

RANDOM WALKS AND TREES

ZHAN SHI¹

Abstract. These notes provide an elementary and self-contained introduction to branching random walks.

Section 1 gives a brief overview of Galton–Watson trees, whereas Section 2 presents the classical law of large numbers for branching random walks. These two short sections are not exactly indispensable, but they introduce the idea of using size-biased trees, thus giving motivations and an avant-goût to the main part, Section 3, where branching random walks are studied from a deeper point of view, and are connected to the model of directed polymers on a tree.

Tree-related random processes form a rich and exciting research subject. These notes cover only special topics. For a general account, we refer to the St-Flour lecture notes of Peres [47] and to the forthcoming book of Lyons and Peres [42], as well as to Duquesne and Le Gall [23] and Le Gall [37] for continuous random trees.

CONTENTS

1. Galton–Watson trees	2
1.1. Galton–Watson trees and extinction probabilities	2
1.2. Size-biased Galton–Watson trees	6
1.3. Proof of the Kesten–Stigum theorem	9
1.4. The Seneta–Heyde norming	10
1.5. Notes	11
2. Branching random walks and the law of large numbers	11
2.1. Warm-up	11
2.2. Law of large numbers	12
2.3. Proof of the law of large numbers: lower bound	14
2.4. Size-biased branching random walk and martingale convergence	15
2.5. Proof of the law of large numbers: upper bound	18
2.6. Notes	19
3. Branching random walks and the central limit theorem	20
3.1. Central limit theorem	20

¹ Université Paris VI, Laboratoire de Probabilités et Modèles Aléatoires, 4 place Jussieu, 75252 Paris Cedex 05, France.
e-mail: zhan.shi@upmc.fr

3.2. Directed polymers on a tree	22
3.3. Small moments of partition function: upper bound in Theorem 3.5	23
3.4. Small moments of partition function: lower bound in Theorem 3.5	25
3.5. Partition function: all you need to know about exponents $-\frac{3\beta}{2}$ and $-\frac{1}{2}$	27
3.6. Central limit theorem: the $\frac{3}{2}$ limit	28
3.7. Central limit theorem: the $\frac{1}{2}$ limit	29
3.8. Partition function: exponent $-\frac{\beta}{2}$	30
3.9. A pathological case	30
3.10. The Seneta–Heyde norming for the branching random walk	30
3.11. Branching random walks with selection, I	31
3.12. Branching random walks with selection, II	33
3.13. Notes	35
4. Solution to the exercises	35
References	38

1. GALTON–WATSON TREES

We start by studying a few basic properties of supercritical Galton–Watson trees. The main aim of this section is to introduce the notion of size-biased trees. In particular, we see in Subsection 1.3 how this allows us to prove the well-known Kesten–Stigum theorem. This notion of size-biased trees will be developed in forthcoming sections to study more complicated models.

1.1. Galton–Watson trees and extinction probabilities

We are interested in processes involving (rooted) trees. The simplest rooted tree is the regular rooted tree, where each vertex has a fixed number (say m , with $m > 1$) of offspring. For example, here is a rooted binary tree:

Let Z_n denote the number of vertices (also called particles or individuals) in the n -th generation, then $Z_n = m^n$, $\forall n \geq 0$.

In probability theory, we often encounter trees where the number of offspring of a vertex is random. The easiest case is when these random numbers are i.i.d., which leads to a Galton–Watson tree¹.

A Galton–Watson tree starts with one initial ancestor (sometimes, it is possible to have several or even a random number of initial ancestors, in which case it will be explicitly stated). It produces a certain number of offspring according to a given probability distribution. The new particles form the first generation. Each of the new particles produces offspring according to the same probability distribution, independently of each

¹Sometimes also referred to as a Bienaymé–Galton–Watson tree.

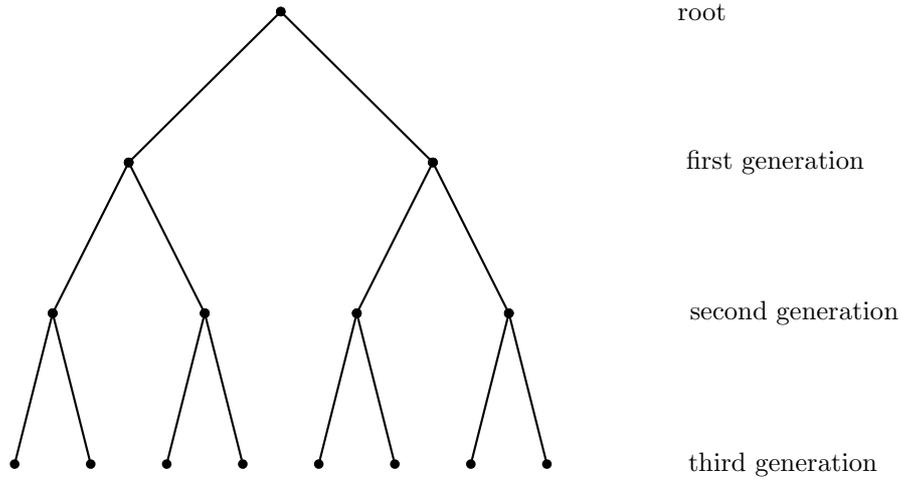


Figure 1: First generations in a rooted binary tree

other and of everything else in the generation. And the system regenerates. We write p_i for the probability that a given particle has i children, $i \geq 0$; thus $\sum_{i=0}^{\infty} p_i = 1$. [In the case of a regular m -ary tree, $p_m = 1$ and $p_i = 0$ for $i \neq m$.]

