

## On Weighted Heights of Random Trees

Rick Durrett,<sup>1</sup> Harry Kesten,<sup>1</sup> and Ed Waymire<sup>2</sup>

*Received October 15, 1989; revised December 27, 1989*

---

Consider the family tree  $T$  of a branching process starting from a single progenitor and conditioned to have  $v = v(T)$  edges (total progeny). To each edge  $\langle e \rangle$  we associate a weight  $W(e)$ . The weights are i.i.d. random variables and independent of  $T$ . The weighted height of a self-avoiding path in  $T$  starting at the root is the sum of the weights associated with the path. We are interested in the asymptotic distribution of the maximum weighted path height in the limit as  $v = n \rightarrow \infty$ . Depending on the tail of the weight distribution, we obtain the limit in three cases. In particular if  $y^2 P(W(e) > y) \rightarrow 0$ , then the limit distribution depends strongly on the tree and, in fact, is the distribution of the maximum of a Brownian excursion. If the tail of the weight distribution is regularly varying with exponent  $0 \leq \alpha < 2$ , then the weight swamps the tree and the answer is the asymptotic distribution of the maximum edge weight in the tree. There is a borderline case, namely,  $P(W(e) > y) \sim cy^{-2}$  as  $y \rightarrow \infty$ , in which the limit distribution exists but involves both the tree and the weights in a more complicated way.

---

**KEY WORDS:** Branching process; Brownian excursion; random tree.

### 1. INTRODUCTION AND STATEMENT OF RESULTS

Let  $Z_0, Z_1, \dots$ , be a discrete parameter Bienaymé-Galton-Watson branching process, starting from a single particle, the "progenitor" (i.e.,  $Z_0 = 1$ ). As has been pointed out by many authors (e.g., Harris,<sup>(11)</sup> Jagers<sup>(12)</sup>), the structure of such a process is conveniently described by its (random) family tree  $T$ , which we shall regard as a rooted and labeled tree, (see below for details). We write  $v = v(T)$  for the total number of vertices in  $T$ ;  $v$  is the total progeny of the branching process. To each edge  $\langle e \rangle$  of the tree  $T$  we associate a random "weight"  $W(e)$ , such that (conditionally on  $T$ ) the ran-

---

<sup>1</sup> Department of Mathematics, Cornell University, Ithaca, New York 14853.

<sup>2</sup> Department of Mathematics, Oregon State University, Corvallis, Oregon 97331.

dom variables  $\{W(e): \langle e \rangle \in T\}$  are independent and identically distributed (i.i.d.). We are interested in limit laws of certain sums of the weights under the condition  $\nu(T) = n$ , as  $n \rightarrow \infty$ . Specifically, if  $\mathbf{s}$  is a self-avoiding path in  $T$  starting at the root, and consisting of the edges  $\langle e_1 \rangle, \langle e_2 \rangle, \dots, \langle e_k \rangle$ , then we define the *height* of the path  $\mathbf{s}$  as

$$H(\mathbf{s}) = H_w(\mathbf{s}) = \sum_{i=1}^k W(e_i)$$

The *weighted tree height* is

$$H(T) = H_w(T) = \max_{\mathbf{s} \subset T} H(\mathbf{s}) \quad (1.1)$$

Our theorems give the conditional limit distribution of  $c_n H(T)$ , given  $\nu(T) = n$ , as  $n \rightarrow \infty$ , for suitable choices of  $c_n$ .

As is clear from definition (1.1), calculation of the conditional limit distribution of  $H(T)$  may be viewed as an extreme value problem for certain types of overlapping sums of i.i.d. random variables. There is one such sum corresponding to each path  $\mathbf{s}$  from the root in  $T$ . The classical problem of the distribution of the maximum of  $n$  i.i.d. random variables corresponds to a tree consisting only of  $n$  branches emanating from the root (or an arrangement of  $n$  i.i.d. weighted segments in parallel).

Conditioning on the total progeny  $\nu$  is somewhat unnatural from the point of view of branching processes. However, our motivation comes from considerations of channel length statistics in river basin hydrology, where  $T$  is regarded as an idealized model for the network of branches of a river (in the absence of geologic controls); a rather extensive background to the literature on this model can be found in Jarvis and Woldenberg.<sup>(13)</sup> Here  $\nu(T)$  equals the total number of branches of the river plus 1, and we may think of  $W(e)$  as the length of the branch  $\langle e \rangle$ . Some justification for conditioning on  $\nu(T)$  (large) in this model can be found in Gupta and Waymire<sup>(9)</sup> and references therein. We point out that  $\nu(T)$  has also been considered in the context of neutron branching processes, where  $\nu$  represents the total number of neutrons produced; see Harris (Ref. 11, Section IV.6.3). Conditioning on a large  $\nu$  may also be reasonable for this interpretation. Finally, there is a large combinatorial literature on trees with a fixed, but large, number of vertices. Some of this literature and its relation with branching processes is surveyed in Aldous.<sup>(1,2)</sup>

To state our results precisely, we introduce some further notation. Following Harris<sup>(11)</sup> and Jagers,<sup>(12)</sup> we label vertices of  $T$  as follows. The root, corresponding to the progenitor, is denoted by  $\langle 0 \rangle$ . It forms the *zeroth generation* of  $T$ . Other labels are assigned recursively. The vertices of

