when their parent node at level $k-1$ is filled, but this does not affect our argument.)

We use the notation $|s_1 - s_2|_t = |\min(s_1, t) - \min(s_2, t)|$, so that

$$
\begin{aligned}
|s_1 - s_2|_t &= |s_1 - s_2|, && \text{if } \max(s_1, s_2) \leqq t \\
&= 0, && \text{if } \min(s_1, s_2) \geqq t.
\end{aligned}
$$

Lemma 7. *Suppose $c > 2^{\frac{1}{2}}$. Then*

$$
E \sum_m |Q_k(m) - q_k^*(m)|_t \leqq 4t + B\, c^k\, 2^{k/2}, \qquad t \geqq 0, k \geqq 1,
$$

where B depends only on c.

Proof. Define $v_k(t) = P(R_k \leqq t)$. Let $V_k(t)$ be the proportion of level k nodes that become available before time t, so that $v_k(t) = E V_k(t)$. Fix t. Define

$$
A = \{(s, y): 0 \leqq s \leqq t, \ y \text{ is between } V_k(s) \text{ and } v_k(s)\},
$$

and consider A as a random region in the plane. Clearly

$$
\mathrm{area}(A) = \int_0^t |V_k(s) - v_k(s)|\, ds.
$$

Now we may represent A as the union of disjoint regions A_m, $m \geqq 0$, defined by

$$
\begin{aligned}
A_m &= A \cap \{(s, y): m\, 2^{-k} \leqq y < (m+1)\, 2^{-k}\} \\
&= \{(s, y): 0 \leqq s \leqq t, \ m\, 2^{-k} \leqq y < (m+1)\, 2^{-k}, \\
&\qquad y \text{ is between } V_k(s) \text{ and } v_k(s)\}.
\end{aligned}
$$

By drawing a diagram and recalling the definitions of $Q_k(m)$ and $q_k(m)$, it is easy to verify that

$$
\mathrm{area}(A_m) = 2^{-k} \{|Q_k(m) - q_k(m)|_t + \varepsilon(m)\},
$$

where

$$
|\varepsilon(m)| \leqq |Q_k(m+1) - Q_k(m)|_t + |q_k(m+1) - q_k(m)|_t.
$$

Since area $(A) = \sum_m \mathrm{area}(A_m)$, we find that

$$
\left| 2^k \int_0^t |V_k(s) - v_k(s)|\, ds - \sum_m |Q_k(m) - q_k(m)|_t \right| \leqq \sum_m |\varepsilon(m)|.
$$

But using the fact that $Q_k(m)$ and $q_k(m)$ are nondecreasing in m, we see that $\sum_m |\varepsilon(m)| \leq 2t$. Rearranging gives

$$\sum_m |Q_k(m) - q_k(m)|_t \leq 2t + 2^k \int_0^\infty |V_k(s) - v_k(s)| \, ds.$$

Writing $v_k^*(s) = P(R_k^* \leq s)$, the same argument shows that

$$\sum_m |q_k^*(m) - q_k(m)|_t \leq 2t + 2^k \int_0^\infty |v_k^*(s) - v_k(s)| \, ds.$$

Adding and taking expectation, we obtain

$$E \sum_m |Q_k(m) - q_k^*(m)|_t \leq 4t + 2^k \int_0^\infty |v_k^*(s) - v_k(s)| \, ds + 2^k \int_0^\infty E |V_k(s) - v_k(s)| \, ds. \qquad (4.14)$$

Recalling the definitions of R_k and R_k^*, we may write $R_k^* = R_0^* + R_k$, so the first integral in (4.14) is

$$\int_0^\infty \{P(R_k \leq s) - P(R_k^* \leq s)\} \, ds = \int_0^\infty \{P(R_k^* > s) - P(R_k > s)\} \, ds$$

$$= E R_k^* - E R_k = E R_0 = \frac{c}{c-1}.$$

Since $c > 2^{\frac{1}{2}}$ we have $2^k \leq c^k 2^{k/2}$, so the first integral term of (4.14) has the correct bound. Next, since $v_k(s) = E V_k(s)$ we can bound the second integral by

$$\int_0^\infty \{\text{var}(V_k(s))\}^{\frac{1}{2}} \, ds \leq B^{\frac{1}{2}} 2^{-k/2} \int_0^\infty \exp\left(\frac{-s}{4c^{k+1}}\right) ds \leq (4cB^{\frac{1}{2}}) 2^{-k/2} c^k,$$

where we used Lemma 5A to bound the variance of $V_k(s)$. Thus the second integral term in (4.14) has the correct bound and the proof of Lemma 10 is completed.

We now introduce the idea of "controlling the process at height k". This is best understood using the percolation description of remark 2, Sect. 1. Consider the 2^k edges from height $k-1$ nodes to height k nodes. In the original "uncontrolled" process, percolation along these starts at random times which, when arranged in increasing order, are the times $Q_k(1), Q_k(2), \ldots, Q_k(2^m)$, considered above. The controlled process is defined to make percolation along these edges start at the deterministic times $q_k(1), q_k(2), \ldots, q_k(2^m)$ defined above; the subsequent evolution of the process following the original rule. Formally, let H_γ denote the time at which node γ becomes available. Consider a node α at height k; this is the M_α'th node at height k to become available, for some random M_α, $1 \leq M_\alpha \leq 2^k$. Let $\Delta_\alpha = q_k^*(M_\alpha) - H_\alpha$. For each node γ that is a descendant of α (or α itself), define $\tilde{H}_\gamma = H_\gamma + \Delta_\alpha$. If this is done for each node α of height k then $\tilde{H}_\gamma$ is defined for each node γ of height $\geq k$. The family $\{\tilde{H}_\gamma\}$ then defines

a randomly-growing tree-process on nodes at heights $i \geq k$, namely, the process in which node γ becomes available at time $\tilde{H}_\gamma$. Call this "the process controlled at level k".

