POISSON-DIRICHLET BRANCHING RANDOM WALKS

LOUIGI ADDARIO-BERRY, KEVIN FORD

ABSTRACT. We determine, to within $O(1)$, the expected minimal position at level n in certain branching random walks. The walks under consideration have displacement vector $(v_1, v_2, \ldots)$ where each v_j is the sum of j independent Exponential(1) random variables and the different v_i need not be independent. In particular, our analysis applies to the Poisson-Dirichlet branching random walk and to the Poisson-weighted infinite tree. As a corollary, we also determine the expected height of a random recursive tree to within O(1).

1. INTRODUCTION

A branching random walk starts from an initial particle, the *root*, with position 0. The root produces some number of children, who are randomly displaced from their parent according to some displacement law. Each child in turn produces some number of children, who are displaced from the position of their parent according to the same law; and so on. In general, the displacements of siblings relative to their parent may be dependent, but for distinct particles v and w, the displacements of the children of v and of the children of w must be independent. When the displacments are non-negative, this is often called an age-dependant branching process, and the displacements are thought of as "times to birth".

There is a natural tree associated with a branching random walk, where the vertices correspond to particles, and an edge from parent to child is weighted with the child's displacement from its parent. More precisely, let T be the *Ulam-Harris* tree, which has vertex set $V = \bigcup_{n=0}^{\infty} \mathbb{N}^n$ (we think of elements of $\mathbb{N}^n$ as concatenations of n integers, and take $\mathbb{N}^0 = \{\emptyset\}$), is rooted at $\emptyset$, and has an edge from v to vi for each $v \in V$ and each $i \in \mathbb{N}$. We call $\mathbb{N}^n$ the n'th generation of T, and for $v = v_1 \ldots v_n \in \mathbb{N}^n$, we say that v has *parent* $p(v) = v_1 \ldots v_{n-1}$ and *children* vi, $i \in \mathbb{N}$. (We will usually write T_n in place of $\mathbb{N}^n$ for readability.)

Now suppose $\mathbf{X} = (X_i : i \in \mathbb{N})$ is a random vector, where each $X_i \in \mathbb{R} \cup \{+\infty\}$. We do *not* require that the entries of $\mathbf{X}$ are independent of one another – this will be important below. Then we form a branching random walk by marking each vertex $v \in V$ with an independent copy $\mathbf{X}^v = (X_i^v : i \in \mathbb{N})$ of $\mathbf{X}$. Write $\mathcal{T}$ for the pair $(T, \{\mathbf{X}_v : v \in V\})$; then $\mathcal{T}$ is our branching random walk. We call $\mathbf{X}$ the *displacement vector* of $\mathcal{T}$. [1] For each $v \in V$ and $i \in \mathbb{N}$, we regard X_i^v as the displacement from v to vi, and let $S(v) = S(v, \mathcal{T})$ be the sum of the displacements on the path from the root to v (formally, if $v = v_1 \ldots v_n$ then $S(v) = \sum_{i=1}^{n} X_{v_i}^{p(v_1 \ldots v_i)}$, and this sum is taken to be $+\infty$ if any of its elements are $+\infty$). We say $\mathcal{T}$ has *finite branching* if almost surely all but finitely many coordinates of $\mathbf{X}$ are equal to $+\infty$.

Date: December 12, 2010.

2000 *Mathematics Subject Classification.* 60J80.

The first author was supported by an NSERC Discovery Grant. The second author was supported by NSF Grant DMS-0901339. The research was conducted in part while the second author was visiting the Institute for Advanced Study, supported by grants from the Ellentuck Fund and The Friends of the Institute For Advanced Study. He thanks the IAS for its hospitality and excellent working conditions.

[1] For the formal details of a probabilistic construction of branching random walks, see, e.g., [Har63].

For $n \in \mathbb{N}$, let $M_n = \inf(S(v) : v \in \mathbb{N}^n)$. In all situations we consider in this paper, this infimum is attained, so M_n is the minimal displacement of any individual in the n'th generation. The minimal displacement is one of the most well-studied parameters associated with branching random walks. It has been known since the 1970's [Ham74, Kin75, Big76] that under quite general conditions, M_n grows asymptotically linearly with lower-order corrections. Recently there have been substantial developments in understanding the finer behavior of M_n on two fronts: first, convergence results for the lower order corrections [ABR09, AS10, HS09]; and second, the concentration of M_n about its mean (or median) [ABR09, BZ09, CD06]. We refer to these as the global behavior and the local behavior of M_n, respectively. Under suitable conditions, M_n generally seems to exhibit the following behavior: for some constants $\alpha \in \mathbb{R}$ and $\beta > 0$, $\mathrm{median}(M_n) = \alpha n + \beta \log n + O(1)$, and furthermore, $M_n/n \to \alpha$ almost surely and $(M_n - \alpha n)/\log n \to \beta$ in probability (but *not* almost surely [HS09]). Also, under sufficiently strong moment conditions for the displacements, $\mathbb{E}\{\exp(\gamma|M_n - \mathbb{E}\, M_n|)\} < \infty$ for some $\gamma > 0$ and all n. (In fact, in some cases the upper tail of $M_n - \mathbb{E}\, M_n$ is even known to decay doubly-exponentially quickly [Bac00, FKL10].)

To date, however, all the results of the kind described in the preceding paragraph that we are aware of, require that the branching random walk has finite branching. In this paper we study the global behavior of M_n for a class of branching random walks which *do not* have finite branching. The class we consider is rather restricted but nonetheless contains at least two interesting special cases, one related to the factorization of random integers, and one related to the analysis of algorithms. Say that $\mathbf{X}$ has *exponential steps* if for all i, X_i is distributed as the sum of i independent Exponential(1) random variables. The main result of this paper is the following theorem. For short, we denote

$$\widetilde{M}_n = \mathrm{median}(M_n) := \sup\{x : \mathbb{P}\{M_n < x\} < 1/2\}.$$

Theorem 1.1. *If* $\mathbf{X}$ *has exponential steps, then*

$$\widetilde{M}_n = \frac{n}{e} + \frac{3}{2e}\log n + O(1).$$

Remark: The $O(1)$ term is uniform over n and over all BRW for which $\mathbf{X}$ has exponential steps.

Using methods from [FKL10], we can deduce from Theorem 1.1 uniform exponential tails for M_n. In the next theorem and at other points throughout the paper, we will use the Vinogradov notation $f \ll g$ which means $f = O(g)$, with subscripts indicating dependence on any parameter, e.g. $f \ll_k g$ means the constant implied by the $\ll$ symbol may depend on k but not on any other variable.

Theorem 1.2. *If* $\mathbf{X}$ *has exponential steps, then for any* $c_1 < e$*, we have*

$$\mathbb{P}\{M_n \leqslant \widetilde{M}_n - x\} \ll_{c_1} e^{-c_1 x} \qquad (n \geqslant 1, x \geqslant 0),$$

and for any $c_2 < 1$*,*

$$\mathbb{P}\{M_n \geqslant \widetilde{M}_n + x\} \ll_{c_2} e^{-c_2 x} \qquad (n \geqslant 1, x \geqslant 0).$$

Again, the above estimates are uniform over all BRW under consideration. Also, Theorem 1.2 implies that $\widetilde{M}_n = \mathbb{E}\, M_n + O(1)$, and so both Theorems 1.1 and 1.2 hold with $\widetilde{M}_n$ replaced by $\mathbb{E}\, M_n$.

The simplest example of a displacement vector with exponential steps is obtained by taking $\mathbf{X} = (E_1, E_1 + E_2, \ldots)$ where $\{E_i\}_{i\in\mathbb{N}}$ are iid Exponential(1) random variables. In this case $\mathcal{T}$ is called the *Poisson-weighted infinite tree* [AS04] and has been used very effectively in probabilistic combinatorial optimization. It also arises in the analysis of an important tree-based data structure in the following way. Order the elements of $\mathcal{T}$ in increasing order of displacement as $\{w_i\}_{i\in\mathbb{N}}$, so in particular we have $w_1 = \emptyset, w_2 = 1 \in \mathbb{N}^1$, and either $w_3 = 2 \in \mathbb{N}^1$ or $w_3 = 11 \in \mathbb{N}^2$. Now for each m let Z_m be the subtree of $\mathcal{T}$ induced by $w_1, \ldots, w_m$. By the memoryless property of the exponential, it follows that the parent of w_{m+1} is a uniformly random element of Z_m – in other words, Z_m is a *random recursive tree* for all m. This connection is well known [Pit94].)

Z_m is also the subtree of $\mathcal{T}$ induced by the set of nodes of displacement at most $S(w_m)$. (Also, it is straightforwardly shown by induction and the memoryless property of the exponential that the families $(Z_m)_{m\in\mathbb{N}}$ and $(S(w_m))_{m\in\mathbb{N}}$ are independent, but we will not need this.) Let H_m be the *height* of Z_m – the largest generation containing a node of Z_m. In other words, $H_m = \max\{n : M_n \leqslant S(w_m)\}$, which is the representation that will be useful below. Devroye [Dev87] showed that $H_m/\log m \to e$ almost surely and in expectation, and Pittel [Pit94] provided a different proof of the almost sure convergence. As a straightforward consequence of Theorems 1.1 and 1.2, we obtain the following more precise information.

Corollary 1.3. *The height H_m of a random recursive tree on m nodes satisfies $\mathbb{E}\,H_m = e\log m - \frac{3}{2}\log\log m + O(1)$. Furthermore, for all $c' < \frac{1}{2e}$, all $m \geqslant 1$, $k \geqslant 1$,*

$$\mathbb{P}\{|H_m - \mathbb{E}\,H_m| \geqslant k\} \ll_{c'} e^{-c'k}.$$

Since the proof of this corollary is very short, we include it in the introduction. In the proof we write $\mathrm{har}(s) = \sum_{i=1}^{s} 1/i$.

