Adv. Appl. Prob. **38,** 1047–1070 (2006)
Printed in Northern Ireland

ON THE ASYMPTOTIC BEHAVIOUR OF RANDOM RECURSIVE TREES IN RANDOM ENVIRONMENTS

K. A. BOROVKOV,* *The University of Melbourne*
V. A. VATUTIN,** *Steklov Mathematical Institute*

Abstract

We consider growing random recursive trees in random environments, in which at each step a new vertex is attached (by an edge of random length) to an existing tree vertex according to a probability distribution that assigns the tree vertices masses proportional to their random weights. The main aim of the paper is to study the asymptotic behaviour of the distance from the newly inserted vertex to the tree's root and that of the mean numbers of outgoing vertices as the number of steps tends to ∞. Most of the results are obtained under the assumption that the random weights have a product form with independent, identically distributed factors.

Keywords: Random recursive tree; random environment; Spitzer's condition; distance to the root; outdegree

2000 Mathematics Subject Classification: Primary 05C80
Secondary 60G50; 05C05; 60F99

1. Introduction

We consider the following random recursive tree model. A recursive tree is constructed incrementally, by attaching a new vertex to a randomly chosen existing tree vertex at each step. Initially, the tree consists of a single vertex, $v(0)$, that has weight $w(0) = 1$ and label 0. At the first step, a new vertex, $v(1)$, is added to the tree as a child of the initial vertex. It is labelled 1, and a random weight, $w(1) > 0$, and a random length, $Y(1) \geq 0$, are respectively assigned to $v(1)$ and to the edge connecting the vertices $v(0)$ and $v(1)$. It is assumed that the edge is directed from $v(0)$ to $v(1)$. At step $j > 1$, given all the weights $w(0), w(1), \dots, w(j-1)$, first a node $v(j^*)$ is chosen at random from the nodes $v(0), v(1), \dots, v(j-1)$ according to the distribution with probabilities proportional to the nodes' weights, and then a new vertex $v(j)$ is added to the tree as a child of the node $v(j^*)$. The new vertex has label j, and a random weight $w(j) > 0$ and a random length $Y(j) \geq 0$ are respectively assigned to it and to the edge connecting the vertices $v(j^*)$ and $v(j)$. As at the initial step (where, for consistency, we will put $1^* = 0$), the edge is directed from $v(j^*)$ to its child vertex $v(j)$. We assume that $\{Y(j)\}_{j\geq 1}$ is a sequence of independent random variables (RVs) which is independent of the sequence of the (generally speaking random) weights $\{w(j)\}_{j\geq 0}$. Interpreting the sequence of weights as a 'random environment' in which our recursive tree is growing, and appealing to an analogy with random walks and branching processes in random environments, it is not unnatural to refer to such a model as a random recursive tree in a random environment.

Received 28 February 2006; revision received 15 September 2006.
* Postal address: Department of Mathematics and Statistics, The University of Melbourne, Parkville, VIC 3010, Australia. Email address: kostya@ms.unimelb.edu.au
** Postal address: Steklov Mathematical Institute RAS, Gubkin St. 8, 119991 Moscow, Russia.

Let

$$D_0 := 0, \qquad D_n := D_{n^*} + Y(n), \quad n \geq 1,$$

be the distance from the vertex $v(n)$ to the root (i.e. the sum of the lengths of the edges connecting $v(n)$ with $v(0)$). Our main aim in this paper is to study, as $n \to \infty$, the asymptotic behaviour of D_n under various assumptions on the random weights $w(j)$ and lengths $Y(j)$, and also that of the mean values of the outgoing degrees

$$N_n(j) := \sum_{k=j+1}^{n} \mathbf{1}_{\{v(k^*)=v(j)\}}, \qquad j \leq n, \tag{1}$$

where $\mathbf{1}_A$ is the indicator of the event A.

Observe that if $w(j) \equiv Y(j) \equiv 1$ for all j, then we obtain the standard random recursive tree ([11]; see also [15]). If $w(j) = a^j,\ j \geq 0$, where $a > 0$ is a constant and $Y(j),\ j \geq 1$, are RVs whose distributions satisfy certain mild conditions, we obtain the recursive tree considered in [10] (in fact, the model of [10] assumed that at each step a fixed number, $k \geq 1$, of children are attached to one of the existing tree vertices, and also that the $Y(j)$ are vector valued).

Here we should also mention other, related models where the weights of the vertices can *change* at each step. Thus, if, after the completion of the kth step of the tree construction, the weight of the vertex $v(j),\ j \leq k$, is $w(j) \equiv w(j,k) = 1 + \beta N_k(j)$, where $\beta \geq 0$ is constant and $Y(j) \equiv 1$, we obtain the linear recursive tree studied in [16] and [5] (see the bibliographies therein for further references). The case where $w(j) \equiv w(j,k) = 1 + N_k(j)$ was considered in [3]; the power-tail limiting behaviour of the degree distribution for this model that had been guessed in [3] was established in [8].

If $w(j) = a_1 \cdots a_j,\ j \geq 1$ (where $a_1, \dots, a_j$ are independent, identically distributed (i.i.d.) RVs), and $Y(j) \equiv 1$, we obtain a version of a weighted recursive tree. It is this last model and its generalizations that will be of the most interest to us in the present paper.

From now on we assume that the weight, $w(j)$, of the vertex $v(j)$ is, generally speaking, random and, once assigned, remains unchanged forever.

Section 2 of the paper is devoted to studying the asymptotic behaviour of the distribution of D_n. Theorems 1 and 2 present general convergence results for the conditional distribution of D_n in the cases where the random weights $w(j)$ tend to 'prescribe' new attachments to vertices close to the root of the tree and, respectively, where the new attachments are 'more dispersed' across the tree. Corollary 2 covers the special case where $w(j) \equiv 1$. The results of Section 2 also show that, for any $\alpha \in (0, 1]$, we can construct a random recursive tree such that D_n behaves like n^α as $n \to \infty$. Theorem 3 implies that, in the case of the 'product-form' weights $w(j) = a_1 \cdots a_j,\ j \geq 1$, with a_j being nondegenerate i.i.d. RVs satisfying the moment conditions $\mathrm{E} \ln a_j = 0$ and $\mathrm{E}\,|\ln a_j|^{2+\delta} < \infty$ for some $\delta > 0$, the limiting distribution of $D_n/\sqrt{n}$ coincides with the law of the maximum of the Brownian motion process on a finite time interval.

Section 3 deals with the expectations of the numbers of outgoing degrees in the case of product-form weights, under the assumption that the random walk generated by the i.i.d. sequence $\{\ln a_j\}$ satisfies Spitzer's condition. Theorem 4 gives the asymptotic behaviour of the unconditional expectations $\mathrm{E}\, N_n(k)$ as $n \to \infty$ when either $k = j$ or $k = n - j$ for a fixed value $j \geq 0$ (in both cases it is given by a regularly varying function of n). Theorem 5 complements it by covering the case where $\min\{j, n-j\} \to \infty$. Here the answer has the form of a product of regularly varying functions of j and $n - j$, respectively; in particular, in the case when $\ln a_j$ has zero mean and finite variance, we obtain $\mathrm{E}\, N_n(j) \sim 2\pi^{-1}(n-j)^{1/2} j^{-1/2}$.

Theorem 6 describes, for a range of j-values, the asymptotic behaviour of the distribution of the conditional expectation $\mathrm{E}_w N_n(j)$ given the sequence of weights $w(1), w(2), \ldots$.

2. The distribution of D_n

2.1. The basic properties of D_n

Let

$$W_n := \sum_{j=0}^{n} w(j), \quad p_n(j) := \frac{w(j)}{W_n}, \qquad j = 0, 1, \ldots, n.$$

Set $f_0(t) := 1$ and $f_j(t) := \mathrm{E}\,\mathrm{e}^{\mathrm{i}tY(j)}$, $j \geq 1$, and let

$$\varphi_0(t) := 1, \qquad \varphi_n(t) := \mathrm{E}_w\,\mathrm{e}^{\mathrm{i}tD_n} := \mathrm{E}[\mathrm{e}^{\mathrm{i}tD_n} \mid w(1), \ldots, w(n-1)], \quad n \geq 1,$$
$$\Psi_n(t) := \mathrm{E}\,\varphi_n(t) = \mathrm{E}\,\mathrm{e}^{\mathrm{i}tD_n}, \qquad n \geq 1$$

(here and in what follows, E_w and P_w respectively denote the conditional expectation and probability given the sequence of weights $\{w(j)\}$).

It is easy to see that

$$\begin{aligned}
\varphi_{n+1}(t) &= \sum_{j=0}^{n} p_n(j)\varphi_j(t) f_{n+1}(t) \\
&= \frac{W_{n-1}}{W_n} \sum_{j=0}^{n-1} p_{n-1}(j)\varphi_j(t) f_{n+1}(t) + p_n(n)\varphi_n(t) f_{n+1}(t) \\
&= (1 - p_n(n)) \frac{f_{n+1}(t)}{f_n(t)} \varphi_n(t) + p_n(n)\varphi_n(t) f_{n+1}(t) \\
&= [1 + (f_n(t) - 1)p_n(n)] \frac{f_{n+1}(t)}{f_n(t)} \varphi_n(t) \\
&= \cdots \\
&= f_{n+1}(t) \prod_{j=1}^{n} [1 + (f_j(t) - 1)p_j(j)].
\end{aligned} \tag{2}$$

Remark 1. Observe that (2) in fact means that, given the environment, the RV D_{n+1} admits a representation of the form of a sum of independent RVs, as follows:

$$D_{n+1} \stackrel{\mathrm{D}}{=} I_1 Y(1) + \cdots + I_n Y(n) + Y(n+1). \tag{3}$$

Here $\{I_j\}$ is a sequence of random indicators that are independent of each other and also of $\{Y(j)\}$, with $\mathrm{P}(I_j = 1) = p_j(j)$, $j \geq 1$. In the special case where $Y(j) \equiv w(j) \equiv 1$, this representation is equivalent to the correspondence between the quantity D_n and the numbers of records in an i.i.d. sequence that was used in [11] (see also Section 3.6 of [17], for a discussion of a somewhat more general situation where the representation (3) with $Y(j) \equiv 1$ holds). Note, however, that in [11] a probabilistic argument that works only in that special case was used to derive representation (3), which is actually the main tool for studying D_n, whereas our approach leads directly to (3) and is much more general.

https://doi.org/10.1017/S0001867800001440 Published online by Cambridge University Press

From the recursive relation (2), we can derive a number of interesting results on the limiting behaviour of D_n. Note that (2) was first derived in the case where $w(j) = a^j,\ j \geq 0$, $Y(j) \in \mathbb{R}^d$, in [10] (one can easily see that this recursive formula and the statements of Theorems 1 and 2 below remain true in the multivariate case as well). In particular, relation (2) immediately implies the following assertion, describing the limiting behaviour of the conditional distribution of D_n (given the weights) when the weight sequence $\{w(j)\}$ 'suggests' new children to attach not too far from the tree's root.