$T$  which are adjacent to  $\langle 0 \rangle$  form the first generation of  $T$ . They are labeled  $\langle 0, \varepsilon_i \rangle$ ,  $1 \leq \varepsilon_i \leq N(0)$ , in some order, and are called the *children of*  $\langle 0 \rangle$ .  $N(0)$  is the number of children of  $\langle 0 \rangle$ ; this number can be zero, in which case  $T$  consists of  $\langle 0 \rangle$  only. A generic label is of the form  $\langle 0, \varepsilon_1, \varepsilon_2, \dots, \varepsilon_k \rangle$ , with  $\varepsilon_i \geq 1$ ,  $1 \leq i \leq k$ . The label is assigned to one of the vertices adjacent to  $\langle 0, \varepsilon_1, \dots, \varepsilon_{k-1} \rangle$ . The vertices other than  $\langle 0, \varepsilon_1, \dots, \varepsilon_{k-1} \rangle$ , which are adjacent to  $\langle 0, \varepsilon_1, \dots, \varepsilon_k \rangle$  (if any) are labeled  $\langle 0, \varepsilon_1, \dots, \varepsilon_{k+1} \rangle$ .  $\varepsilon_{k+1} = 1, 2, \dots, N$ , where  $N = N(0, \varepsilon_1, \dots, \varepsilon_k)$  is the total number of such vertices of  $T$ . These vertices are called the *children of*  $\langle 0, \varepsilon_1, \dots, \varepsilon_k \rangle$ . All vertices which receive a label of the form  $\langle 0, \varepsilon_1, \dots, \varepsilon_k \rangle$  make up the  $k$ th generation of  $T$ . They are precisely the vertices of  $T$  which are connected to the root  $\langle 0 \rangle$  by a self-avoiding path in  $T$  of exactly  $k$  edges (see Fig. 1). The height  $h$  of such a vertex is taken to be  $k$ . Thus

$$h(0, \varepsilon_1, \dots, \varepsilon_k) = k \tag{1.2}$$

The total size of the  $k$ th generation is  $Z_k$  ( $Z_0 = 1$ ). The *extinction time* is

$$\tau = \tau(T) := \min\{k \geq 1: Z_k = 0\} \tag{1.3}$$

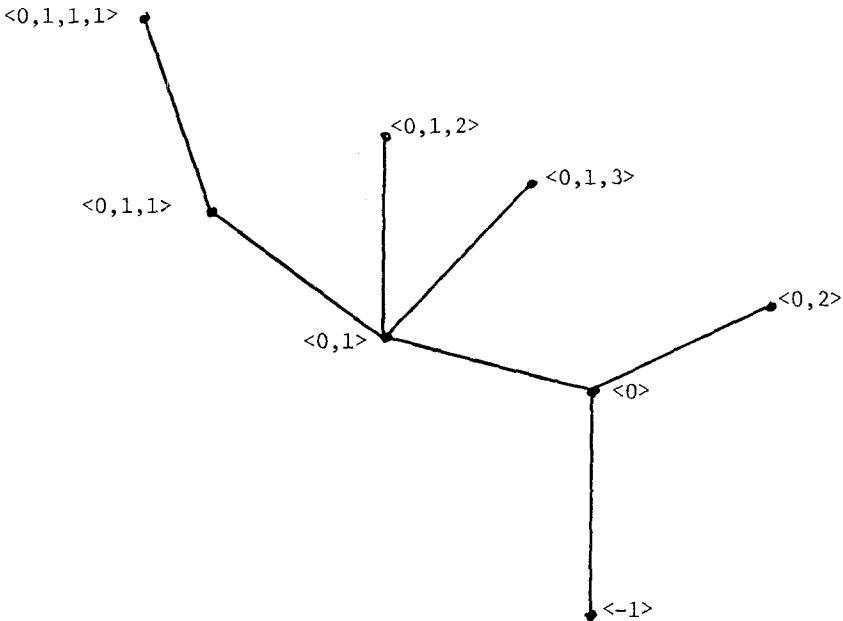


Fig. 1

and the total progeny is

$$v = v(T) := \sum_{k=0}^{\infty} Z_k = \sum_{k=0}^{\tau-1} Z_k \tag{1.4}$$

The *height* of  $T$  is

$$h(T) = \tau - 1 = \max_{\langle \mathbf{x} \rangle \in T} h(\mathbf{x}) \tag{1.5}$$

We shall also use  $\langle 0, \varepsilon_1, \dots, \varepsilon_k \rangle$  as a label for the unique edge between  $\langle 0, \varepsilon_1, \dots, \varepsilon_{k-1} \rangle$  and  $\langle 0, \varepsilon_1, \dots, \varepsilon_k \rangle$ , so that references to edges and vertices of the graph for  $T$  may be used interchangeably. An edge (stem) connecting the root  $\langle 0 \rangle$  to an artificial vertex  $\langle -1 \rangle$  may be added in this regard.

Next we choose an i.i.d. family of random variables  $\{W(0), W(0, \varepsilon_1, \dots, \varepsilon_k) : \varepsilon_i \geq 1, 1 \leq i \leq k, k \geq 1\}$ , which is independent of  $T$ .  $W(\mathbf{x})$  is the weight associated with  $\langle \mathbf{x} \rangle$  if  $\langle \mathbf{x} \rangle = \langle 0, \varepsilon_1, \dots, \varepsilon_k \rangle$  is one of the edges of  $T$ . The *weighted height* of  $\langle \mathbf{x} \rangle = \langle 0, \varepsilon_1, \dots, \varepsilon_k \rangle$  then becomes

$$H(\mathbf{x}) = H_w(\mathbf{x}) = \sum_{j=1}^k W(0, \varepsilon_1, \dots, \varepsilon_j) \quad [H(0) = 0] \tag{1.6}$$

and the *weighted height of the tree*  $T$  is

$$H(T) = H_w(T) := \max_{\langle \mathbf{x} \rangle \in T} H_w(\mathbf{x}) \tag{1.7}$$

The regular height, as given in (1.5), is obtained in the special case when  $P(W(0) = 1) = 1$ .