Proof. The random variable $S(w_m)$ is distributed as the sum, $F_1 + \ldots + F_{m-1}$, of independent random variables with F_i having $\mathrm{Exponential}(i)$ distribution for $i = 1, \ldots, m-1$. Equivalently, $S(w_m)$ is distributed as the maximum of $m-1$ iid $\mathrm{Exponential}(1)$ random variables. Thus, $\mathbb{E}\,S(w_m) = \mathrm{har}(m-1)$ and for all $x > 0$,

$$\mathbb{P}\{S(w_m) \geqslant \mathrm{har}(m-1) + x\} \leqslant (m-1)e^{-(\mathrm{har}(m-1)+x)} \leqslant e^{-x}, \tag{1.1}$$

$$\mathbb{P}\{S(w_m) \leqslant \mathrm{har}(m-1) - x\} = \left(1 - e^{-(\mathrm{har}(m-1)-x)}\right)^{m-1}. \leqslant e^{-e^{x-1}}, \tag{1.2}$$

Now write

$$d(m) = \max\{n : \widetilde{M}_n \leqslant \mathrm{har}(m-1)\} = e\log m - \frac{3}{2}\log\log m + O(1),$$

and note that $\widetilde{M}_{d(m)} = \mathrm{har}(m-1) + O(1)$ by Theorem 1.1. It follows that for $k \geqslant 1$, if $H_m \geqslant d(m) + k$ then either

$$M_{d(m)+k} \leqslant \mathrm{har}(m-1) + \frac{k}{2e} \leqslant \widetilde{M}_{d(m)+k} - \frac{k}{2e} + O(1),$$

or

$$S(w_m) \geqslant \mathrm{har}(m-1) + \frac{k}{2e}.$$

By Theorem 1.2 and (1.1), it follows that $\mathbb{P}\{H_m \geqslant d(m) + k\} \ll_{c_1} e^{-c_1 k/(2e)}$ for each $c_1 < e$. A similar argument using Theorem 1.2 and (1.2) shows the bound $\mathbb{P}\{H_m \leqslant d(m) - k\} \ll_{c_2} e^{-c_2 k/(2e)}$ for each $c_2 < 1$. ☐

Another important example of a displacement vector with exponential steps arises from a discrete time random fragmentation process. Let $U_1, U_2, \ldots$ be independent uniform $[0,1]$ random variables. Set $G_1 = U_1$ and for $i > 1$ set $G_i = (1-U_1)\cdot\ldots\cdot(1-U_{i-1})U_i$. The distribution of the sequence

$$\mathbf{G} = (G_1, G_2, \ldots)$$

was first studied, in greater generality, in [Hal44]. (One motivation for Halmos' paper was a problem about loss of energy of neutrons after many collisions; after each collision the neutron loses a random fraction of its current energy.) $\mathbf{G}$ is also a special case of the *Griffiths-Engen-McCloskey* GEM distribution. Further, $(G_{\sigma(1)}, G_{\sigma(2)}, \dots)$ has the Poisson-Dirichlet (or PD) distribution, where $\sigma : \mathbb{N} \to \mathbb{N}$ is the permutation that arranges the terms of $(G_1, G_2, \dots)$ in decreasing order. (We remark that both the GEM and the PD distributions as defined above are in fact special cases from a more general two-parameter family of distributions [Pit06] – in the standard notation, we are considering the $\mathrm{GEM}(0,1)$ and $\mathrm{PD}(0,1)$ distributions.) The PD distribution arises in a number of natural decomposition situations, such as factorization of large random integers [Bil72, DG93] and cycle lengths of random permutations [Pit06].

Letting $X_k = -\log G_k$ for each k yields a vector $(X_1, X_2, \dots)$ with exponential steps. We refer to the resulting branching random walk as a *Poisson-Dirichlet branching random walk*. This example has more complicated dependence between the X_i than the first example. Since $\sum_{i=1}^{\infty} G_i = 1$ almost surely, there is another way to think of the branching random walk. Imagine that an object of mass m is placed at the root $\emptyset$. The root divides this mass into pieces according to the vector $\mathbf{G}^{\emptyset}$ and sends the pieces to its children, sending a mass $mG_k^{\emptyset}$ to its k'th child. This rule is repeated recursively, so each node v sends proportion G_k^v of the mass it receives to its k'th child vk. This structure is variously called a *multiplicative cascade* or, more commonly at the moment, a *fragmentation process* [Ber06]. The special case of Theorem 1.1 when $\mathcal{T}$ is a Poisson-Dirichlet branching random walk is used in [FKL10] to analyze a tree model related to primality testing, proving heuristic evidence for the behavior of the distribution of tree heights. In this special case of a PD branching random walk, a much stronger estimate for the right tail of M_n was proved in [FKL10], namely for any $c_3 < 1$,

$$\mathbb{P}\left\{ M_n \geqslant \widetilde{M}_n + x \right\} \leqslant \exp\{-e^{c_3 x - c_4}\} \qquad (n \geqslant 1, x \geqslant 0),$$

where c_4 is a constant depending on c_3. Such a right tail bound cannot hold in general; for example, for the case of $\mathcal{T}$ being a Poisson-weighted infinite tree, we have $\mathbb{P}\{M_1 \geqslant x\} = e^{-x}$. (It seems likely that among branching random walks with exponential steps, the Poisson-weighted infinite tree and the Poisson-Dirichlet branching random walk are extremal examples, with the former having the heaviest tails for $M_n - \widetilde{M}_n$ and the latter the strongest tail bounds for $M_n - \widetilde{M}_n$. However, we do not have a precise conjecture in this direction.)

The *Pratt tree* for a prime p has root p whose children are the prime factors of $p-1$; the subtrees of the children of the root are recursively constructed in the same fashion (stopping when $p = 2$). We let $H(p)$ be the height of the Pratt tree for p. It is easily seen that the height is always at most $(\log p)/(\log 2) + 1$. Such trees were used by Pratt [Pra75] to show that if p is prime, then there exists a certificate (formal proof) of the primality of p, of length $O(H(p)\log p) = O((\log p)^2)$. It is then of interest to understand the "typical" behavior of $H(p)$. [FKL10] uses Theorems 1.1 and 1.2 to support the following conjecture.

Conjecture 1.4 ([FKL10], Conjecture 3)**.** *There exist constants $c, c' > 0$ and real numbers $\{E(p) : p \text{ prime}\}$ such that*

- $H(p) = e \log p - \frac{3}{2} \log\log p + E(p)$,
- *for all $z \geqslant 0$, and $x \geqslant 0$,*

$$e^{-c'z}\pi(x) \ll |\{\text{primes } p \leqslant x : E(p) \geqslant z\}| \ll e^{-cz}\pi(x),$$

and

$$|\{\text{primes } p \leqslant x : E(p) \leqslant -z\}| \ll \exp(-e^{cx})\pi(x).$$

Here $\pi(x)$ is the number of primes which are at most x.

The structure of the remainder of the paper is as follows. In Section 2 we introduce a little additional notation. In Section 3 we use straightforward calculations to prove weak bounds on the likely value of M_n, and to "reduce the search space" of nodes in T_n which have a chance of attaining the minimal displacement M_n. Section 4 studies the sample path properties of a uniformly random element of certain "homogeneous" subsets of T_n, and forms a key step of the proof. In Section 5 we prove the lower bound of Theorem 1.1, and in Section 6 we prove the upper bound. Finally, the details of the proof of Theorem 1.2 are found in Section 7.

2. Notation

Given $v = v_1 v_2 \dots v_n \in V$, we let $h(v) = \sum_{i=1}^n v_i$, and remark that $S(v)$ has distribution Gamma$(h(v))$. If $v \in T_n$, we write $k(v) = h(v) - n$, and write $T_{n,k}$ for the set of nodes $v \in T_n$ with $k(v) = k$. We denote by $T_n(x)$ (resp. $T_{n,k}(x)$) the set of nodes of T_n (resp. $T_{n,k}$) with displacement at most x.

The Bachmann-Landau notations $o()$ and $O()$ have their usual meaning. As mentioned earlier, we use the Vinogradov notation $f \ll g$ which means $f = O(g)$. We also use the Hardy notation $f \asymp g$ which means $f = O(g)$ and $g = O(f)$. Constants implied by these symbols are absolute unless otherwise indicated, e.g. by a subscript.

3. Some basic expectations

In order to restrict the set of nodes we need to consider when searching for the precise location of M_n, we first assert the following two straightforward facts, whose proofs are forthcoming.

Lemma 3.1. *(a) The expected number of nodes $v \in T_n$ with $|h(v) - (1+1/e)n| \leqslant \sqrt{n}$ and with $S(v) \leqslant n/e + \log n/(2e)$ is $\gg 1$.*

(b) The expected number of nodes $v \in T_n$ with $S(v) \leqslant n/e + (2/e)\log n$ and with $|h(v) - (1+1/e)n| > \sqrt{6n \log n}$ is $O(n^{-1/2})$.

Together, (a) and (b) suggest that in order to find M_n, it should suffice to look at nodes in T_n satisfying $h(v) = (1+1/e)n + O(\sqrt{n})$, as will indeed be the case. In proving (a) and (b), we will in fact prove more general bounds that will be useful throughout the paper.

We first remark that for $v \in V$ with $h(v) = h$, $S(v)$ has density function

$$\lim_{dx \downarrow 0} \frac{\mathbb{P}\{S(v) \in dx\}}{dx} = \gamma_h(x) = \frac{x^{h-1} e^{-x}}{(h-1)!} \qquad (x \geqslant 0).$$

For all $n \geqslant 1$, $k \geqslant 0$, we have

$$|T_{n,k}| = \binom{n+k-1}{k}, \tag{3.1}$$

so the sum of the density functions for nodes $v \in T_{n,k}$ is

$$f_{n,k}(x) = \binom{n+k-1}{k} \gamma_{n+k}(x) = \frac{x^{n+k-1} e^{-x}}{k!(n-1)!} = \frac{x^k}{k!} \cdot \frac{x^{n-1} e^{-x}}{(n-1)!}.$$

This function will play a significant role, and we now derive bounds on its value for a variety of ranges of k and x. We remark that assertions (a) and (b), above, state in particular that to find M_n we should take both k and x near n/e. Thus, writing $k = (n+r)/e$ and $x = (n+y)/e$, by Stirling's formula we have

$$f_{n,k}(x) = \frac{\left(1 + O\left(\frac{1}{n} + \frac{1}{k}\right)\right)}{n+y} \sqrt{\frac{n}{n+r}} e^{(r-y)/e} \left(1 - \frac{r-y}{n+r}\right)^{(n+r)/e} \left(1 + \frac{y}{n}\right)^n \frac{e^{3/2}}{2\pi}, \tag{3.2}$$

When $r = O(\sqrt{n})$, $y = O(\sqrt{n})$, we have $(1 + y/n)^n \asymp e^y$ and

$$(1 - (r - y)/(n + r))^{(n+r)/e} \asymp e^{-(r-y)/e},$$

and so obtain the simpler approximation

$$f_{n,k}(x) \asymp \frac{e^y}{n}.$$

Consequently,

$$\mathbb{E}\,|\{v \in T_{n,k} : S(v) \leqslant (n + \tfrac{1}{2}\log n)/e\}| \asymp \int_0^{(\log n)/2} \frac{e^y}{n} \asymp n^{-1/2} \tag{3.3}$$

for any fixed $k = n/e + O(\sqrt{n})$ – where the constants implicit in $O(\sqrt{n})$ and in (3.3) may depend on each other – and so we obtain

$$\mathbb{E}\,|\{v \in T_{n,k} : S(v) \leqslant (n + \tfrac{1}{2}\log n)/e, |k - n/e| \leqslant \sqrt{n}\}| \gg 1.$$

This justifies claim (a) of Lemma 3.1, and we now turn to Lemma 3.1 (b). The next lemma is [FKL10, Lemma 5.1], and we give a different proof below.