To avoid trivial discussions, we assume throughout that $p_0 + p_1 < 1$.

As before, we write Z_n for the number of particles in the n -th generation. It is clear that if $Z_n = 0$ for a certain n , then $Z_j = 0$ for all $j \geq n$.

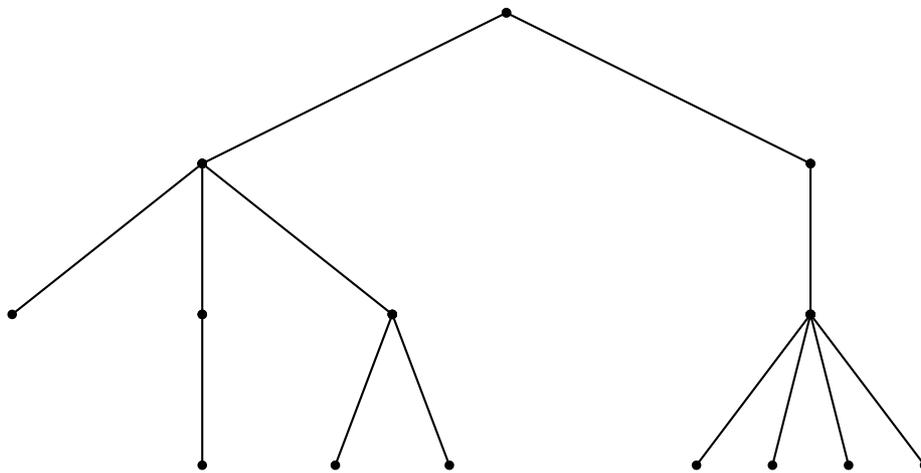


Figure 2: First generations in a Galton–Watson tree

In Figure 2, we have $Z_0 = 1$, $Z_1 = 2$, $Z_2 = 4$, $Z_3 = 7$.

One of the first questions we ask is about the extinction probability

$$q := \mathbf{P}\{Z_n = 0 \text{ eventually}\}. \tag{1.1}$$

It turns out that the expected number of offspring plays an important role. Let

$$m := \mathbf{E}(Z_1) = \sum_{i=0}^{\infty} ip_i \in (0, \infty]. \quad (1.2)$$

Theorem 1.1. *Let q be the extinction probability defined in (1.1).*

(i) *The extinction probability q is the smallest root of the equation $f(s) = s$ for $s \in [0, 1]$, where*

$$f(s) := \sum_{i=0}^{\infty} s^i p_i, \quad (0^0 := 1)$$

is the generating function of the reproduction law.

(ii) *In particular, $q = 1$ if $m \leq 1$, and $q < 1$ if $m > 1$.*

Proof. By definition, $f(s) = \mathbf{E}(s^{Z_1})$. Conditioning on Z_{n-1} , Z_n is the sum of Z_{n-1} i.i.d. random variables having the common distribution which is that of Z_1 ; thus $\mathbf{E}(s^{Z_n} | Z_{n-1}) = f(s)^{Z_{n-1}}$, which implies $\mathbf{E}(s^{Z_n}) = \mathbf{E}(f(s)^{Z_{n-1}})$. By induction, $\mathbf{E}(s^{Z_n}) = f_n(s)$ for any $n \geq 1$, where f_n denotes the n -th fold composition of f . In particular, $\mathbf{P}(Z_n = 0) = f_n(0)$.

The event $\{Z_n = 0\}$ being non-decreasing in n (i.e., $\{Z_n = 0\} \subset \{Z_{n+1} = 0\}$), we have

$$q = \mathbf{P}\left(\bigcup_n \{Z_n = 0\}\right) = \lim_{n \rightarrow \infty} \mathbf{P}(Z_n = 0) = \lim_{n \rightarrow \infty} f_n(0).$$

Let us look at the graph of the function f on $[0, 1]$. The function is (strictly) increasing and strictly convex, with $f(0) = p_0 \geq 0$ and $f(1) = 1$. In particular, it has at most two fixed points.