We write

$$p_r = P(Z_1 = r | Z_0 = 1), \quad f(s) = \sum_{r=0}^{\infty} p_r s^r$$

for the offspring distribution and its generating function. Throughout we shall assume that this offspring distribution satisfies the following conditions: There exists a number  $\alpha > 0$  for which

$$f(\alpha) = \alpha f' \in (\alpha) \quad \text{and} \quad \sum_{r=0}^{\infty} r^2 p_r \alpha^r < \infty \tag{1.8}$$

$$p_1 < 1 \tag{1.9}$$

Condition (1.9) merely rules out a trivial degenerate case. As we shall indicate, the typical case in which (1.8) holds is the critical branching

process whose offspring distribution has a finite second moment ( $\alpha = 1$  in this case). A minor nuisance which needs to be discussed is periodicity. Let

$$d = \text{g.c.d.}\{r: p_r > 0\} \tag{1.10}$$

Then all  $Z_n, n \geq 1$ , have to be divisible by  $d$ , and  $v$  can take only values congruent to  $1 \pmod d$ .

The following theorem is our main result.

**Theorem 1.** Let the branching process  $\{Z_n\}$  satisfy (1.8) and (1.9). Assume further that

$$0 < \mu = EW(0) < \infty \tag{1.11}$$

and

$$y^2 P(|W(0)| > y) \rightarrow 0, \quad \text{as } y \rightarrow \infty \tag{1.12}$$

Then,

$$\begin{aligned} \lim_{\substack{n \rightarrow \infty \\ n \equiv 1(d)}} P\left(\frac{C_\alpha}{\mu\sqrt{n}} H_w(T) \leq x \mid v(T) = n\right) \\ = \sum_{k=-\infty}^{\infty} \{1 - (kx)^2\} e^{-k^2x^2} = P(\sqrt{2}M^*_+ \leq x) \end{aligned} \tag{1.13}$$

where

$$C_\alpha = \left\{ \frac{\alpha^2}{f(\alpha)} \sum_{r=0}^{\infty} r(r-1) p_r \alpha^r \right\}^{1/2}$$

and  $M^*_+$  is the maximum of a Brownian excursion.

The special case of this theorem when  $W(0) = 1$  w.p.1 (i.e., when  $H_w$  is replaced by  $h$ , and  $\mu = 1$ ) is due to Kolchin.<sup>(17)</sup> Our contribution is to show that under conditions (1.11) and (1.12) the particular form of the distribution has no influence, but that the limit distribution in (1.13) is entirely governed by  $T$ . In fact  $H_w(T)$  behaves like  $\mu h(T)$  for our purposes. This is made explicit in the following proposition in which  $T$  does not have to come from a branching process at all.

**Proposition 1.** Let  $T_n$  be any sequence of trees with

$$v(T_n) = n \quad \text{and} \quad h(T_n) = O(\sqrt{n})$$

If (1.11) and (1.12) hold, then

$$\frac{H(T_n) - \mu h(T_n)}{\sqrt{n}} \rightarrow 0 \quad \text{in probability} \tag{1.14}$$

It seems somewhat strange that the limit distribution in (1.13) would be that of  $\sqrt{2}M_+^*$  [cf. Durrett and Iglehart<sup>(5)</sup> for the last equality in (1.13)]. For this reason we discuss in Section 2 the special case of a geometric offspring distribution (and constant weights). In this case we can use a simple mapping which goes back to Harris<sup>(10)</sup> to equate the height  $h(T)$  with a functional of a simple symmetric random walk. Although it was not used there, the Harris coding was noted in de Bruijn, Knuth, and Rice<sup>(4)</sup> in connection with the expected tree height. In any case, we shall see precisely how this coding explains why the distribution of  $M_+^*$  crops up.

Almost the same method of proof as used for Proposition 1 allows us to deal with the case where  $W(0)$  has heavier tails than in (1.2). We restrict ourselves to regularly varying tails where (1.11) and (1.12) are replaced by

$$P(W(0) > y) \sim y^{-\alpha}L(y), \quad y \rightarrow \infty \tag{1.15}$$

for a slowly varying function  $L(\cdot)$  and  $0 \leq \alpha < 2$ . (The case  $\alpha = 2$  is a boundary case; see Remark 3 below.) We shall see that, in contrast to Theorem 1, under (1.15) the branching tree  $T$  has no influence on the conditional limit distribution of  $H_w(T)$ . This limit distribution is simply the same as for the maximum of  $n$  i.i.d. weights associated with the edges of  $T$ . Specifically we prove the following theorem.

**Theorem 2.** Let the branching process  $\{Z_n\}$  satisfy (1.8) and (1.9). If  $W(0)$  satisfies (1.15) for some  $0 \leq \alpha < 2$ , and  $c_n$  satisfies

$$nc_n^{-\alpha}L(c_n) \sim 1, \quad n \rightarrow \infty$$

then

$$\lim_{\substack{n \rightarrow \infty \\ n \equiv 1(d)}} P(c_n H_w(T) \leq y \mid v(T) = n) = \exp\{-y^{-\alpha}\}, \quad (y \geq 0)$$

*Remark 1.* Under condition (1.8) one can introduce the branching process  $\{Z_n^{(\alpha)}\}$  with  $Z_0^{(\alpha)} = 1$  and offspring distribution

$$p_r^{(\alpha)} = P(Z_1^{(\alpha)} = r \mid Z_0^{(\alpha)} = 1) = \frac{p_r \alpha^r}{f(\alpha)}$$

This branching process is critical, and its family tree has the same conditional distribution, given  $\sum Z_1^{(\alpha)} = n$ , as the tree  $T$ , given  $v(T) = n$ ; cf.

