TIGHTNESS FOR THE MINIMAL DISPLACEMENT OF BRANCHING RANDOM WALK

MAURY BRAMSON* OFER ZEITOUNI§

Abstract. Recursion equations have been used to establish weak laws of large numbers for the minimal displacement of branching random walk in one dimension. Here, we use these equations to establish the tightness of the corresponding sequences after appropriate centering. These equations are special cases of recursion equations that arise naturally in the study of random variables on tree-like structures. Such recursion equations are investigated in detail, in [BZ06], in a general context. Here, we restrict ourselves to investigating the more concrete setting of branching random walk, and provide motivation for the rigorous arguments that are given in [BZ06]. We also discuss briefly the cover time of symmetric simple random walk on regular binary trees, which is another application of the more general recursion equations.

1. Introduction and Statement of Results

Consider the following branching random walk (BRW) on $\mathbb{R}$. A particle starting at 0 is assumed to move randomly to a site according to a given distribution function $G(\cdot)$. At this time, it dies and gives birth to k offspring with probability p_k, independently of the previous motion. Each of these offspring, in turn, moves independently according to the same distribution $G(\cdot)$ over the next time step, then dies and gives birth to k offspring according to the distribution $\{p_k\}$. This procedure is repeated at all integer times, with the movement of all particles and the number of offspring all being assumed to be independent of one another.

To avoid the possibility of extinction and trivial special cases, we assume that $p_0 = 0$ and $p_1 < 1$. This implies that the mean number of offspring $m_1 = \sum_{k=1}^{\infty} kp_k > 1$, that is, the branching process is *supercritical.* The special case where the branching is binary, that is, where $p_2 = 1$, will exhibit the same basic behavior as the general case, and so the reader may wish to concentrate on it instead.

Let Z_n denote the number of particles at time n of the BRW, with $\mathfrak{x}_k(n),\ k = 1, \ldots, Z_n$, being the positions of these particles when ordered

Date: February 6, 2007. Revised April 20, 2007.
*Research partially supported by NSF grant number DMS-0226245.
§Research partially supported by NSF grant number 0503775.

in some fashion. We write

$$\mathcal{M}_n = \min_{1\le k\le Z_n} \mathfrak{x}_k(n) \tag{1.1}$$

for the *minimal displacement* of the BRW at time n. When $G(0) = 0$, one can alternatively interpret $G(\cdot)$ as the lifetime distribution of individual particles of the branching process. In this setting, $\mathcal{M}_n$ becomes the first birth time of the n-th generation of particles.

Let $F_n(\cdot)$ denote the distribution function of $\mathcal{M}_n$. One can express $F_n(\cdot)$ recursively in terms of $G(\cdot)$ and

$$Q(u) = 1 - \sum_k p_k(1-u)^k \quad \text{for } u \in [0,1]. \tag{1.2}$$

One has

$$F_{n+1}(x) = (Q(F_n) * G)(x) = \int_{y\in\mathbb{R}} Q(F_n(x-y))dG(y), \tag{1.3}$$

with $F_0(x) = 1_{\{x\ge 0\}}$. Here, $*$ denotes the standard convolution.

Equation (1.3) is the backwards equation for $F_{n+1}(\cdot)$ in terms of $F_n(\cdot)$. It is simple to derive by relating it to the minimal displacement of the n-th generation offspring for each of the first generation offspring of the original particle. The composite function $Q(F_n)$ gives the distribution of the minimum of these individual minimal displacements (relative to their parents), with convolution by $G(\cdot)$ then contributing the common displacement due to the movement of the original particle. In the special case where $p_2 = 1$, (1.2) reduces to $Q(u) = 2u - u^2$. We note that $Q : [0,1] \to [0,1]$ is strictly concave, in general, with

$$Q(0) = 0, \quad Q(1) = 1 \quad \text{and} \quad Q'(0) = m_1 > 1. \tag{1.4}$$

One can equally well assume that branching for the BRW occurs at the beginning of each time step, before the particles move rather than after. This alternative format has often been employed in the literature. The corresponding distribution functions $F_n^r(\cdot)$ then satisfy the analog of (1.3),

$$F_{n+1}^r = Q(G * F_n^r). \tag{1.5}$$

Since $F_1 = G * F_0$, one has $F_n^r = Q(F_n)$ for all n; $\{F_n\}$ and $\{F_n^r\}$ will therefore have the same asymptotic behavior. The distribution functions $F_n^r(\cdot)$ of the minimal displacement of this BRW were studied in [H74].

It follows from [H74,Th.2], that for appropriate γ_0,

$$F_n^r(\gamma n) \to \begin{cases} 0 & \text{for } \gamma < \gamma_0, \\ 1 & \text{for } \gamma > \gamma_0, \end{cases} \tag{1.6}$$

as $n \to \infty$, provided $G(\cdot)$ has finite mean and its support is bounded below. (Related results were proved in [Ki75] and [Ki76], and by H. Kesten (unpublished).) In his debate with Kingman on the proper postulates for subadditivity, Hammersley incorrectly stated that the minimal displacement $\mathcal{M}_n^r$ was subadditive in the sense of his postulates S_1, S_2' and S_3'. (See the correction in Remark 9 of the appendix in [H74].) Sufficiently broad postulates that include $\mathcal{M}_n^r$ were given in [Li85], whose subadditive ergodic theorem demonstrated the strong law analog of (1.6). Analogous laws of large numbers hold for $F_n(\cdot)$ and $\mathcal{M}_n$. Here, we will investigate the refined behavior of $F_n(\cdot)$.