Lemma 3.2. *For all n and $x \geqslant 0$,*

$$\mathbb{E}\,|T_n(x)| = \frac{x^n}{n!}.$$

Proof. We have

$$\mathbb{E}\,|T_n(x)| = \sum_{k\geqslant 0}\sum_{v\in T_{n,k}} \mathbb{P}\{S(v) \leqslant x\} = \sum_{k\geqslant 0}\int_0^x f_{n,k}(t)\,dt = \frac{x^n}{n!}.$$

□

It follows immediately from Lemma 3.2 and Stirling's formula that the median of M_n is $\geqslant \frac{n}{e} + \frac{1}{2e}\log n + O(1)$.

We next obtain bounds on the probability that k is very different from x when $x \geqslant n/(2e)$. First we quote easy bounds for the tails of the Poisson distribution.

Proposition 3.3. *If $z > 0$ and $0 < \alpha \leqslant 1 \leqslant \beta$ then*

$$\sum_{k\leqslant\alpha z} \frac{z^k}{k!} < \left(\frac{e}{\alpha}\right)^{\alpha z}, \qquad \sum_{k\geqslant\beta z} \frac{z^k}{k!} < \left(\frac{e}{\beta}\right)^{\beta z}.$$

Proof. We have

$$\sum_{k\leqslant\alpha z} \frac{z^k}{k!} = \sum_{k\leqslant\alpha z} \frac{(\alpha z)^k}{k!}\left(\frac{1}{\alpha}\right)^k \leqslant \left(\frac{1}{\alpha}\right)^{\alpha z} \sum_{k\leqslant\alpha z} \frac{(\alpha z)^k}{k!} < \left(\frac{e}{\alpha}\right)^{\alpha z}.$$

The second inequality follows in the same way. □

An easy corollary is the following.

Lemma 3.4. *For $0 \leqslant t \leqslant x^{1/6}$,*

$$\sum_{\{k:|k-x|\geqslant t\sqrt{x}\}} f_{n,k}(x) \ll e^{-t^2/2}\frac{x^{n-1}}{(n-1)!}.$$

Taking $t = \lceil \sqrt{5 \log n} \rceil$ and integrating the above bound over $n/e \leqslant x \leqslant n/e + (2/e)\log n$, we obtain the bound

$$\mathbb{E}\left|\left\{v \in T_n : \frac{n}{e} \leqslant S(v) \leqslant \frac{n + 2\log n}{e}, |h(v) - S(v)| \geqslant \sqrt{5n\log n}\right\}\right| = O\left(\frac{1}{n^{1/2}}\right).$$

Since $\sqrt{5n\log n} + (2/e)\log n < \sqrt{6n\log n}$ for n large, combining the preceding expectation bound with Lemma 3.2 (applied with $x = n/e$) and Stirling's formula it follows that

$$\mathbb{E}\left\{\left|\bigcup_{\{k:|k-n/e|\geqslant\sqrt{6n\log n}\}} T_{n,k}((n + 2\log n)/e)\right|\right\} = O\left(\frac{1}{n^{1/2}}\right),$$

which establishes Lemma 3.1 (b).

4. Randomly sampled random walk

For integers $n \geqslant 1$, $k \geqslant 0$ and a vertex $v = v_1 \cdots v_n \in T_{n,k}$, let $h_i = h(v_1 \cdots v_i)$ and $W_i(v) = S(v_1 \ldots v_i)$ for $1 \leqslant i \leqslant n$, and write $\mathbf{W}(v) = (W_1(v), \ldots, W_n(v))$. We write $\mathbf{W}$, W_i and h_i in place of $\mathbf{W}(v)$, $W_i(v)$ and $h_i(v)$ when v is clear from context. We will always write $\mathbf{v}_{n,k}$ for a uniformly random element of $T_{n,k}$, independent of $\mathbf{v}_{n',k'}$ for $(n, k) \neq (n', k')$, and write $\mathcal{W}_{n,k}$ for the distribution of the sequence $\mathcal{W}(\mathbf{v}_{n,w}) = (W_1(\mathbf{v}_{n,k}), \ldots, W_n(\mathbf{v}_{n,k}))$. Although the sequence $0, W_1, \ldots, W_n$ is not a random walk, it useful to think of it as such for the purposes of estimating various probabilities.

Denote by $\mathcal{H}_{n,k}$ the set of vectors $(h_1, \ldots, h_n)$ of positive integers with $0 < h_1 < \cdots < h_n = n + k$ and note that $|\mathcal{H}_{n,k}| = \binom{n+k-1}{k}$. The sequence $(h_1(\mathbf{v}_{n,k}), \ldots, h_n(\mathbf{v}_{n,k}))$ is distributed as a uniformly random element of $\mathcal{H}_{n,k}$.

For $v \in T_n$, let $L_a = L_a(v)$ denote the event $\{W_i \geqslant (i/n)W_n - a\,(i \leqslant n)\}$. A vertex v is called *leading* if $L_0(v)$ holds, and – informally – *near–leading* if $L_a(v)$ holds for some small a. (We also will need to consider the event $R_a(v) = \{W_i \leqslant (i/n)W_n + a\,(i \leqslant n)\}$, and when this event occurs we say v is "near trailing".)

If M_n is not much larger than normal, v is the vertex at level n with minimal $S(v)$ and $W_i \leqslant (i/n)W_n - c$ for a large c, then M_i will be smaller than normal and this is rare. Hence, with high probability v will be a near-leading vertex On the other hand, near-leading vertices are rare – a given vertex in T_n is near leading with probability $O(f(a)/n)$ for some function f. It will turn out, as in prior work [ABR09], that $\mathbb{E}\, M_n$ is within $O(1)$ of the smallest x such that the expected number of leading nodes with displacement at most x is at least 1.

In this section, we develop estimates for the probability that vertices of $T_{n,k}$ are near–leading. As in [ABR09], we also show that for a near-leading vertex v, it is rare for $W_i(v) - (i/n)W_n(v)$ to be small if i is far away from 0 and far from n. This useful fact will play an important role in the proof of Theorem 1.1.

The next proposition, stated without proof, follows from the well-known fact that a Poisson sample becomes a uniform sample once conditioned on the position of the nth point.

Proposition 4.1. *For any positive real numbers $b_1, \ldots, b_n$ and B, and any $v \in T_n$,*

$$\mathbb{P}\left(W_i \geqslant b_i\,(i < n) | W_n = B\right) = \mathbb{P}\left(\frac{W_i}{W_n} \geqslant \frac{b_i}{B}\,(i < n)\right), \quad (4.1)$$

$$\mathbb{P}\left(W_i \leqslant b_i\,(i < n) | W_n = B\right) = \mathbb{P}\left(\frac{W_i}{W_n} \leqslant \frac{b_i}{B}\,(i < n)\right), \quad (4.2)$$

Proposition 4.1 allows us to rescale the values W_i to choose a convenient value for W_n: for given B', letting $b'_i = b_i \cdot B'/B$, the proposition implies that

$$\mathbb{P}\{W_i \geqslant b_i\,(i<n)|W_n = B\} = \mathbb{P}\{W_i \geqslant b'_i\,(i<n)|W_n = B'\}.$$

We will use this fact rather casually in what follows. We will also use the following variant of a well-known fact about cyclically exchangeable sequences.

Proposition 4.2. *For any $S > 0$,*

$$\mathbb{P}\left\{L_0(\mathbf{v}_{n,k})|W_n = S\right\} = \mathbb{P}\left\{R_0(\mathbf{v}_{n,k})|W_n = S\right\} = \frac{1}{n}.$$

Proof. For $0 \leqslant l < n$, let $W_{n+l} = W_n + W_l$. Then, for each $0 \leqslant l < n$ and all $0 < j \leqslant n$, let $W_j^{(l)} = W_{j+l} - W_l$. Then for all l, $W_n^{(l)} = W_n$. Furthermore, each sequence $\mathbf{W}^{(l)} = (W_1^{(l)}, \ldots, W_n^{(l)})$ has distribution $\mathcal{W}_{n,k}$ and a.s. exactly one of them is leading by the Cycle Lemma [DM47]. Similarly, exactly one of the sequences $\mathbf{W}^{(l)}$ is "trailing". □

The next two lemmas are analogs of Lemmas 11 and 12 in [ABR09], and are proved using some of the same ideas. Whereas lemmas in [ABR09] use heavily the fact that a random walk $0, S_1, \ldots, S_n$ can be broken into independent sub-walks $0, S_1, \ldots, S_j$ and $0, S_{j+1}-S_j, \ldots, S_n-S_j$, in our situation the analogous subsequences $0, W_1, \ldots, W_j$ and $0, W_{j+1} - W_j, \ldots, W_n - W_j$ are not independent. However, we do have the following straightforward fact, which essentially says that conditioning on any subset of the differences $h_1 - h_0, \ldots, h_n - h_{n-1}$ breaks the sequence into independent subsequences with distributions from the same family. The proof is omitted.

Fact 4.3. Fix integers $n \geqslant 1$, $k \geqslant 0$, and let $(W_1, \ldots, W_n)$ have law $\mathcal{W}_{n,k}$. Then for any integers $1 \leqslant i \leqslant m \leqslant n$, and $1 = n_0 < n_1 < \ldots < n_m = n$, conditional upon $h_i - h_{i-1}$, the sequence

$$(W_{n_{i-1}+1} - W_{n_{i-1}}, \ldots, W_{n_i} - W_{n_i-1})$$

has law $\mathcal{W}_{n_i-n_{i-1},(h_i-h_{i-1})-(n_i-n_{i-1})}$, and is mutually independent of $(h_1, \ldots, h_n)$, of $(W_1, \ldots, W_{n_{i-1}})$, and of $(W_{n_i+1} - W_{n_i}, \ldots, W_n - W_{n-1})$.

Lemma 4.4. *Uniformly for $S > 0$, $0 \leqslant k \leqslant n$ and $a \geqslant 0$,*

$$\mathbb{P}\left\{L_a(\mathbf{v}_{n,k})|W_n(\mathbf{v}_{n,k}) = S\right\} \ll \frac{(an/S)^6 + 1}{n},$$

$$\mathbb{P}\left\{R_a(\mathbf{v}_{n,k})|W_n(\mathbf{v}_{n,k}) = S\right\} \ll \frac{(an/S)^6 + 1}{n}.$$

Remark: Most likely, the exponent "6" can be replaced with "2", in analogy with results from [ABR09] about ballot theorems for random walks.

Given that $L_a(\mathbf{v}_{n,k})$ holds, it is likely that $W_i - (i/n)W_n$ remains large when i is far from 1 and far from n. It is also likely that h_j is not too large when j is small, and similarly $h_n - h_j$ is not large when j is near n. The next two lemmas make this very precise.