Theorem 1. *If*

$$\sum_{j=1}^{\infty} p_j(j) < \infty \quad \textit{almost surely (a.s.)}$$

and the distribution of $Y(n)$ *has a weak limit as* $n \to \infty$, *i.e. for a characteristic function* $f(t)$,

$$\lim_{n\to\infty} f_n(t) = f(t),$$

then there exists the limit

$$\lim_{n\to\infty} \varphi_n(t) = \varphi_\infty(t) := f(t) \prod_{j=1}^{\infty} [1 + (f_j(t) - 1) p_j(j)] \quad \textit{a.s.}$$

This result, in turn, implies that $D_n \xrightarrow{\mathrm{D}} D_\infty$ as $n \to \infty$, where D_∞ is a proper RV with the characteristic function $\mathrm{E}\,\varphi_\infty(t)$.

The next assertion applies in situations where the attachment preferences are spread 'more uniformly' across the tree.

Theorem 2. *Let the sequence of RVs* $\{Y(j)\}_{j\geq 1}$ *be uniformly integrable, and let there exist both a sequence* $\{h_n\}$ *such that* $h_n \to \infty$ *as* $n \to \infty$ *and an RV* ζ *such that the following convergence in distribution occurs as* $n \to \infty$:

$$\zeta_n := \frac{1}{h_n} \sum_{j=1}^{n} p_j(j)\, \mathrm{E}\, Y(j) \xrightarrow{\mathrm{D}} \zeta. \tag{4}$$

Then, for any t,

$$\varphi_n\left(\frac{t}{h_n}\right) \xrightarrow{\mathrm{D}} \mathrm{e}^{\mathrm{i}t\zeta}.$$

Remark 2. We can easily see that if, instead of (4), we have $\zeta_n \to \zeta$ a.s. for some RV ζ, then

$$\lim_{n\to\infty} \varphi_n\left(\frac{t}{h_n}\right) = \mathrm{e}^{\mathrm{i}t\zeta} \quad \text{a.s.},$$

uniformly in t from any compact set.

Proof of Theorem 2. It is not difficult to see that, due to the uniform integrability condition, as $n \to \infty$,

$$f_j\left(\frac{t}{h_n}\right) - 1 = \mathrm{E}\exp\left\{\frac{\mathrm{i}t Y(j)}{h_n}\right\} - 1 = \frac{1}{h_n}(\mathrm{i}t\,\mathrm{E}\,Y(j) + o(1))$$

https://doi.org/10.1017/S0001867800001440 Published online by Cambridge University Press

uniformly in $j \geq 1$ and in t from any compact set. Hence, as $p_j(j) \leq 1$, by (2) we have

$$\varphi_n\left(\frac{t}{h_n}\right) = f_n\left(\frac{t}{h_n}\right)\prod_{j=1}^{n-1}\left[1 + \left(f_j\left(\frac{t}{h_n}\right) - 1\right)p_j(j)\right]$$
$$= (1 + \varepsilon_n(t))\exp\left\{\frac{\mathrm{i}t}{h_n}\sum_{j=1}^{n} p_j(j)\,\mathrm{E}\,Y(j)\right\},$$

where $\varepsilon_n(t) = o_{\mathrm{P}}(1)$ as $n \to \infty$. This clearly implies the assertion of the theorem.

Corollary 1. *Under the conditions of Theorem* 2*,*

$$\lim_{n\to\infty} \Psi_n\left(\frac{t}{h_n}\right) = \mathrm{E}\,\mathrm{e}^{\mathrm{i}t\zeta},$$

so $D_n/h_n \xrightarrow{\mathrm{D}} \zeta$ *as* $n \to \infty$*.*

From Theorem 2 we can also easily deduce the following result obtained in [10] (note that, in the special case where $Y(j) \equiv 1$, the result was originally established in [11]).

Corollary 2. *If* $w(j) \equiv 1,\ j = 0, 1, 2, \ldots,$ *the family of RVs* $\{Y(j)\}_{j\geq 1}$ *is uniformly integrable, and*

$$\frac{1}{n}\sum_{j=1}^{n} \mathrm{E}\,Y(j) \to \mu \in \mathbb{R} \quad \text{as } n \to \infty,$$

then $D_n/\ln n \xrightarrow{\mathrm{P}} \mu$*.*

Proof. In this case clearly $p_j(j) = 1/(j+1)$ and, as was shown in Lemma 1(i) of [10], under the above conditions,

$$\zeta_n = \frac{1}{\ln n}\sum_{j=1}^{n}\frac{1}{j+1}\,\mathrm{E}\,Y(j) \to \mu.$$

The assertion of the corollary thus follows from Theorem 2.

We also obtain the same asymptotics for D_n when the weights are random but remain the same 'on average'.

Corollary 3. *If* $Y(j) \equiv 1,\ j \geq 1,$ *and the sequence of random weights* $\{w(j)\}$ *satisfies the strong law of large numbers, i.e. as* $n \to \infty$*,*

$$\frac{1}{n}\sum_{j=1}^{n} w(j) \to a > 0 \quad \text{a.s.},$$

then $D_n/\ln n \xrightarrow{\mathrm{P}} 1$*.*

Proof. It again suffices to apply (a slightly modified version of) Lemma 1(i) of [10] (this time to the sequences $\{y_n := anW_n^{-1}\}$ and $\{x_n := w(n)/a\}$) and use our Theorem 2.

https://doi.org/10.1017/S0001867800001440 Published online by Cambridge University Press

Remark 3. To obtain a faster-than-logarithmic growth rate for D_n (assuming that $Y(j) \equiv 1$), the weights $w(j)$ should grow faster than any power function. Indeed, if, say,

$$w(j) = j^{\alpha} l(j), \qquad \alpha \in \mathbb{R},$$

is a regularly varying function, then clearly

$$\sum_{j=1}^{\infty} p_j(j) < \infty \quad \text{if } \alpha < -1$$

(meaning that in this case Theorem 1 is applicable), and, by Karamata's theorem, $W_n \sim (\alpha+1)^{-1} n^{\alpha+1} l(n)$ if $\alpha > -1$, meaning that $p_j(j) \sim 1/(\alpha+1)j$ and, hence,

$$\sum_{j=1}^{n} p_j(j) \sim \frac{\ln n}{\alpha+1} \quad \text{if } \alpha > -1.$$

Thus, in the latter case, $D_n/\ln n \xrightarrow{\mathrm{P}} 1/(\alpha+1)$.

On the other hand, for, say,

$$w(j) = \alpha j^{\alpha-1} \mathrm{e}^{j^{\alpha}}, \qquad \alpha \in (0, 1],$$

we obtain $W_n \sim \mathrm{e}^{n^{\alpha}}$ and, hence,

$$\sum_{j=1}^{n} p_j(j) \sim n^{\alpha}.$$

This example thus shows that, for any $\alpha \in (0, 1]$, we can construct a random recursive tree with $D_n/n^{\alpha} \xrightarrow{\mathrm{P}} 1$ as $n \to \infty$.

2.2. The case of product-form random weights

In this subsection we will construct and study recursive trees with random vertex weights of the form $w(j) = a_1 \cdots a_j$, $j \geq 1$ (where the a_j are i.i.d. RVs), and unit edge lengths. As will be clearly seen from the proofs below, the main results will also hold in the case of random i.i.d. edge lengths with finite mean (Remark 4). Thus, restricting our attention to the case of unit edge lengths leads to no loss of generality, but makes the exposition more compact and transparent.

Denote by $\mathcal{T}_n$, $n = 0, 1, 2, \ldots$, the set of all rooted recursive trees having n nonrooted vertices and unit edge lengths (that is, $\mathcal{T}_n$ consists of the rooted trees whose roots are labelled 0 and whose nonrooted vertices are labelled by numbers $1, 2, \ldots, n$ in such a way that, for any nonrooted vertex labelled, say, j, the shortest path leading from it to the root traverses only the vertices labelled by numbers less than j). For a tree $t_n \in \mathcal{T}_n$, let $t_n(j) \in \mathcal{T}_{n+1}$ be the recursive tree which is obtained from t_n by adding a vertex labelled $n+1$ as a child of the vertex with the label $j \in \{0, 1, \ldots, n\}$.

We can describe the construction of our random recursive tree as follows. First, we run a random walk

$$S_0 = 0, \qquad S_j = \theta_1 + \cdots + \theta_j, \quad j \geq 1,$$

where the $\theta_j \stackrel{\mathrm{D}}{=} \theta$, $j = 1, 2, \ldots, n$, are i.i.d. RVs. Second, given S_j, $j = 0, 1, \ldots, n$, we construct a (conditional) Markov chain $T_0, T_1, \ldots, T_n$, with $T_k \in \mathcal{T}_k$, $k = 0, 1, \ldots, n$, by assigning the weight $w(j) := \mathrm{e}^{-S_j}$ to the vertex labelled $j \geq 0$ (so $w(j) = a_1 \cdots a_j$, $j \geq 1$, in the notation of Section 1, with the $a_j := \mathrm{e}^{-\theta_j}$ being i.i.d. RVs), so that now we have, for $r = 0, 1, \dots, n$,

$$W_r = \sum_{q=0}^{r} w(q) = \sum_{q=0}^{r} \mathrm{e}^{-S_q}, \tag{5}$$

$$p_r(j) = \frac{\mathrm{e}^{-S_j}}{W_r} = \frac{\mathrm{e}^{-S_j}}{\sum_{q=0}^{r} \mathrm{e}^{-S_q}}, \qquad j = 0, 1, \dots, r, \tag{6}$$

and then letting, for any $t_r \in \mathcal{T}_r$,

$$\mathrm{P}_w(T_{r+1} = t_r(j) \mid T_r = t_r) \equiv \mathrm{P}(T_{r+1} = t_r(j) \mid T_r = t_r;\ w(0), w(1), \dots, w(r)) := p_r(j), \quad j = 0, 1, \dots, r.$$

The main result of this subsection is the following theorem.

Theorem 3. *If*

$$\mathrm{E}\,\theta = 0, \qquad \sigma^2 := \mathrm{E}\,\theta^2 > 0, \qquad \mathrm{E}\,|\theta|^{2+\delta} < \infty \quad \textit{for some } \delta > 0, \tag{7}$$

then, as $n \to \infty$,

$$\zeta_n := \frac{1}{\sqrt{n}} \sum_{j=1}^{n} p_j(j) \xrightarrow{\mathrm{D}} \sigma_m \max_{0 \le u \le 1} B(u), \qquad \sigma_m := \sigma \int_0^\infty \frac{m(\mathrm{d}y)}{y} < \infty,$$

where $\{B(u)\}_{u \ge 0}$ *is the standard Brownian motion process and the measure* m *is specified in the proof (see (11)).*

Together with Corollary 1, the above assertion immediately yields the following result.

Corollary 4. *Under the conditions of Theorem 3,*

$$\frac{D_n}{\sqrt{n}} \xrightarrow{\mathrm{D}} \sigma_m \max_{0 \le u \le 1} B(u) \quad \textit{as } n \to \infty.$$

In other words, for any $x > 0$,

$$\mathrm{P}(D_n > \sigma_m \sqrt{n} x) \to 2(1 - \Phi(x)),$$

where Φ *is the standard normal distribution function.*

Remark 4. It is obvious that the assertion of Corollary 4 remains true in the case of i.i.d. random edge lengths, $Y(j) \ge 0$, with finite, positive mean, the only difference being that σ_m should be replaced by $\sigma_m\,\mathrm{E}\,Y(1)$ in its formulation.