There is an older, related theory of branching Brownian motion (BBM). Individual particles are assumed to execute standard Brownian motion on $\mathbb{R}$. Starting with a single particle at 0, particles die after independent rate-1 exponentially distributed holding times, at which point they give birth to k offspring with distribution $\{p_k\}_{k\geq 1}$. All particles are assumed to move independently of one another and of the number of offspring at different times, which are themselves independent.

The minimal displacement

$$\mathcal{M}_t = \min_{1\leq k\leq Z_t} \mathfrak{x}_k(t) \tag{1.7}$$

is the analog of (1.1), where, as before, Z_t and $\mathfrak{x}_k(t)$, $k = 1, \dots, Z_t$, are the number of particles and their positions at time t. It is not difficult to show (see, e.g., [M75]), that the distribution function $u(t,x) = P(\mathcal{M}_t \leq x)$ satisfies

$$u_t = \frac{1}{2}u_{xx} + f(u), \tag{1.8}$$

with

$$f(u) = Q(u) - u \tag{1.9}$$

and $u(0,x) = 1_{\{x\geq 0\}}$. When the branching is binary, $f(u) = u(1-u)$. The literature for BBM often treats the *maximal displacement* $\mathcal{M}_n^{\max} = \max_{1\leq k\leq Z_n} \mathfrak{x}_k(n)$ rather than the minimal displacement. (Questions about $\mathcal{M}_n$ or $\mathcal{M}_n^{\max}$ can be rephrased as questions about the other by substituting $-x$ for the coordinate x and reflecting $G(\cdot)$ accordingly.) Here, we choose to employ the minimal displacement for comparison with BRW.

When $f(\cdot)$ is continuously differentiable and satisfies the more general equation

$$f(0) = f(1) = 0, \quad f(u) > 0, \quad f'(u) \leq f'(0), \quad \text{for } u \in (0,1), \tag{1.10}$$

(1.8) is typically either referred to as the *K-P-P equation* or the *Fisher equation*. For solutions $u(t,x)$ of (1.8) with $u(0,x) = \mathbf{1}_{\{x\geq 0\}}$, $u(t,\cdot)$ will be a distribution function for each t. In both [KPP37] and [F37], (1.8) was employed to model the spread of an advantageous gene through a population.

In [KPP37], it was shown that, under (1.10) and $u(0,x) = \mathbf{1}_{\{x\geq 0\}}$, the solution of (1.8) converges to a travelling wave $w(x)$, in the sense that for appropriate $m(t)$,

$$u(t, x + m(t)) \to w(x) \quad \text{as } t \to \infty \tag{1.11}$$

uniformly in x, where $w(\cdot)$ is a distribution function for which $\tilde{u}(t,x) = w(x+\sqrt{2}\,t)$ satisfies (1.8). Moreover,

$$m(t)/t \to -\sqrt{2} \quad \text{as } t \to \infty. \tag{1.12}$$

(The centering term $m(t)$ can be chosen so that $u(t,m(t)) = c$, for any given $c \in (0,1)$, on $t > 0$.) In particular,

$$u(t,\gamma t) \to \begin{cases} 0 & \text{for } \gamma < -\sqrt{2}, \\ 1 & \text{for } \gamma > -\sqrt{2}, \end{cases} \tag{1.13}$$

as $t \to \infty$, which is the analog of (1.6).

A crucial step in the proof of (1.11) consists of showing, for $m(t)$ chosen so $u(t,m(t)) = 1/2$, that

$$\begin{aligned} &u(t,x+m(t)) \text{ is increasing in } t \text{ for } x < 0, \\ &u(t,x+m(t)) \text{ is decreasing in } t \text{ for } x > 0. \end{aligned} \tag{1.14}$$

That is, $v(t,\cdot) = u(t,\cdot + m(t))$ "stretches" as t increases. A direct consequence of (1.11) and (1.14) is that the family $v(t,\cdot)$ is *tight*, that is, for each $\varepsilon > 0$, there is an $A_\varepsilon > 0$, so that for all t,

$$v(t,+A_\varepsilon) - v(t,-A_\varepsilon) > 1 - \varepsilon. \tag{1.15}$$

One can give detailed asymptotics on $m(\cdot)$ ([Br78]). In particular, in the binary case,

$$m(t) = -2^{1/2}t + 3 \cdot 2^{-3/2} \log t + O(1). \tag{1.16}$$

One can also analyze the convergence of $v(t,\cdot)$ under more general initial data ([Br83]).