For $v \in T_n$, define the events

$$B_a(v) = \{\exists m \in [a^{40}, n - a^{40}] : W_m(v) \leqslant (m/n)W_n(v) + \min(m, n-m)^{1/40}\}$$

and

$$D_a(v) = \{\exists j : h_j(v) > 3aj \text{ or } h_n(v) - h_j(v) > 3a(n-j)\}.$$

Lemma 4.5. *Uniformly for $0 \leqslant k \leqslant n/2$, $n/10 \leqslant S \leqslant n$ and $a \geqslant 1$,*

$$\mathbb{P}\left\{L_a(\mathbf{v}_{n,k}), B_a(\mathbf{v}_{n,k})|W_n(\mathbf{v}_{n,k}) = S\right\} \ll \frac{1}{na^7}.$$

Lemma 4.6. *Uniformly for* $0 \leqslant k \leqslant n/2$, $n/10 \leqslant S \leqslant n$ *and* $a \geqslant 0$,

$$\mathbb{P}\left\{L_a(\mathbf{v}_{n,k}), D_a(\mathbf{v}_{n,k}) | W_n(\mathbf{v}_{n,k}) = S\right\} \ll \frac{e^{-a}}{n}.$$

Proof of Lemma 4.4. It suffices to prove the lemma when $a \geqslant 10$. We also assume $a \leqslant n^{1/6}$, or else the conclusion is trivial. Finally, in light of Proposition 4.1 we may assume without loss of generality that $S = n + k$, so that $n \leqslant S \leqslant 2n$.

Let

$$m = a^2, \quad l = \lceil km/n \rceil,\ n' = n + 2m,\ k' = k + 2l,\ \lambda = \frac{n' + k'}{n'},\ a' = \frac{an\lambda}{S}.$$

We remark that $m, l \leqslant n^{1/3}$, $a\lambda/2 \leqslant a' \leqslant a\lambda$, and for n large enough $1 \leqslant \lambda \leqslant 3$.

By Proposition 4.2,

$$A := \mathbb{P}\left\{L_0(\mathbf{v}_{n',k'}) | W_{n'}(\mathbf{v}_{n',k'}) = \lambda n'\right\} = \frac{1}{n'}. \tag{4.3}$$

Now let $\mathbf{W}' = (W'_1, \ldots, , W'_{n'})$ be a sequence with law $\mathcal{W}_{n',k'}$. We bound A from below by counting only sequences with $h'_m = m + l$ and $h'_{n-m} = (n' + k') - (m + l)$. In this way, we can break $\mathbf{W}'$ into three subsequences, namely

$\widetilde{\mathbf{W}}$, where $\widetilde{W}_j = W'_j, \tilde{h}_j = h'_j$ $(1 \leqslant j \leqslant m)$,
$\mathbf{W}$, where $W_j = W'_{j+m} - W'_m, h_j = h'_{j+m} - h'_j$ $(1 \leqslant j \leqslant n)$,
$\widehat{\mathbf{W}}$, where $\widehat{W}_j = W'_{n'} - W'_{n'-j}, \hat{h}_j = n' + k' - h'_{n'-j}$ $(1 \leqslant j \leqslant m)$.

That is, $\widetilde{\mathbf{W}}$ captures the first m steps, $\mathbf{W}$ the next n steps, and $\widehat{\mathbf{W}}$ the last m steps taken in reverse order.

We'll work with four events:

$$\begin{aligned}
E_1 &= \{h'_m = m + l,\ h'_{n'-m} = (n' + k') - (m + l)\},\\
E_2 &= \{\widetilde{W}_j \geqslant \lambda j\ (j \leqslant m),\ \widetilde{W}_m - \lambda m \in [a', 2a']\},\\
E_3 &= \{\widehat{W}_j \leqslant \lambda j\ (j \leqslant m),\ \widehat{W}_m - \lambda m \in [-3a', -2a']\},\\
E_4(x, y) &= \{W_j \geqslant \lambda j - x\ (j < n)\}.
\end{aligned}$$

Given E_1, $\widetilde{\mathbf{W}}$ and $\widehat{\mathbf{W}}$ have law $\mathcal{W}_{m,l}$, and $\mathbf{W}$ has law $\mathcal{W}_{n,k}$, and all three are independent. Also given E_1, the events E_2, E_3 and $E_4(x, y)$ are independent. Thus,

$$\begin{aligned}
A \geqslant\ & \mathbb{P}\{E_1 \mid W'_{n'} = \lambda n'\}\, \mathbb{P}\{E_2 | E_1, W'_{n'} = \lambda n'\}\, \mathbb{P}\{E_3 | E_1, W'_{n'} = \lambda n'\}\\
& \cdot \inf_{\substack{a' \leqslant x \leqslant 2a' \\ -3a' \leqslant y \leqslant -2a'}} \mathbb{P}\{E_4(x, y) | E_1, W'_{n'} = \lambda n', W_n = \lambda n - x - y\}.
\end{aligned} \tag{4.4}$$

Since $m + l = O(n^{1/3})$, if $k > 0$ then a slightly tedious but routine computation with Stirling's formula and (3.1) gives

$$\mathbb{P}\{E_1 \mid W'_{n'} = \lambda n'\} = \frac{\binom{m+l-1}{l}^2 \binom{n+k-1}{k}}{\binom{n+2m+2l+k-1}{k+2l}} \asymp \binom{m + l - 1}{l}^2 \frac{k^{2l} n^{2m}}{(n + k)^{2m+2l}} \asymp \frac{1}{l}. \tag{4.5}$$

When $k = 0$, trivially $\mathbb{P}\{E_1 \mid W'_{n'} = \lambda n'\} = 1$. For the remainder of the proof we write $\mathbb{P}^{\mathrm{c}}\{\cdot\}$ to mean $\mathbb{P}\{\cdot \mid E_1, W'_{n'} = \lambda n'\}$. Next,

$$\mathbb{P}^{\mathrm{c}}\{E_2\} = \mathbb{P}^{\mathrm{c}}\left\{\widetilde{W}_j \geqslant \lambda j\ (j < m) | \widetilde{W}_m - \lambda m \in [a', 2a']\right\} \mathbb{P}^{\mathrm{c}}\left\{\widetilde{W}_m - \lambda m \in [a', 2a']\right\}.. \tag{4.6}$$

Given that $W'_{n'} = \lambda n'$ and $h'_m = m + l$, $\widetilde{W}_m$ has distribution

$$\lambda n' \cdot \text{Beta}(m + l, n + k - m - l),$$

and in particular has mean

$$\lambda n'(m + l)/(n + k) = \lambda m + O\left(\frac{m^2}{n}\right) = \lambda m + O(1)$$

and variance

$$(\lambda n')^2 \frac{(m + l)(n + k - m - l)}{(n + k)^2(n + k + 1)} = O(m).$$

Since $a' \geqslant \frac{a}{2} \geqslant \frac{1}{2}\sqrt{m}$, it follows that the second probability on the right side of (4.6) is $\gg 1$. Applying Proposition 4.1 followed by Proposition 4.2, the first factor on the right side of (4.6) is

$$\begin{aligned}
&\geqslant \inf_{a' \leqslant x \leqslant 2a'} \mathbb{P}^c\{\widetilde{W}_j \geqslant \lambda j \ (j < m) | \widetilde{W}_m = \lambda m + x\} \\
&\geqslant \inf_{a' \leqslant x \leqslant 2a'} \mathbb{P}\left\{L_0(\mathbf{v}_{m,l}) | W_m(\mathbf{v}_{m,l}) = \lambda m + x\right\} = \frac{1}{m}.
\end{aligned}$$

Therefore,

$$\mathbb{P}^c\{E_2\} \gg \frac{1}{m} \gg \frac{1}{a^2}. \tag{4.7}$$

Similarly,

$$\mathbb{P}^c\{E_3\} \gg \inf_{-3a' \leqslant y \leqslant -2a'} \mathbb{P}\left\{L_0(\mathbf{v}_{m,l}) | W_m(\mathbf{v}_{m,l}) = \lambda m + y\right\} = \frac{1}{m} \gg \frac{1}{a^2}. \tag{4.8}$$

Lastly, for $a' \leqslant x \leqslant 2a'$ and $-3a' \leqslant y \leqslant -2a'$, Proposition 4.1 yields

$$\begin{aligned}
\mathbb{P}^c\{E_4(x, y) | W_n = \lambda n - x - y\} &= \mathbb{P}\left\{\frac{W_j}{W_n} \geqslant \frac{\lambda j - x}{\lambda n - x - y} \ (j \leqslant n)\right\} \\
&\geqslant \mathbb{P}\left\{\frac{W_j}{W_n} \geqslant \frac{j}{n} - \frac{a}{S} \ (j \leqslant n)\right\} \\
&= \mathbb{P}\left\{L_a(\mathbf{v}_{n,k}) | W_n(\mathbf{v}_{n,k}) = S\right\}. \qquad (4.9)
\end{aligned}$$

Together, (4.3)–(4.9) imply

$$\frac{1}{n} \gg \frac{1}{a^6} \mathbb{P}\left\{L_a(\mathbf{v}_{n,k}) | W_n(\mathbf{v}_{n,k}) = S\right\},$$

which proves the first assertion of the lemma. The proof of the second part is identical. □

Proof of Lemma 4.5. Fix k, S, and a as in the statement of the lemma. We write $W_m = W_m(\mathbf{v}_{n,k})$, $h_m = h_m(\mathbf{v}_{n,k})$ and so on. If $a^{40} > n/2$ then there is nothing to prove so we assume $a^{40} \leqslant n/2$. For $a^{40} \leqslant m \leqslant n - a^{40}$ and $l \geqslant 0$, let

$$A_{m,l} = \mathbb{P}\left\{L_a(\mathbf{v}_{n,k}), W_m \leqslant \frac{m}{n} S + \min(m, n - m)^{1/40} \,\middle|\, W_n = S, h_m = m + l\right\}.$$

Break $(W_1, \ldots, W_n)$ into two sequences: $\widetilde{W}_j = W_j$ for $j \leqslant m$, and $\widehat{W}_j = W_n - W_{n-j}$ for $j \leqslant n - m$ (the latter being the final $n - m$ steps taken in reverse). Given $h_m = h_m - h_0$, these sequences are independent by Fact 4.3. We write $\mathbb{P}^c\{\cdot\}$ for the conditional probability measure $\mathbb{P}\{\cdot | h_m = m + l\}$, and $\mathbb{E}^c\{\cdot\}$ for the corresponding expectation operator. Also, let $\lambda = \frac{n+k}{n}$.