Proof of Theorem 3. Let

$$L_n := \min_{0 \le k \le n} S_k.$$

Using the proof of Theorem 4.1 of [2], we will show that

$$\frac{1}{|L_n|} \sum_{j=1}^{n} p_j(j) \to \int_0^\infty \frac{m(\mathrm{d}y)}{y} < \infty \quad \text{a.s.} \tag{8}$$

https://doi.org/10.1017/S0001867800001440 Published online by Cambridge University Press

Since, by the invariance principle,

$$\frac{|L_n|}{\sqrt{n}} \xrightarrow{\text{D}} \sigma \max_{0\le u\le 1} B(u) \quad \text{as } n\to\infty, \tag{9}$$

the assertion of the theorem will then immediately follow.

First, denote by

$$\gamma_0 := 0, \qquad \gamma_{j+1} := \min\{n > \gamma_j : S_n < S_{\gamma_j}\}, \quad j \ge 0,$$

the strict descending ladder epochs of the random walk $\{S_n\}_{n\ge 0}$. All the RVs introduced are finite a.s., as, in view of (7), $\{S_n\}_{n\ge 0}$ is recurrent.

Let $\{X_n\}_{n\ge 0}$ be a Markov chain defined, for $n = 1, 2, \dots$, by

$$X_n := \mathrm{e}^{\theta_n} X_{n-1} + 1.$$

When $X_0^x = x > 0$ is a fixed value, we will use the notation $\{X_n^x\}_{n\ge 0}$. Clearly,

$$X_n^x = x\mathrm{e}^{S_n} + \sum_{q=1}^{n} \mathrm{e}^{S_n - S_q} = \mathrm{e}^{S_n}(x - 1 + W_n). \tag{10}$$

Set $\gamma := \gamma_1$. Under our assumptions in (7), the expectation $\mathrm{E}\, S_\gamma < 0$ is finite (see, e.g. Corollary 10, Section 17, of [9]), and the Markov chain $\{X_{\gamma_n}\}_{n\ge 1}$ with transition kernel

$$M_\gamma(x, \cdot) := \mathrm{P}(X_\gamma^x \in \cdot), \qquad x > 0,$$

has a unique invariant probability measure, m_γ, satisfying (see, e.g. Lemma 5.49 of [13] and page 481 of [2]):

$$m_\gamma(A) = \int_0^\infty m_\gamma(\mathrm{d}x) M_\gamma(x, A).$$

Moreover, the measure m defined by

$$m(f) := \frac{1}{\mathrm{E}[-S_\gamma]} \int_0^\infty \mathrm{E}\Bigg[\sum_{k=0}^{\gamma-1} f(X_k^x)\Bigg] m_\gamma(\mathrm{d}x) \tag{11}$$

is an invariant measure for the Markov chain $\{X_n\}_{n\ge 0}$ (see [2]).

Now note that, by virtue of (6) and (10),

$$\zeta_n = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} p_j(j) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \frac{1}{X_j^1}.$$

Let P_{δ_y} be the distribution of the two-dimensional random walk

$$Z_n := (X_n, \mathrm{e}^{S_n}), \qquad n \ge 0$$

(on the group of transformations $x \mapsto ax + b$ of the real line with the composition law $(b_1, a_1)(b_2, a_2) = (b_1 + a_1 b_2, a_1 a_2)$), when $X_0 = y$. It was shown in the proof of Theorem 4.1 of [2] that if $f \in L^1(m)$ then

$$\lim_{n\to\infty} \frac{1}{|L_n|} \sum_{j=1}^{n} f(X_j) = \int_0^\infty f(y) m(\mathrm{d}y) \quad P_{m_\gamma}\text{-a.s.}, \tag{12}$$

where

$$P_{m_\gamma} := \int_0^\infty P_{\delta_y} m_\gamma(\mathrm{d}y)$$

is the law of the two-dimensional random walk $\{Z_n\}_{n\geq 1}$ when the distribution of X_0 is m_γ.

For $N = 1, 2, \ldots$ and $x > 0$, let

$$g_N(x) := \frac{1}{x} \mathbf{1}_{\{N^{-1} \leq x \leq N\}} \leq \frac{1}{x} =: g(x).$$

Clearly, for all $x > 0$,

$$g_N(x) \nearrow g(x) \quad \text{as } N \to \infty, \tag{13}$$

and $g_N(x) \in L^1(m)$ for each $N = 1, 2, \ldots$. Therefore, by (12),

$$\lim_{n\to\infty} \frac{1}{|L_n|} \sum_{j=1}^n g_N(X_j) = \int_0^\infty g_N(y) m(\mathrm{d}y) \quad P_{m_\gamma}\text{-a.s.} \tag{14}$$

On the other hand, for each $N \geq 1$ and any $x > 0$,

$$\begin{aligned}
\frac{1}{|L_n|} \sum_{j=1}^n g_N(X_j^x) &\leq \frac{1}{|L_n|} \sum_{j=1}^n g(X_j^x) = \frac{1}{|L_n|} \sum_{j=1}^n \frac{\mathrm{e}^{-S_j}}{x - 1 + W_j} \\
&\leq \frac{1}{|L_n|} \sum_{j=1}^n \int_{x-1+W_{j-1}}^{x-1+W_j} \frac{\mathrm{d}y}{y} = \frac{1}{|L_n|} \int_{x-1+W_0}^{x-1+W_n} \frac{\mathrm{d}y}{y} \\
&= \frac{1}{|L_n|} [\ln(x - 1 + W_n) - \ln x] \leq \frac{1}{|L_n|} [\ln(x + n\mathrm{e}^{|L_n|}) - \ln x] \\
&\leq \frac{1}{|L_n|} \left[\ln n\mathrm{e}^{|L_n|} + \frac{x}{n\mathrm{e}^{|L_n|}} - \ln x \right] = 1 + \frac{1}{|L_n|} [\ln n + O(1)] \\
&\xrightarrow{\mathrm{P}} 1 \quad \text{as } n \to \infty,
\end{aligned} \tag{15}$$

by the invariance principle (see, e.g. [6, Section 2.10]).

Combining (14) with (15) shows that

$$\sup_{N\geq 1} \int_0^\infty g_N(y) m(\mathrm{d}y) \leq 1,$$

which, together with (13), yields

$$\int_0^\infty g(y) m(\mathrm{d}y) \leq 1.$$

Therefore, by (12),

$$\lim_{n\to\infty} \frac{1}{|L_n|} \sum_{j=1}^n g(X_j) = \int_0^\infty g(y) m(\mathrm{d}y) = \int_0^\infty \frac{\mathrm{d}m(y)}{y} \quad P_{m_\gamma}\text{-a.s.}$$

To see that this convergence holds for all starting points $x > 0$, it suffices to observe that $g(z)$ is monotone in $z > 0$ and that $X_j^{x_1} > X_j^{x_2}$, $j \geq 1$, for x_1 and x_2 with $x_1 > x_2 > 0$.

This, in view of (8) and (9), completes the proof of Theorem 3.

3. The expectations of the outdegrees of vertices

Let $N_n(j)$ be the outdegree of the vertex $v(j)$, $j = 0, 1, \dots, n$, in $\mathcal{T}_n$, i.e. the number of edges coming out of $v(j)$ in a tree having n nonrooted vertices. Clearly, the RV $N_n(j)$ admits the representation (1) and, therefore,

$$\begin{aligned}\mathrm{E}_w N_n(j) = \mathrm{E}[N_n(j) \mid w(1), \dots, w(n-1)] &= \sum_{k=j+1}^{n} \mathrm{E}_w \mathbf{1}_{\{v(k^*)=v(j)\}} \\ &= \sum_{k=j+1}^{n} p_{k-1}(j) = \mathrm{e}^{-S_j} \sum_{k=j}^{n-1} W_k^{-1} \qquad (16)\end{aligned}$$

and

$$\mathrm{E}\, N_n(j) = \sum_{k=j}^{n-1} \mathrm{E}\, \mathrm{e}^{-S_j} W_k^{-1}. \qquad (17)$$

Our aim in this section is to investigate the asymptotic behaviour (as $n \to \infty$) of the expectations $\mathrm{E}\, N_n(j)$ and that of the distributions of the RVs $\mathrm{E}_w N_n(j)$ in different ranges of the parameter j.

3.1. The asymptotic behaviour of $\mathrm{E}\, N_n(j)$

In this section we impose weaker restrictions (compared to conditions (7), used in Section 2) on the random walk $S_n = \theta_1 + \cdots + \theta_n$, $n \geq 1$, where $\theta_j \stackrel{\mathrm{D}}{=} \theta$ are i.i.d. RVs. Namely, we assume only that Spitzer's condition holds:

There exists a $\rho \in (0, 1)$ *such that*

$$\frac{1}{n} \sum_{k=1}^{n} \mathrm{P}(S_k > 0) \to \rho \quad \text{as } n \to \infty. \qquad (18)$$

It is known [12] that this condition is equivalent to Doney's condition,

$$\mathrm{P}(S_n > 0) \to \rho \quad \text{as } n \to \infty$$

(for a further discussion of condition (18), see, e.g. Section 8.9 of [7]).

We will need a number of auxiliary results concerning the random walk $\{S_n\}_{n\geq 0}$. Let

$$\Gamma_0 := 0, \qquad \Gamma_{j+1} := \inf\{n > \Gamma_j : S_n > S_{\Gamma_j}\}, \quad j \geq 0,$$

be the strict ascending ladder epochs of the random walk $\{S_n\}_{n\geq 0}$. Recall that the γ_i, $0 = \gamma_0 < \gamma_1 < \gamma_2 < \cdots$, denote the strict descending ladder epochs of the walk. Introduce the two renewal functions

$$U(x) := 1 + \sum_{j=1}^{\infty} \mathrm{P}(S_{\Gamma_j} < x), \quad x > 0, \qquad U(0) = 1, \qquad U(x) = 0, \quad x < 0,$$

$$V(x) := \sum_{j=0}^{\infty} \mathrm{P}(S_{\gamma_j} \geq -x), \qquad x > 0, \qquad V(0) = 1, \qquad V(x) = 0, \quad x < 0,$$

https://doi.org/10.1017/S0001867800001440 Published online by Cambridge University Press

and set

$$M_n := \max_{0\le k\le n} S_k, \qquad \tilde{M}_n := \max_{1\le k\le n} S_k.$$

It is known (see, e.g. Lemma 1 of [14] and Lemma 1 of [19]) that, under condition (18),

$$\mathrm{E}\, U(-\theta)\, \mathbf{1}_{\{-\theta>0\}} = \mathrm{e}^{-\phi}, \qquad \mathrm{E}\, U(x-\theta)\, \mathbf{1}_{\{x-\theta>0\}} = U(x), \quad x>0, \tag{19}$$

where

$$\phi := \sum_{j=1}^{\infty} \frac{1}{j}\, \mathrm{P}(S_j = 0) < \infty$$

and

$$\mathrm{E}\, V(x+\theta) = V(x), \qquad x \ge 0. \tag{20}$$

By means of $V(x)$ and $U(x)$ we can specify two sequences of probability measures, $\{\mathrm{P}_n^-\}_{n\ge1}$ and $\{\mathrm{P}_n^+\}_{n\ge1}$, on the σ-algebras $\{\Sigma_n := \sigma(S_1,\dots,S_n)\}_{n\ge1}$, with the corresponding expectations $\{\mathrm{E}_n^-\}_{n\ge1}$ and $\{\mathrm{E}_n^+\}_{n\ge1}$, by setting

$$\mathrm{E}_n^-[\psi_n(S_1,\dots,S_n)] := \mathrm{e}^{\phi}\, \mathrm{E}[\psi_n(S_1,\dots,S_n) U(-S_n)\, \mathbf{1}_{\{\tilde{M}_n<0\}}], \tag{21}$$

$$\mathrm{E}_n^+[\psi_n(S_1,\dots,S_n)] := \mathrm{E}[\psi_n(S_1,\dots,S_n) V(S_n)\, \mathbf{1}_{\{L_n\ge0\}}] \tag{22}$$

for each bounded, measurable function $\psi_n(x_1,\dots,x_n)$. It is easy to verify that (19) and (20) imply that each of the sequences $\{\mathrm{P}_n^\pm\}_{n\ge1}$ is consistent, and that by Kolmogorov's extension theorem there therefore exist measures P^- and P^+ on the σ-algebra $\sigma(S_1, S_2, \dots)$ such that their restrictions, $\mathrm{P}^\pm|_{\Sigma_n}$, to Σ_n coincide with $\mathrm{P}_n^\pm$, $n = 1, 2, \dots$.