Kennedy.<sup>(15)</sup> This is the reason why we called a critical branching process the typical case for our purposes.

*Remark 2.* Gupta, Mesa, and Waymire<sup>(8)</sup> consider the special case when  $W(0)$  is exponentially distributed and the tree is binary (i.e.,  $p_0 = p_2 = \frac{1}{2}$ ). The proof for this case is specialized in that it exploits the Markovian structure of the branching process. The proof does not depend on the result for (constant) unit weights. In fact, using the comparison methods of the present paper one may also derive (for binary trees) Kolchin's<sup>(17)</sup> theorem for constant weights, as well as our more general theorems (1.1) and (1.3), from the exponential case.

*Remark 3.* If  $W(0)$  has a tail as in (1.15), but with  $\alpha = 2$ , then Theorem 2 does not apply as stated, while Theorem 1 applies only if  $L(y) \rightarrow 0$  as  $y \rightarrow \infty$ . If (1.15) holds with  $\alpha = 2$  and  $L(y) \rightarrow \infty$  as  $y \rightarrow \infty$ , then Theorem 2 still holds. However, there is a "middle weight" case, namely, the borderline case, where

$$P(W(0) > y) \sim cy^{-2}, \quad y \rightarrow \infty$$

for some  $c > 0$ . In this case the limit in the left-hand side of (1.13) still exists, but it can no longer be easily calculated explicitly. It involves both  $n^{-1/2} \max_{\langle \mathbf{x} \rangle \in T} W(\mathbf{x})$  and the structure of  $T$ .

*Remark 4.* Recently some other interesting connections between trees, branching processes, and excursions have been found with applications to *Brownian local times* and to the construction of *superprocesses*; see Neveu and Pitman<sup>(20),(21)</sup> and LeGall.<sup>(18),(19)</sup> However, the result described in Section 2 is the only case known to us which exploits such structure to solve the extremal problem. In particular, the binary case is still not known as a direct consequence of a sample path analysis. These and a variety of other ideas are considered more generally in recent papers of Aldous.<sup>(1),(2)</sup>

## 2. GEOMETRIC OFFSPRING DISTRIBUTION

This section is not required for the main results described in the introduction, and can be skipped by readers interested only in the proofs of those results. As pointed out in the introduction, a general calculation of the limit distribution for the case of constant weights is given in Kolchin.<sup>(17)</sup>

We suppose that the offspring distribution is given by  $p_j := P(Z_1 = j | Z_0 = 1) = r^j(1 - r)$ ,  $j = 0, 1, 2, \dots$  ( $0 < r < 1$ ), and that  $W(x) = 1$  w.p.1. Let  $\{S_n\}$  be the simple symmetric random walk starting at 0, and let  $\gamma = \min\{k \geq 1 : S_k = 0\}$ . Define a (*signed*) *random walk excursion*,

$\{X_t^{(n)}: 0 \leq t \leq 1\}$ , as the process  $\{n^{-1/2}S_{[nt]}: 0 \leq t \leq 1\}$  conditioned on  $\{\gamma = n\}$ . Let  $P_n$  denote the distribution of  $\{X_t^{(n)}\}$ . For a random variable  $X$  and an event  $A$ , let  $(X|A)$  represent a random variable distributed according to the conditional distribution of  $X$  given  $A$ . We will see that the limit distribution of Theorem 1 is easily obtained by an application of the encoding of trees as random walks given by Harris (Ref. 10, Theorem 5, p. 485). Note that our trees are *ordered* by the subtree structure in the sense that a tree in which, for example, the first-generation offspring  $\langle 0, 1 \rangle$  has 5 children and first-generation offspring  $\langle 0, 2 \rangle$  has 7 children is distinct from the tree in which  $\langle 0, 1 \rangle$  has 7 children and  $\langle 0, 2 \rangle$  has 5. Here is a description of an encoding which can be used. Add an artificial vertex  $\langle -1 \rangle$  to  $T$ , which is connected by an edge to  $\langle 0 \rangle$  but not to any other vertex. Order the vertices lexicographically with  $\langle -1 \rangle$  at the beginning. For example, the tree of Fig. 1 yields the list  $\langle -1 \rangle, \langle 0 \rangle, \langle 0, 1 \rangle, \langle 0, 1, 1 \rangle, \langle 0, 1, 1, 1 \rangle, \langle 0, 1, 2 \rangle, \langle 0, 1, 3 \rangle, \langle 0, 2 \rangle$ . Now assign to  $T$  a random walk path by drawing a closed polygonal path going around  $T$  clockwise as in Fig. 1. Start this polygonal path at  $\langle -1 \rangle$  and go to  $\langle 0 \rangle$ . If the path is at  $\langle 0, \varepsilon_1, \varepsilon_2, \dots, \varepsilon_k \rangle$  and  $\langle 0, \varepsilon_1, \dots, \varepsilon_k, j \rangle$  is the first child of  $\langle 0, \varepsilon_1, \dots, \varepsilon_k \rangle$  (in the lexicographical ordering) not yet "surrounded" by the path, then move to  $\langle 0, \varepsilon_1, \varepsilon_2, \dots, \varepsilon_k, j \rangle$ . If there are no such children, then the path moves back to  $\langle 0, \varepsilon_1, \varepsilon_2, \dots, \varepsilon_{k-1} \rangle$ . We continue until all vertices of  $T$  have been surrounded and the path returns to  $\langle 0 \rangle$ . From there we go to  $\langle -1 \rangle$ . The random walk path corresponding to  $T$  will take a step of  $+1$  (resp.  $-1$ ) whenever the polygonal path moves from some vertex  $\langle 0, \varepsilon_1, \varepsilon_2, \dots, \varepsilon_k \rangle$  to a child  $\langle 0, \varepsilon_1, \varepsilon_2, \dots, \varepsilon_{k+1} \rangle$  (resp. to its parent  $\langle 0, \varepsilon_1, \varepsilon_2, \dots, \varepsilon_{k-1} \rangle$ ). The first and last steps of the random walk are a  $+1$  and a  $-1$  corresponding to the moves of the polygonal path from  $\langle -1 \rangle$  to  $\langle 0 \rangle$  and back from  $\langle 0 \rangle$  to  $\langle -1 \rangle$ . One easily sees that there are two steps in the polygonal path for each edge of  $T$ , one giving rise to a step of  $+1$  by the random walk and one to a step of  $-1$ . Thus the image will be a random walk path of length  $2v(T)$ , which has  $\gamma$  equal to  $2v(T)$  and is strictly positive at times  $1, 2, \dots, 2v(T) - 1$ . Each such random walk path is the image of a unique labeled and rooted tree. From this it is not hard to obtain the next result.