Suppose first that $a^{40} \leqslant m \leqslant n/2$. Put $b = m^{1/40}\frac{n+k}{S}$ and $a' = a\frac{n+k}{S}$. Note that $\mathbb{E}^{c}\left\{\widetilde{W}_m \mid W_n = S\right\} = S \cdot (m+l)/(n+k)$. Rescaling by $(n+k)/S$ (this is allowed by the comment just after Proposition 4.1), by the definitions of b and a' we have

$$\begin{aligned} A_{m,l} \leqslant\ & \mathbb{P}^{c}\left\{\widetilde{W}_m - \lambda m \in [-a', b] | W_n = \lambda n\right\} \\ & \times \sup_{-a' \leqslant x \leqslant b} \mathbb{P}^{c}\left\{\widetilde{W}_j \geqslant \lambda j - a'\ (j < m) | \widetilde{W}_m = \lambda m + x\right\} \\ & \times \sup_{-b \leqslant x \leqslant a'} \mathbb{P}^{c}\left\{\widehat{W}_j \leqslant \lambda j + a'\ (j < n-m) | \widehat{W}_{n-m} = \lambda(n-m) + x\right\}. \end{aligned} \tag{4.10}$$

Given that $W_n = \lambda n$ and $h_m = m + l$, $\widetilde{W}_m$ has distribution $\lambda n \cdot \mathrm{Beta}(m+l, k+n-m-l)$ and so the first factor on the RHS of (4.10) is $O(b/\sqrt{m})$ uniformly in l. Applying Proposition 4.1 and the first inequality of Lemma 4.4, the second factor on the RHS of (4.10) is

$$\begin{aligned} & \leqslant \mathbb{P}^{c}\left\{\frac{\widetilde{W}_j}{\widetilde{W}_m} \geqslant \frac{j}{m} - \frac{a' + bj/m}{m\lambda + b}\ (j < m)\right\} \\ & \leqslant \mathbb{P}^{c}\left\{\frac{\widetilde{W}_j}{\widetilde{W}_m} \geqslant \frac{j}{m} - \frac{a' + b}{m\lambda + b}\ (j < m)\right\} \\ & = \mathbb{P}\left\{L_{a'+b}(\mathbf{v}_{m,l}) | W_m(\mathbf{v}_{m,l}) = m\lambda + b\right\} \ll \frac{(a'+b)^6 + 1}{m} \ll \frac{b^6}{m}, \end{aligned}$$

so the product of the first two factors on the right-hand side of (4.10) is $O(b^7 m^{-3/2})$. Similarly, by Proposition 4.1 and the second inequality of Lemma 4.4, the third factor on the RHS of (4.10) is is

$$\leqslant \mathbb{P}\left\{\frac{\widehat{W}_j}{\widehat{W}_{n-m}} \leqslant \frac{j}{n-m} + \frac{a'+b}{\lambda(n-m) - b}\ (j < n-m)\right\} \ll \frac{b^6}{n-m} \ll \frac{b^6}{n}.$$

Combining these bounds, we obtain that when $a^{40} \leqslant m \leqslant n/2$, $A_{m,l} \ll b^{13}/(nm^{3/2})$. The estimation of $A_{m,l}$ with $m > n/2$ is identical, by reversing the roles of $\widetilde{\mathbf{W}}$ and $\widehat{\mathbf{W}}$. Therefore,

$$\begin{aligned} \mathbb{P}\left(L_a(\mathbf{v}_{n,k}), B_a(\mathbf{v}_{n,k}) | W_n = S\right) & \ll \sum_{a^{40} \leqslant m \leqslant n/2} \sum_{l \geqslant 0} \mathbb{P}\left\{h_m = m + l\right\} A_{m,l} \\ & \ll \frac{1}{n} \sum_{a^{40} \leqslant m \leqslant n/2} \frac{m^{13/40}}{m^{3/2}} \ll \frac{1}{na^7}. \end{aligned}$$

□

Proof of Lemma 4.6. As before, we write $W_n = W_n(\mathbf{v}_{n,k})$, $h_j = h_j(\mathbf{v}_{n,k})$ and so on. We may assume $a \geqslant 10$, or else the conclusion follows from Lemma 4.4. We also assume $k \geqslant 1$, or else $h_j = j$ for every j and $D_a(\mathbf{v}_{n,k})$ is impossible. For fixed j, given h_j the sequence $(W_1, \dots, W_n)$ breaks into two independent sequences $\widetilde{\mathbf{W}}$, consisting of the first j steps, and $\widehat{\mathbf{W}}$, consisting of the last $n-j$ steps taken in reverse. If $W_n = S$ and $L_a(\mathbf{v}_{n,k})$ holds, then there is an integer $b \geqslant -a-1$ so that $\widetilde{W}_j - \frac{j}{n}S \in [b, b+1]$. Consequently, $\widehat{W}_{n-j} - \frac{n-j}{n}S \in [-b-1, -b]$.

Fix $j > 3aj$ – note that in this case $j < \frac{n+k}{3a} \leqslant \frac{n}{20}$ – and suppose that $h_j = h$. Given that $h_n = h$ and $W_n = S$, tW_j has distribution $S \cdot \mathrm{Beta}(h, n+k-h)$. Since $k \leqslant n/2$ and $S \leqslant n$, it is then straightforward to check that $\mathbb{P}\left\{\widetilde{W}_j \geqslant b \mid h_n = h, W_n = s\right\} \leqslant e^{-b/4}$ for $b \geqslant 4h$. We also have

$$\begin{aligned}
&\mathbb{P}\left\{\widehat{W}_i \geqslant \frac{i}{n}S - (a+b)\,(i \leqslant n-j) \,\middle|\, \widehat{W}_{n-j} - \frac{n-j}{n}S \in [-b-1,-b], h_j = h, W_n = S\right\} \\
&\quad\leqslant \mathbb{P}\left\{L_{2a+b}(\mathbf{v}_{n-j,k-h+j}) \,\middle|\, W_{n-j}(\mathbf{v}_{n-j,k-h+j}) - \frac{n-j}{n}S \in [-b-1,-b]\right\} \\
&\quad\ll \frac{(2a+b)^6}{n}
\end{aligned}$$

by Lemma 4.4 if $b \leqslant n^{1/6}$, and trivially otherwise. Summing on b, we find that

$$\mathbb{P}\left\{L_a(\mathbf{v}_{n,k}) | h_j = h, W_n = S\right\} \ll \sum_{-a-1 \leqslant b \leqslant 4h} \frac{(2a+b)^6}{n} + \sum_{b > 4h} \frac{(2a+b)^6}{ne^{b/4}} \ll \frac{h^7}{n}.$$

Note that $(h_1, \dots, h_n)$ is independent of W_n and so $\mathbb{P}\{h_j = h \mid W_n = S\} = \mathbb{P}\{h_j = h\}$. Since $h - j \leqslant k \leqslant n/2$, by Stirling's formula

$$\begin{aligned}
\mathbb{P}\{h_j = h\} &= \binom{h-1}{h-j} \frac{\binom{n+k-h-1}{k-h+j}}{\binom{n+k-1}{k}} \\
&\leqslant \frac{h^j}{j!} \cdot \frac{(n-1)\cdots(n-j)\cdot k \cdots (k-h+j+1)}{(n+k-1)\cdots(n+k-h)} \\
&\leqslant \left(\frac{eh}{j}\right)^j \left(\frac{k}{n}\right)^{h-j} \leqslant (6ae)^{h/(3a)}\, 2^{-h} < e^{-h/2}, \qquad (4.11)
\end{aligned}$$

the last inequality holding at least for $a \geqslant 5$ (which we have assumed). Summing over $h > 3aj$, then over j, we find that

$$\mathbb{P}\left\{L_a(\mathbf{v}_{n,k}), \exists j : h_j > 3aj \,|W_n = S\right\} \ll \frac{1}{n} \sum_{j \geqslant 1} \sum_{h > 3aj} h^7 e^{-h/2} \ll \frac{e^{-a}}{n}. \qquad (4.12)$$

Next, suppose $h = h_n - h_j > 3a(n-j)$, in which case $n - j < \frac{n}{20}$. Let $b' = W_j - \frac{j}{n}S$. Since $W_{i+1} \geqslant W_i$ for all i, $W_j \leqslant S$ and so $b' \leqslant \frac{n-j}{n}S \leqslant n-j$. Also, in order for $L_{n,k}(a)$ to occur we must have $b' \geqslant -a$. Thus, writing $\mathcal{I} = [-a, n-j]$, and ignoring the last $n-j$ steps of W for an upper bound, we have

$$\begin{aligned}
&\mathbb{P}\left\{L_a(\mathbf{v}_{n,k}) | h_n - h_j = h, W_n = S\right\} \\
&\quad\leqslant \sup_{b' \in \mathcal{I}} \mathbb{P}\left\{\widetilde{W}_i \geqslant \frac{i}{n}S - a\,(i \leqslant j) \,\middle|\, \widetilde{W}_j = \frac{j}{n}S + b', h_n - h_j = h\right\} \\
&\quad\leqslant \sup_{b' \in \mathcal{I}} \mathbb{P}\left\{L_{a+b'}(\mathbf{v}_{j,n+k-h}) \,\middle|\, W_j(\mathbf{v}_{j,n+k-h}) = \frac{j}{n}S + b'\right\},
\end{aligned}$$

Note that $a \leqslant n/2$ (or else $3a > 3n/2 > n+k$ and $D_a(\mathbf{v}_{n,k})$ is impossible). Since $j \geqslant \frac{19}{20}n$ and $b' \geqslant -a \geqslant -n/2$, by Lemma 4.4 and straightforward manipulations the last probability is $O(\frac{1}{n}(a+b')^6) = O(\frac{1}{n}(a+n-j)^6)$. Also, $\mathbb{P}\{h_n - h_j = h\} = \mathbb{P}\{h_{n-j} = h\} < e^{-h/2}$ by the same calculation as in (4.11). Summing over $h > 3a(n-j)$ and $j \leqslant n-1$ gives

$$\mathbb{P}\left\{L_a(\mathbf{v}_{n,k}), \exists j : h_n - h_j > 3a(n-j) \,|W_n = S\right\} \ll \frac{e^{-a}}{n}.$$

Together with (4.12), this completes the proof. ∎

5. THE LOWER BOUND IN THEOREM 1.1

We continue to adopt the notational conventions from the previous section. Let c be a sufficiently large positive constant, and $b = e^{c/3}$. Let

$$Y_n = \bigcup_{|k-n/e| \leqslant \sqrt{6n\log n}} T_{n,k},$$

and put $m_n = \frac{n}{e} + \frac{3\log n}{2e}$. If $M_n \leqslant m_n - c$, then one of the following must occur:

(i) For some $v \in T_n$, $S(v) \leqslant m_n - \log n$;
(ii) For some k satisfying $|k - n/e| > \sqrt{6n\log n}$ and some $v \in T_{n,k}$, $S(v) \leqslant m_n$;
(iii) For some $v \in Y_n$, $m_n - \log n \leqslant S(v) \leqslant m_n$ and $W_i \leqslant (i/n)W_n - \log n$ for some i;
(iv) For some $v \in Y_n$, $m_n - \log n \leqslant S(v) \leqslant m_n - c$ and $W_i \geqslant (i/n)W_n - b$ for all i;
(v) For some $v \in Y_n$ and some integer $a \in [b, \log n + 1]$, $m_n - \log n \leqslant S(v) \leqslant m_n$, $W_i \geqslant (i/n)W_n - a$ for all i and $W_j < (j/n)W_n - (a-1)$ for some j (write $F_{a,j}$ for the event that this occurs for a given a and j with j minimal, and note that these events are disjoint).