Now switch back to describing tree-processes in terms of available nodes. Let $X_i^k(t)$, $i \geq k$, be the proportion of height i nodes which are available at time t, in the process controlled at level k. Except at level k, the controlled process evolves in the same way as the uncontrolled process, and in particular the calculations (4.2) of the conditional incremental means and covariances remain true for $X_i^k(t)$ for $i \geq k+1$. At level k we find

(a) $E(\Delta X_k^k | X^k(t)) \sim -c^{-k} X_k^k(t) \Delta t + 2^{-(k-1)} \Delta F_k(t)$,

(b) $\text{var}(\Delta X_k^k | X^k(t)) \sim (2c)^{-k} X_k^k(t) \Delta t$

where $F_k(t)$ is the integer part of $2^{k-1} P(R_k^* \leq t)$. For the purposes of estimating conditional means and variances as $k \to \infty$ in the sequel, we may replace $2^{-(k-1)} F_k(t)$ in (a) by its continuous approximation $P(R_k^* \leq t)$. Since $d/dt\, P(R_k^* \leq t) = c^{-(k-1)} x_{k-1}(t)$, we may suppose

(a') $E(\Delta X_k^k(t) | X^k(t)) \sim -c^{-k} X_k^k(t) \Delta t + c^{-(k-1)} x_{k-1}(t) \Delta t$.

We use the controlling idea as follows. Fix i_0, and for each n consider the process controlled at level $n - i_0$, that is, the process $X_i^{n-i_0}(t)$, $i \geq n - i_0$. Rescale as at (4.3) to define

$$Z_i^{n, i_0}(t) = 2^{n/2} \{ X_{n+i}^{n-i_0}(t\, c^n) - x_{n+i}(t\, c^n) \}, \qquad i \geq -i_0.$$

As was indicated just after (4.6), define $Z_i^{\infty, i_0}(t)$, $i \geq -i_0$, to be the solution of the stochastic differential equations (4.5) satisfying $\bar{Z}_i(t) \equiv W_i(t) \equiv 0$, for $i < -i_0$; $Z_i(0) = 0$, for all i. Fix i_1. We want to show that

$$(Z_i^{n, i_0}(t);\ -i_0 \leq i \leq i_1,\ t \geq 0) \xrightarrow{\mathscr{D}} (Z_i^{\infty, i_0}(t);\ -i_0 \leq i \leq i_1,\ t \geq 0) \qquad (4.16)$$

in the sense of weak convergence of $(i_0 + i_1 + 1)$-dimensional processes. To do this, we verify the hypotheses of Proposition 1. Hypothesis (iv) holds because all the processes are zero at $t = 0$. Hypothesis (iii) holds because the tree grows only one node at a time. To verify hypotheses (i) and (ii), we must match the conditional means and variances. For $i \geq -i_0 + 1$, this is done by (4.3) and Theorem 1. For $i = -i_0$, (a') and (b) give

$$E(\Delta Z_{i_0}^{n, i_0} | Z^{n, i_0} = z) \sim -c^{-i_0} z_{i_0} \Delta t,$$

$$E([\Delta Z_{i_0}^{n, i_0}]^2 | Z^{n, i_0} = z) \sim (2c)^{-i_0} X_{n-i_0}(t\, c^{n-i_0}) \Delta t$$

whereas the boundary conditions for Z^{∞, i_0} give

$$E(\Delta Z_{i_0}^{\infty, i_0} | Z^{\infty, i_0} = z) \sim -c^{-i_0} z_{i_0} \Delta t,$$

$$E([\Delta Z_{i_0}^{\infty, i_0}]^2 | Z^{\infty, i_0} = z) \sim (2c)^{-i_0} x_{i_0}(t) \Delta t$$

and again Theorem 1 matches these as $n \to \infty$. Thus we can apply Proposition 1 to obtain (4.16). Moreover, letting $i_1 \to \infty$ and setting $Z_i^{n, i_0}(t) \equiv 0$, for $i < -i_0$, we may regard (4.16) as holding for the infinite vector processes indexed by

$-\infty < i < \infty$, since the topology on R^Z is coordinatewise convergence. This establishes part (b) of the outline (4.4).

For the proof of part (c), we first use Lemma 7 to bound the effect of controlling.

Lemma 8. $E|X_i^k(t) - X_i(t)| \leq c^{1+k-i} 2^{-k/2} B(1 + t\, c^{-k} 2^{2-k/2})$, $k \leq i$, $t \geq 0$, *where B is the bound of Lemma 7.*

Proof. Consider a node α at height k. Condition on M_α and $H_\alpha = Q_k(M_\alpha)$. Let γ be a descendant of α at height $i \geq k$. Conditionally,

$P(\gamma$ available at t for uncontrolled process$) = p_{k,i}(t - Q_k(M_\alpha))$,

$P(\gamma$ available at t for controlled process$) = p_{k,i}(t - q_k^*(M_\alpha))$.

By Lemma 1(a), the difference between these conditionsal probabilities is bounded by $c^{1-i}|Q_k(M_\alpha) - q_k^*(M_\alpha)|$. Since $p_{k,i}(s) = 0$ for $s \leq 0$, we can replace this bound by $c^{1-i}|Q_k(M_\alpha) - q_k^*(M_\alpha)|_t$, in the notation of Lemma 7. Uncondition and sum over γ to obtain

$$E|X_i^k(t) - X_i(t)| \leq 2^{-i} c^{1-i} \sum_\gamma E|Q_k(M_\alpha) - q_k^*(M_\alpha)|_t$$

$$\leq 2^{-k} c^{1-i} \sum_\alpha E|Q_k(M_\alpha) - q_k^*(M_\alpha)|_t$$

since each γ has 2^{i-k} ancestors α of height k. Lemma 7 now completes the proof of Lemma 8.

Now set $b = c^{-1} 2^{\frac{1}{2}} < 1$. Lemma 8 and the definition of Z^{n,i_0} give

$$E|Z_i^{n,i_0}(t) - Z_i^n(t)| \leq b^{i_0} c^{1-i} B(1 + t\, c^{i_0} 2^{2+i_0/2 - n/2}).$$

In particular, for fixed i and t,

$$\lim_{i_0 \to \infty} \limsup_{n \to \infty} E|Z_i^{n,i_0}(t) - Z_i^n(t)| = 0. \tag{4.17}$$

Now refer back to the outline of the proof, (4.4). The Z^n are random elements of the space of maps $z: [0, \infty) \to R^Z$. We have already proved (a) and (b) of (4.4). to prove (c) for this space we would want, in place of (4.17), a similar result for $E \sup_{t \leq t_0} |Z_i^{n,i_0}(t) - Z_i^n(t)|$. This seems hard to obtain directly; we will finesse the problem as follows. Fix t_1 and consider $Z^n(t_1)$ as a random element of R^Z. Then (4.17) certainly implies (4.4)(c) in this one-dimensional setting, so we have proved $Z^n(t_1) \xrightarrow{\mathscr{D}} Z(t_1)$. Similarly, for fixed $(t_1, t_2, \ldots, t_q)$ we have convergence of finite-dimensional distributions. To complete the proof it suffices (because convergence on R^Z is coordinatewise) to fix a coordinate i and to prove

$$(Z_i^n(t); t \geq 0) \xrightarrow{\mathscr{D}} (Z_i(t); t \geq 0) \quad \text{in } D([0, \infty), R.$$