Although BRW is the discrete time analog of branching Brownian motion, with (1.3) corresponding to (1.8), more refined results on the asymptotic behavior of $F_n(\cdot)$ corresponding to those of $u(t,\cdot)$ in (1.11) have, except in special cases, remained elusive. When $G(\cdot)$ is logarithmically concave, that is, $G(\cdot)$ satisfies

$$G'(x) = e^{-\varphi(x)}, \qquad \text{where } \varphi(x) \in (-\infty,\infty] \text{ is convex}, \tag{1.17}$$

one can show that the analog of (1.14) holds for $F_0(x) = \mathbf{1}_{\{x \geq 0\}}$. As in [Lu82] and [Ba00], the analog of (1.11) follows from this. Results of this nature for general $G(\cdot)$ are not known. In fact, without some modification, the analog of (1.11) will be false in general. For example, when $G(\cdot)$ is concentrated on the integers and $\gamma_0 \notin \mathbb{Z}$, it is easy to see that the analog of (1.11) cannot hold. On the positive side, in [HS07], the analog of the second term in (1.16), as well as more refined sample path asymptotics, is derived for typical BRW.

There has recently also been some interest in related problems that arise in the context of sorting algorithms, for which the movement of offspring of a common parent will be dependent. [D03] showed the analog of (1.11) for a specific choice of $G(\cdot)$. In [R03] and in [A-B07] (in the latter paper, for general $G(\cdot)$ having bounded support), $m(t)$ is calculated for related models. [CD06] treats a generalization of the model in [D03].

In [BZ06], it is shown that, after appropriate centering, the random sequence $\{\mathcal{M}_n^{\max}\}_{n\geq 0}$ corresponding to the maximal displacement of BRW is tight. In keeping with previous results given here, we restate this result in terms of $\{\mathcal{M}_n\}_{n\geq 0}$. For this, we employ the following notation. The shifted sequence $\{\mathcal{M}_n^s\}_{n\geq 0}$ is defined as

$$\mathcal{M}_n^s = \mathcal{M}_n - \mathrm{Med}(F_n), \tag{1.18}$$

where $\mathrm{Med}(F_n) = \inf\{x : F_n(x) \geq 1/2\}$ and $F_n(\cdot)$ is the distribution function of $\mathcal{M}_n$. The function $F_n^s(\cdot)$ denotes the distribution function of $\mathcal{M}_n^s$. The sequence $\{\mathcal{M}_n^s\}_{n\geq 0}$, or equivalently $\{F_n^s\}_{n\geq 0}$, is *tight* if for any $\varepsilon > 0$, there is an $A_\varepsilon > 0$ such that $F_n^s(A_\varepsilon) - F_n^s(-A_\varepsilon) > 1 - \varepsilon$ for all n. This is the analog of (1.15).

Rather than (1.17) as the main condition on $G(\cdot)$, it is assumed in [BZ06] that for some $a > 0$ and $\mathcal{M}_0 > 0$, $G(\cdot)$ satisfies

$$G(x - M) \leq e^{-aM} G(x) \quad \text{for all } x \leq 0,\ M \geq M_0. \tag{1.19}$$

In addition to specifying that $G(\cdot)$ has an exponentially decreasing left tail, (1.19) requires that $G(\cdot)$ be "flat" on no interval $[x-M, x]$, for x and M chosen as above. It follows with a little work from [H74,(3.97)], that in order for $\gamma_0 > -\infty$ to hold, the left tail of $G(\cdot)$ needs to be exponentially decreasing. The flatness condition included in (1.19) is needed for the method of proof in [BZ06]; this additional condition will be satisfied for most distributions that one encounters in practice. One also requires that the branching law for the BRW satisfy $p_1 < 1$ and

$$\sum_{k=1}^{\infty} k^\theta p_k = m_\theta < \infty \quad \text{for some } \theta \in (1, 2], \tag{1.20}$$

that is, the branching law has a finite θ-th moment.

Employing the above conditions, we now state the main result on branching random walks from [BZ06]:

Theorem 1.1. *Assume that the random walk increments $G(\cdot)$ of a BRW satisfy (1.19) and that the branching law $\{p_k\}_{k\geq 1}$ satisfies $p_1 < 1$ and (1.20). Then, the sequence of random variables $\{\mathcal{M}_n^s\}_{n\geq 0}$ is tight.*

In [BZ06], Theorem 1.1 follows quickly from a more general result there, Theorem 2.5. Theorem 2.5 assumes a more general version of the recursion equation (1.3), and is formulated so as to be applicable to other problems involving random walks on tree-like structures.

The other specific problem studied in [BZ06] is the cover time of random walk for regular binary trees. The regular binary tree $\mathbf{T}_n$ of depth n consists of the first n generations, or *levels*, of a regular binary tree. The root o denotes the original ancestor, and the 2^k vertices at the k-th level, for $k \leq n$, are its k-th generation descendants. We consider each level $k-1$ vertex to be an immediate neighbor of the two level k vertices that are immediately descended from it.