By Lemma 3.2 and Stirling's forumula, the probability of (i) is at most $\mathbb{E}\{T_n(m_n - \log n)\} = O(n^{1-e})$. The probability of (ii) is $O(n^{-1/2})$ by Lemma 3.1 (b). If (iii) occurs, then $M_i \leqslant (i/n)m_n - \log n$, and this happens with probability at most $\mathbb{E}\{T_i((i/n)m_n - \log n)\}$, which is $O(n^{3/2-e}i^{-1/2})$ by Lemma 3.2. Summing on i, we find that (iii) occurs with probability $O(n^{2-e})$.

To bound the probability of the event in (iv), we write E_k for the event that there is $v \in T_{n,k}$ for which $m_n - \log n \leqslant S(v) \leqslant m_n - c$ and $W_i \geqslant (i/n)W_n - b$ for all i, so that by a union bound and Lemma 4.4, the probability of (iv) is at most

$$\begin{aligned}
& \sum_{|k-n/e| \leqslant \sqrt{6n\log n}} \mathbb{P}\{E_k\} \\
\leqslant \quad & \sum_{|k-n/e| \leqslant \sqrt{6n\log n}} |T_{n,k}| \, \mathbb{P}\{m_n - \log n \leqslant S(\mathbf{v}_{n,k}) \leqslant m_n - c, L_b(\mathbf{v}_{n,k})\} \\
\ll \quad & \sum_{|k-n/e| \leqslant \sqrt{6n\log n}} |T_{n,k}| \, \mathbb{P}\{m_n - \log n \leqslant S(\mathbf{v}_{n,k}) \leqslant m_n - c\} \cdot \frac{b^6}{n} \\
\leqslant \quad & \frac{b^6}{n} \cdot \mathbb{E}\{T_n(m_n - c)\} \\
\ll \quad & e^{(2-e)c}.
\end{aligned} \tag{5.1}$$

(This line of argument will arise again in bounding (v), and we will omit the details.)

Finally, we bound (v). To do so, we are forced to separately treat j in three different ranges. First suppose $j \leqslant a^{40}$. If $F_{a,j}$ occurs then $M_j \leqslant (j/n)m_n - a$, the probability of which is $O(j^{-1/2}e^{-ea})$ by Lemma 3.2. Summing on a and on $j \leqslant a^{40}$ gives a total probability of $O(e^{-2b})$ for this range of parameters.

Next suppose that $a^{40} < j < n - a^{40}$, so that $(\min(j, n-j))^{1/40} \geqslant a$. If $F_{a,j}$ occurs then for some $v \in Y_n$, $L_a(v)$ and $B_a(v)$ both occur. Note that for n large enough $n/10 \leqslant m_n - \log n \leqslant m_n \leqslant n$, and for all k for which $T_{n,k} \subseteq Y_n$ we have $0 \leqslant k \leqslant n/2$. Thus, for such n, k, and a, we may apply Lemma 4.5 to see that

$$\mathbb{P}\{L_a(\mathbf{v}_{n,k}), B_a(\mathbf{v}_{n,k}) \mid m_n - \log n \leqslant W_n(\mathbf{v}_{n,k}) \leqslant m_n\} \ll \frac{1}{na^7}.$$

Further, the expected number of $v \in Y_n$ with $S(v) \leqslant m_n$ is $O(n)$ by Lemma 3.2. By these two bounds and a reprise of the argument leading to (5.1), we see that for a given a, the probability of

$\bigcup_{j\in[a^{40},n-a^{40}]} F_{a,j}$ is $O(1/a^7)$ and summing over integers $a \in [b, \log n+1]$, gives a total probability of $O(1/b^6) = O(e^{-2c})$.

Now suppose $F_{a,j}$ holds with $j \in [n-a^{40}, n]$ and $a \in [b, \log n+1]$. By the definition of $F_{a,j}$, letting w be the unique ancestor of v in T_j, the event $L_{a-1}(w)$ also occurs. Since $j \geqslant n-(\log n+1)^{40}$, for n sufficiently large $|m_j - (j/n)m_n| \leqslant 1$ and hence $S(w) \leqslant m_j + 1 - (a-1)$. On the other hand, for any integer $k' \geqslant 1$, by Lemma 4.4 we have

$$\mathbb{P}\left\{W_j(\mathbf{v}_{j,k'}) \leqslant m_j + 2 - a, L_{a-1}(\mathbf{v}_{j,k'})\right\} \ll \frac{a^6}{j}\,\mathbb{P}\left\{W_j(\mathbf{v}_{j,k'}) \leqslant m_j + 2 - a\right\}$$

By Lemma 3.2, it follows that

$$\mathbb{P}\left\{G_{a,j}\right\} \ll \frac{a^6}{j}\,\mathbb{E}\,|T_j(m_j+2-a)| \ll a^6 e^{-ea}.$$

Summing first over $j \in [n-a^{40}, n]$, then over $a \in [b, \log n+1]$, we see that the probability $G_{a,j}$ occurs for any a and j in this range is

$$\ll b^{46}e^{-eb} = \exp\{(46/3)c - e^{1+c/3}\} < e^{-2c},$$

as long as c is large enough. Combining the three ranges, we obtain that (v) occurs with probability $\ll e^{-2c}$. Altogether, the probability that one of (i)-(v) holds is $\ll e^{(2-e)c}$, which is less than $1/2$ if c is chosen large enough. Hence, $\widetilde{M}_n \geqslant m_n - c$.

6. THE UPPER BOUND IN THEOREM 1.1

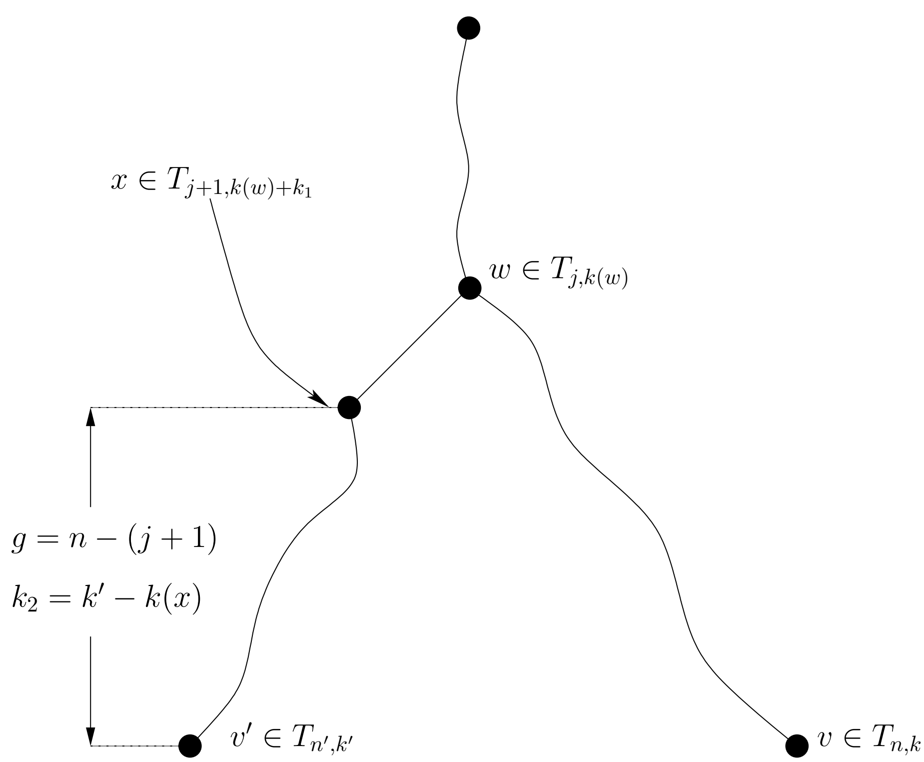

FIGURE 1

For the upper bound for median(M_n), we use a second-moment method. By the Cauchy-Schwarz inequality, for any non-negative random variable X,

$$\mathbb{P}\{X > 0\} \geqslant \frac{[\mathbb{E}\,X]^2}{\mathbb{E}\,X^2}. \tag{6.1}$$

When X is the size of some random subset $\mathcal{X}$ of a ground set V, we may rewrite (6.1) using the fact that

$$\mathbb{E}\,X^2 = \sum_{v,w\in V} \mathbb{P}\{v \in \mathcal{X}, w \in \mathcal{X}\} = \sum_{v\in V} \mathbb{E}\,[X \mid v \in \mathcal{X}]\,\mathbb{P}\{v \in \mathcal{X}\},$$

so that

$$\mathbb{P}\{X > 0\} \geqslant \frac{[\mathbb{E}\,X]^2}{\sum_{v\in V} \mathbb{E}\,[X \mid v \in \mathcal{X}]\,\mathbb{P}\{v \in \mathcal{X}\}} \geqslant \frac{\mathbb{E}\,X}{\sup_{v\in V} \mathbb{E}\,[X \mid v \in \mathcal{X}]}. \tag{6.2}$$

Let a be a large positive constant. Let $V = Y$, where Y is defined as in the previous section, and let $\mathcal{X}$ be the set of nodes in $v \in Y$ satisfying

(i) $m_n - 1 \leqslant S(v) \leqslant m_n$,
(ii) $L_a(v)$,
(iii) neither $B_a(v)$ nor $D_a(v)$.

Taking $X = |\mathcal{X}|$, by Lemmas 4.2, 4.5 and 4.6

$$\mathbb{E}\,X \geqslant \mathbb{E}[T_n(m_n) - T_n(m_n - 1)]\left(\frac{1}{n} - O\left(\frac{1}{a^7 n}\right) - O\left(\frac{e^{-a}}{n}\right)\right) \gg 1 \tag{6.3}$$

if a is chosen large enough.

Recall that for all $v \in Y$, $|k(v) - n/e| \leqslant \sqrt{6n\log n}$. For fixed $v \in Y$, we need to estimate $\mathbb{E}\{X \mid v \in \mathcal{X}\}$.

The definitions of the coming two paragraphs are for the most part depicted in Figure 1. Write $j = j(v, v')$ for the integer $0 \leqslant j < n$ such that v and v' are descendants of two distinct children of some node $w = w(v, v') \in T_j$ (and let $j(v, v') = n$ if $v = v'$). Supposing $0 \leqslant j(v, v') \leqslant n - 1$, let x be the unique child of w on the path from w to v'.

Also, write $\mathbf{W} = \mathbf{W}(v)$ and $\mathbf{W}' = (W'_1, \ldots, W'_{n'}) = \mathbf{W}(v')$. Let $g = n - (j(v, v') + 1)$, let $\widetilde{W}_i = \widetilde{W}_i(v, v') = W'_{j+i+1} - W'_{j+1}$ for $1 \leqslant i \leqslant g$, and let $\widetilde{\mathbf{W}} = (\widetilde{W}_1, \ldots, \widetilde{W}_g)$, so in particular $W'_{n'} = W'_{j+1} + \widetilde{W}_g$.

Finally, let let $k' = k(v')$, let $k_1 = k_1(v, v') = k(x) - k(w)$ and let $k_2 = k' - k(x)$, so $k_1 + k_2 = k' - k(w)$. Note that once g and k_2 are fixed, $\widetilde{\mathbf{W}}$ is independent of $\mathbf{W}$ and has law $\mathcal{W}_{g,k_2}$.