From Lemma 1 and Lemma 5 we have

$$E[Z_i^n(t)]^2 \leq A c^{-2i} + B 2^{-i}, \quad A, B \text{ constant,}$$

so that, for each fixed i the random variables $Z_i^n(t)$ are uniformly integrable; here and below, *uniform* is with respect to $n \geq 1$ and $0 \leq t \leq t_0$. Fix i. Convergence of the finite dimensional distributions $Z_i^n(\cdot) \to Z_i(\cdot)$ implies weak convergence on the function space $L_0[0, \infty)$, that is, convergence in Lebesgue measure on bounded intervals, see, e.g., Grinblat [9]. Uniform integrability extends this to weak convergence on $L_1[0, \infty)$. Next the conditional mean calculation at the beginning of this section

$$E(\Delta Z_i^n | Z^n(t) = z) \sim \{c^{-(i-1)} z_{i-1} - c^{-i} z_i\} \Delta t,$$

implies that $M_i^n(t)$ is a martingale, where

$$
\begin{aligned}
M_i^n(t) &= Z_i^n(t) - I_i^n(t) + J_i^n(t), \\
I_i^n(t) &= \int_0^t c^{-(i-1)} Z_{i-1}^n(s)\, ds, \quad J_i^n(t) = \int_0^t c^{-i} Z_i^n(s)\, ds.
\end{aligned}
\tag{4.18}
$$

Now weak convergence of $Z_i^n(\cdot)$ on $L_1[0, \infty)$ implies weak convergence of $J_i^n(\cdot)$ on $C[0, \infty)$; and similarly for $I_i^n(\cdot)$. Thus $M_i^n(\cdot)$ converges, in the sense of finite-dimensional distributions, to the limit martingale $M_i(\cdot)$ defined similarly in terms of Z. Moreover $M_i^n(t)$ is uniformly integrable, since $Z_i^n(t)$ is. Applying the result (4.19), below, $M_i^n(\cdot)$ converges weakly in $D[0, \infty)$, and therefore, $Z_i^n(t) = M_i^n(t) + I_i^n(t) - J_i^n(t)$ converges weakly in $D[0, \infty)$. This final step made use of the following result, which is given in [1]. It extends a result of Loynes, [11], which required an extra technical condition.

Let $\{M_n(t)\}$ be a sequence of martingales that is uniformly integrable for each t, and suppose $M_n \to M_\infty$ in the sense of finite-dimensional distributions. If M_∞ is continuous then $M_n \to M_\infty$ weakly in $D[0, \infty)$.   (4.19)

5. The Entropy Algorithm

The results of Sect. 4, in the special case $c = 2$, lead to a description of the second-order properties of the Ziv entropy algorithm discussed in Remark 1, Sect. 1. Let

$$N(t) = \sum 2^i X_i(t); \quad B(t) = \sum (i - 2) 2^i X_i(t).$$

Thus, as noted in (1.1), $N(t)$ is essentially the number of nodes filled at time t and $B(t)$ is essentially the number of bits used at time t, in the random tree grown by applying the entropy algorithm to fair coin-tossing. Let $W(b)$ be the value of $N(t)$ at the first time t such that $B(t) \geq b$. Ziv showed in [17] that

$$W(b) \cdot b^{-1} \log_2 b \to 1 \text{ a.s.}, \quad \text{as } b \to \infty.$$

To state the corresponding central limit theorem, let us put

$$h(t) = EN(t) = \sum_{i=0}^{\infty} 2^i p_i(t); \quad b(t) = EB(t) = \sum_{i=0}^{\infty} (i-2) 2^i p_i(t).$$

The forward equations, (2.1), give

$$h'(t) = \sum_{i=0}^{\infty} p_i(t); \quad b'(t) = \sum_{i=0}^{\infty} i \, p_i(t),$$

so $b(t)$ is an increasing function and $h(t) = t + 1$. Let $t(b)$ be the inverse of $b(t)$.

Theorem 3. $\dfrac{W(b) - t(b)}{\sigma(b)} \xrightarrow{\mathcal{D}} \mathrm{Normal}(0, 1)$ *as* $b \to \infty$, *where* $\sigma(b) = b^{\frac{1}{2}} (\log_2 b)^{-\frac{3}{2}} f(b)$, *and* f *is bounded away from* 0 *and* ∞.

Our first step in the proof of Theorem 3 is to show that we can replace $t(b)$ by $t^*(b)$, the inverse function of

$$b^*(t) = \sum_{i=-\infty}^{\infty} (i-2) 2^i x_i(t).$$

The forward equations, (2.1), show that $db^*/dt = \sum_{i=-\infty}^{\infty} i \, x_i(t)$, hence, the lower tail estimate, Lemma 2, together with the shift-periodic property $x_i(t) = x_{i+1}(2t)$ then guarantees the existence of positive constants α and K so that

$$db^*/dt \geq \alpha, \quad \text{for } t \geq K. \tag{5.1}$$

In particular, when $t \geq K$, $b^*(t)$ will have an inverse function $t^*(b)$. Note also that $\sum 2^i x_i(t) = t$, since the forward equations yield the derivative $\sum x_i(t)$, which is identically 1 for $t > 0$. Thus, for any j we can write

$$b^*(t) = j \sum_{i=-\infty}^{\infty} 2^i x_i(t) + 2^j \sum_{i=-\infty}^{\infty} (i-j-2) 2^{i-j} x_i(t)$$

$$= j \, t + 2^j b^*(t \, 2^{-j}).$$

Equivalently, $b^*(t_0 2^j) = 2^j (j t_0 + b^*(t_0))$, which shows that

$$b^*(t) \sim t \log_2 t \quad \text{as } t \to \infty$$

and hence that

$$t^*(b) \sim b / \log_2 b \quad \text{as } b \to \infty.$$

We now prove the following estimate

$$|b(t) - b^*(t)| = o(t^{1/4}) \quad \text{as } t \to \infty. \tag{5.2}$$

Note that

$$b(t) - b^*(t) = \sum_{i=-\infty}^{\infty} (i-2) \, 2^i (p_i(t) - x_i(t)).$$

The sum over the indices $i < 0$ is uniformly bounded since $p_i = 0$ and $x_i \leq 1$ for those indices. The sum over the indices $i > 2i_0$ is bounded for $2^{i_0} \leq t \leq 2^{i_0+1}$, uniformly in i_0, from our upper tail estimate, Lemma 3, while Lemma 1(b) shows that the sum over the indices $0 \leq i \leq 2i_0$ is bounded by $8i_0^2$. This proves (5.2).