We consider a particle, which starts at the root, and executes a symmetric simple random walk on $\mathbf{T}_n$. That is, the corresponding Markov process $\{X_j\}_{j\geq 0}$ satisfies $X_0 = o$, with each neighbor being chosen with equal probability at each time step. The *cover time* $\mathcal{C}_n$ of $\mathbf{T}_n$ is the time required for the particle to visit every site in $\mathbf{T}_n$, and is given by

$$\mathcal{C}_n = \min\left\{ j \geq 0 : \bigcup_{j=0}^{j} \{X_j\} = \mathbf{T}_n \right\}.$$

The following weak law of large numbers was shown in [A91]:

$$\mathcal{C}_n / 4(\log 2) n^2 2^n \to_{n\to\infty} 1 \quad \text{in probability.} \tag{1.21}$$

A natural question to ask is how $\mathcal{C}_n$ should be scaled so that the resulting random variables, after shifting by their medians, are tight. In [BZ06], it is shown that the correct scaling is given by

$$\mathcal{E}_n = \sqrt{\mathcal{C}_n / 2^n}. \tag{1.22}$$

More precisely, defining the shift $\mathcal{E}_n^s = \mathcal{E}_n - \mathrm{Med}(\mathcal{E}_n)$ similarly to (1.18), one has:

Theorem 1.2. *The sequence of random variables $\{\mathcal{E}_n^s\}_{n\geq 0}$ for the regular binary tree is tight. Furthermore, it is non-degenerate in the sense that there exists a constant $V > 0$ such that*

$$\limsup_{n\to\infty} P(|\mathcal{E}_n^s| < V) < 1. \tag{1.23}$$

Theorem 1.2 also follows as a special case of Theorem 2.5 in [BZ06], although a non-negligible amount of additional work is required in applying the theorem (see Section 4 of [BZ06]). We note that the statement and the proof of Theorem 1.2 extend to regular k-ary trees, although the extension to Galton-Watson trees is not automatic.

Our interest in the cover time of $\mathbf{T}_n$ is motivated partially by the analogous problem on the lattice tori $\mathbb{Z}_n^2 = \mathbb{Z}^2/n\mathbb{Z}^2$. Let $\mathbf{C}_n$ denote the number of steps required for a simple random walk to cover $\mathbb{Z}_n^2$. Confirming a conjecture in [A89], it was proved in [DPRZ04], that $\pi \mathbf{C}_n / 4n^2(\log n)^2 \to 1$ in probability. The intuition (although not the details) behind the proof in [DPRZ04] draws heavily from the covering of the regular binary tree by simple random walk. We thus expect that a result similar to Theorem 1.2 should hold for $\mathbf{C}_n$, and put forward the following conjecture.

Conjecture 1.3. *The sequence of random variables*

$$\mathbf{E}_n = \sqrt{\frac{\mathbf{C}_n}{n^2}} - \mathrm{Med}\left(\sqrt{\frac{\mathbf{C}_n}{n^2}}\right) \tag{1.24}$$

is tight and non-degenerate.

The remaining two sections of this paper are devoted to showing the tightness of BRW after appropriate centering. Section 2 provides a quick proof of Theorem 1.1, under the restriction that the incremental distribution $G(\cdot)$ of the random walk have no left tail. The argument is taken from Remark 4.2 of [BZ06]. Section 3 sketches the proof of the general case of Theorem 1.1, providing the motivation behind the main steps. This restriction to Theorem 1.1 here, rather than including the more general Theorem 2.5 of [BZ06], is done with the goal of minimizing the rather cumbersome notation needed for the latter. For details of the general case, the reader is referred to Sections 2 and 3 of [BZ06].

2. Proof of Theorem 1.1 for $G(\cdot)$ having no left tail

We assume in this section that $G(\cdot)$ has no left tail, that is,

$$G(B_1) = 0 \quad \text{for some } B_1 \in \mathbb{R}. \tag{2.1}$$

For such $G(\cdot)$, there is a simple direct proof of Theorem 1.1 The basic point is that the minimal displacement of such a BRW satisfies

$$\mathcal{M}_{n+1} \geq \mathcal{M}_n + B_1 \tag{2.2}$$

pathwise, and therefore

$$F_{n+1}(x + B_1) \leq F_n(x) \quad \text{for all } x. \tag{2.3}$$

This will provide a contradiction if $\{F_n\}_{n\geq 0}$ is not tight, as we now show. Another proof is given in [DH91]. There, in addition to (2.1), it is assumed, in essence, that $G(\cdot)$ has finite mean.