For integers j, $0 \leqslant j \leqslant n$, let $\mathcal{F}_j = \mathcal{F}_j(v) = \{v' \in \mathcal{X}, j(v, v') = j\}$ and let $F_j = F_j(v) = \mathbb{E}\{|\mathcal{F}_j| \mid v \in \mathcal{X}\}$. Clearly, $F_n = 1$, as $j = n$ implies $v = v'$.

Now fix v'. If $v' \in \mathcal{X}$, then by (i), (ii) and (iii), we have

$$k_1 + k_2 \leqslant \min\left(3a(j + 1), 3a(g + 1)\right) \tag{6.4}$$

and if $v \in \mathcal{X}$ then, with $j = j(v, v')$, we have

$$W_j \geqslant \frac{j}{n}(m_n - 1) + \begin{cases} (-a) & \text{whatever the value of } j \\ \min(j, n - j)^{1/40} & \text{if } a^{40} \leqslant j \leqslant n - a^{40}. \end{cases} \tag{6.5}$$

Consider separately four ranges of j. First, if $n - a^{40} \leqslant j \leqslant n - 1$, then for sufficiently large n, (6.4) implies that $k_1 + k_2 \leqslant 3a(n - j)$, so F_j is deterministically at most

$$\sum_{l\leqslant 3a(n-j)} |T_{n-j,l}| = \sum_{l\leqslant 3a(n-j)} \binom{n - j + l - 1}{l} \leqslant 3a^{41}(a^{40} + 3a^{41})^{a^{40}}.$$

Hence, recalling that a is now a fixed, large constant,

$$\sum_{n-a^{40}\leqslant j\leqslant n} F_j \ll 1. \tag{6.6}$$

Next, let $r=(2\log n)^{40}$. If $n-r<j\leqslant n-a^{40}$ then for n sufficiently large, $j\geqslant n-j=g+1$, and (6.5) implies that in order to have $W_n'\leqslant m_n$ we must have

$$\widetilde{W}_g \leqslant \frac{g+1}{n}m_n - g^{1/40}+1 \leqslant g/e - g^{1/40}+2,$$

the second inequality holding for sufficiently large n. For fixed k_1, by Lemma 3.2 we thus have

$$\begin{aligned}\mathbb{E}\left\{|\{v'\in\mathcal{X}, j(v,v')=j, k_1(v,v')=k_1\}|\mid v\in\mathcal{X}\right\} &\leqslant \mathbb{E}\,T_g(g/e-g^{1/40}+2)\\ &\ll \exp[-eg^{1/40}].\end{aligned}$$

Using (6.4) to bound k_1 and summing over j yields

$$\sum_{n-r<j\leqslant n-a^{40}} F_j \ll \sum_{a^{40}\leqslant g\leqslant r} a(g+1)\exp[-eg^{1/40}] \ll 1. \tag{6.7}$$

Next, suppose $r\leqslant j\leqslant n-r$. By (6.5), in order to have $W_n'\leqslant m_n$ it must be that

$$\widetilde{W}_g \leqslant \frac{g+1}{n}m_n - \min(j,n-j)^{1/40}+1 \leqslant \frac{g}{e}-\log n.$$

Since we also require $k_1(v,v')\leqslant 3an$ by (6.4), we have $F_j\leqslant 3an\,\mathbb{E}\,T_g(g/e-\log n)\ll 1/n^2$ for this range of j, and hence

$$\sum_{r\leqslant j\leqslant n-r} F_j \ll \frac{1}{n}. \tag{6.8}$$

Finally, suppose $0\leqslant j\leqslant r$. Here $g\geqslant n-r-1=n+O((\log n)^{40})$, and since $L_a(v)$ holds by assumption, if $v\in\mathcal{X}$ then

$$W_j \geqslant \frac{j}{n}W_n - a > \frac{j}{n}m_n-(a+1).$$

For each integer $b\in[-(a+1),2\log n)$, let E_b be the event that $W_j-(j/n)m_n\in[b,b+1)$. Also, let E^* be the event that $W_j-(j/n)m_n\geqslant\lceil 2\log n\rceil$. The events $\{E_b : -(a+1)\leqslant b<2\log n\}$ and E^* together partition the event $\{v'\in\mathcal{F}_j(v)\}$, so by conditioning

$$F_j \leqslant \max\left(\mathbb{E}\left\{|\mathcal{F}_j|\mid v\in\mathcal{X},E^*\right\}, \max_{-(a+1)\leqslant b<2\log n}\mathbb{E}\left\{|\mathcal{F}_j|\mid v\in\mathcal{X},E_b\right\}\right) \tag{6.9}$$

If $W_j\geqslant (j/n)m_n+2\log n$, then to have $v'\in\mathcal{F}_j$, we must have $\widetilde{W}_g(v')\leqslant g/e-\log n$ so, as in the case $r\leqslant j\leqslant n-r$, we have

$$\mathbb{E}\left\{|\mathcal{F}_j|\mid v\in\mathcal{X},E^*\right\} \ll \frac{1}{n^2}.$$

Now suppose $W_j-(j/n)m_n\in[b,b+1]$, where b is an integer satisfying $-a-1\leqslant b\leqslant 2\log n$. Note that if $b<(j^{1/40}-2)$ and $a^{40}\leqslant j\leqslant r$ then $\mathcal{F}_j(v)$ is necessarily empty due to $B_a(v)$, so for such j and b, $\mathbb{E}\left\{|\mathcal{F}_j|\mid v\in\mathcal{X},E_b\right\}=0$. For the rest, we further subdivide $\mathcal{F}_j$, writing $\mathcal{F}_{j,l}=\{v'\in\mathcal{X}, j(v,v')=j, k_1(v,v')=l\}$. By (6.4) we have

$$\mathbb{E}\left\{|\mathcal{F}_j|\mid v\in\mathcal{X},E_b\right\} = \sum_{l\leqslant 3(j+1)}\mathbb{E}\left\{|\mathcal{F}_{j,l}|\mid v\in\mathcal{X},E_b\right\}.$$

Suppose additionally that $W'_{j+1}(v') - W'_j(v') \in [\Delta, \Delta+1]$, where Δ is a non-negative integer. Since $W_j(v') = W_j(v)$, in order to have $v' \in \mathcal{F}_j$, by (i) we require[2]

$$\widetilde{W}_g - \frac{g}{n} m_n \in [m_n/n - (b+\Delta+3), m_n/n - (b+\Delta)].$$

Since $0 \leqslant m_n/n < 1$ and, for n sufficiently large, $m_g - 1 \leqslant (g/n)m_n \leqslant m_g$, this implies that, writing $\mathcal{I} = [-(b+\Delta+4), -(b+\Delta-1)]$, we must have

$$\widetilde{W}_g - m_g \in \mathcal{I}.$$

By (i) and (ii), we also require

$$\widetilde{W}_i \geqslant \frac{i}{n} W_n - b - \Delta - a - 2 \geqslant \frac{i}{g} m_g - b - \Delta - a - 3 \qquad (i \leqslant g)$$

This implies that for all $i \leqslant g$,

$$\widetilde{W}_i \geqslant \frac{i}{g} W_g - \max(b+\Delta+a-3, a-2).$$

None of this depends on l, so for any $0 \leqslant l \leqslant 3(j+1)$, writing $m = \max(b+\Delta+a-3, a-2)$,

$$\begin{aligned}
\mathbb{E}\left\{|\mathcal{F}_{j,l}| \mid v \in \mathcal{X}, E_b\right\} &\leqslant \sum_{\substack{1 \leqslant k_2 \leqslant 3a(j+1) \\ \Delta \geqslant 0}} \mathbb{E}\left|\{v \in T_{g,k_2} \ : \ S(v) - (g/n)m_n \in \mathcal{I}, L_m(v)\}\right| \\
&\leqslant \sum_{\substack{1 \leqslant k_2 \leqslant 3a(j+1) \\ \Delta \geqslant 0}} \Big(\mathbb{E}\left\{T_{g,k_2}(m_g - (b+\Delta-1)\right\} \\
&\qquad \cdot \sup_{x \in \mathcal{I}} \mathbb{P}\left\{L_m(\mathbf{v}_{g,k_2}) \mid W_g(\mathbf{v}_{g,k_2}) = m_g + x\right\} \Big) \\
&\ll \sum_{\Delta \geqslant 0} \mathbb{E}\left\{T_g(m_g - (b+\Delta-1))\right\} \cdot j \cdot \frac{\max(b+\Delta+a-3, a-2)^6}{n} \\
&\ll \sum_{\Delta \geqslant 0} n e^{-e(b+\Delta)} \cdot j \cdot \frac{\max(b+\Delta+a-3, a-2)^6}{n} \\
&\ll j e^{-eb} (a+|b|)^6,
\end{aligned}$$

the third-to-last line by Lemma 4.4 and the second-to-last by Lemma 3.2. Summing over $0 \leqslant l \leqslant 3(j+1)$, it follows that

$$\mathbb{E}\left\{|\mathcal{F}_j| \mid v \in \mathcal{X}, E_b\right\} \ll j^2 e^{-eb} (a+|b|)^6.$$

For $j \leqslant a^{40}$ this is $O(1)$ uniformly in b. When $j > a^{40}$ we also have $b \geqslant j^{1/40} - 2$ and for such j, the above bound is $O(j^3 e^{-ej^{1/40}})$. By (6.9) it follows that for such $0 \leqslant j \leqslant r$,

$$F_j \ll \begin{cases} 1 & \text{if } j \leqslant a^{40} \\ \max(n^{-2}, j^3 \exp(-ej^{1/40})) & \text{if } j > a^{40} \end{cases}$$

Summing on j, we find that

$$\sum_{0 \leqslant j \leqslant r} F_j \ll 1. \tag{6.10}$$

[2]The m_n/n terms come from the "skipped step" from W'_j to W'_{j+1}, and the $(b+\Delta+3)$ comes from $b+1$, $\Delta+1$, and the requirement that $S(v) \geqslant m_n - 1$.

Together, (6.6),(6.7),(6.8), and (6.10) imply that for every $v \in T$,

$$\mathbb{E}[X : v \in \mathcal{X}] = O(1).$$

Combining this estimate with (6.2) and (6.3), shows that $\mathbb{P}\{X > 0\} \gg 1$, and if $X > 0$ then $M_n \leqslant m_n$, so there exists an absolute constant $\epsilon > 0$ such that for all n,

$$\mathbb{P}\{M_n \leqslant m_n\} \geqslant \epsilon.$$

From here it is straightforward to show that $M_n \leqslant m_n + O(1)$, and we now do so. The next two lemmas, taken from [FKL10], are standard bounds for BRW. As the proofs are short, we include them here.