The result (5.2) and the derivative bound (5.1) give the estimate

$$|t^*(b) - t(b)| \leq \frac{1}{\alpha} |b(t(b)) - b^*(t(b))| \leq \frac{1}{\alpha} t^{\frac{1}{4}}(b),$$

which holds for b large. The result (5.2) and the fact that $b^*(t) \sim t \log_2 t$ together imply that $b(t) \sim t \log_2 t$, which then implies $t(b) \sim b/\log_2 b$ so that $|t^*(b) - t(b)| = O(b^{\frac{1}{4}})$. This completes the proof that it is enough to establish Theorem 3 with $t^*(b)$ in place of $t(b)$.

Our next step in proving Theorem 3 is a time-change argument, which can be expressed in the following purely analytic form.

Lemma 9. *Let $L_1^n(t)$ and $L_2^n(t)$ be right-continuous on the interval $I = [a, \infty)$, where a is a fixed positive number. Let $L_1(t)$, and $L_2(t)$ be continuous functions on I, let $u(t)$ be a bounded continuously differentiable function such that $u'(t) > 0$ on I. Let δ_n and ε_n both decrease to 0 such that the following hold.*

(a) $\lim_{n \to \infty} L_i^n = L_i$, *uniformly in any compact subset of I, $i = 1, 2$.*

(b) *The maximum jump of $L_1^n(t)$ is $o(\delta_n)$.*

For each $b > 0$ let $t_n(b)$ be a solution to $t + \delta_n u(t) = b$. Define the scalar functions

$$y_1^n(t) = t + \varepsilon_n L_1^n(t)$$
$$y_2^n(t) = t + \delta_n u(t) + \varepsilon_n L_1^n(t) + \varepsilon_n \delta_n L_2^n(t),$$

and define $s_n(b)$ to be the smallest t such that $y_2^n(t) \geq b$. Then, as $n \to \infty$

(i) $s_n(b)$ *and* $t_n(b) \to b$,

(ii) $\dfrac{y_1^n(s_n(b)) - t_n(b)}{\varepsilon_n \delta_n} \to u'(b) L_1(b) - L_2(b).$

Proof. Since u is strictly increasing the existence and uniqueness of $t_n(b)$ is guaranteed, while right continuity guarantees that $y_2^n(s_n(b)) \geq b$. Write t_n for $t_n(b)$ and s_n for $s_n(b)$. By assumption (a), $y_i^n(t) \to t$, uniformly on bounded intervals, $i = 1, 2$, so t_n and s_n both converge to b. Also, assumption (a) implies that the maximum jump of L_2^n is $o(1)$, so assumption (b) implies that

$$y_2^n(s_n) = b + o(\varepsilon_n \delta_n).$$

Hence,

$$\frac{y_1^n(s_n(b)) - t_n(b)}{\varepsilon_n \delta_n} = -L_2^n(s_n) + \varepsilon_n^{-1}(u(t_n) - u(s_n)) + o(1).$$

Certainly $L_2^n(s_n) \to L_2(b)$ and $u(t_n) - u(s_n) \sim (t_n - s_n) u'(b)$, so it suffices to prove

$$\varepsilon_n^{-1}(t_n - s_n) \to L_1(b). \tag{5.3}$$

Now $y_2^n(t_n) = b + \varepsilon_n L_1^n(t_n) + O(\varepsilon_n \delta_n)$, so the fact that $y_2^n(s_n) = b + O(\varepsilon_n \delta_n)$ implies that

$$\varepsilon_n^{-1}(y_2^n(t_n) - y_2^n(s_n)) \to L_1(b).$$

From the definition of y_2^n, and the fact that $t_n, s_n \to b$, we get

$$\varepsilon_n^{-1}(y_2^n(t_n) - y_2^n(s_n)) = \varepsilon_n^{-1}(t_n - s_n) + \varepsilon_n^{-1} \delta_n(u(t_n) - u(s_n)) + o(1)$$
$$= \varepsilon_n^{-1}(t_n - s_n)(1 + o(1)) + o(1),$$

and so our desired result (5.3) follows from the relation $\varepsilon_n^{-1}(y_2^n(t_n) - y_2^n(s_n)) \to L_1(b)$. The lemma is proved.

To apply Lemma 9 we use the expressions

$$X_{n+i}(t\,2^n) = x_{n+i}(t\,2^n) + 2^{-n/2} Z_i^n(t), \qquad b^*(t) = \sum (i-2)\, 2^i x_i(t),$$

in the definitions of $N(t)$ and $B(t)$ and a little algebra, including the fact that $\sum 2^i x_i(t) \equiv t$, to obtain

$$2^{-n} N(t\,2^n) = t + 2^{-n/2} \sum 2^i Z_i^n(t),$$
$$n^{-1} 2^{-n} B(t\,2^n) = t + n^{-1} b^*(t) + 2^{-n/2} \sum 2^i Z_i^n(t) + n^{-i} 2^{-n/2} \sum (i-2)\, 2^i Z_i^n(t).$$

Define $\varepsilon_n = 2^{-n/2}$, $\delta_n = 1/n$, $u(t) = b^*(t)$, and

$$L_1^n(t) = \sum 2^i Z_i^n(t), \qquad L_2^n(t) = \sum (i-2)\, 2^i Z_i^n(t),$$
$$L_1(t) = \sum 2^i Z_i(t), \qquad L_2(t) = \sum (i-2)\, 2^i Z_i(t),$$

where the vector process $Z(t) = (Z_i(t))$ is the limit, in distribution, of the processes, $Z^n(t) = (Z_i^n(t))$, as given by Theorem 2. The forward equations, (2.1), give $u'(t) = \sum i x_i(t)$, hence the lower tail estimate, Lemma 2, implies that $u(t)$ satisfies the conditions of Lemma 9 for sufficiently large a. The L_i are continuous, the L_i^n are right continuous, and a direct calculation, using the transition rates for the Markov chain $X(t)$, shows that the maximum jump of L_1^n is $o(\delta_n)$. Thus to apply Lemma 9 we need only show how the L_i^n and L_i can be joined on

an appropriate space so that uniform convergence of L_i^n to L_i on compact subsets of $(0, \infty)$ will hold. Theorem 2 implies that, for any fixed I_0,

$$\sum_{|i| \leq I_0} 2^i Z_i^n(t) \overset{\mathscr{D}}{\to} \sum_{|i| \leq I_0} 2^i Z_i^n(t),$$

$$\sum_{|i| \leq I_0} (i-2) 2^i Z_i^n(t) \overset{\mathscr{D}}{\to} \sum_{|i| \leq I_0} (i-2) 2^i Z_i^n(t).$$

Let us suppose the following lemma is true.