Proof of Theorem 1.1 for $G(\cdot)$ satisfying (2.1). Suppose that for given $\varepsilon > 0$, n and x_1,

$$\varepsilon \leq F_n(x_1) \leq 1 - \varepsilon. \tag{2.4}$$

Then, since Q is strictly concave with $Q(0) = 0$ and $Q(1) = 1$,

$$Q(F_n(x_1)) \geq F_n(x_1) + \delta \tag{2.5}$$

for some $\delta = \delta(\varepsilon) > 0$. Choosing $B_2 = B_2(\varepsilon)$ so that $G(B_2) \geq 1 - \delta/3$, one also has

$$F_{n+1}(x + B_2) \geq Q(F_n(x)) - \delta/3 \quad \text{for all } x. \tag{2.6}$$

Together, (2.5) and (2.6) imply that

$$F_{n+1}(x_1 + B_2) \geq F_n(x_1) + 2\delta/3. \tag{2.7}$$

Assume now that $\{F_n^s\}_{n\geq 0}$ is not tight. Then, $\varepsilon > 0$ may be chosen so that for arbitrarily large A, the right side of (2.7) is at least $F_n(x_1+A)+\delta/3$ for some n and x_1 satisfying (2.4), where δ is chosen as above. Hence, setting $A = -B_1 + B_2$,

$$F_{n+1}(x_1 + B_2) \geq F_n(x_1 - B_1 + B_2) + \delta/3. \tag{2.8}$$

Setting $x = x_1 - B_1 + B_2$, (2.8) contradicts (2.3). So, $\{F_n^s\}_{n\geq 0}$ is in fact tight. □

3. Sketch of the proof of Theorem 1.1

Let $\mathcal{D}$ denote the set of distribution functions on $\mathbb{R}$. The following Lyapunov function $L : \mathcal{D} \to \mathbb{R}$ is central to the proof of Theorem 1.1. For $u \in \mathcal{D}$ and given $\delta_0, \varepsilon_1, M > 0$ and $b > 1$, we set

$$L(u) = \sup_{\{x: u(x)\in(0,\delta_0]\}} \ell(u; x), \tag{3.1}$$

where

$$\ell(u; x) = \log\left(\frac{1}{u(x)}\right) + \log_b \left(1 + \varepsilon_1 - \frac{u(x+M)}{u(x)}\right)_+. \tag{3.2}$$

Here, we let $\log 0 = -\infty$ and $(x)_+ = x \vee 0$, where $a \vee b = \max(a, b)$. If the set on the right side of (3.1) is empty, we let $L(u) = -\infty$. In particular, $L(F_0) = -\infty$.

Most of the work in demonstrating Theorem 1.1 is contained in the following result.

Theorem 3.1. *For each BRW as in Theorem 1.1, there is a choice of parameters $\delta_0, \varepsilon_1, M > 0$ and $b > 1$, such that*

$$\sup_n L(F_n) < \infty. \tag{3.3}$$

Theorem 3.1 implies that, with the given choice of parameters, for all n and all x with $0 < F_n(x) \leq \delta_0$,

$$\log\left(1 + \varepsilon_1 - \frac{F_n(x+M)}{F_n(x)}\right)_+ \leq (\log b)(C_0 + \log F_n(x)), \tag{3.4}$$

where $C_0 = \sup_{n\geq 0} L(F_n) < \infty$. In particular, by taking $\delta_1 > 0$ small enough so that the right hand side in the last inequality is sufficiently negative when $F_n(x) \leq \delta_1$, one obtains the following corollary.

Corollary 3.2. *For each BRW as in Theorem 1.1, there exists $\delta_1 = \delta_1(C_0, \delta_0, \varepsilon_1, b, M) > 0$ such that, for all n,*

$$F_n(x) \leq \delta_1 \quad \text{implies} \quad F_n(x+M) \geq \left(1 + \frac{\varepsilon_1}{2}\right) F_n(x). \tag{3.5}$$

The inequality (3.5) implies that $\{F_n^s\}_{n\geq 0}$ is "tight at values less than δ_1". In order to demonstrate Theorem 1.1, we need to show that the sequence is also "tight at values larger than δ_1". Once (3.5) is known, this will follow without any conditions on G or $\{p_k\}$, other than that $p_1 < 1$. This is the content of Proposition 3.3. Theorem 1.1 follows directly from Proposition 3.3 and Corollary 3.2.

Proposition 3.3. *Suppose that $p_1 < 1$, and that (3.5) holds for all n and some choice of $\delta_1, M, \varepsilon_1 > 0$. Then, the sequence of distributions $\{F_n^s\}_{n\geq 0}$ is tight.*

The proof of Proposition 3.3 requires a couple pages of computation; details are given in the proofs of Proposition 2.9 and Lemma 2.10 in [BZ06]. The idea, in spirit, is to show that $F_n(x)$ must grow at a uniform multiplicative rate through successive iterations, until reaching a value within distance η of 1, for some $\eta > 0$, at coordinates not changing by much after each individual iteration. (This is easy to show if G is bounded.) It follows from this, that if F_n is "relatively flat" (in a multiplicative sense) somewhere away from 1, then F_{n-1} must be "almost as flat" at some nearby location, where its value is also smaller by a fixed factor $\gamma < 1$ (which depends on η). Iterating backwards in n, it follows that F_0 is relatively flat somewhere away from 1. Since $F_0(x) = \mathbf{1}_{\{x \geq 0\}}$, this is in fact not the case, and so for no n can F_n be relatively flat anywhere away from 1. This bound on flatness will be uniform in n, which will imply $\{F_n^s\}_{n\geq 0}$ is tight.

We need to justify Theorem 3.1. Rather than demonstrate Theorem 3.1 directly, it suffices to demonstrate the following variant.