**Proposition 2.** Given  $v(T) = n$  each of the  $(1/n) \binom{2n-2}{n-1}$  sample realizations of  $T$  have equal probability. Moreover,

$$(n^{-1/2}h(T) | v(T) = n) = \sqrt{2} \left( \max_{0 \leq t \leq 1} \{|X_t^{(2n)}|\} | \gamma = 2n \right)$$

where equality is in distribution.



The proof of the next theorem will now follow as an application of the above proposition and some weak convergence theory for conditioned random walks.

**Theorem 3.** For the geometric offspring distribution with  $0 < r < 1$ ,

$$\lim_{n \rightarrow \infty} P(n^{-1/2}h(T) \leq y \mid v(T) = n) = P\left(M_+^* \leq \frac{y}{\sqrt{2}}\right)$$

*Proof.* We simply use the fact that,

$$P_n \Rightarrow P^* \tag{2.1}$$

where  $\Rightarrow$  denotes weak convergence and  $P^*$  is the distribution of a signed Brownian excursion  $\{B_t^*\}$ ; e.g., see Kaigh,<sup>(14)</sup> Kennedy,<sup>(16)</sup> or Durrett, Iglehart, and Miller.<sup>(6)</sup> The result now follows from Proposition 2 since  $\max |\cdot|$  is a continuous functional of the excursion.  $\square$

### 3. LIGHT WEIGHTS (PROOF OF THEOREM 1.1)

Throughout this section  $\{W(\mathbf{x})\}$  are i.i.d. random variables, independent of  $T$  (i.e., the branching process), which satisfy (1.11) and (1.12). We begin with the following:

*Proof of Proposition 1.* Let  $\langle \mathbf{x}_n \rangle$  be a vertex of  $T_n$  such that

$$h(\mathbf{x}_n) = H(T_n) = \max_{\mathbf{x} \in T_n} h(\mathbf{x})$$

If  $\mathbf{x}_n = \langle 0, \varepsilon_1, \dots, \varepsilon_h \rangle$  with  $h = h(T_n)$ , then

$$H_w(T_n) \geq \sum_{j=1}^h W(0, \varepsilon_1, \dots, \varepsilon_j)$$

Since  $h(T_n)$  is assumed to be  $O(\sqrt{n})$ , it follows from the weak law of large numbers that

$$P(H_w(T_n) - \mu h(T_n) \leq -\varepsilon\sqrt{n}) \leq P\left(\sum_{j=1}^h \{W(0, \varepsilon_1, \dots, \varepsilon_j) - \mu\} \leq -\varepsilon\sqrt{n}\right) \rightarrow 0 \tag{3.1}$$

for every  $\varepsilon > 0$ .

This proves a lower bound for  $H(T_n)$ . To obtain an upper bound we will use a version of Chebyshev's inequality. As a warmup, suppose the distribution of  $W(0)$  has a finite eighth moment. Then

$$P(H_w(T_n) - \mu h(T_n) \geq \varepsilon\sqrt{n}) \leq \sum_{\mathbf{x} \in T_n} P([H(\mathbf{x}) - \mu h(\mathbf{x})] \geq \varepsilon\sqrt{n}) \tag{3.2}$$

This sum contains  $\nu(T_n) = n$  terms, and for  $\mathbf{x} = \langle 0, \varepsilon_1, \dots, \varepsilon_k \rangle$  we have for suitable constants  $C_i < \infty$

$$\begin{aligned}
 P(H_w(\mathbf{x}) - \mu h(\mathbf{x}) \geq \varepsilon \sqrt{n}) &= P\left(\sum_{j=1}^k [W(0, \varepsilon_1, \dots, \varepsilon_j) - \mu] \leq \varepsilon \sqrt{n}\right) \\
 &\leq \varepsilon^{-8} n^{-4} E \left\{ \left( \sum_{j=1}^k [W(0, \varepsilon_1, \dots, \varepsilon_j) - \mu] \right)^8 \right\} \\
 &\leq C_1 \varepsilon^{-8} n^{-4} k^4 E[W(0) - \mu]^8 \\
 &\leq C_2 \varepsilon^{-8} n^{-4} \{h(T_n)\}^4 \\
 &\leq C_3 \varepsilon^{-8} n^{-2}
 \end{aligned} \tag{3.3}$$

Thus (3.2) tends to zero in this case.