Lemma 10. *If $t_0 > 0$ then*

$$\lim_{I_0 \to \infty} \limsup_{n \to \infty} \sum_{|i| > I_0} |i| \cdot 2^i \operatorname{Emax}_{t \leq t_0} |Z_i^n(t)| = 0.$$

We can then obtain

$$L_1^n(t) = \sum_{i=-\infty}^{\infty} 2^i Z_i^n(t) \overset{\mathscr{D}}{\to} L_1(t) = \sum_{i=-\infty}^{\infty} 2^i Z_i(t),$$

$$L_2^n(t) = \sum_{i=-\infty}^{\infty} (i-2) 2^i Z_i^n(t) \overset{\mathscr{D}}{\to} L_2(t) = \sum_{i=-\infty}^{\infty} 2^i (i-2) Z_i(t),$$

and, moreover, $(L_1^n, L_2^n) \overset{\mathscr{D}}{\to} (L_1, L_2)$. This will be enough to establish Theorem 3, for by the Skorohod representation theorem, together with Lemma 10, we can assume that the convergence (L_1^n, L_2^n) to (L_1, L_2) is also almost surely uniform on bounded intervals. We can then fix an ω for which such uniform convergence holds and apply Lemma 9 to obtain, for fixed b,

$$\frac{2^{-n} N(S_n(b)) - t_n(b)}{n^{-1} 2^{-n/2}} \overset{\mathscr{D}}{\to} u'(b) \sum 2^i Z_i(b) - \sum (i-2) 2^i Z_i(b). \tag{5.4}$$

where $t_n + \delta_n u(t_n) = b$, and S_n is the smallest $t 2^n$ such that $n^{-1} 2^{-n} B(t 2^n) \geq b$. In the notation of the beginning of this section $t_n = 2^{-n} t^*(n b 2^n)$ and $N(S_n) = W(n b 2^n)$. Moreover, since the limit process $Z(t)$ is Gaussian, mean 0, the right side of (5.4) is Normal$(0, g^2(b))$, for some $g(b)$, such that $0 < g(b) < \infty$. (See (d), Sect. 7.) Thus (5.4) becomes

$$\frac{W(b n 2^n) - t^*(b n 2^n)}{n^{-1} 2^{n/2} g(b)} \overset{\mathscr{D}}{\to} \operatorname{Normal}(0, 1), \quad \text{as } n \to \infty. \tag{5.5}$$

By modifying the statement and proof of Lemma 9, it can be shown that 5.4 and (5.5) remain true if b is replaced by a sequence b_n that converges to a finite, positive value b. Then by considering subsequences, (5.5) remains true for any sequence $\{b_n\}$, bounded away from 0 and ∞. Now let $m \to \infty$ and define

sequences $b = b_m$ and $n = n_m$ so that $m = b \, n \, 2^n$ and $\{b_m\}$ is bounded away from 0 and ∞. Then $n \sim \log_2 m$, $2^n \sim m(b \cdot \log_2 m)^{-1}$, and so

$$n^{-1} 2^{n/2} g(b) \sim m^{1/2} (\log^2 m)^{-3/2} f(m)$$

where $f(m) = b^{-1/2} g(b)$ is bounded away from 0 and ∞. Now (5.5) gives Theorem 3.

Thus it is enough to establish Lemma 10. We shall write out in detail the proof of the weaker result in which "$E \max$" is replaced by "$\max E$":

$$\lim_{I_0 \to \infty} \limsup_{n \to \infty} \sum_{|i| > I_0} |i| \cdot 2^i \, E \max_{t \le t_0} E|Z_i^n(t)| = 0. \tag{5.6}$$

To go from here to Lemma 10 involves the same issues as were in the argument at the end of Sect. 4: We use the decomposition (4.18)

$$Z_i^n(t) = M_i^n(t) + I_i^n(t) - J_i^n(t),$$

and then appeal to maximal inequalities for the martingales M^n and smoothness properties for the integrated processes I^n, J^n. We omit the details of this argument.

To establish (5.6) we first prove two lemmas.

Lemma 11. *There is a constant α such that for all $n \ge 1$, $i \ge 1$, and $0 \le t \le 1$,*

$$2^{n+i} \operatorname{var} X_{n+i}(t \, 2^n) \le \alpha \, 2^{-3i/2}.$$

Proof. By our variance bound, Lemma 4, it is enough to show that

$$p_{n+i}(t \, 2^n) + \sum_{k = -(n-1)}^{i} 2^{i-k} \operatorname{var} p_{n+k, n+i}(t \, 2^n - R_{n+k}) \le \alpha \, 2^{-3i/2}. \tag{5.7}$$

From our basic upper tail estimate, Lemma 3, the first term has the correct bound. Now consider the sum over $k \le -i/2$. We noted in the proof of Lemma 5 that $\operatorname{var} p_{i,j}(t - R_i) \le (1 - c^{-2})^{-1} c^{2i-2j}$. In our case, $c = 2$ and we obtain

$$\sum_{k \le -i/2} 2^{i-k} \operatorname{var} p_{n+k, n+i}(t \, 2^n - R_{n+k}) \le \tfrac{4}{3} \sum_{k \le -i/2} 2^{-(i-k)} \le \tfrac{8}{3} \cdot 2^{-3i/2},$$

and this bound has the right form.

Now consider the sum in (5.7) over $-i/2 \le k \le i - 1$. We have

$$\sup_{s \le t 2^n} p_{n+k, n+i}(s) \le \sup_{s \le t 2^{-k}} p_{i-k}(s) \le t^{i-k} 2^{-\frac{1}{2}(i-k)(i+k-1)},$$

where the first inequality comes from the general scaling fact that $p_{i,j}(u) = p_{j-i}(u \, 2^{-i})$, and the second inequality from the upper tail estimate of Lemma 3. For any nonnegative random variable Y, $\operatorname{var} Y \le (\max Y)^2$, so

$$\operatorname{var} p_{n+k, n+i}(t \, 2^n - R_{n+k}) \le t^{2(i-k)} 2^{-(i-k)(i+k-1)}.$$

Summing over $-i/2 \le k \le i-1$ then gives an upper bound of the desired form.