Theorem 3.4. *For each BRW as in Theorem 1.1, there is a choice of parameters $\delta_0, \varepsilon_1, M, C_1 > 0$ and $b > 1$, with the property that if $L(F_{n+1}) \geq C$ for some n and some $C > C_1$, then $L(F_n) \geq C$.*

Proof of Theorem 3.1 assuming Theorem 3.4. If $\sup_n L(F_n) = \infty$, then for any C, one can choose n such that $L(F_n) \geq C$. For $C > C_1$, it follows by Theorem 3.4, that $L(F_0) \geq C$. This contradicts $L(F_0) = -\infty$. □

The demonstration of Theorem 3.4 is rather involved, and is done, in the more general setting of Theorem 2.5, in Section 3 of [BZ06]. In the remainder of this section, we provide heuristics for the main steps. These are somewhat simpler in the setting of Theorem 1.1.

The iteration of F_n in (1.3) involves two steps, which consist of first applying Q to F_n, and then convoluting the combined quantity by G. We will show Theorem 3.4 by bounding the change in L over each of these steps. This involves analyzing the contribution of each of the two components of ℓ with $\ell = \ell_1 + \ell_2$, for

$$\ell_1(u;x) = \log(1/u(x)) \tag{3.6}$$

and

$$\ell_2(u;x) = \log_b\left(1 + \varepsilon_1 - \frac{u(x+M)}{u(x)}\right)_+ . \tag{3.7}$$

The first step of the iteration, where Q is applied to F_n, is relatively easy to analyze. As $F_n(x) \to 0$,

$$\ell_1(Q(F_n);x) - \ell_1(F_n;x) = -\log(Q(F_n(x))/F_n(x)) \to -\log m_1, \tag{3.8}$$

since $Q'(0) = m_1$. Also, as $F_n(x) \to 0$,

$$\ell_2(Q(F_n);x) - \ell_2(F_n;x) = \log_b\left[\frac{1 + \varepsilon_1 - Q(F_n(x+M))/Q(F_n(x))}{1 + \varepsilon_1 - F_n(x+M)/F_n(x)}\right] \to 0 \tag{3.9}$$

if $F_n(x+M)/F_n(x)$ is bounded away from $1+\varepsilon_1$, since Q is nearly linear near 0. Together, (3.8) and (3.9) imply that as $F_n(x) \to 0$,

$$\ell(Q(F_n);x) - \ell(F_n;x) \to -\log m_1, \tag{3.10}$$

if $F_n(x+M)/F_n(x)$ is bounded away from $1+\varepsilon_1$.

More careful reasoning along the above lines shows that for small enough δ_0 in (3.1) and $L(Q(F_n)) > 0$,

$$L(Q(F_n)) - L(F_n) \le -\frac{1}{2}\log m_1. \tag{3.11}$$

One obtains the terms involving L, on the left side of (3.11), by taking the supremum of ℓ over $F_n(x) \in (0,\delta_0]$, respectively, over $Q(F_n(x)) \in (0,\delta_0]$, as in (3.1). By employing $L(Q(F_n)) > 0$ and the condition (1.20) on $\{p_k\}_{k\ge 1}$, one can avoid the constraint on $F_n(x+M)/F_n(x)$ in (3.10).

We still need to analyze the second step of the iteration, where $Q(F_n)$ is convoluted with G. Since L decreases over the first step, as in (3.11), to demonstrate Theorem 3.4, it suffices to show that for some C_1,

$$L(F_{n+1}) - L(Q(F_n)) \le \frac{1}{2}\log m_1 \tag{3.12}$$

when $L(F_{n+1}) > C_1$.

The reasoning that is required for (3.12) is more involved. The part requiring the most effort is the proof of the following proposition, which is taken from Proposition 3.2 of [BZ06]. It states that for $x_2 = x_1 + M$, if (3.15) holds, then $u(x_2-y)/u(x_1-y)$ must either satisfy the upper bound in (3.16) somewhere on $[-M \vee r(u,\varepsilon',x_1),\infty)$, or the stronger upper bound in (3.17) somewhere on $[r(u,\varepsilon',x_1),-M)$. The definition of $r(u,\varepsilon',x_1)$ is a bit cumbersome, with

$$r = r(u,\varepsilon',x_1) = \begin{cases} q & \text{if } u(x_2-q) \ge u(x_1-q+M/2)/(1-4\varepsilon'), \\ q+\frac{M}{2} & \text{otherwise,} \end{cases} \tag{3.13}$$

where

$$q = q(u,\varepsilon',x_1) = \sup\{y<0 : u(x_2-y) \ge (1+8\varepsilon')u(x_1-y)\} \le 0. \tag{3.14}$$

If $q(u,\varepsilon',x_1) = -\infty$, one sets $r(u,\varepsilon',x_1) = -\infty$. Intuitively, $x_1 - q$ is the first point to the right of x_1 where u is "very non-flat", where we interpret u to be "very non-flat" at x_1 if $u(x_1+M)/u(x_1)$ is not close to 1. The point r is chosen so that u is "very non-flat" at all points in $(x_1-r, x_1-r+M/2]$.