In the general case we shall first use a truncation before we apply Chebyshev's inequality. Choose  $b_n$  such that

$$\frac{b_n}{\sqrt{n}} \rightarrow 0, \quad nP(|W(0)| > b_n) \rightarrow 0 \tag{3.4}$$

Such constants  $b_n$  exist by virtue of (1.12). We then have [with  $\mathbf{1}(A)$  denoting the indicator function of  $A$ ]

$$P(W(\mathbf{x}) \neq W(\mathbf{x}) \mathbf{1}(|W(\mathbf{x})| \leq b_n) \text{ for some } \mathbf{x} \in T_n) \rightarrow 0 \tag{3.5}$$

so that one may replace  $W(\mathbf{x})$  by  $W(\mathbf{x}) \mathbf{1}(|W(\mathbf{x})| \leq b_n)$  for the proof of (1.14). This, however, will not be a good enough truncation. We shall have to replace the  $W(\mathbf{x})$  by

$$W^{(n)}(\mathbf{x}) := W(\mathbf{x}) \mathbf{1}(|W(\mathbf{x})| \leq a_n)$$

for a sequence of constant  $a_n \leq b_n$ ,  $a_n \rightarrow \infty$ , which in general does not satisfy the second requirement of (3.4). Consequently (3.5) with  $b_n$  replaced by  $a_n$  will not be valid. However, by using the fact that  $T_n$  is a tree we will be able to obtain

$$P(\exists \langle 0, \varepsilon_1, \dots, \varepsilon_k \rangle \in T_n, j_1 < j_2 \leq k \text{ s.t. } |W(0, \varepsilon_1, \dots, \varepsilon_{j_i})| > a_n, i = 1, 2) \rightarrow 0 \tag{3.6}$$

But

$$\sum_{j=1}^k W(0, \varepsilon_1, \dots, \varepsilon_j) \leq \sum_{j=1}^k W^{(n)}(0, \varepsilon_1, \dots, \varepsilon_j) + \max_{\mathbf{x} \in T} W(\mathbf{x}) \tag{3.7}$$

for any  $\langle 0, \varepsilon_1, \dots, \varepsilon_k \rangle$  for which there is at most one  $j \leq k$  with  $|W(0, \varepsilon_1, \dots, \varepsilon_j)| > a_n$ . Therefore,

$$\begin{aligned}
 P(H(T_n) - \mu h(T_n) \geq \varepsilon \sqrt{n}) &\leq P(|W(\mathbf{x})| > b_n, \text{ for some } \mathbf{x} \in T_n) \\
 &+ P(\exists \langle 0, \varepsilon_1, \dots, \varepsilon_k \rangle \in T_n, j_1 < j_2 \leq k \text{ s.t. } |W(0, \varepsilon_1, \dots, \varepsilon_{j_i})| > a_n, i = 1, 2) \\
 &+ \sum_{\langle 0, \varepsilon_1, \dots, \varepsilon_k \rangle} P\left(\sum_{j=1}^k [W^{(n)}(0, \varepsilon_1, \dots, \varepsilon_j) - \mu] \geq \varepsilon \sqrt{n} - b_n\right) \\
 &\leq o_n(1) + nP\left(\sum_{j=0}^k [W_j^{(n)} - E\{W_j^{(n)}\}] \geq \frac{\varepsilon}{2} \sqrt{n}\right) \tag{3.8}
 \end{aligned}$$

where  $W_1^{(n)}, W_2^{(n)}, \dots$  are i.i.d. with the distribution of  $W(0) \mathbf{1}(|W(0)| \leq a_n)$ . [In the last inequality we used the fact that

$$E\{W_i^{(n)} - \mu\} \rightarrow 0$$

and  $k \leq h(T_n) = O(\sqrt{n})$ ,  $b_n = o(\sqrt{n})$ .] In view of (3.8) it therefore suffices to show that for  $a_n = n^{3/8}$ , (3.6) is valid, and the right-hand side of (3.8) tends to 0.

To start with (3.6), its left-hand side is bounded by

$$\begin{aligned}
 P(\exists \langle 0, \varepsilon_1, \dots, \varepsilon_{j_2} \rangle \in T_n, j_1 < j_2, \text{ s.t. } |W(0, \varepsilon_1, \dots, \varepsilon_{j_i})| > a_n, i = 1, 2) \\
 \leq (h(T_n) + 1) n(P(|W(0)| > a_n))^2 \\
 = o(n^{3/2} a_n^{-4})
 \end{aligned}$$

because there are at most  $v(T_n) = n$  choices for  $\langle 0, \varepsilon_1, \dots, \varepsilon_{j_2} \rangle \in T_n$ , and once we have picked  $\langle 0, \varepsilon_1, \dots, \varepsilon_{j_2} \rangle$ , there are at most  $(j_2 + 1) \leq h(T_n) + 1$  choices for  $j_1$ . Thus (3.6) holds.

To estimate the last term in the right-hand side of (3.7), we slightly sharpen (3.3). Note that

$$\begin{aligned}
 \xi_r(n) &:= E\{|W^{(n)}(0) - E\{W^{(n)}(0)\}|^r\} \\
 &\leq C(r) \int_0^{a_n} y^{r-1} P(|W(0)| \geq y) dy \\
 &= O(a_{r-2} \log a_n) \tag{3.9}
 \end{aligned}$$

for  $r \geq 2$  and suitable constants  $C(r)$ . It follows that for some  $C < \infty$

$$\begin{aligned}
 E\left\{\sum_{j=1}^k [W^{(n)}(0, \varepsilon_1, \dots, \varepsilon_j) - E\{W^{(n)}(0)\}]^8\right\} \\
 \leq C\{k^4 \xi_4^4(n) + k^3(\xi_2^2(n) \xi_4(n) + \xi_3^2(n) \xi_2(n)) \\
 + k^2(\xi_4^2(n) + \xi_6(n) \xi_2(n) + \xi_5(n) \xi_3(n)) + k \xi_8(n)\} \tag{3.10}
 \end{aligned}$$