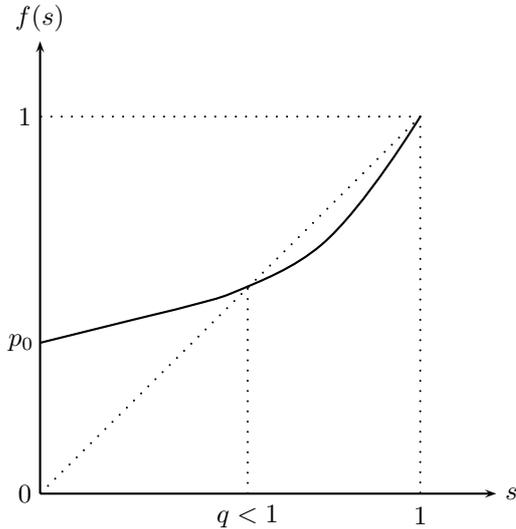


Figure 3a: $m > 1$

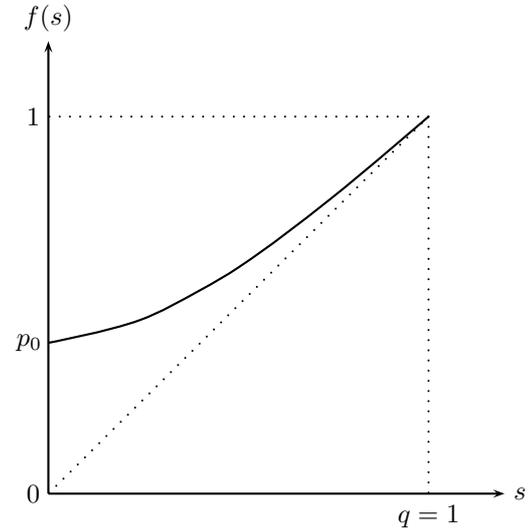


Figure 3b: $m \leq 1$

If $m \leq 1$, then $p_0 > 0$, and $f(s) > s$ for all $s \in [0, 1)$, which implies $f_n(0) \rightarrow 1$. In other words, $q = 1$ is the unique root of $f(s) = s$.

Assume now $m \in (1, \infty]$. This time, $f_n(0)$ converges increasingly to the unique root of $f(s) = s$, $s \in [0, 1)$. In particular, $q < 1$. □

Theorem 1.1 tells us that in the subcritical case (i.e., $m < 1$) and in the critical case ($m = 1$), the Galton–Watson process dies out with probability 1, whereas in the supercritical case ($m > 1$), the Galton–Watson process survives with (strictly) positive probability. In the rest of the text, we will be mainly interested in the supercritical case $m > 1$. But for the time being, let us introduce

$$W_n := \frac{Z_n}{m^n}, \quad n \geq 0,$$

which is well-defined as long as $m < \infty$. It is clear that W_n is a martingale (with respect to the natural filtration of (Z_n) , for example). Since it is non-negative, we have

$$W_n \rightarrow W, \quad \text{a.s.},$$

for some non-negative random variable W . [We recall that if (ξ_n) is a sub-martingale with $\sup_n \mathbb{E}(\xi_n^+) < \infty$, then ξ_n converges almost surely to a finite random variable. Apply this to $(-W_n)$.]

By Fatou’s lemma, $\mathbf{E}(W) \leq \liminf_{n \rightarrow \infty} \mathbf{E}(W_n) = 1$. It is, however, possible that $W = 0$. So it is important to know when W is non-degenerate.

We make the trivial remark that $W = 0$ if the system dies out. In particular, by Theorem 1.1, we have $W = 0$ a.s. if $m \leq 1$. What happens if $m > 1$?

We start with two simple observations. The first says that in general, $\mathbf{P}(W = 0)$ equals q or 1, whereas the second tells us that W is non-degenerate if the reproduction law admits a finite second moment.

Proposition 1.2. *Assume $m < \infty$. Then $\mathbf{P}(W = 0)$ equals either q or 1.*

Proof. There is nothing to prove if $m \leq 1$. So let us assume $1 < m < \infty$.

By definition,² $Z_{n+1} = \sum_{i=1}^{Z_n} Z_n^{(i)}$, where $Z_n^{(i)}$, $i \geq 1$, are copies of Z_n , independent of each other and Z_1 . Dividing on both sides by m^n and letting $n \rightarrow \infty$, it follows that mW is distributed as $\sum_{i=1}^{Z_1} W^{(i)}$, where $W^{(i)}$, $i \geq 1$, are copies of W , independent of each other and Z_1 . In particular, $\mathbf{P}(W = 0) = \mathbf{E}[\mathbf{P}(W = 0)^{Z_1}] = f(\mathbf{P}(W = 0))$, i.e., $\mathbf{P}(W = 0)$ is a root of $f(s) = s$ for $s \in [0, 1]$. In words, $\mathbf{P}(W = 0) = q$ or 1. □

Exercise 1.3. *If $\mathbf{E}(Z_1^2) < \infty$ and $m > 1$, then $\sup_n \mathbf{E}(W_n^2) < \infty$.*

Exercise 1.4. *If $\mathbf{E}(Z_1^2) < \infty$ and $m > 1$, then $\mathbf{E}(W) = 1$, and $\mathbf{P}(W = 0) = q$.*

It turns out that the second moment condition in Exercise 1.4 can be weakened to an $X \log X$ -type integrability condition. Let $\log^+ x := \log \max\{x, 1\}$.

²Usual notation: $\sum_{\emptyset} := 0$.

Theorem 1.5. (Kesten and Stigum [35]) *Assume $1 < m < \infty$. Then*

$$\mathbf{E}(W) = 1 \Leftrightarrow \mathbf{P}(W > 0 | \text{non-extinction}) = 1 \Leftrightarrow \mathbf{E}(Z_1 \log^+ Z_1) < \infty.$$

Remark. (i) The conclusion in the Kesten–Stigum theorem can also be stated as $\mathbf{E}(W) = 1 \Leftrightarrow \mathbf{P}(W = 0) = q \Leftrightarrow \sum_{i=1}^{\infty} p_i i \log i < \infty$.