It is known (see Lemma 2.7 of [1]) that, under condition (18),

$$\eta_1 := \sum_{k=1}^{\infty} \mathrm{e}^{S_k} < \infty \quad \mathrm{P}^-\text{-a.s.}, \qquad \eta_2 := \sum_{k=0}^{\infty} \mathrm{e}^{-S_k} < \infty \quad \mathrm{P}^+\text{-a.s.} \tag{23}$$

Finally, it is not difficult to deduce from Lemma 3 of [19] that, if we let

$$H_n^-(x) := \mathrm{P}\biggl(\sum_{k=1}^{n} \mathrm{e}^{S_k} \le x \biggm| \tilde{M}_n < 0\biggr), \qquad H_n^+(x) := \mathrm{P}\biggl(\sum_{k=0}^{n} \mathrm{e}^{-S_k} \le x \biggm| L_n \ge 0\biggr)$$

and

$$H^-(x) := \mathrm{P}^-(\eta_1 < x), \qquad H^+(x) := \mathrm{P}^+(\eta_2 < x),$$

then, under condition (18),

$$H_n^\pm(x) \Rightarrow H^\pm(x) \quad \text{as } n\to\infty, \tag{24}$$

where the symbol '$\Rightarrow$' denotes convergence at all continuity points of the limiting function.

In what follows we will often use the following result (see, e.g. Lemma 2.1 of [1], Theorem 8.9.12 of [7], and Lemma 2 of [19]). Let

$$\lambda_n(x) := \mathrm{P}(L_n \ge -x), \quad \tilde{\mu}_n(x) := \mathrm{P}(\tilde{M}_n < x), \qquad x \ge 0.$$

https://doi.org/10.1017/S0001867800001440 Published online by Cambridge University Press

Lemma 1. *Under Spitzer's condition, (18), there exist functions $l_1(n)$ and $l_2(n)$, related by $l_1(n)l_2(n) \sim \pi^{-1} \sin \pi\rho$, $n \to \infty$, that are slowly varying at infinity and such that*

$$\mathrm{P}(L_n \geq 0) \sim n^{\rho-1} l_1(n) \quad \text{and} \quad \mathrm{P}(\tilde{M}_n < 0) \sim n^{-\rho} l_2(n), \quad \text{as } n \to \infty. \tag{25}$$

Moreover, there exist absolute constants $C_1 > 0$ and $C_2 > 0$ such that, for all $n \geq 1$ and $x \geq 0$,

$$\lambda_n(x) \leq C_1 V(x)\,\mathrm{P}(L_n \geq 0), \qquad \tilde{\mu}_n(x) \leq C_2 U(x)\,\mathrm{P}(\tilde{M}_n < 0). \tag{26}$$

In (25) and in the rest of the paper, by $a_n \sim b_n$ we mean that $a_n/b_n \to 1$ as $n \to \infty$.

Let $\{S_n^-\}_{n\geq 0}$ and $\{S_n^+\}_{n\geq 0}$ be two independent copies of $\{S_n\}_{n\geq 0}$, and let

$$L_n^+ := \min_{0\leq r\leq n} S_r^+, \qquad \tilde{M}_n^- := \max_{1\leq l\leq n} S_l^-.$$

Introduce the probability distributions

$$\mathrm{P}_{-,+} := \mathrm{P}^- \times \mathrm{P}^+, \qquad \mathrm{P}_{\cdot,+} := \mathrm{P} \times \mathrm{P}^+, \qquad \mathrm{P}_{-,\cdot} := \mathrm{P}^- \times \mathrm{P}$$

on the sample space $\mathbb{R}^\infty \times \mathbb{R}^\infty$ of the pair $(\{S_n^-\}_{n\geq 0}, \{S_n^+\}_{n\geq 0})$, where P is the distribution of the original sequence $\{S_n\}_{n\geq 0}$ and the measures $\mathrm{P}^\pm$ are specified by (21) and (22), and let $\mathrm{E}_{-,+}$, $\mathrm{E}_{\cdot,+}$, and $\mathrm{E}_{-,\cdot}$ be the expectation operators under the respective measures.

We will call an array of RVs $\{G_{l,r} : l, r \in \mathbb{N}\}$ *adapted* if, for any pair of indices $l, r \in \mathbb{N}$, the RV $G_{l,r}$ is measurable with respect to the σ-algebra $\sigma(S_1^-, \dots, S_l^-) \otimes \sigma(S_1^+, \dots, S_r^+)$. The following result is contained in Lemma 3 of [19].

Lemma 2. *Let Spitzer's condition, (18), hold, and let $\{G_{l,r} : l, r \in \mathbb{N}\}$ be an adapted array of uniformly bounded RVs. If the limit*

$$\lim_{l,r\to\infty} G_{l,r} =: G \quad \mathrm{P}_{-,+}\text{-a.s.}$$

exists then

$$\lim_{l,r\to\infty} \mathrm{E}[G_{l,r} \mid \tilde{M}_l^- < 0,\ L_r^+ \geq 0] = \mathrm{E}_{-,+}\, G.$$

The next statement is a slight modification of Lemma 2.5 of [1] and can be proved using the same arguments used there.

Lemma 3. *Let Spitzer's condition, (18), hold, and let $\{G_{l,r} : l, r \in \mathbb{N}\}$ be an adapted array of uniformly bounded RVs. If the limit*

$$\lim_{r\to\infty} G_{l,r}\, \mathbf{1}_{\{\tilde{M}_l^- < 0\}} =: G_l^+\, \mathbf{1}_{\{\tilde{M}_l^- < 0\}} \quad \mathrm{P}_{\cdot,+}\text{-a.s.}$$

exists, then

$$\lim_{r\to\infty} \mathrm{E}[G_{l,r}\, \mathbf{1}_{\{\tilde{M}_l^- < 0\}} \mid L_r^+ \geq 0] = \mathrm{E}_{\cdot,+}\, G_l^+\, \mathbf{1}_{\{\tilde{M}_l^- < 0\}},$$

and if the limit

$$\lim_{l\to\infty} G_{l,r}\, \mathbf{1}_{\{L_r^+ \geq 0\}} =: G_r^-\, \mathbf{1}_{\{L_r^+ \geq 0\}} \quad \mathrm{P}_{-,\cdot}\text{-a.s.}$$

exists, then

$$\lim_{l\to\infty} \mathrm{E}[G_{l,r}\, \mathbf{1}_{\{L_r^+ \geq 0\}} \mid \tilde{M}_l^- < 0] = \mathrm{E}_{-,\cdot}\, G_r^-\, \mathbf{1}_{\{L_r^+ \geq 0\}}.$$

The following result was proved in Lemma 2.2 of [1]. Denote by

$$\tau(n) := \min\{k \geq 0 \colon S_k \leq S_l,\ l \in [0, n]\}$$

the left-most point at which the random walk $\{S_n\}$ attains its minimum value on the time interval $[0, n]$.

Lemma 4. *Let Spitzer's condition, (18), hold, and let $u(x) \geq 0$, $x \geq 0$, be a nonincreasing function such that $\int_0^\infty u(x)\,\mathrm{d}x < \infty$. Then, for every $\varepsilon > 0$, there exists an integer J such that, for all $n \geq J$,*

$$\sum_{p=J}^{n} \mathrm{E}[u(-S_p);\ \tau(p) = p]\,\mathrm{P}(L_{n-p} \geq 0) \leq \varepsilon\,\mathrm{P}(L_n \geq 0).$$

Introduce the RVs

$$G_r^+(j) := \frac{\mathrm{e}^{-S_{j-r}^+}\,\mathbf{1}_{\{j \geq r\}} + \mathrm{e}^{S_{r-j}^-}\,\mathbf{1}_{\{j<r\}}}{\sum_{p=1}^{r} \mathrm{e}^{S_p^-} + \eta_2^+},$$

$$G_r^-(j) := \frac{\mathrm{e}^{S_{j-r}^-}\,\mathbf{1}_{\{j>r\}} + \mathrm{e}^{-S_{r-j}^+}\,\mathbf{1}_{\{j \leq r\}}}{\eta_1^- + \sum_{p=0}^{r} \mathrm{e}^{-S_p^+}},$$

where η_1^- and η_2^+ are defined as in (23), but for the random walks $\{S_n^-\}_{n\geq 0}$ and $\{S_n^+\}_{n\geq 0}$, respectively. Note that $0 < G_r^{\pm} \leq 1$ and, in view of (23), that $G_r^+(j)$ and $G_r^-(j)$ are a.s. positive under the measures $\mathrm{P}_{\cdot,+}$ and $\mathrm{P}_{-,\cdot}$, respectively. Set

$$\tilde{L}_n^+ := \min_{1 \leq p \leq n} S_p^+ \tag{27}$$

and let

$$c_j := \sum_{l=0}^{\infty} \mathrm{E}_{\cdot,+}\,G_l^+(j)\,\mathbf{1}_{\{\tilde{M}_l^- < 0\}}, \qquad d_j := \sum_{q=1}^{j}\sum_{r=0}^{\infty} \mathrm{E}_{-,\cdot}\,G_r^-(q)\,\mathbf{1}_{\{\tilde{L}_r^+ > 0\}}. \tag{28}$$

We can easily verify that c_j and d_j are finite for any $j = 0, 1, \ldots$. Thus,

$$\begin{aligned} c_j \leq j + 1 + \sum_{l=j+1}^{\infty} \mathrm{E}_{\cdot,+}\,\mathrm{e}^{S_{l-j}^-}\,\mathbf{1}_{\{\tilde{M}_l^- < 0\}} &= j + 1 + \sum_{p=1}^{\infty} \mathrm{E}\,\mathrm{e}^{S_p}\,\mathbf{1}_{\{\tilde{M}_p < 0\}} \\ &= j + 1 + \sum_{p=1}^{\infty} \mathrm{E}\,\mathrm{e}^{S_p}\,\mathbf{1}_{\{S_1 < 0, \ldots, S_p < 0\}} \\ &< \infty \end{aligned}$$

(see D2, Section 17, of [18]).