Finally consider the case when $k=i$. Note that if $0 \le Y \le 1$, then var $Y \le P(Y \ne 0)$, so that

$$\text{var } p_{n+i,n+i}(t\,2^n - R_{n+i}) \le P(R_{n+i} \le t\,2^n),$$

and this has the correct upper bound, by Lemma 3. This establishes Lemma 11.

Our next lemma is another variation on Lemmas 3 and 1(b).

Lemma 12. *For fixed* t_0,

$$2^{n/2} \max_{t \le t_0} \sum_{i \ge 1} i\,2^i |x_{n+i}(t\,2^n) - p_{n+i}(t\,2^n)| \to 0, \quad \text{as } n \to \infty.$$

Proof. Recall that $p_{n+i}(t\,2^n) = P(R_{n+i} < t\,2^n \le R_{n+i+1})$ and

$$x_{n+i}(t\,2^n) = P(R_{n+i}^* < t\,2^n \le R_{n+i+1}^*) = P(R_{n+i} + R_0^* < t\,2^n \le R_{n+i+1} + R_0^*).$$

Thus

$$|p_{n+i}(t\,2^n) - x_{n+i}(t\,2^n)|$$

$$\le P(R_{n+i} < t\,2^n \le R_{n+i} + R_0^*) + P(R_{n+i+1} < t\,2^n \le R_{n+i+1} + R_0^*). \quad (5.8)$$

Now, for $0 < x < t\,2^n$, conditioning on R_{n+i-1} gives

$$P(R_{n+i} < t\,2^n \le R_{n+i} + R_0^* \,|\, R_{n+i-1} = x) = P(x + 2^{n+i}\xi_{n+i} < t\,2^n \le x + 2^{n+i}\xi_{n+i} + R_0^*)$$

$$\le ER_0^*/2^{n+i} = 2^{1-n-i},$$

where the inequality holds by conditioning on R_0^* and noting that the density of $2^{n+i}\xi_{n+i}$ is bounded by $1/2^{n+i}$. So

$$P(R_{n+i} < t\,2^n \le R_{n+i} + R_0^*) \le 2^{1-n-i} P(R_{n+i-1} \le t\,2^n).$$

Applying the same argument to R_{n+i+1} and using (5.8) we get,

$$|p_{n+i}(t\,2^n) - x_{n+i}(t\,2^n)| \le 2^{2-n-i} P(R_{n+i-1} \le t\,2^n) \le 2^{2-n-i}\,t^{i-1}\,2^{-(i-1)(i-2)/2},$$

by the proof of Lemma 3. Thus the quantity of interest to us

$$2^{n/2} \max_{t \le t_0} \sum_{i \ge 1} i\,2^i |x_{n+i}(t\,2^n) - p_{n+i}(t\,2^n)|$$

is bounded by

$$2^{-n/2} \sum_{i \ge 1} i\,t_0^{i-1}\,2^{2-(i-1)(i-2)/2}.$$

Since this sum is finite the lemma is established.

Now we are ready to prove (5.6). For the sum over $i < -I_0$, we write

$$\sum_{i < -I_0} |i| \cdot 2^i \max_{t \le t_0} E|Z_i^n(t)|$$

$$= \sum_{i=I_0}^{n} i\, 2^{-i} \max_{t \le t_0} E|Z_{-i}^n(t)| + \sum_{i=n+1}^{\infty} i\, 2^{-i} \max_{t \le t_0} E|Z_{-i}^n(t)|.$$

In the sum over $i > n$, $X_{n-i}(t) = 0$ and $x_{n-i}(t) \le 1$, so the sum is bounded above by $2^{n/2} \sum_{i=n+1}^{\infty} i\, 2^{-i}$, which goes to 0. For the sum over the range $I_0 \le i \le n$, we have the upper bound

$$2^{n/2} \sum_{i=I_0}^{n} i\, 2^{-i} \max_{t \le t_0} [\operatorname{var} X_{n-i}(t\, 2^n)]^{\frac{1}{2}} + 2^{n/2} \sum_{i=I_0}^{n} i\, 2^{-i} \max_{t \le t_0} |x_{n-i}(t\, 2^n) - p_{n-i}(t\, 2^n)|.$$

The first sum is bounded by $B^{\frac{1}{2}} \sum_{i=I_0}^{\infty} i\, 2^{-i/2}$, from Lemma 5(c), while Lemma 1(b) can be used to bound the second sum by $4 \cdot 2^{-n/2} \sum_{i=I_0}^{n} i$. This establishes (5.6) in the case when $i < -I_0$.

Now consider the case when $i > I_0$. We may take $t_0 = 2^q$, say. For $i \ge q + 1$ we then have,

$$\operatorname{var} Z_i^n(t) = 2^n \operatorname{var} X_{n+i}(t\, 2^n)$$

$$= 2^{-i} 2^{n+i} \operatorname{var} X_{(n+q)+(i-q)}(t\, 2^{-q} 2^{n+q})$$

$$\le 2^{-i} \alpha\, 2^{-3(i-q)/2}, \quad \text{for } t \le 2^q,$$

by Lemma 11, since $t\, 2^{-q} \le 1$. Thus

$$\max_{t \le 2^q} [\operatorname{var} Z_i^n(t)]^{\frac{1}{2}} \le \alpha^{\frac{1}{2}} 2^{3q/4} 2^{-5i/4}.$$

This implies

$$\lim_{I_0 \to \infty} \limsup_{n \to \infty} \sum_{i > I_0} i\, 2^i \max_{t \le 2^q} [\operatorname{var} Z_i^n(t)]^{\frac{1}{2}} = 0. \tag{5.9}$$

Now

$$E|Z_i^n(t)| \le [\operatorname{var} Z_i^n(t)]^{\frac{1}{2}} + |EZ_i^n(t)|$$

and

$$|EZ_i^n(t)| = 2^{n/2} |p_{n+i}(t\, 2^n) - x_{n+i}(t\, 2^n)|,$$

so the upper bound for (5.6) follows from (5.9) and Lemma 12.