(ii) The condition $\mathbf{E}(Z_1 \log^+ Z_1) < \infty$ may look technical. We will see in the next paragraph why this is a natural condition. \square

1.2. Size-biased Galton–Watson trees

In order to introduce size-biased Galton–Watson processes, we need to view the tree as a random element in a probability space $(\Omega, \mathcal{F}, \mathbf{P})$.

Let $\mathcal{U} := \{\emptyset\} \cup \bigcup_{k=1}^{\infty} (\mathbb{N}^*)^k$, where $\mathbb{N}^* := \{1, 2, \dots\}$.

If $u, v \in \mathcal{U}$, we denote by uv the concatenated element, with $u\emptyset = \emptyset u = u$.

A tree ω is a subset of \mathcal{U} satisfying: (i) $\emptyset \in \omega$; (ii) if $uj \in \omega$ for some $j \in \mathbb{N}^*$, then $u \in \omega$; (iii) if $u \in \omega$, then $uj \in \omega$ if and only if $1 \leq j \leq N_u(\omega)$ for some non-negative integer $N_u(\omega)$.

In the language of trees, if $u \in \mathcal{U}$ is an element of the tree ω , u is a vertex of the tree, and $N_u(\omega)$ the number of children. Vertices of ω are labeled by their line of descent: if $u = i_1 \dots i_n \in \mathcal{U}$, then u is the i_n -th child of the i_{n-1} -th child of \dots of the i_1 -th child of the initial ancestor \emptyset .

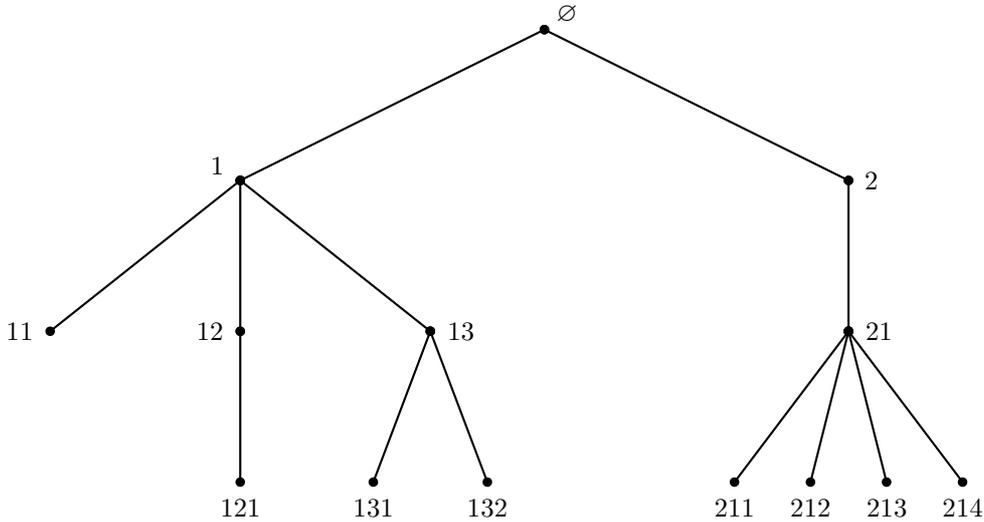


Figure 4: Vertices of a tree as elements of \mathcal{U}

Let Ω be the space of all trees. We now endow it with a sigma-algebra. For any $u \in \mathcal{U}$, let $\Omega_u := \{\omega \in \Omega : u \in \omega\}$ denote the subspace of Ω consisting of all the trees containing u as a vertex. [In particular, $\Omega_{\emptyset} = \Omega$

because all the trees contain the root as a vertex, according to part (i) of the definition.] The promised sigma-algebra associated with Ω is defined by

$$\mathcal{F} := \sigma\{\Omega_u, u \in \mathcal{U}\}.$$

Let $\mathbb{T} : \Omega \rightarrow \Omega$ be the identity application.

Let $(p_k, k \geq 0)$ be a probability. According to Neveu [44], there exists a probability \mathbf{P} on Ω such that the law of \mathbb{T} under \mathbf{P} is the law of the Galton–Watson tree with reproduction distribution (p_k) .

Let $\mathcal{F}_n := \sigma\{\Omega_u, u \in \mathcal{U}, |u| \leq n\}$, where $|u|$ is the length of u (representing the generation of the vertex u in the language of trees). Note that \mathcal{F} is the smallest sigma-field containing every \mathcal{F}_n .

For any tree $\omega \in \Omega$, let $Z_n(\omega)$ be the number of individuals in the n -th generation.³ It is easily checked that for any n , Z_n is a random variable taking values in $\mathbb{N} := \{0, 1, 2 \dots\}$.