Now we are ready to formulate and prove the following statement.

Theorem 4. *Let Spitzer's condition, (18), hold. Then, for any fixed $j \geq 0$,*

$$\lim_{n\to\infty} \frac{\mathrm{E}\,N_n(j)}{n\,\mathrm{P}(L_n \geq 0)} = \frac{c_j}{\rho}, \tag{29}$$

$$\lim_{n\to\infty} \frac{\mathrm{E}\,N_n(n-j)}{\mathrm{P}(\tilde{M}_n < 0)} = d_j. \tag{30}$$

https://doi.org/10.1017/S0001867800001440 Published online by Cambridge University Press

Remark 5. In view of (25), the relations (29) and (30) can be rewritten as follows:

$$\mathrm{E}\, N_n(j) \sim c_j \rho^{-1} n^{\rho} l_1(n) \quad \text{and} \quad \mathrm{E}\, N_n(n-j) \sim d_j n^{-\rho} l_2(n), \quad \text{as } n \to \infty.$$

Proof of Theorem 4. To prove the theorem, we have to evaluate the sum, (17), of expectations of the form

$$\mathrm{E}\, \mathrm{e}^{-S_j} W_k^{-1} = \sum_{l=0}^{k} \mathrm{E}\, \mathrm{e}^{-S_j} W_k^{-1} \mathbf{1}_{\{\tau(k)=l\}}. \tag{31}$$

The key idea both in this proof and in that of Theorem 5 is quite similar to that of the Laplace method: the main contribution to the expectation (31) comes from the event where j is close to $\tau(k)$ (for other values of $j \le k$, the quantity e^{-S_j} will typically be quite small compared with W_k).

First, we will show that, for each fixed $\varepsilon > 0$, there exists a $J \equiv J(\varepsilon)$ such that, for all $j \ge 0$ and all $k \ge J + j$,

$$\mathrm{E}\, \mathrm{e}^{-S_j} W_k^{-1} \mathbf{1}_{\{\tau(k) \ge J+j\}} \le \varepsilon\, \mathrm{P}(L_{k-j} \ge 0). \tag{32}$$

Indeed, as $W_k \ge \mathrm{e}^{-S_{\tau(k)}}$, we have

$$\begin{aligned}
\mathrm{E}\, \mathrm{e}^{-S_j} W_k^{-1} \mathbf{1}_{\{\tau(k) \ge J+j\}} &\le \mathrm{E}\, \mathrm{e}^{S_{\tau(k)}-S_j} \mathbf{1}_{\{\tau(k) \ge J+j\}} \\
&= \sum_{p=J}^{k-j} \mathrm{E}\, \mathrm{e}^{S_{p+j}-S_j} \mathbf{1}_{\{\tau(k)=p+j\}} \\
&\le \sum_{p=J}^{k-j} \mathrm{E}\, \mathrm{e}^{S_p} \mathbf{1}_{\{\tau(k-j)=p\}} \\
&= \sum_{p=J}^{k-j} \mathrm{E}[\mathrm{e}^{S_p} \mathbf{1}_{\{\tau(p)=p\}}]\, \mathrm{P}(L_{k-j-p} \ge 0),
\end{aligned}$$

and to obtain the result required it remains to apply Lemma 4 with $u(x) = \mathrm{e}^{-x}$.

The next step is to demonstrate that, for any fixed $j \ge 0$ and $l \ge 1$,

$$\lim_{k\to\infty} \frac{\mathrm{E}\, \mathrm{e}^{-S_j} W_k^{-1} \mathbf{1}_{\{\tau(k)=l\}}}{\mathrm{P}(L_k \ge 0)} = \mathrm{E}_{\cdot,+}\, G_l^{+}(j)\, \mathbf{1}_{\{\tilde{M}_l^{-} < 0\}}. \tag{33}$$

However, this is an easy consequence of Lemma 3. Indeed, assume first that $j \ge l$. Then, for the RVs $G_{l,r}(j)$ defined for $r \ge j - l$ by

$$G_{l,k-l}(j) := \frac{\mathrm{e}^{-S_{j-l}^{+}}}{\sum_{p=1}^{l} \mathrm{e}^{S_p^{-}} + \sum_{q=0}^{k-l} \mathrm{e}^{-S_q^{+}}} \le 1, \qquad k \ge j$$

(for $r < j - l$ we can set $G_{l,r}(j) \equiv 1$), we have

$$\begin{aligned}
\mathrm{E}\, \mathrm{e}^{-S_j} W_k^{-1} \mathbf{1}_{\{\tau(k)=l\}} &= \mathrm{E}\, \frac{\mathrm{e}^{S_{\tau(k)}-S_j}}{\sum_{p=0}^{k} \mathrm{e}^{S_{\tau(k)}-S_p}} \mathbf{1}_{\{\tau(k)=l\}} \\
&= \mathrm{E}\, G_{l,k-l}(j)\, \mathbf{1}_{\{\tilde{M}_l^{-} < 0,\, L_{k-l}^{+} \ge 0\}} \\
&= \mathrm{E}[G_{l,k-l}(j)\, \mathbf{1}_{\{\tilde{M}_l^{-} < 0\}} \mid L_{k-l}^{+} \ge 0]\, \mathrm{P}(L_{k-l} \ge 0)
\end{aligned}$$

(here the second relation follows from the duality principle: we use the 'time-reversed random walk' on $[0, l]$).

It is evident that, as $k \to \infty$,

$$G_{l,k-l}(j)\, \mathbf{1}_{\{\tilde{M}_l^- < 0\}} \to G_l^+(j)\, \mathbf{1}_{\{\tilde{M}_l^- < 0\}} \quad \mathrm{P}_{\cdot,+}\text{-a.s.}$$

and, therefore, by Lemma 3, that

$$\lim_{k\to\infty} \mathrm{E}[G_{l,k-l}(j)\, \mathbf{1}_{\{\tilde{M}_l^- < 0\}} \mid L_{k-l}^+ \geq 0] = \mathrm{E}_{\cdot,+}\, G_l^+(j)\, \mathbf{1}_{\{\tilde{M}_l^- < 0\}}. \tag{34}$$

On the other hand, in view of (25), for each fixed l we have

$$\lim_{k\to\infty} \frac{\mathrm{P}(L_{k-l} \geq 0)}{\mathrm{P}(L_k \geq 0)} = 1. \tag{35}$$

Combining this with (34) gives (33). The case where $j < l$ can be treated in a similar way.

Now everything is ready to complete the proof of the first part of the theorem. It follows from (31), (32), and (33) that, for each fixed $j \geq 0$,

$$\mathrm{E}\, \mathrm{e}^{-S_j} W_k^{-1} \sim c_j\, \mathrm{P}(L_k \geq 0) \quad \text{as } k \to \infty. \tag{36}$$

Therefore, for a fixed $\varepsilon > 0$ there exists a $K(\varepsilon) < \infty$ such that, for all $K \geq K(\varepsilon)$ and $n > K$,

$$\begin{aligned}(1-\varepsilon)c_j \sum_{k=K+1}^{n-1} \mathrm{P}(L_k \geq 0) \leq \mathrm{E}\, N_n(j) &= \sum_{k=j+1}^{K} \mathrm{E}\, \mathrm{e}^{-S_j} W_k^{-1} + \sum_{k=K+1}^{n-1} \mathrm{E}\, \mathrm{e}^{-S_j} W_k^{-1} \\ &\leq (K-j) + (1+\varepsilon)c_j \sum_{k=K+1}^{n-1} \mathrm{P}(L_k \geq 0). \end{aligned} \tag{37}$$

By (25) and Karamata's theorem (see, e.g. Section 1.6 of [7]),

$$\sum_{k=K+1}^{n-1} \mathrm{P}(L_k \geq 0) \sim \frac{n}{\rho} \mathrm{P}(L_n \geq 0) \quad \text{as } n \to \infty.$$

This together with (37) completes the proof of (29).

Now we will prove (30). Let $\{S_n^*\}_{n\geq 0} \stackrel{\mathrm{D}}{=} \{-S_n\}_{n\geq 0}$ be the 'reflected' random walk. By the duality principle, for each fixed $q \leq j$,

$$\begin{aligned}\mathrm{E}\, \mathrm{e}^{-S_{n-j}} W_{n-q}^{-1} &= \mathrm{E}\, \frac{\mathrm{e}^{-S_{n-j}}}{\sum_{p=0}^{n-q} \mathrm{e}^{-S_{n-q-p}}} = \mathrm{E}\, \frac{\mathrm{e}^{S_{n-q}-S_{n-j}}}{\sum_{p=0}^{n-q} \mathrm{e}^{S_{n-q}-S_{n-q-p}}} \\ &= \mathrm{E}\, \frac{\mathrm{e}^{-S_{j-q}^*}}{\sum_{p=0}^{n-q} \mathrm{e}^{-S_p^*}} = \mathrm{E}\, \mathrm{e}^{-S_{j-q}^*} (W_{n-q}^*)^{-1} \end{aligned} \tag{38}$$

(with an obvious definition for W_{n-q}^*).

Next we set

$$L_n^* := \min_{0\leq k\leq n} S_k^*, \qquad \tilde{M}_n^* := \max_{1\leq k\leq n} S_k^*$$

https://doi.org/10.1017/S0001867800001440 Published online by Cambridge University Press

and observe that, as $n \to \infty$,

$$\mathrm{P}(L_n^* \geq 0) = \mathrm{P}(M_n \leq 0) \sim \mathrm{e}^{\phi}\, \mathrm{P}(\tilde{M}_n < 0). \tag{39}$$

Indeed, by setting

$$\chi := \inf\{k \geq 1\colon S_k \geq 0\}, \qquad \tilde{\chi} := \inf\{k \geq 1\colon S_k > 0\},$$

we find from the factorization identities that, for $|z| < 1$,

$$1 - \mathrm{E}\, z^{\tilde{\chi}} = \exp\bigg\{\sum_{n=0}^{\infty} \frac{z^n}{n}\, \mathrm{P}(S_n > 0)\bigg\}, \qquad 1 - \mathrm{E}\, z^{\chi} = \exp\bigg\{\sum_{n=0}^{\infty} \frac{z^n}{n}\, \mathrm{P}(S_n \geq 0)\bigg\}$$

(see, e.g. Corollary 4, Section 16, of [9]). Dividing both sides of these identities by $1 - z = \mathrm{e}^{\ln(1-z)}$ yields

$$\begin{aligned}\sum_{n=0}^{\infty} z^n\, \mathrm{P}(M_n \leq 0) = \sum_{n=0}^{\infty} z^n\, \mathrm{P}(\tilde{\chi} > n) &= \frac{1 - \mathrm{E}\, z^{\tilde{\chi}}}{1 - z} \\ &= \exp\bigg\{-\sum_{n=1}^{\infty} \frac{z^n}{n}\, \mathrm{P}(S_n > 0) + \sum_{n=1}^{\infty} \frac{z^n}{n}\bigg\} \\ &= \exp\bigg\{\sum_{n=1}^{\infty} \frac{z^n}{n}\, \mathrm{P}(S_n \leq 0)\bigg\}\end{aligned}$$

and, similarly,

$$\sum_{n=0}^{\infty} z^n\, \mathrm{P}(\tilde{M}_n < 0) = \sum_{n=0}^{\infty} z^n\, \mathrm{P}(\chi > n) = \exp\bigg\{\sum_{n=1}^{\infty} \frac{z^n}{n}\, \mathrm{P}(S_n < 0)\bigg\}.$$

Therefore,

$$\sum_{n=0}^{\infty} z^n\, \mathrm{P}(M_n \leq 0) = \mathrm{e}^{\phi(z)} \sum_{n=0}^{\infty} z^n\, \mathrm{P}(\tilde{M}_n < 0), \qquad \phi(z) := \sum_{n=1}^{\infty} \frac{z^n}{n}\, \mathrm{P}(S_n = 0).$$

To prove (39), it remains to use (25) and Karamata's Tauberian theorem (see, e.g. Corollary 1.7.3 of [7]), noting that $\phi(z) \to \phi$ as $z \nearrow 1$.