6. The Height of the Tree

Let $M(t) - 1$ denote the height of the tree at time t. As noted in (1.1), $M(t) = \max\{i : X_i(t) > 0\}$. We shall prove the following

Theorem 4. *For* $c > 1$,

$$\frac{M(t) - \log_c t}{(\log_c t)^{\frac{1}{2}}} \to \left\{ \frac{2}{\log_2 c} \right\}^{\frac{1}{2}}, \quad \text{a.s.}$$

Remark. By analogy with standard theory for maxima of i.i.d. discrete variables with faster than geometric tails, one would expect the stronger result that a.s. $M(t) = m(t)$ or $m(t+1)$ ultimately, for some deterministic $m(t)$. H. Kesten (private communication) has results in this direction.

Proof. Write $\beta_0 = \{2/\log_2 c\}^{\frac{1}{2}}$. Let α be a node at height $i + k$. Our upper tail estimate, Lemma 3, implies

$$P(\alpha \text{ is available before time } c^i) \leq c^{-k(k-1)/2},$$

so

$$P(M(c^i) \geq i + k) \leq 2^{i+k} c^{-k(k-1)/2}.$$

Now fix $\beta > \beta_0$ and let $k = k(i) = [\beta i^{\frac{1}{2}}]$, where $[\cdot]$ denotes the greatest integer function. Then

$$\sum_i 2^{i+k} c^{-k(k-1)/2} < \infty,$$

so by the Borel-Cantelli lemma we have a.s.

$$M(c^i) < i + k(i), \quad \text{ultimately}.$$

This gives the desired upper bound. For the lower bound, fix $\beta < \beta_0$ and again let $k = k(i) = [\beta i^{\frac{1}{2}}]$. We then have the following.

$$2^i P\left(\sum_{j=1}^{k} c^j \xi_j \leq 1 \right) \to \infty, \quad \text{geometrically fast as } i \to \infty. \tag{6.1}$$

To prove this, note that

$$P\left(\sum_{j=1}^{k} c^j \xi_j \leq 1 \right) \geq P(c^j \xi_j \leq k^{-1}, \quad \text{for } 1 \leq j \leq k)$$

$$= \prod_{j=1}^{k} \{ 1 - \exp(-c^{-j}/k) \}$$

$$\geq \frac{1}{2} \prod_{j=1}^{k} (c^{-j}/k)$$

$$= \frac{1}{2} k^{-k} c^{-k(k+1)/2}.$$

Multiplying by 2^i then gives a lower bound of the form $\frac{1}{2} k^{-k} \theta^i$, where $\theta = 2c^{-\beta^2/2} > 1$, and since $k = O(i^{\frac{1}{2}})$ this grows geometrically, proving (6.1).

Continuing with the proof of Theorem 4, for each node $\tilde{\alpha}$ of height i associate one descendant node α of height $i+k$. Let R_α and $R_{\tilde{\alpha}}$ be the respective times that α and $\tilde{\alpha}$ first become available. Then

$$P(R_\alpha - R_{\tilde{\alpha}} \leq c^i) = P\left(\sum_{j=i+1}^{i+k} c^j \xi_j \leq c^i \right) = P\left(\sum_{j=1}^{k} c^j \xi_j \leq 1 \right),$$

and then the result (6.1) implies

$$q_i \equiv P(R_\alpha - R_{\tilde{\alpha}} \leq c^i, \quad \text{for some } \tilde{\alpha}) \to 1,$$

geometrically fast. Since $E R_{\tilde{\alpha}} = B c^i$, for some constant B, we have

$$P(R_{\tilde{\alpha}} \leq i^2 c^i) \geq 1 - B/i^2,$$

so that

$$P(M(i^2 c^i + c^i) \geq i+k) \geq P(R_{\tilde{\alpha}} \leq i^2 c^i, R_\alpha - R_{\tilde{\alpha}} \leq c^i, \text{for some } \tilde{\alpha})$$
$$\geq q_i - B/i^2.$$

The Borel-Cantelli lemma then implies that a.s.

$$M(i^2 c^i + c^i) \geq i + k(i), \quad \text{ultimately},$$

and this leads to the lower bound.

Now write $\hat{M}(n)$ for the height of the discrete-time tree when it has n nodes. Embedding into the continuous-time tree gives $\hat{M}(n) = M(S_n)$, where, as in Sect. 3, S_n is the time at which the n'th node is filled. Theorem 4 implies

$$\frac{\hat{M}(n) - \log_c S_n}{(\log_c S_n)^{\frac{1}{2}}} \to (2/\log_2 c)^{\frac{1}{2}}, \quad \text{a.s.}$$

By (3.3) and (3.4), $\log_c S_n - \log_2 n$ is bounded a.s. as $n \to \infty$, so we can rearrange to obtain the following corollary to Theorem 4.

$$\frac{\hat{M}(n) - \log_c n}{(2 \cdot \log_c n)^{\frac{1}{2}}} \to 1, \quad \text{a.s.} \tag{6.2}$$

7. Further Remarks and Connections with Other Work

(a) For the continuous-time model the rate at which new nodes are being added at time t is $\sum 2^i c^{-i} X_i(t)$. In the special case $c = 2$ this rate is identically 1 so that the times at which new nodes are added is a Poisson process of rate 1. For $c > 2$ the rate slows down as the tree grows, while for $c < 2$ the rate speeds up. These features are artifacts of the continuous-time formulation and have no direct interpretation for the shape of the tree $\hat{T}(n)$ with n nodes.

(b) The case $c = 2^{\frac{1}{2}}$ is a critical case in our proof of Theorem 2. We conjecture that a Gaussian limit holds here also, but that for $c < 2^{\frac{1}{2}}$ there is some non-Gaussian limit process obtained by scaling by c^n instead of $2^{n/2}$. Partial results on this latter conjecture have been obtained by H. Kesten, (private communication).

(c) The condition $c > 1$ means that available nodes of lesser height are likely to be filled before available nodes at greater height. This leads to a strong "stable growth" property of the random tree, in that heights of different terminal nodes at time t do not vary much, and the typical variation stays bounded as $t \to \infty$. Our results show this "stability" property in that the vector $\hat{X}(2^n)$ looks like the fixed vector $\hat{x}(1)$ shifted n units to the right.

(d) Good asymptotic estimates of the standard deviation of $W(b)$ are needed to make Theorem 3 into a useful hypothesis testing tool. This is tantamount to estimating the variance $g^2(b)$ of the right-hand side of (5.4). In principle this can be evaluated as follows. First, note that $u'(b) = \sum i \, x_i(b)$. Second, use the integral representation (4.6) for $Z_i(b)$ in terms of the Brownian motions $W_i(t)$, then collect terms in each W_j. This eventually leads to a complicated expression for $g(b)$ in terms of x_i and p_i. In particular, it can be shown that $g(b)$ is bounded away from 0 and ∞ on bounded intervals.