Proposition 3.5. *Assume that G satisfies (1.19), and that for some $u \in \mathcal{D}$, $n, x_1 \in \mathbb{R}$ and $\varepsilon' \in (0,1/8)$, and for $x_2 = x_1 + M$,*

$$(u*G)(x_2) < (1+\varepsilon')(u*G)(x_1). \tag{3.15}$$

Then, at least one of the following two statements holds for each $\delta > 0$:

$$u(x_2-y) \le (1+\varepsilon'+\delta)u(x_1-y) \quad \text{for some } y \ge -M \vee r, \tag{3.16}$$

$$u(x_2-y) \le (1+\varepsilon'-\delta e^{-ay/4})u(x_1-y) \quad \text{for some } y \in [r,-M). \tag{3.17}$$

Proposition 3.5 can be motivated as follows. The inequality in (3.15) can be rewritten as

$$\int_{-\infty}^{\infty} u(x_2 - y)dG(y) < (1+\varepsilon') \int_{-\infty}^{\infty} u(x_1 - y)dG(y). \tag{3.18}$$

Because of the drop in $u(x_1 - y)$ "at" $y = r$, it will follow from this that, in fact,

$$\int_{[r,\infty)} u(x_2 - y)dG(y) < (1+\varepsilon') \int_{[r,\infty)} u(x_1 - y)dG(y). \tag{3.19}$$

Inequality (3.19) requires some work, and is shown in Lemma 3.5 of [BZ06]. On the other hand, if the inequality in (3.16) fails everywhere on $[-M \vee r, \infty)$, then

$$\int_{[-M\vee r,\infty)} u(x_2 - y)dG(y) > (1+\varepsilon'+\delta) \int_{[-M\vee r,\infty)} u(x_1 - y)dG(y) \tag{3.20}$$

must hold. Because of this, the integral of $(1+\varepsilon')u(x_1-y)$ over $y \in [r, -M \vee r)$ needs to exceed that of $u(x_2 - y)$ sufficiently, in order for (3.19) to hold. So, $(1+\varepsilon')u(x_1-y)/u(x_2-y)$ needs to be "large" somewhere on $[r, -M \vee r)$. Because of the condition (1.19) on the left tail of G, this ratio must become increasingly large in the sense of (3.17), as y decreases.

One can analyze the behavior of $\ell(F_{n+1}; x) - \ell(Q(F_n); x)$ separately under the scenarios (3.16) and (3.17); the condition (3.15) with $\varepsilon' < \varepsilon_1$ will always be satisfied when $\ell(F_{n+1}; x) > -\infty$. Precise results are given in Proposition 3.3 of [BZ06]. Here, we summarize this behavior by considering the terms $\ell_i(F_{n+1}; x_1) - \ell_i(Q(F_n); x_1 - y)$, $i = 1, 2$, in each case, where y is chosen as in Proposition 3.5. We recall that $F_{n+1} = Q(F_n) * G$.

Suppose that (3.15) and (3.16) both hold for a given x_1, with $u = Q(F_n)$. One can show that for y chosen as in (3.16),

$$\ell_i(F_{n+1}; x_1) - \ell_i(Q(F_n); x_1 - y) \le \varepsilon_2, \tag{3.21}$$

$i = 1, 2$, where $\varepsilon_2 > 0$ depends on ε_1, and can be chosen as close to 0 as desired, for ε_1 close to 0. The case $i = 1$ employs an elementary truncation argument involving the right tail of G, and $i = 2$ employs (3.16). Summing (3.21), over $i = 1, 2$, gives

$$\ell(F_{n+1}; x_1) - \ell(Q(F_n); x_1 - y) \le 2\varepsilon_2. \tag{3.22}$$

Suppose, on the other hand, that (3.15) and (3.17) hold for a given x_1. For y chosen as in (3.17), one can again estimate each of the differences $\ell_i(F_{n+1}; x_1) - \ell_i(Q(F_n); x_1 - y)$, $i = 1, 2$. One can show that for small enough ε_1 and b close to 1, the difference for $i = 2$ is more negative than the difference for $i = 1$ is positive, with

$$\ell(F_{n+1}; x_1) - \ell(Q(F_n); x_1 - y) \le \frac{a}{2} y < 0. \tag{3.23}$$

This is a consequence of the contribution of the term $-\delta e^{-ay/4}$ in (3.17) to $\ell_2(Q(F_n); x_1 - y)$, and of the upper bound $y \geq r$ places on $\ell_1(F_{n+1}; x_1) - \ell_1(Q(F_n); x_1 - y)$. Choosing ε_2 in (3.21) so that $\varepsilon_2 \leq \frac{1}{4}\log m_1$, it follows from (3.22) and (3.23), that

$$\ell(F_{n+1}; x_1) - \ell(Q(F_n); x_1 - y) \leq \frac{1}{2}\log m_1 \tag{3.24}$$

for all x_1 satisfying (3.15).