Now, from (36) and (39) we find that, as $n \to \infty$,

$$\mathrm{E}\, \mathrm{e}^{-S_{j-q}^*} (W_{n-q}^*)^{-1} \sim c_{j-q}^*\, \mathrm{P}(L_{n-q}^* \geq 0) \sim c_{j-q}^* \mathrm{e}^{\phi}\, \mathrm{P}(\tilde{M}_n < 0),$$

where, with a natural definition of $\mathrm{E}_{\cdot,+}^*$ and with $\tilde{L}_r^+$ as defined in (27), from definitions (21) and (22), we have

$$\mathrm{e}^{\phi} c_{j-q}^* = \mathrm{e}^{\phi} \sum_{l=0}^{\infty} \mathrm{E}_{\cdot,+}^*\, G_l^{*+}(j-q)\, \mathbf{1}_{\{\tilde{M}_l^{*-} < 0\}} = \sum_{r=0}^{\infty} \mathrm{E}_{-,\cdot}\, G_r^{-}(j-q)\, \mathbf{1}_{\{\tilde{L}_r^{+} > 0\}}\,.$$

Therefore, from (17) and (38), as $n \to \infty$ we have

$$\begin{aligned}\mathrm{E}\, N_n(n-j) &= \sum_{k=n-j}^{n-1} \mathrm{E}\, \mathrm{e}^{-S_{n-j}} W_k^{-1} = \sum_{q=1}^{j} \mathrm{E}\, \mathrm{e}^{-S^*_{j-q}} (W^*_{n-q})^{-1} \\ &\sim \mathrm{P}(L^*_n \geq 0) \sum_{q=1}^{j} c^*_{j-q} \sim \mathrm{P}(L^*_n \geq 0) \mathrm{e}^{-\phi} \sum_{q=1}^{j} \sum_{r=0}^{\infty} \mathrm{E}_{-,\cdot}\, G^-_r(j-q)\, \mathbf{1}_{\{\tilde{L}^+_r > 0\}} \\ &\sim d_j\, \mathrm{P}(\tilde{M}_n < 0),\end{aligned}$$

as desired. Theorem 4 is thus proved.

The next theorem describes the asymptotic behaviour of the expectation $\mathrm{E}\, N_n(j)$ as

$$\min\{j, n-j\} \to \infty.$$

Theorem 5. *Let Spitzer's condition, (18), be satisfied. Then*

$$\lim_{j, n-j \to \infty} \frac{\mathrm{E}\, N_n(j)}{(n-j)\, \mathrm{P}(\tilde{M}_j < 0)\, \mathrm{P}(L_{n-j} \geq 0)} = \frac{1}{\rho}. \tag{40}$$

Remark 6. In view of (25), the assertion of the theorem can be rewritten as follows:

$$\mathrm{E}\, N_n(j) \sim \rho^{-1} j^{-\rho} l_2(j) (n-j)^{\rho} l_1(n-j) \quad \text{as } j, n-j \to \infty.$$

It follows that, for any fixed $\varepsilon \in (0, \frac{1}{2})$, we have, for $t \in [\varepsilon, 1-\varepsilon]$,

$$\mathrm{E}\, N_n(\lfloor nt \rfloor) \sim \frac{\sin \pi\rho}{\pi\rho} \left(\frac{1-t}{t}\right)^{\rho} \quad \text{as } n \to \infty,$$

where $\lfloor x \rfloor$ denotes the integer part of x. It is interesting to compare this with the corresponding (obvious) asymptotics for the case $w(j) \equiv 1$, where $\mathrm{E}\, N_n(\lfloor nt \rfloor) \sim -\ln t$ (of course, the functions of t on the right-hand sides of the respective relations are densities on $(0, 1)$).

In the case where $\mathrm{E}\, \theta = 0$ and $\mathrm{E}\, \theta^2 < \infty$, we do not even need to bound the value j/n away from 0 and 1: in that case, from the asymptotic behaviour of the denominators in (40) (see, e.g. page 94 of [9]), we have

$$\mathrm{E}\, N_n(j) \sim \frac{2}{\pi} \left(\frac{n-j}{j}\right)^{1/2} \quad \text{as } j, n-j \to \infty.$$

Note also that assertions (29) and (30) (Theorem 4) can be viewed as the 'boundary cases' of (40): there is a 'smooth transition' between these asymptotics. To make the meaning of the last statement more precise, we also have to show that $c_j \sim \mathrm{P}(\tilde{M}_j < 0)$ as $j \to \infty$, which is a separate, nontrivial problem that is somewhat beyond the scope of the present paper. This claim, however, is more than plausible, as can be seen from the following simple (but not quite rigorous) argument. When j is large, the main contribution to the sum in c_j in (28) comes from terms with l close to j. Indeed, for $l - j \gg 1$, the lth term in that sum is

$$\mathrm{E}_{\cdot,+}\left[\frac{\mathrm{e}^{S^-_{l-j}}}{\sum_{p=1}^{l} \mathrm{e}^{S_p} + \eta_2^+} \,\middle|\, \tilde{M}^-_l < 0\right] \mathrm{P}(\tilde{M}^-_l < 0) \sim \mathrm{E}_{-,+}\left[\frac{\mathrm{e}^{S^-_{l-j}}}{\eta_1^- + \eta_2^+}\right] \mathrm{P}(\tilde{M}^-_l < 0), \tag{41}$$

https://doi.org/10.1017/S0001867800001440 Published online by Cambridge University Press

where the equivalence is due to the following interpretation: using P^- corresponds to conditioning our random walk to stay negative; see, e.g. [4]. This interpretation also implies that the contribution of all such terms to the sum c_j will be relatively small, as, for large $l-j$, the sums S^-_{l-j} typically assume fairly large negative values, given that the walk stays negative. A similar assertion holds when $j-l \gg 1$. On the other hand, for the terms with indices l that are 'close' to j (assume, say, that $|l-j| \le K$ for an arbitrarily large, fixed K), we could use the fact that $\mathrm{P}(\tilde{M}^-_l < 0) \sim \mathrm{P}(\tilde{M}^-_j < 0)$ as $j \to \infty$ and $l/j \to 1$ (due to regular variation), and then observe that the sum of the conditional expectations that appear on the left-hand side of (41) will be close to 1. This is because there we are basically conditioning on the random walk $\{S^-_k\}$ staying negative, and, under $\mathrm{E}_{-,+}$, for large K and $j \gg K$ we have

$$\sum_{|l-j|\le K} \frac{\mathrm{e}^{-S^+_{j-l}}\mathbf{1}_{\{j\ge l\}} + \mathrm{e}^{S^-_{l-j}}\mathbf{1}_{\{j<l\}}}{\sum_{p=1}^{l}\mathrm{e}^{S^-_p} + \eta^+_2} \approx \frac{1}{\eta^-_1 + \eta^+_2}\Bigg[\sum_{k=0}^{K}\mathrm{e}^{-S^+_k} + \sum_{k=1}^{K}\mathrm{e}^{S^-_k}\Bigg] \approx 1.$$

We split the proof of the theorem into several steps. As said above, the main contribution to the expectation $\mathrm{E}\,\mathrm{e}^{-S_j}W_k^{-1}$ from the sum (17) comes from the event where j is close to $\tau(k)$. So first we will show that the contribution from the complementary event is negligibly small.

Lemma 5. *Under Spitzer's condition, (18), for any $\varepsilon > 0$ there exists a $J \equiv J(\varepsilon) < \infty$ such that, for all $j \ge J$ and $k-j \ge J$,*

$$\mathrm{E}[\mathrm{e}^{S_{\tau(k)}-S_j};\ |\tau(k)-j| \ge J] \le \varepsilon\,\mathrm{P}(\tilde{M}_j < 0)\,\mathrm{P}(L_{k-j} \ge 0). \tag{42}$$

Proof. Fix a $J > 0$ and choose a $j \ge J$ and a $k \ge j+J$. We have

$$\mathrm{E}[\mathrm{e}^{S_{\tau(k)}-S_j};\ |\tau(k)-j| \ge J] = R_1 + R_2,$$

where

$$R_1 := \sum_{t=0}^{j-J}\mathrm{E}[\mathrm{e}^{S_{\tau(k)}-S_j};\ \tau(k)=t], \qquad R_2 := \sum_{t=j+J}^{k}\mathrm{E}[\mathrm{e}^{S_{\tau(k)}-S_j};\ \tau(k)=t].$$

First consider R_2. For $t \ge j$, we obtain

$$\begin{aligned}\mathrm{E}[\mathrm{e}^{S_{\tau(k)}-S_j};\ \tau(k)=t] &= \mathrm{E}\Big[\mathrm{e}^{S_t-S_j};\ \min_{0\le p\le t-1} S_p > S_t,\ \min_{t\le p\le k} S_p \ge S_t\Big]\\ &= \mathrm{E}\Big[\mathrm{e}^{S_t-S_j};\ \min_{0\le p\le t-1} S_p > S_t\Big]\mathrm{P}(L_{k-t}\ge 0)\\ &= \mathrm{E}\Big[\mathrm{e}^{S_{t-j}};\ \max_{1\le p\le t} S_p < 0\Big]\mathrm{P}(L_{k-t}\ge 0),\end{aligned}$$

by the duality principle. Defining for each $l \ge 0$ the shifted random walk

$$\{S^{(l)}_p := S_{l+p} - S_l\}_{p\ge 0},$$

from (26) we obtain

$$\begin{aligned}\mathrm{E}\Big[\mathrm{e}^{S_{t-j}};\ \max_{1\le p\le t} S_p < 0\Big] &= \mathrm{E}\Big[\mathrm{e}^{S_{t-j}}\,\mathrm{P}\Big(\max_{1\le p\le j} S^{(t-j)}_p < -S_{t-j}\ \Big|\ S_{t-j}\Big);\ \max_{1\le p\le t-j} S_p < 0\Big]\\ &= \mathrm{E}[\mathrm{e}^{S_{t-j}}\tilde{\mu}_j(-S_{t-j});\ \tilde{M}_{t-j} < 0]\\ &\le C_2\,\mathrm{P}(\tilde{M}_j < 0)\,\mathrm{E}[\mathrm{e}^{S_{t-j}}U(-S_{t-j});\ \tilde{M}_{t-j} < 0].\end{aligned}$$

https://doi.org/10.1017/S0001867800001440 Published online by Cambridge University Press