(e) A model equivalent to the case $c = 2$ of our discrete-time tree has been discussed by Bradley and Strenski, who call it "directed diffusion-limited aggregation", [3], and obtain results weaker than our Theorem 1D and (6.2). They show, in our notation of Sect. 6, that if $G_h(n) = P(\hat{M}(n) < h)$ then

$$G_{h+1}(n+1) = 2^{-n} \sum_{m=0}^{n} \binom{n}{m} G_h(m) \, G_n(n-m).$$

They show essentially that $M(n) - \log_2 n \to \infty$, which is of course weaker than our (6.2); but it is possible that a more refined analysis of their recursion might lead to stronger results in the spirit of our results. They also consider the quantity

$$\bar{\rho}(h) = 2^{-h} E \min \{ n : M(n) = h \}.$$

Our (6.2) implies that $\bar{\rho}(h)$ is roughly of the order of $2^{-(2h)^{\frac{1}{2}}}$, confirming their conjecture that $\bar{\rho}(h) \to 0$, slower than exponentially.

(f) A tree can be identified with a partition of the unit interval into intervals of length 2^{-h}. Thus a process of randomly-growing trees can be identified with a process of randomly splitting intervals in half. Brennan and Durrett have studied more general processes where intervals are split into parts according to some continuous distribution V, obtaining results related to some of ours, [4, 5]. Unfortunately their results do not cover the degenerate distribution which is discussed here, although some of their arguments are the same as ours.

(g) Randomly-growing trees are also of interest in computer science. One model closely related to ours is the binary search tree described in Knuth, [10, Sect. 6.22]. Such trees are structures for storing numbers x_n (regarded as labels of files) so that a given number x can be rapidly located. The first number x_0 is stored at the root. Subsequent numbers are then stored at available nodes

by making comparisons with numbers already stored and moving along the
0-branch if smaller and the 1-branch if larger. To be precise, suppose
$x_0, x_1, \ldots, x_n$ have already been stored at the nodes of a tree $t(n)$. The next
number x_{n+1} is stored at the first available node of the branch $b_1 b_2 b_3 \ldots$ defined
by the rule:

$$\text{If } .b_1 b_2 \ldots b_m \text{ contains } y, \text{ then } b_{m+1} = 0, \ x < y, \text{ else } b_{m+1} = 1.$$

If the numbers $X_0, X_1, X_2, \ldots$ are independent with a common continuous distri-
bution, then the random tree $T(n)$ obtained is our discrete-time tree for the
case when $c = 1$. In this context, quantities like the mean number of nodes which
must be examined to locate a given random x are of interest, and these are
functionals of $T(n)$. Pittel, [12], contains the deepest known results, partly using
the same embedding into the continuous-time process that we exploit. The trees
$T(n)$ are rather less "balanced" in the $c = 1$ case; for instance, the average height
and maximum height of occupied nodes are asymptotic to $2 \cdot \log_2 n$ and
$4.311 \cdot \log_2 n$, respectively; whereas in the case $c > 1$ both are asymptotic to $\log_2 n$.
Other related results are contained in Pittel, [13, 14].

(h) Though our results concern only the case $c > 1$, the models make sense
for $0 < c \leq 1$ also. The case $c = 1$ is mentioned in the preceding remark. Let
us mention some known results and open problems for the case $c < 1$. First,
we note that the "stability" property mentioned above doesn't hold, as shown
by Brennan and Durrett, [4, 5]. In this case the long branches tend to grow
faster than the short branches. In the continuous-time model, the height reaches
infinity at some random finite time L, hence our technique of studying the
discrete-time model by embedding into the continuous-time model does not
work. By considering the time required to add the first node, we get the relation,

$$L \overset{\mathscr{D}}{\to} \xi + c \cdot \min(L_1, L_2) \tag{7.1}$$

where ξ has exponential, mean 1, distribution; L_1 and L_2 are distributed as
L; and ξ, L_1, and L_2 are independent. Athreya, [2], studies a more general
problem of this type: his results imply that (7.1) has a unique solution. For
the discrete-time model, the height $N(n)$ of the tree with n nodes has $M(n) < n$;
of course, the behavior of $M(n)$ as $n \to \infty$ is an interesting open problem.

References

1. Aldous, D.: Stopping times and tightness, II. Tech. report No. 124, Dept. of Statistics, University
   of California, Berkeley, 1987
2. Athreya, K.B.: Discounted branching random walks. Adv. Appl. Probab. **17**, 53–66 (1985)
3. Bradley, R.M., Strenski, P.N.: Directed diffusion-limited aggregation on the Bethe lattice: exact
   results. Phys. Rev. B **30**, 6788–6790 (1984)
4. Brennan, M.D., Durrett, R.: Splitting intervals. Ann. Probab. **14**, 1023–1036 (1986)

5. Brennan, M.D., Durrett, R.: Splitting intervals II: Limit laws for lengths. Probab. Th. Rel. Fields **75**, 109–127 (1987)
6. Durrett, R.: Oriented percolation in two dimensions. Ann. Probab. **12**, 999–1040 (1984)
7. Elliott, R.: Stochastic calculus and its applications. New York: Springer 1982
8. Gihman, I.I., Skorohod, A.V.: The theory of stochastic processes, vol. 3. New York: Springer 1979
9. Grinblatt, L.S.: A limit theorem for measurable random processes and its applications. Proc. Amer. Soc. **61**, 371–376 (1976)
10. Knuth, D.: The art of computer programming, vol. 3. Reading, Mass.: Addison-Wesley 1973
11. Loynes, R.M.: A criterion for tightness for a sequence of martingales. Ann. Probab. **4**, 859–862 (1976)
12. Pittel, B.: On growing random binary trees. J. Math. Anal. Appl. **103**, 461–480 (1984)
13. Pittel, B.: Asymptotic growth of a class of random trees. Ann. Probab. **13**, 414–427 (1985)
14. Pittel, B.: Paths in a random digital tree: limiting distributions. Adv. Appl. Probab. **18**, 139–155 (1986)
15. Ross, S.M.: Stochastic processes. New York: Wiley 1983
16. Wignarajah, G.: Complexity tests for statistical independence. M.S. thesis, University of Toledo, 1985
17. Ziv, J.: Coding theorems for individual sequences. IEEE Trans. Inf. Theory **24**, 405–412 (1978)
18. Ziv, J., Lempel, A.: A universal algorithm for sequential data compression. IEEE Trans. Inf. Theory **23**, 337–343 (1977)