Since  $k \leq h(T_n) = O(\sqrt{n})$  and  $a_n = n^{3/8}$ , the right-hand side of (3.10) is  $o(n^{11/4} \log n)$ . As in (3.3) we now obtain that the last term in the right-hand side of (3.7) is  $o(n^{-3+11/4} \log n) \rightarrow 0$ . Thus Proposition 1 follows.  $\square$

Theorem 1 is now immediate from Proposition 1 and the special case of Theorem 1 when  $P(W(0) = 1) = 1$  (which amounts to replacing  $H$  by  $h$  and  $\mu = 1$ ) which was proved by Kolchin.<sup>(17)</sup>

#### 4. HEAVY WEIGHTS (PROOF OF THEOREM 1.3)

Throughout this section  $\{W(\mathbf{x})\}$  are i.i.d. random variables, independent of  $T$  (i.e., the branching process), which satisfy (1.15) for some  $0 \leq \alpha < 2$  and a slowly varying function  $L(\cdot)$ , i.e.,  $L(ty)/L(t) \rightarrow 1$  as  $t \rightarrow \infty$ . For the appropriate selection of scale coefficients  $c_n$  such that

$$nP(W(\mathbf{x}) > c_n) = nc_n^{-\alpha}L(c_n) \rightarrow 1, \quad \text{as } n \rightarrow \infty \tag{4.1}$$

one has (see Feller,<sup>(7)</sup> p. 278)

$$c_n^{-1} \max_{\langle \mathbf{x} \rangle \in T} W(\mathbf{x}) \Rightarrow F_\infty(y) := \exp\{-y^{-\alpha}\} \quad (y \geq 0), \quad \text{as } y \rightarrow \infty \tag{4.2}$$

where  $v(T) = n$ .

We will show, as asserted by Theorem 2, that in this case the distribution of the heaviest path in the tree is distributed as the largest weight in the tree. In the following estimates the specific values of the slowly varying factor may change from one line to another; however, to make the notation manageable we shall recycle the symbol  $L$  whenever convenient.

Again it will be more convenient to consider an arbitrary sequence of trees  $T_n$  satisfying

$$v(T_n) = n \quad \text{and} \quad h(T_n) = O(\sqrt{n}) \tag{4.3}$$

We base our proof on (3.6) and (3.7) here also. The strategy is to select truncation levels  $a_n \rightarrow \infty$  in such a way that the trimmed height coincides with the truncated height of a path with (high) probability  $1 - o(1)$  as  $n \rightarrow \infty$ . The truncation is also chosen so that for trees of height  $h(T_n) = O(\sqrt{n})$ , the largest of these truncated heights is  $o(c_n)$  in probability, whereas the maximum weight over the tree is of order  $c_n$ , and scales according to (4.2). In particular, as already noted in Section 3,

$$\begin{aligned} P(\exists \langle 0, \varepsilon_1, \dots, \varepsilon_{j_i} \rangle \in T_n, j_1 < j_2, \text{ s.t. } |W(0, \varepsilon_1, \dots, \varepsilon_{j_i})| > a_n, i = 1, 2) \\ \leq (h(T_n) + 1) n(P(|W(0)| > a_n))^2 \\ = O(n^{3/2}[a_n^{-\alpha}L(a_n)]^2) \end{aligned}$$

Thus, by virtue of Feller (Ref. 7, Lemma 2, p. 277), (3.6) holds for

$$a_n = n^\rho \quad \text{when} \quad \frac{3}{4\alpha} < \rho < \frac{1}{\alpha} \tag{4.4}$$

The next thing to check is that the largest of the truncated heights  $H_w^{(n)}(T_n) := \max_{\langle 0, \varepsilon_1, \dots, \varepsilon_k \rangle \in T_n} \sum_{j=1}^k W^{(n)}(0, \varepsilon_1, \dots, \varepsilon_j)$ , is  $o(c_n)$  for this truncation level. Again using Lemma 2 on p. 277 of Feller,<sup>(7)</sup> it is easy to check that the tree height  $h(T_n)$  [being  $O(n^{1/2})$ ] is  $o(c_n)$ , since  $c_n = n^{1/\alpha}L(n)$ ; use Theorem 1.5.12, p. 28, of Bingham, Goldie, Teugels.<sup>(3)</sup> Also, one has  $\mu_n := EW^{(n)}(0)$  is bounded for  $\alpha > 1$ , while  $\mu_n$  is regularly varying (in  $a_n$ ) with exponent  $1 - \alpha$  for  $0 \leq \alpha \leq 1$ . In either case, therefore,  $\mu_n h(T_n)$  is  $o(c_n)$  for  $3/4\alpha < \rho < 1/\alpha$ . With this we again need only estimate the centered values  $H_w^{(n)}(T_n) - \mu_n h(T_n)$ ; i.e., if the centered truncated height is  $o(c_n)$ , then so will be the truncated height. The following lemma provides the suitable choice of  $\rho$  for this, [consistent with (4.4)].

**Lemma 1.** There is a  $0 < \theta < 1/4\alpha$  such that for  $\rho = 3/4\alpha + \theta < 1/\alpha$ ,

$$P(|H_w^{(n)}(T_n) - \mu_n h(T_n)| > \varepsilon c_n) = o(1)$$

for each  $\varepsilon > 0$ .