Hence,

$$R_2 \le C_2\,\mathrm{P}(\tilde{M}_j < 0) \sum_{t=j+J}^{k} \mathrm{E}[\mathrm{e}^{S_{t-j}} U(-S_{t-j});\ \tilde{M}_{t-j} < 0]\,\mathrm{P}(L_{k-t} \ge 0)$$
$$= C_2\,\mathrm{P}(\tilde{M}_j < 0) \sum_{p=J}^{k-j} \mathrm{E}[\mathrm{e}^{S_p} U(-S_p);\ \tilde{M}_p < 0]\,\mathrm{P}(L_{k-j-p} \ge 0).$$

Since $U(x)$ is a renewal function, we have $U(x) = O(x),\ x \to \infty$. Thus, there exists a constant C_3 such that $\mathrm{e}^{-x}U(x) \le u(x) := C_3\mathrm{e}^{-x/2}$ for all $x > 0$. Since $\int_0^\infty u(x)\,\mathrm{d}x < \infty$, it follows from Lemma 4 and the duality principle that, for every $\varepsilon > 0$, there exists a $J_1 \equiv J_1(\varepsilon) < \infty$ such that, for all $k - j > J_1$,

$$\sum_{p=J_1}^{k-j} \mathrm{E}[\mathrm{e}^{S_p} U(-S_p);\ \tilde{M}_p < 0]\,\mathrm{P}(L_{k-j-p} \ge 0) \le \frac{\varepsilon}{2C_2}\,\mathrm{P}(L_{k-j} \ge 0).$$

Thus, for $k - j > J \ge J_1$,

$$R_2 \le \frac{\varepsilon}{2}\,\mathrm{P}(\tilde{M}_j < 0)\,\mathrm{P}(L_{k-j} \ge 0). \tag{43}$$

Now we will evaluate R_1. For $t < j$, we obtain

$$\mathrm{E}[\mathrm{e}^{S_{\tau(k)}-S_j};\ \tau(k) = t] = \mathrm{E}\Big[\mathrm{e}^{S_t-S_j};\ \min_{0\le p\le t-1} S_p > S_t;\ \min_{t\le p\le k} S_p \ge S_t\Big]$$
$$= \mathrm{E}\Big[\mathrm{e}^{S_t-S_j};\ \min_{t\le p\le k} S_p \ge S_t\Big]\,\mathrm{P}(\tilde{M}_t < 0)$$
$$= \mathrm{E}\Big[\mathrm{e}^{-S_{j-t}};\ \min_{0\le p\le k-t} S_p \ge 0\Big]\,\mathrm{P}(\tilde{M}_t < 0),$$

where to obtain the second relation we have again used the duality principle. Arguing as before, we see that

$$\mathrm{E}\Big[\mathrm{e}^{-S_{j-t}};\ \min_{0\le p\le k-t} S_p \ge 0\Big]$$
$$= \mathrm{E}\Big[\mathrm{e}^{-S_{j-t}}\,\mathrm{P}\Big(\min_{0\le p\le k-j} S_p^{(j-t)} \ge -S_{j-t}\ \Big|\ S_{j-t}\Big);\ \min_{0\le p\le j-t} S_p \ge 0\Big]$$
$$= \mathrm{E}[\mathrm{e}^{-S_{j-t}}\lambda_{k-j}(S_{j-t});\ L_{j-t} \ge 0]$$
$$\le C_1\,\mathrm{P}(L_{k-j} \ge 0)\,\mathrm{E}[\mathrm{e}^{-S_{j-t}}V(S_{j-t});\ L_{j-t} \ge 0].$$

Hence,

$$R_1 \le C_1\,\mathrm{P}(L_{k-j} \ge 0) \sum_{t=0}^{j-J} \mathrm{E}[\mathrm{e}^{-S_{j-t}}V(S_{j-t});\ L_{j-t} \ge 0]\,\mathrm{P}(\tilde{M}_t < 0)$$
$$= C_1\,\mathrm{P}(L_{k-j} \ge 0) \sum_{p=J}^{j} \mathrm{E}[\mathrm{e}^{-S_p}V(S_p);\ L_p \ge 0]\,\mathrm{P}(\tilde{M}_{j-p} < 0).$$

https://doi.org/10.1017/S0001867800001440 Published online by Cambridge University Press

From this bound we can deduce, using Lemma 4 and the same argument as that employed to evaluate R_2, that for every $\varepsilon > 0$ there exists a $J_2(\varepsilon) < \infty$ such that, for all $j > J \geq J_2$,

$$R_1 \leq \frac{\varepsilon}{2}\,\mathrm{P}(\tilde{M}_j < 0)\,\mathrm{P}(L_{k-j} \geq 0). \tag{44}$$

Combining (43) with (44) and setting $J := \max\{J_1, J_2\}$ completes the proof of Lemma 5.

Next we evaluate the contributions to the expectations of interest from the events where $\tau(k)$ is equal to a fixed number close to j.

Lemma 6. *Under Spitzer's condition, (18), for any fixed* $r \in \mathbb{Z}$*,*

$$\lim_{j,k-j\to\infty} \frac{\mathrm{E}[\mathrm{e}^{-S_j} W_k^{-1};\ \tau(k) = j+r]}{\mathrm{P}(\tilde{M}_j < 0)\,\mathrm{P}(L_{k-j} \geq 0)} = \mathrm{E}_{-,+}\,\frac{\mathrm{e}^{S_r^-}\,\mathbf{1}_{\{r\geq 0\}} + \mathrm{e}^{-S_{-r}^+}\,\mathbf{1}_{\{r<0\}}}{\eta_1^- + \eta_2^+}, \tag{45}$$

where η_1^- *and* η_2^+ *are independent RVs defined as in (23), but for the independent random walks* $\{S_n^-\}_{n\geq 0}$ *and* $\{S_n^+\}_{n\geq 0}$*, respectively.*

Proof. For $0 \leq r \leq k-j$, let

$$G_{j+r,k-j-r} := \frac{\mathrm{e}^{S_r^-}}{\sum_{p=1}^{j+r} \mathrm{e}^{S_p^-} + \sum_{p=0}^{k-j-r} \mathrm{e}^{-S_p^+}}.$$

Then

$$\begin{aligned}
&\mathrm{E}[\mathrm{e}^{-S_j} W_k^{-1};\ \tau(k) = j+r]\\
&\quad= \mathrm{E}\bigg[\frac{\mathrm{e}^{S_{j+r}-S_j}}{\sum_{p=0}^{k} \mathrm{e}^{S_{j+r}-S_p}};\ \min_{0\leq p\leq j+r-1} S_p > S_{j+r};\ \min_{j+r\leq p\leq k} S_p \geq S_{j+r}\bigg]\\
&\quad= \mathrm{E}[G_{j+r,k-j-r};\ \tilde{M}_{j+r}^- < 0,\ L_{k-j-r}^+ \geq 0]\\
&\quad= \mathrm{E}[G_{j+r,k-j-r} \mid \tilde{M}_{j+r}^- < 0,\ L_{k-j-r}^+ \geq 0]\,\mathrm{P}(\tilde{M}_{j+r} < 0)\,\mathrm{P}(L_{k-j-r} \geq 0).
\end{aligned}$$

Clearly, $0 < G_{j+r,k-j-r} \leq 1$ and

$$\lim_{j,k-j\to\infty} G_{j+r,k-j-r} = \frac{\mathrm{e}^{S_r^-}}{\eta_1^- + \eta_2^+} \qquad \mathrm{P}_{-,+}\text{-a.s.}$$

Hence, by applying Lemma 2 and recalling (25) and the properties of regularly varying functions (cf. (35)), we obtain (45) for $r \geq 0$. The proof of (45) for $r < 0$ is almost identical. Lemma 6 is thus proved.

Proof of Theorem 5. For a fixed $\varepsilon > 0$, let $J \equiv J(\varepsilon)$ be such that (42) holds. For $j \geq J$ and $n - j \geq J + 1$, from (17) we have

$$\mathrm{E}\,N_n(j) = R_3 + R_4 + R_5,$$

where

$$R_3 := \sum_{k=j}^{j+J-1} \mathrm{E}\,\mathrm{e}^{-S_j} W_k^{-1}, \qquad R_4 := \sum_{k=j+J}^{n-1} \mathrm{E}[\mathrm{e}^{-S_j} W_k^{-1};\ |\tau(k) - j| < J],$$

$$R_5 := \sum_{k=j+J}^{n-1} \mathrm{E}[\mathrm{e}^{-S_j} W_k^{-1};\ |\tau(k) - j| \geq J].$$

https://doi.org/10.1017/S0001867800001440 Published online by Cambridge University Press

We evaluate the quantities R_i, $i = 3, 4, 5$, separately. First observe that, in view of (30) (with n replaced by k), there exists a constant C_3 such that, for all sufficiently large j,

$$R_3 \le C_3 J\,\mathrm{P}(\tilde{M}_j < 0).$$

Thus, since

$$(n-j)\,\mathrm{P}(L_{n-j} \ge 0) \sim (n-j)^{\rho} l_1(n-j) \to \infty \quad \text{as } n-j \to \infty,$$

it follows that

$$R_3 = o((n-j)\,\mathrm{P}(\tilde{M}_j < 0)\,\mathrm{P}(L_{n-j} \ge 0)) \quad \text{as } n-j \to \infty.$$

Furthermore, using the obvious inequality $W_k \ge \mathrm{e}^{-S_{\tau(k)}}$ and the bound (42) together with (25) and Karamata's theorem, for $j \ge J$ and some constant $C_5 > 0$ we have

$$\begin{aligned} R_5 &\le \varepsilon\,\mathrm{P}(\tilde{M}_j < 0) \sum_{k=j+J}^{n-1} \mathrm{P}(L_{k-j} \ge 0) \\ &= \varepsilon\,\mathrm{P}(\tilde{M}_j < 0) \sum_{p=J}^{n-j-1} \mathrm{P}(L_p \ge 0) \\ &\le \varepsilon C_5 (n-j)\,\mathrm{P}(\tilde{M}_j < 0)\,\mathrm{P}(L_{n-j} \ge 0) \end{aligned}$$

and, therefore,

$$\frac{R_5}{(n-j)\,\mathrm{P}(\tilde{M}_j < 0)\,\mathrm{P}(L_{k-j} \ge 0)} \le \varepsilon C_5.$$

Finally, set

$$E_J := \mathrm{E}_{-,+} \frac{1 + \sum_{r=1}^{J-1} (\mathrm{e}^{S_r^-} + \mathrm{e}^{-S_r^+})}{\eta_1^- + \eta_2^+}.$$

Using Lemma 6, (25), and the properties of regularly varying functions, we see that, as $\min\{j, n-j\} \to \infty$,

$$\begin{aligned} R_4 &\sim E_J\,\mathrm{P}(\tilde{M}_j < 0) \sum_{k=j+J}^{n-1} \mathrm{P}(L_{k-j} \ge 0) \\ &\sim E_J\,\mathrm{P}(\tilde{M}_j < 0) \sum_{p=J}^{n-j-1} \mathrm{P}(L_p \ge 0) \\ &\sim E_J\,\mathrm{P}(\tilde{M}_j < 0) \rho^{-1} (n-j)\,\mathrm{P}(L_{n-j} \ge 0). \end{aligned}$$

Since $\lim_{J\to\infty} E_J = 1$ by the dominated convergence theorem, the assertion of Theorem 5 immediately follows from the above relation for R_4 and the bounds for R_3 and R_5.