Let $\widehat{\mathbf{P}}$ be the probability on (Ω, \mathcal{F}) such that for any n ,

$$\widehat{\mathbf{P}}|_{\mathcal{F}_n} = W_n \bullet \mathbf{P}|_{\mathcal{F}_n},$$

i.e., $\widehat{\mathbf{P}}(A) = \int_A W_n d\mathbf{P}$ for any $A \in \mathcal{F}_n$. Here, $\mathbf{P}|_{\mathcal{F}_n}$ and $\widehat{\mathbf{P}}|_{\mathcal{F}_n}$ are the restrictions of \mathbf{P} and $\widehat{\mathbf{P}}$ on \mathcal{F}_n , respectively. Since W_n is a martingale, the existence of $\widehat{\mathbf{P}}$ is guaranteed by Kolmogorov’s extension theorem.

For any n ,

$$\widehat{\mathbf{P}}(Z_n > 0) = \mathbf{E}[\mathbf{1}_{\{Z_n > 0\}} W_n] = \mathbf{E}[W_n] = 1.$$

Therefore, $\widehat{\mathbf{P}}(Z_n > 0, \forall n) = 1$. In other words, there is almost surely non-extinction of the Galton–Watson tree \mathbb{T} under the new probability $\widehat{\mathbf{P}}$. The Galton–Watson tree \mathbb{T} under $\widehat{\mathbf{P}}$ is called a size-biased Galton–Watson tree. Let us give a description of its paths.

Let $N := N_\emptyset$. If $N \geq 1$, then there are N individuals in the first generation. We write $\mathbb{T}_1, \mathbb{T}_2, \dots, \mathbb{T}_N$ for the N subtrees rooted at each of the N individual in the first generation.⁴

Exercise 1.6. Let $k \geq 1$. If A_1, A_2, \dots, A_k are elements of \mathcal{F} , then

$$\begin{aligned} & \widehat{\mathbf{P}}(N = k, \mathbb{T}_1 \in A_1, \dots, \mathbb{T}_k \in A_k) \\ &= \frac{kp_k}{m} \frac{1}{k} \sum_{i=1}^k \mathbf{P}(A_1) \cdots \mathbf{P}(A_{i-1}) \widehat{\mathbf{P}}(A_i) \mathbf{P}(A_{i+1}) \cdots \mathbf{P}(A_k). \end{aligned} \tag{1.3}$$

Equation (1.3) tells us the following fact about the size-biased Galton–Watson tree: The root has the biased distribution, i.e., having k children with probability $\frac{kp_k}{m}$; among the individuals in the first generation, one of them is chosen randomly (according to the uniform distribution) such that the subtree rooted at this

³The rigorous definition is $Z_n(\omega) := \#\{u \in \mathcal{U} : u \in \omega, |u| = n\}$.

⁴The rigorous definition of $\mathbb{T}_u(\omega)$ if $u \in \omega$ is a vertex of ω : $\mathbb{T}_u(\omega) := \{v \in \mathcal{U} : uv \in \omega\}$.

vertex is a size-biased Galton–Watson tree, whereas the subtrees rooted at all other vertices in the first generation are independent copies of the usual Galton–Watson tree.

Iterating the procedure, we obtain a decomposition of the size-biased Galton–Watson tree with an (infinite) spine and with i.i.d. copies of the usual Galton–Watson tree: The root $\emptyset =: w_0$ has the biased distribution, i.e., having k children with probability $\frac{k p k}{m}$. Among the children of the root, one of them is chosen randomly (according to the uniform distribution) as the element of the spine in the first generation (denoted by w_1). We attach subtrees rooted at all other children; these subtrees are independent copies of the usual Galton–Watson tree. The vertex w_1 has the biased distribution. Among the children of w_1 , we choose at random one of them as the element of the spine in the second generation (denoted by w_2). Independent copies of the usual Galton–Watson tree are attached as subtrees rooted at all other children of w_1 , whereas w_2 has the biased distribution. And so on. See Figure 5.

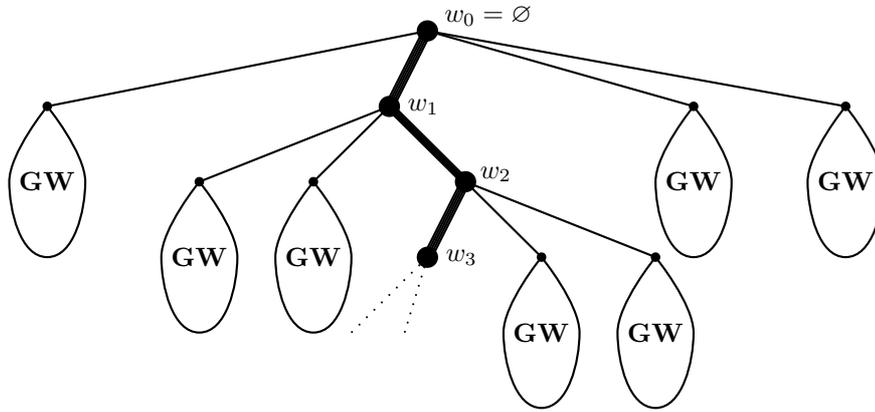


Figure 5: A size-biased Galton–Watson tree

From technical point of view, it is more convenient to connect size-biased Galton–Watson trees with Galton–Watson branching processes with immigration, described as follows.