We now combine (3.22) and (3.24) to demonstrate (3.12). For $L(F_{n+1}) = C$ and any $\varepsilon_3 > 0$, one can choose x_1 so that

$$\ell(F_{n+1}; x_1) \geq C - \varepsilon_3. \tag{3.25}$$

Suppose that C_1 in Theorem 3.1 is chosen so that $C_1 \geq \log(1/\delta_0) + 1$, and that $C > C_1$. Then, for x_1 chosen as in (3.25), it follows from (3.22) that $Q(F_n(x_1 - y)) \leq \delta_0$, if ε_2 and ε_3 satisfy $2\varepsilon_2 + \varepsilon_3 \leq 1$. Since $\varepsilon_2, \varepsilon_3 > 0$ are arbitrary, it follows from this and (3.24), that

$$L(F_{n+1}) - L(Q(F_n)) \leq \frac{1}{2}\log m_1, \tag{3.26}$$

which is the desired inequality (3.12). This implies Theorem 3.4, which in turn implies Theorem 3.1 and Theorem 1.1.

References

[A-B07] D. Addario-Berry, *Ballot Theorems and the Height of Trees*, Ph.D. thesis, McGill University (2007).

[A89] D. Aldous, *Probability Approximations via the Poisson Clumping Heuristic*, Springer, New York (1989).

[A91] D. Aldous, Random walk covering of some special trees, *J. of Math. Analysis and Applications* **157** (1991), pp. 271-283.

[Ba00] M. Bachmann, Limit theorems for the minimal position in a branching random walk with independent logconcave displacements, *Adv. Appl. Prob.* **32** (2000), pp. 159-176.

[Bi90] J.D. Biggins, The central limit theorem for the supercritical branching random walk, and related results, *Stoch. Proc. Appl.* **34** (1990), pp. 255-274.

[Br78] M. Bramson, Maximal displacement of branching Brownian motion, *Comm. Pure Appl. Math.* **31** (1978), pp. 531-581.

[Br83] M. Bramson, Convergence of solutions of the Kolmogorov equation to travelling waves, *Mem. Amer. Math. Soc.* **44** (1983), No. 285.

[BZ06] M. Bramson and O. Zeitouni, Tightness for a family of recursion equations, preprint (2006). arXiv:math/0612382v1 [math.PR].

[CD06] B. Chauvin and M. Drmota, The random multisection problem, travelling waves, and the distribution of the height of m-ary search trees, *Algorithmica* (2006) (Online First).

[D03] M. Drmota, An analytic approach to the height of binary search trees, II, *J. ACM* **50** (2003), pp. 333–374.

[DH91] F. M. Dekking and B. Host, Limit distributions for minimal displacement of branching random walks, *Probab. Theory Relat. Fields* **90** (1991), pp. 403–426.

[DPRZ04] A. Dembo, Y. Peres, J. Rosen, and O. Zeitouni, Cover times for planar Brownian motion and random walks, *Ann. Math.* **160** (2004), pp. 433-464.

[F37] R.A. Fisher, The advance of advantageous genes, *Ann. of Eugenics* **7** (1937), pp. 355-369.

[H74] J.M. Hammersley, Postulates for subadditive processes, *Ann. Probab.* **2** (1974), pp. 652-680.

[HS07] Y. Hu and Z. Shi, Minimal position and critical martingale convergence in branching random walks, and directed polymers on disordered trees, preprint (2007). arXiv:math/0702799v2 [math.PR].

[Ki75] J.F.C. Kingman, The first birth problem for an age-dependent branching process, *Ann. Probab.* **3** (1975), pp. 790-801.

[Ki76] J.F.C. Kingman, Subadditive processes, in École d'Été de Probabilités de Saint-Flour V - 1975. *Lecture Notes in Mathematics* **539**, Springer, Berlin (1976), pp 167-223.

[KPP37] A. Kolmogorov, I. Petrovsky and N. Piscounov, Étude de l'équation de la diffusion avec croissance de la quantité de matière et son application à un problème biologique, *Moscou Universitet Bull. Math.* **1** (1937), pp. 1-25.

[Li85] T. Liggett, An improved subadditive ergodic theorem, *Ann. Probab.* **13** (1985), pp. 1279-1285.

[Lu82] R. Lui, A nonlinear integral operator arising from a model in population genetics I. Monotone initial data, *SIAM J. Math. Anal.* **13** (1982), pp. 913-937.

[M75] H. McKean, Application of Brownian motion to the equation of Kolmogorov-Petrovskii-Piskunov, *Comm. Pure Appl. Math.* **28** (1975), pp. 323-331.

[R03] B. Reed, The height of a random binary search tree, *J. ACM* **50** (2003), pp. 306–332.

Maury Bramson
School of Mathematics
University of Minnesota
206 Church St. SE
Minneapolis, MN 55455
bramson@math.umn.edu

Ofer Zeitouni
School of Mathematics
University of Minnesota
206 Church St. SE
Minneapolis, MN 55455
and
Depts. of EE and of Math.
Technion
Haifa 32000, Israel
zeitouni@math.umn.edu