3.2. The asymptotic behaviour of the distribution of $\mathrm{E}_w N_n(j)$

Unfortunately, our description of the asymptotic behaviour of $\mathrm{E}_w N_n(j)$ will be less detailed than that for $\mathrm{E}\, N_n(j)$. We will be able to describe the distribution of the RV $\mathrm{E}_w N_n(j)$ only for values of j located either to the right or in a small left-hand vicinity of the random epoch $\tau(n)$.

Theorem 6. *Let Spitzer's condition, (18), be satisfied and let $j \equiv j(n)$ be an arbitrary (random) sequence with the property that $(\tau(n) - j)_+ = o(n)$ in probability as $n \to \infty$. Then*

$$\mathrm{P}\bigg(\frac{\mathrm{e}^{S_j - S_{\tau(n)}}}{n-j}\,\mathrm{E}_w\, N_n(j) < x\bigg) \Rightarrow \mathrm{P}_{-,+}\bigg(\frac{1}{\eta_1^- + \eta_2^+} < x\bigg), \tag{46}$$

where η_1^- and η_2^+ are RVs defined as in (23), but for the independent random walks $\{S_n^-\}_{n\ge0}$ and $\{S_n^+\}_{n\ge0}$, respectively.

Proof. Since the RVs W_n (see (5)) are increasing in n, from (16) we have the following lower bound:

$$\mathrm{E}_w\, N_n(j) \ge (n-j)\mathrm{e}^{-S_j} W_n^{-1} = \frac{(n-j)\mathrm{e}^{S_{\tau(n)} - S_j}}{\sum_{k=0}^n \mathrm{e}^{S_{\tau(n)} - S_k}}.$$

Now we will derive an upper bound for $\mathrm{E}_w\, N_n(j)$. To this end observe that, according to (23), for any fixed $\varepsilon > 0$ and $\delta > 0$ there exists a $J < \infty$ such that

$$\mathrm{P}^+\bigg(\sum_{k=J}^{\infty} \mathrm{e}^{-S_k} > \delta\bigg) \le \varepsilon. \tag{47}$$

Clearly, for any $j \in [\tau(n), n-1]$,

$$\begin{aligned}\mathrm{E}_w\, N_n(j) &\le \mathrm{e}^{S_{\tau(n)} - S_j}(\tau(n) + J - j)_+ + \mathrm{e}^{-S_j}(n-j) W_{\tau(n)+J}^{-1} \\ &= \mathrm{e}^{S_{\tau(n)} - S_j}\bigg[(\tau(n) + J - j)_+ + (n-j)\bigg(\sum_{k=0}^{\tau(n)+J} \mathrm{e}^{S_{\tau(n)} - S_k}\bigg)^{-1}\bigg].\end{aligned}$$

Hence, we obtain

$$\begin{aligned}\bigg(\sum_{k=0}^{n} \mathrm{e}^{S_{\tau(n)} - S_k}\bigg)^{-1} &\le \frac{\mathrm{e}^{S_j - S_{\tau(n)}}}{n-j}\,\mathrm{E}_w\, N_n(j) \\ &\le \frac{(\tau(n) + J - j)_+}{n-j} + \bigg(\sum_{k=0}^{\tau(n)+J} \mathrm{e}^{S_{\tau(n)} - S_k}\bigg)^{-1}.\end{aligned} \tag{48}$$

Evidently, for $y > 0$,

$$\begin{aligned}\mathrm{P}\bigg(\sum_{k=0}^{n} \mathrm{e}^{S_{\tau(n)} - S_k} < y\bigg) &= \sum_{p=0}^{n} \mathrm{P}\bigg(\sum_{k=0}^{n} \mathrm{e}^{S_{\tau(n)} - S_k} < y;\ \tau(n) = p\bigg) \\ &= \sum_{p=0}^{n} \mathrm{P}\bigg(\sum_{l=1}^{p} \mathrm{e}^{S_l^-} + \sum_{r=0}^{n-p} \mathrm{e}^{-S_r^+} < y;\ \tilde{M}_p^- < 0,\ L_{n-p}^+ \ge 0\bigg).\end{aligned} \tag{49}$$

Furthermore, from (23) and (24) note that, as $\min\{p, n-p\} \to \infty$,

$$\mathrm{P}\bigg(\sum_{l=1}^{p} \mathrm{e}^{S_l^-} + \sum_{r=0}^{n-p} \mathrm{e}^{-S_r^+} < y \,\bigg|\, \tilde{M}_p^- < 0,\ L_{n-p}^+ \ge 0\bigg) \Rightarrow \mathrm{P}_{-,+}(\eta_1^- + \eta_2^+ < y). \tag{50}$$

If condition (18) is met, then the generalized arcsine law holds (see, e.g. Theorems 8.9.9 and 8.9.5 of [7]):

$$\lim_{n\to\infty} \mathrm{P}\left(\frac{\tau(n)}{n} \le x\right) = \frac{\sin \pi\rho}{\pi} \int_0^x t^{\rho-1}(1-t)^{-\rho}\,\mathrm{d}t, \qquad x \in [0,1]. \tag{51}$$

Thus, for any $\varepsilon_1 > 0$, there exists a $\delta_1 \in (0, \frac{1}{2})$ such that

$$\mathrm{P}(\tau(n) \notin (n\delta_1, n(1-\delta_1))) \le \varepsilon_1, \tag{52}$$

which, combined with (49) and (50), shows that, as $n \to \infty$,

$$\mathrm{P}\left(\sum_{k=0}^{n} \mathrm{e}^{S_{\tau(n)}-S_k} < y\right) \Rightarrow \mathrm{P}_{-,+}(\eta_1^- + \eta_2^+ < y). \tag{53}$$

A similar argument combined with (47) shows that

$$\mathrm{P}\left(\sum_{k=0}^{\tau(n)+J} \mathrm{e}^{S_{\tau(n)}-S_k} < y\right) \Rightarrow \mathrm{P}_{-,+}(\eta_1^- + \eta_2^+ < y) \tag{54}$$

as first $n \to \infty$ and then $J \to \infty$. On the other hand, again using (51), we conclude that, for $j \in [\tau(n), n-1]$,

$$\begin{aligned} \frac{(\tau(n)+J-j)_+}{n-j} &\le \mathbf{1}_{\{\tau(n)+J>j\}} \frac{J}{n-j} \mathbf{1}_{\{\tau(n)\ge n-\sqrt{n}\}} \\ &\quad + \mathbf{1}_{\{\tau(n)+J>j\}} \frac{J}{\sqrt{n}-J} \mathbf{1}_{\{\tau(n)<n-\sqrt{n}\}} \\ &\le J\,\mathbf{1}_{\{\tau(n)\ge n-\sqrt{n}\}} + \frac{J}{\sqrt{n}-J} \qquad (55) \\ &\xrightarrow{\mathrm{P}} 0 \qquad (56) \end{aligned}$$

as first $n \to \infty$ and then $J \to \infty$.

Using (53) on the left-hand side of (48), and (54) and (56) on the right-hand side of (48), proves (46) for $j \in [\tau(n), n-1]$.

For $\tau(n) - j > 0$, we can use similar arguments. The only difference is that, in this case,

$$(\tau(n) + J - j)_+ = \tau(n) + J - j,$$

and for $j < \tau(n)$ (varying with n in such a way that $(\tau(n) - j)_+ = o(n)$) the conclusion (56) still holds, by (52). Theorem 6 is thus proved.

Acknowledgements

This research was supported by the ARC Centre of Excellence for Mathematics and Statistics of Complex Systems. The second author was also supported by the Russian Foundation for Basic Research (grant 05-01-00035) and by the program 'Contemporary Problems of Theoretical Mathematics' of the Russian Academy of Sciences. He is grateful to the Department of Mathematics and Statistics of the University of Melbourne for its hospitality during his visit there.

The authors are very grateful to the referee for his valuable comments, which improved the exposition of the paper.

References

[1] Afanasyev, V. I., Geiger, J., Kersting, G. and Vatutin, V. A. (2005). Criticality for branching processes in random environment. *Ann. Prob.* **33,** 645–673.

[2] Babilot, M., Bougerol, P. and Élie, L. (1997). The random difference equation $X_n = A_n X_{n-1} + B_n$ in the critical case. *Ann. Prob.* **25,** 478–493.

[3] Barabási, A. L. and Albert, R. (1999). Emergence of scaling in random networks. *Science* **286,** 509–512.

[4] Bertoin, J. and Doney, R. A. (1994). On conditioning a random walk to stay nonnegative. *Ann. Prob.* **22,** 2152–2167.

[5] Biggins, J. D. and Grey, D. R. (1997). A note on the growth of random trees. *Statist. Prob. Lett.* **32,** 339–342.

[6] Billingsley, P. (1968). *Convergence of Probability Measures*. John Wiley, New York.

[7] Bingham, N. H., Goldie, C. M. and Teugels, J. L. (1987). *Regular Variation*. Cambridge University Press.

[8] Bollobás, B., Riordan, O., Spencer, J. and Tusnády, G. (2001). The degree of a scale-free random graph process. *Random Structures Algorithms* **18,** 279–290.

[9] Borovkov, A. A. (1976). *Stochastic Processes in Queueing Theory*. Springer, New York.

[10] Borovkov, K. A. and Motyer, A. (2005). On the asymptotic behavior of a simple growing point process model. *Statist. Prob. Lett.* **72,** 265–275.

[11] Devroye, L. (1988). Applications of the theory of records in the study of random trees. *Acta Informatica* **26,** 123–130.

[12] Doney, R. A. (1995). Spitzer's condition and ladder variables in random walks. *Prob. Theory Relat. Fields* **101,** 577–580.

[13] Élie, L. (1982). Compartement asymptotic du noyau potentiel sur les groupes de Lie. *Ann. Sci. École Norm. Sup.* **15,** 257–364.

[14] Hirano, K. (1998). Determination of the limiting coefficient for exponential functionals of random walks with positive drift. *J. Math. Sci. Univ. Tokyo* **5,** 299–332.

[15] Mahmoud, H. (1992). *Evolution of Random Search Trees*. John Wiley, New York.

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

[17] Rüschendorf, L. and Rachev, S. (1995). Probability metrics and recursive algorithms. *Adv. Appl. Prob.* **27,** 770–799.

[18] Spitzer, F. (1964). *Principles of Random Walk*. Van Nostrand, Princeton, NJ.

[19] Vatutin, V. A. and Dyakonova, E. E. (2004). Galton–Watson branching processes in random environment. II. Finite-dimensional distributions. *Theory Prob. Appl.* **49,** 275–308.