A Galton–Watson branching processes with immigration starts with no individual (say), and is characterized by a reproduction law and an immigration law. At generation n (for $n \geq 1$), Y_n new individuals immigrate into the system, while all individuals regenerate independently and following the same reproduction law; we assume that $(Y_n, n \geq 1)$ is a collection of i.i.d. random variables following the same immigration law, and independent of everything else in that generation.

Our description of the size-biased Galton–Watson tree can be reformulated in the following way: $(Z_n - 1, n \geq 0)$ under $\hat{\mathbf{P}}$ is a Galton–Watson branching process with immigration, whose immigration law is that of $\hat{N} - 1$, with $\mathbf{P}(\hat{N} = k) := \frac{k p k}{m}$ for $k \geq 1$.

1.3. Proof of the Kesten–Stigum theorem

We prove the Kesten–Stigum theorem, by means of size-biased Galton–Watson trees. Let us start with a few elementary results.

Exercise 1.7. *Let X, X_1, X_2, \dots be i.i.d. non-negative random variables.*

- (i) *If $\mathbf{E}(X) < \infty$, then $\frac{X_n}{n} \rightarrow 0$ a.s.*
- (ii) *If $\mathbf{E}(X) = \infty$, then $\limsup_{n \rightarrow \infty} \frac{X_n}{n} = \infty$ a.s.*

For the next elementary results, let (\mathcal{F}_n) be a filtration, and let \mathbf{P} and $\widehat{\mathbf{P}}$ be probabilities⁵ on $(\Omega, \mathcal{F}_\infty)$. Assume that for any n , $\widehat{\mathbf{P}}|_{\mathcal{F}_n}$ is absolutely continuous with respect to $\mathbf{P}|_{\mathcal{F}_n}$. Let $\xi_n := \frac{d\widehat{\mathbf{P}}|_{\mathcal{F}_n}}{d\mathbf{P}|_{\mathcal{F}_n}}$, and let $\xi := \limsup_{n \rightarrow \infty} \xi_n$.

Exercise 1.8. *Prove that (ξ_n) is a \mathbf{P} -martingale. Prove that $\xi_n \rightarrow \xi$ \mathbf{P} -a.s., and that $\xi < \infty$ \mathbf{P} -a.s.*

Exercise 1.9. *Prove that*

$$\widehat{\mathbf{P}}(A) = \mathbf{E}(\xi \mathbf{1}_A) + \widehat{\mathbf{P}}(A \cap \{\xi = \infty\}), \quad \forall A \in \mathcal{F}_\infty. \quad (1.4)$$

Hint: You can first prove the identity assuming $\widehat{\mathbf{P}} \ll \mathbf{P}$ and using Lévy’s martingale convergence theorem⁶.

Exercise 1.10. *Prove that*

$$\widehat{\mathbf{P}} \ll \mathbf{P} \Leftrightarrow \xi < \infty, \widehat{\mathbf{P}}\text{-a.s.} \Leftrightarrow \mathbf{E}(\xi) = 1, \quad (1.5)$$

$$\widehat{\mathbf{P}} \perp \mathbf{P} \Leftrightarrow \xi = \infty, \widehat{\mathbf{P}}\text{-a.s.} \Leftrightarrow \mathbf{E}(\xi) = 0. \quad (1.6)$$

The Kesten–Stigum theorem will be a consequence of Seneta [51]’s theorem for branching processes with immigration.

Exercise 1.11. (Seneta’s theorem) *Let Z_n denote the number of individuals in the n -th generation of a branching process with immigration (Y_n) . Assume that $1 < m < \infty$, where m denotes the expectation of the reproduction law.*

- (i) *If $\mathbf{E}(\log^+ Y_1) < \infty$, then $\lim_{n \rightarrow \infty} \frac{Z_n}{m^n}$ exists and is finite a.s.*
- (ii) *If $\mathbf{E}(\log^+ Y_1) = \infty$, then $\limsup_{n \rightarrow \infty} \frac{Z_n}{m^n} = \infty$, a.s.*

Proof of the Kesten–Stigum theorem. Assume $1 < m < \infty$.

If $\sum_{i=1}^\infty p_i i \log i < \infty$, then $\mathbf{E}(\log^+ \widehat{N}) < \infty$. By Seneta’s theorem, $\lim_{n \rightarrow \infty} W_n$ exists $\widehat{\mathbf{P}}$ -a.s. and is finite $\widehat{\mathbf{P}}$ -a.s. By (1.5), $\mathbf{E}(W) = 1$; in particular, $\mathbf{P}(W = 0) < 1$ and thus $\mathbf{P}(W = 0) = q$ (Proposition 1.2).

⁵We denote by \mathcal{F}_∞ , the smallest sigma-algebra containing all \mathcal{F}_n .

⁶That is, if η is \mathbf{P} -integrable, then $\mathbf{E}(\eta | \mathcal{F}_n)$ converges in $L^1(\mathbf{P})$ and \mathbf{P} -almost surely, when $n \rightarrow \infty$, to $\mathbf{E}(\eta | \mathcal{F}_\infty)$.