Lemma 6.1. *For any BRW, positive integers m, n and positive real numbers M, N,*

$$\mathbb{P}\{M_{m+n} \geqslant M + N\} \leqslant \mathbb{E}[(\mathbb{P}\{M_n \geqslant N\})^{T_m(M)}].$$

Proof. Suppose $M_{m+n} \geqslant M + N$ and $T_m(M) = k$. For each of these k individuals, all of their descendants in generation $m + n$ are offset from their generation m ancestor by at least N. ☐

Lemma 6.2. *Let m, n be positive integers and let $M > 0, \varepsilon > 0$ be real. If $\mathbb{E}\{(1-\varepsilon)^{T_m(M)}\} < \frac{1}{2}$, then $\mathbb{P}\{M_n < \widetilde{M}_{n+m} - M\} \leqslant \varepsilon$. In particular, the conclusion holds if $\mathbb{P}\{T_m(M) < 1/\varepsilon\} \leqslant \frac{1}{5}$.*

Proof. Let $q = \sup\{x : \mathbb{P}\{M_n < x\} < \varepsilon\}$; then $\mathbb{P}\{M_n < q\} \leqslant \varepsilon$. By Lemma 6.1,

$$\mathbb{P}\{M_{m+n} \geqslant M + q\} \leqslant \mathbb{E}\left[(\mathbb{P}\{M_n \geqslant q\})^{T_m(M)}\right] < \frac{1}{2}.$$

Therefore, $M + q \geqslant \widetilde{M}_{m+n}$, and thus $\mathbb{P}\{M_n < \widetilde{M}_{m+n} - M\} \leqslant \mathbb{P}\{M_n < q\} \leqslant \varepsilon$. To prove the second part, assume that $\mathbb{P}\{T_m(M) < 1/\varepsilon\} \leqslant \frac{1}{5}$. Then

$$\mathbb{E}\left\{(1-\varepsilon)^{T_m(M)}\right\} \leqslant \mathbb{P}\{T_m(m) < \tfrac{1}{\varepsilon}\} + (1 - \mathbb{P}\{T_m(M) < \tfrac{1}{\varepsilon}\})(1-\varepsilon)^{1/\varepsilon} \leqslant \frac{1}{5} + \frac{4}{5e} < \frac{1}{2}. \quad ☐$$

Now take A such that $\mathbb{P}\{T_1(A) < 1/\varepsilon\} \leqslant \frac{1}{5}$. By Lemma 6.2,

$$\mathbb{P}\left\{M_n \leqslant \widetilde{M}_n - A\right\} \leqslant \mathbb{P}\left\{M_n \leqslant \widetilde{M}_{n+1} - A\right\} \leqslant \varepsilon,$$

and hence $\widetilde{M}_n \leqslant M_n + A$, which completes the proof of the upper bound in Theorem 1.1.

7. Proof of Theorem 1.2

Let $a > 1/e$ and $0 < \eta < ae/2$. By Biggins' analog of Chernoff's inequality for the BRW [Big77, Theorem 2], for large r we have $\mathbb{P}\{T_r(ar) \leqslant (ae-\eta)^r\} \leqslant \frac{1}{5}$. Let r_0 be large enough that, in addition, $\widetilde{M}_{n+r} \geqslant M_n + (1/e - \eta)r$ for all $r \geqslant r_0$ and all n (such an r_0 exists by Theorem 1.1). Now fix $r \geqslant r_0$ and let $M = ar$, let $m = r$, and let $\varepsilon = (ae-\eta)^{-r}$. We then have $\mathbb{P}\{T_m(M) < 1/\varepsilon\} \leqslant 1/5$, so for all n, by the preceding bound for $\widetilde{M}_{n+r}$ and by Lemma 6.2, we obtain that

$$\mathbb{P}\{M_n \leqslant \widetilde{M}_n - (a - 1/e + \eta)r\} \leqslant \mathbb{P}\{M_n \leqslant \widetilde{M}_{n+r} - ar\} \leqslant (ae-\eta)^{-r}.$$

The first estimate follows with $c_1 = \frac{\log(ae-\eta)}{(a-1/e+\eta)}$. Fix a, let $\eta \to 0$, then let $a \to 1/e$, so that $c_1 \to e$. This proves the first part of Theorem 1.2.

For the second part, fix $0 < \varepsilon < 1/50$ and let $\delta = \varepsilon^2$, so that $\delta(1 + \log((1-\varepsilon/5)/\delta)) < \varepsilon/5$. Then choose r_0 sufficiently large that for all $r \geqslant r_0$, we have $(1-\varepsilon/5)r + 2\lceil \log(2r)\rceil < r$, and for all $s \geqslant \log(2r_0)$, we have $\mathbb{P}\{T_s(2s) \leqslant 4^s\} \leqslant e^{-1/\delta}$ (as in the first part, such r exists by [Big77, Theorem 2]).

Recall that if $h \in \mathbb{N}^1 = T_1$ is a child of the root in T then $S(h)$ is $\mathrm{Gamma}(h)$ distributed. Thus, for any positive integer r, by a union bound

$$\begin{aligned}\mathbb{P}\left\{T_1((1-\varepsilon/5)r) \leqslant \delta r\right\} &\leqslant \sum_{h \leqslant \delta r} \mathbb{P}\{S(h) \geqslant (1-\varepsilon/5)r\} \\ &= \sum_{h \leqslant \delta r} \frac{e^{-(1-\varepsilon/5)r}((1-\varepsilon/5)r)^h}{h!} \\ &\leqslant e^{-(1-\varepsilon/5)r} e^{(1+\log((1-\varepsilon/5)/\delta))\delta r} \\ &\leqslant e^{-(1-2\varepsilon/5)r},\end{aligned}$$

the second-to-last inequality by Proposition 3.3.

Write $s = \lceil \log(2r) \rceil$, and let E be the event that there are at least 4^s nodes in T_{s+1} with displacement at most $(1-\varepsilon/5)r + 2s < r$. If $T_1((1-\varepsilon/5)r) > \delta r$ then either E occurs, or else for each $h \leqslant \lceil \delta r \rceil$, the number of $v \in T_{s+1}$ descending from $h \in T_1$ with $S(v) - S(h) \leqslant 2s$ is less than 4^s. The latter event has probability less then $(e^{-(1/\delta)})^{\delta r} = e^{-r}$. It follows that

$$\mathbb{P}\left\{E^c\right\} \leqslant e^{-(1-2\varepsilon/5)r} + e^{-r} \leqslant e^{-(1-\varepsilon/2)r},$$

the last inequality holding for large r. Finally, if $M_n > \widetilde{M}_{n-(s+1)} + r$, then for each node $v \in T_{s+1}$ with $S(v) \leqslant (1-\varepsilon/5)r + 2s$, for all $w \in T_n$ descending from v we must have $S(w) - S(v) \geqslant \widetilde{M}_{n-(s+1)}$. If E occurs then there are at least $4^s \geqslant 2r$ such nodes v, and so

$$\mathbb{P}\left\{M_n > \widetilde{M}_{n-(s+1)} + r\right\} \leqslant e^{-(1-\varepsilon/2)r} + 2^{-2r} < e^{-(1-\varepsilon)r},$$

the last inequality holding for large r. Since $\widetilde{M}_{n-(s+1)} \leqslant \widetilde{M}_n$, the second part of Theorem 1.2 is proved by letting $\varepsilon \to 0$.

Acknowledgements. The authors thank Hugh Montgomery for bringing paper [Hal44] to our attention.

References

[Hal44] P Halmos, *Random alms*, Ann. Math. Stat. **15** (1944), 182–189.
[DM47] A. Dvoretzky and T. Motzkin, *A problem of arrangements*, Duke Mathematical Journal **14** (1947), 305–313.
[Har63] T. E. Harris, *The theory of branching processes*, Springer, Berlin, 1963.
[Bil72] P. Billingsley, *On the distribution of large prime divisors*, Periodica Mathematica Hungarica **2** (1972), 283–289.
[Ham74] J.M. Hammersley, *Postulates for subadditive processes*, The Annals of Probability **2** (1974), no. 4, 652–680.
[Kin75] J.F.C. Kingman, *The first birth problem for an age-dependent branching process*, The Annals of Probability **3** (1975), no. 5, 790–801.
[Pra75] V. Pratt, *Every prime has a succinct certificate*, Siam Journal on Computing **4** (1975), 214–220.
[Big76] J. D. Biggins, *The first- and last-birth problems for a multitype age-dependent branching process*, Advances in Applied Probability **8** (1976), no. 3, 446–459.
[Big77] J. D. Biggins, *Chernoff's theorem in the branching randon walk*, J. Appl. Prob. **14** (1977), 630–636.

[Dev87] L. Devroye, *Branching processes in the analysis of the heights of trees*, Acta Informatica **24** (1987), 277–298.

[DG93] P. Donnelly and G. Grimmett, *On the asymptotic distribution of large prime factors*, Journal of the London Mathematical Society **2** (1993), no. 3, 395–404.

[Pit94] B. Pittel, *Note on the height of recursive trees and m-ary search trees*, Random Structures and Algorithms **5** (1994), 337–347.

[Bac00] M. Bachman, *Limit theorems for the minimal position in a branching random walk with independent logconcave displacements*, Adv. Appl. Prob. **32** (2000), 159–176.

[AS04] D. J. Aldous and J. M. Steele, *The objective method: probabilistic combinatorial optimization and local weak convergence*, Probability on Discrete Structures **110** (2004), 1–72.

[Ber06] J. Bertoin, *Random fragmentation and coagulation processes*, Cambridge Studies in Advanced Mathematics, vol. 102, Cambridge University Press, 2006.

[CD06] B. Chauvin and M. Drmota, *The random multisection problem, travelling waves, and the distribution of the height of m-ary search trees*, Algorithmica **46** (2006), no. 3-4, 299–327.

[Pit06] J. Pitman, *Combinatorial stochastic processes*, Lecture Notes in Mathematics, vol. 1875, Springer-Verlag, 2006.

[ABR09] L. Addario-Berry and B. Reed, *Minima in Branching Random Walks*, Annals of Probability **37** (2009), no. 3, 1044–1079.

[BZ09] M. Bramson and O. Zeitouni, *Tightness for a family of recursive equations*, Annals of Probability **37** (2009), no. 2, 615–653.

[HS09] Y. Hu and Z. Shi, *Minimal position and critical martingale convergence in branching random walks, and directed polymers on disordered trees*, Annals of Probability **37** (2009), no. 2, 742–789.

[AS10] E. Aidekon and Z. Shi, *Weak convergence for the minimal position in a branching random walk: a simple proof*, arXiv:1006.1266v2 [math.PR], 2010+.

[FKL10] K. Ford, S. Konyagin, and F. Luca, *Prime chains and pratt trees*, Geom. Funct. Anal. (2010), 1231–1258.

A-B: McGill University, 1005-805 Sherbrooke West, Montreal, QC, H3A 2K6, Canada
E-mail address: louigi@math.mcgill.ca

F: Department of Mathematics, University of Illinois at Urbana-Champaign Urbana, 1409 West Green St., Urbana, IL 61801, USA
E-mail address: ford@math.uiuc.edu