*Proof.* All of the estimates using Chebyshev's inequality in (3.10) apply with  $\rho$  replacing the  $\frac{3}{8}$  there. In particular one obtains the upper bound

$$\begin{aligned} & \max_{\mathbf{x} \in T_n} P(|H_w^{(n)}(\mathbf{x}) - \mu_n h(\mathbf{x})| > \varepsilon c_n) \\ & \leq n \frac{1}{\varepsilon^8 c_n^8} \max_{\langle 0, \varepsilon_1, \dots, \varepsilon_k \rangle \in T_n} E \left\{ \sum_{j=1}^k [W^{(n)}(\langle 0, \varepsilon_1, \dots, \varepsilon_j \rangle) - \mu_n] \right\}^8 \\ & \leq C \frac{n}{\varepsilon^8 c_n^8} [n^2 \xi_2^4(n) + n^{3/2} \{ \xi_2^2(n) \xi_4(n) + \xi_3^2(n) \xi_2(n) \} \\ & \quad + n \{ \xi_4^2(n) + \xi_6(n) \xi_2(n) + \xi_5(n) \xi_3(n) \} + n^{1/2} \xi_8(n)] \end{aligned}$$

where by standard calculations (see Feller,<sup>(7)</sup> p. 280) one has, for  $r \geq 2$ ,

$$\xi_r(n) \leq K_r a_n^{(r-\alpha)} L(a_n) = K_r n^{(r-\alpha)\rho} L(n)$$

Let  $\eta > 0$ . Again writing  $c_n = n^{1/2}L(n)$  and using the Lemma 2, p. 277, of Feller,<sup>(7)</sup> one obtains that for all  $n$  sufficiently large the Chebyshev bound is dominated for some  $K > 0$  by

$$K n^{3/2 + (8-\alpha)\rho + \eta - 8/\alpha + 8\eta} \leq K n^{3/4 - 2/\alpha + (8-\alpha)\theta + 9\eta} = o(1)$$

for  $\theta$  and  $\eta$  sufficiently small. □

## ACKNOWLEDGMENTS

This research was partially supported by the Army Research Office through the Mathematical Sciences Institute of Cornell University. All three authors are also partially supported by NSF through grants to their respective universities. The third named author is also partially supported by ARO grant 2772-GS.

## REFERENCES

1. Aldous, D. (1991). The Continuum Random Tree I, *Ann. Prob.* **19**(1), (to be published).
2. Aldous, D. (1989). *Discrete and Continuum Random Trees: The Notebook*, The University of California at Berkeley. Preprint.
3. Bingham, N. H., Goldie, C. M., and Teugels, J. L. (1987). *Regular Variation*, Cambridge University Press, New York, pp. 15–22.
4. Brujin, N. G. de, Knuth, D. E., and Rice, S. O. (1972). The average height of planted plane trees. In Read, E. C. (ed.), *Graph Theory and Computing*, Academic Press, New York.
5. Durrett, R., and Iglehart, D. L. (1977). Functions of Brownian meander and Brownian excursion. *Ann. Prob.* **5**, 130–135.
6. Durrett, R., Iglehart, D. L., and Miller, D. (1977). Weak convergence to Brownian meander and Brownian excursion. *Ann. Prob.* **5**, 117–129.
7. Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, Vol. II, John Wiley and Sons, New York.
8. Gupta, V. K., Mesa, O., and Waymire, E. (1990). Tree dependent extreme values: The exponential case. *J. Appl. Prob.* **27**, 124–133.
9. Gupta, V. K., and Waymire, E. (1991). The spatial statistics of random networks and a problem in river basin hydrology. In: *Spatial Statistics and Imaging*, A. Possolo (ed.), *IMS Lectures Notes-Monograph Series*, (to be published).
10. Harris, T. E. (1956). First passage and recurrence distributions. *Trans. Amer. Math. Soc.* **73**, 471–486.
11. Harris, T. E. (1963). *The Theory of Branching Processes*, Springer-Verlag, New York.
12. Jagers, P. (1975). *Branching Processes with Biological Applications*, John Wiley and Sons, New York.
13. Jarvis, R. S., and Woldenberg, M. J. (1984). *River Networks*, Benchmark Papers in Geology, Vol. 80, Hutchinson Ross, Stroudburg, Pennsylvania.
14. Kaigh, W. (1975). A conditional local limit theorem for recurrent random walk. *Ann. Prob.* **3**, 883–888.
15. Kennedy, D. P. (1975). The Galton-Watson process conditioned on the total progeny. *J. Appl. Prob.* **12**, 800–806.
16. Kennedy, D. P. (1976). The distribution of the maximum Brownian excursion. *J. Appl. Prob.* **13**, 371–376.
17. Kolchin, V. (1978). Moment of degeneration of a branching process and height of a random tree. *Math. Notes* **24**, 954–961.
18. LeGall, J.-F. (1989a). Une construction trajectorielle de certains processus de Markov a valeurs mesures. *C. R. Acad. Sci. Paris* **308**, 533–538.

19. LeGall, J.-F. (1989b). Marches aleatoires, mouvement Brownien et processus de branchement. In *Seminaire de Probabilites XXIII*, Springer Lecture Notes in Mathematics, Springer-Verlag, Berlin. Preprint.
20. Neveu, J., and Pitman, J. (1989a). The branching process in a Brownian excursion. In *Seminaire de Probabilites XXIII*, Lecture Notes in Mathematics, Springer-Verlag, Berlin. Preprint.
21. Neveu, J., and Pitman, J. (1989b). Renewal property of the extrema and tree property of a one-dimensional Brownian motion. In *Seminaire de Probabilites XXIII*, Lecture Notes in Mathematics, Springer-Verlag, Berlin. Preprint.