If $\sum_{i=1}^{\infty} p_i i \log i = \infty$, then $\mathbf{E}(\log^+ \widehat{N}) = \infty$. By Seneta's theorem, $\lim_{n \rightarrow \infty} W_n$ exists $\widehat{\mathbf{P}}$ -a.s. and is infinite $\widehat{\mathbf{P}}$ -a.s. By (1.6), $\mathbf{E}(W) = 0$, and thus $\mathbf{P}(W = 0) = 1$. \square

1.4. The Seneta–Heyde norming

In the supercritical case, the Kesten–Stigum theorem (Theorem 1.5) tells us that under the condition $\mathbf{E}(Z_1 \log^+ Z_1) < \infty$, $\frac{Z_n}{m^n}$ converges almost surely to a limit (i.e., W), which vanishes precisely on the set of extinction. It turns out that even without the condition $\mathbf{E}(Z_1 \log^+ Z_1) < \infty$, the conclusion still holds true if we are allowed to modify the normalizing function.

Theorem 1.12. (Seneta [50], Heyde [31]) *Assume $1 < m < \infty$. Then there exists a sequence (c_n) of positive constants such that $\frac{c_{n+1}}{c_n} \rightarrow m$ and that $\frac{Z_n}{c_n}$ has a (finite) almost sure limit vanishing precisely on the set of extinction.*

Proof. We note that f^{-1} is well-defined on $[p_0, 1]$.

Let $s_0 \in (q, 1)$ and define by induction $s_n := f^{-1}(s_{n-1})$, $n \geq 0$. Clearly, $s_n \uparrow 1$.

Conditioning on \mathcal{F}_{n-1} , Z_n is the sum of Z_{n-1} i.i.d. random variables having the common distribution which is that of Z_1 ; thus $\mathbf{E}(s_n^{Z_n} | \mathcal{F}_{n-1}) = f(s_n)^{Z_{n-1}}$, which implies that $s_n^{Z_n}$ is a martingale. Since it is also bounded, it converges almost surely and in L^1 to a limit⁷, say Y , with $\mathbf{E}(Y) = \mathbf{E}(s_0^{Z_0}) = s_0$.

Let $c_n := \frac{1}{\log(1/s_n)}$, $n \geq 0$. By definition, $s_n^{Z_n} = e^{-Z_n/c_n}$, so that $\lim_{n \rightarrow \infty} \frac{Z_n}{c_n}$ exists a.s., and lies in $[0, \infty]$. By l'Hôpital's rule,

$$\lim_{s \rightarrow 1} \frac{\log f(s)}{\log s} = \lim_{s \rightarrow 1} \frac{f'(s) s}{f(s)} = m.$$

This yields $\frac{c_{n+1}}{c_n} \rightarrow m$.

Consider the set $A := \{\lim_{n \rightarrow \infty} \frac{Z_n}{c_n} = 0\}$. Let as before $\mathbb{T}_1, \dots, \mathbb{T}_{Z_1}$ denote the subtrees rooted at each of the individuals in the first generation. Then⁸

$$\mathbf{P}(A) = \mathbf{P}(\mathbb{T} \in A) = \mathbf{E}\{\mathbf{P}(\mathbb{T} \in A | Z_1)\} \leq \mathbf{E}\{\mathbf{P}(\mathbb{T}_1 \in A, \dots, \mathbb{T}_{Z_1} \in A | Z_1)\},$$

the inequality being a consequence of the fact that $\frac{c_{n+1}}{c_n} \rightarrow m$. Since $\mathbb{T}_1, \dots, \mathbb{T}_{Z_1}$ are i.i.d. given Z_1 , we have $\mathbf{P}(\mathbb{T}_1 \in A, \dots, \mathbb{T}_{Z_1} \in A | Z_1) = [\mathbf{P}(A)]^{Z_1}$; therefore, $\mathbf{P}(A) \leq \mathbf{E}\{[\mathbf{P}(A)]^{Z_1}\} = f(\mathbf{P}(A))$.

On the other hand, $\mathbf{P}(A) \geq q$. Thus $\mathbf{P}(A) \in \{q, 1\}$, and $\mathbf{P}(A | \text{non-extinction}) \in \{0, 1\}$. Since $\mathbf{E}(Y) = s_0 < 1$, this yields $\mathbf{P}(A | \text{non-extinction}) = 0$; in other words, $\{\lim_{n \rightarrow \infty} \frac{Z_n}{c_n} = 0\} = \{\text{extinction}\}$ almost surely.

Similarly, we can pose $B := \{\lim_{n \rightarrow \infty} \frac{Z_n}{c_n} < \infty\}$ and check that $\mathbf{P}(B) \leq f(\mathbf{P}(B))$. Since $\mathbf{P}(B) \geq q$, we have $\mathbf{P}(B | \text{non-extinction}) \in \{0, 1\}$. Now, $\mathbf{E}(Y) = s_0 > q$, we obtain $\mathbf{P}(B | \text{non-extinction}) = 1$. \square

⁷We use the fact that if (ξ_n) is a martingale with $\mathbf{E}(\sup_n |\xi_n|) < \infty$, then it converges almost surely and in L^1 .

⁸In the literature, we say that the property A is *inherited*. What we have proved here says that an inherited property has probability either 0 or 1 given non-extinction.

Remark. Of course⁹, in the Seneta–Heyde theorem, we have $c_n \approx m^n$ if and only if $\mathbf{E}(Z_1 \log^+ Z_1) < \infty$.

1.5. Notes

Subsection 1.1 concerns elementary properties of Galton–Watson processes. For a general account, we refer to standard books such as Asmussen and Hering [3], Athreya and Ney [4], Harris [30].

The formulation of branching processes described at the beginning of Subsection 1.2 is due to Neveu [44]; the idea of viewing Galton–Watson branching processes as tree-valued random variables can be found in Harris [30].

The idea of size-biased branching processes, which goes back at least to Kahane and Peyrière [33], has been used by several authors in various contexts. Its presentation in Subsection 1.2, as well as its use to prove the Kesten–Stigum theorem, comes from Lyons, Pemantle and Peres [41]. Size-biased branching processes can actually be used to prove the corresponding results of the Kesten–Stigum theorem in the critical and subcritical cases. See [41] for more details.

The short proof of Seneta’s theorem (Exercise 1.11) is borrowed from Asmussen and Hering [3], pp. 50–51.

The Seneta–Heyde theorem (Theorem 1.12) was first proved by Seneta for convergence in distribution, and then by Heyde for almost sure convergence. The short proof is from Lyons and Peres [42]. For another simple proof, see Grey [25].

2. BRANCHING RANDOM WALKS AND THE LAW OF LARGE NUMBERS

The Galton–Watson branching process simply counts the number of particles in each generation. In this section, we make an extension in the spatial sense by associating each individual of the Galton–Watson process with a random variable. This results to a branching random walk.

We first study a simple example of branching random walk; in particular, the idea of using the Cramér–Chernoff large deviation theorem appears in a natural way. We then put this idea into a general setting, and prove a law of large numbers for the branching random walk. Our basic technique relies, once again, on (a spatial version of) size-biased trees. In particular, this technique also gives a spatial version of the Kesten–Stigum theorem, namely, the Biggins martingale convergence theorem.

2.1. Warm-up

We study a simple example of branching random walk in this paragraph. Let \mathbb{T} be a binary tree rooted at \emptyset . Let $(\xi_x, x \in \mathbb{T})$ be a collection¹⁰ of i.i.d. random variables indexed by all the vertices of \mathbb{T} . To simplify the situation, we assume that the common distribution of ξ_x is the uniform distribution on $[0, 1]$.

For any vertex $x \in \mathbb{T}$, let $[\emptyset, x]$ denote the shortest path connecting \emptyset to x . If x is in the n -th generation, then $[\emptyset, x]$ is composed of $n + 1$ vertices, each one is the parent of the next vertex whereas the last vertex

⁹By $a_n \approx b_n$, we mean $0 < \liminf_{n \rightarrow \infty} \frac{a_n}{b_n} \leq \limsup_{n \rightarrow \infty} \frac{a_n}{b_n} < \infty$.

¹⁰By an abuse of notation, we keep using \mathbb{T} to denote the vertices of \mathbb{T} .

is simply x . Let $\llbracket \emptyset, x \rrbracket := \llbracket \emptyset, x \rrbracket \setminus \{\emptyset\}$. We define $V(\emptyset) := 0$ and

$$V(x) := \sum_{y \in \llbracket \emptyset, x \rrbracket} \xi_y, \quad x \in \mathbb{T} \setminus \{\emptyset\}.$$

Then $(V(x), x \in \mathbb{T})$ is an example of branching random walk which we study in its generality in the next subsection. For the moment, we continue with our example.

For any x in the n -th generation, $V(x)$ is by definition sum of n i.i.d. uniform-[0, 1] random variables, so by the usual law of large numbers, $V(x)$ would be approximatively $\frac{n}{2}$ when n is large. We are interested in the asymptotic behaviour, when $n \rightarrow \infty$, of¹¹

$$\frac{1}{n} \inf_{|x|=n} V(x).$$

If, with probability one, $\gamma := \lim_{n \rightarrow \infty} \frac{1}{n} \inf_{|x|=n} V(x)$ exists and is a constant, then $\gamma \leq \frac{1}{2}$ according to the above discussion. A little more thought will convince us that the inequality would be strict: $\gamma < \frac{1}{2}$. Here is why.

Let $s \in (0, \frac{1}{2})$. For any $|x| = n$, the probability $\mathbf{P}\{V(x) \leq sn\}$ goes to 0 when $n \rightarrow \infty$ (weak law of large numbers); this probability is exponentially small (the Cramér–Chernoff theorem¹²): $\mathbf{P}\{V(x) \leq sn\} \approx \exp(-I(s)n)$, for some constant $I(s) > 0$ depending on s which is explicitly known. Let $N(s, n)$ be the number of x in the n -th generation such that $V(x) \leq sn$. Then $\mathbf{E}(N(s, n))$ is approximatively $2^n \exp(-I(s)n)$. In particular, if $I(s) < \log 2$, then $\mathbf{E}(N(s, n)) \rightarrow \infty$, and one would expect that $N(s, n)$ would be at least one when n is sufficiently large. If indeed $N(s, n) \geq 1$, then $\inf_{|x|=n} V(x) \leq sn$, which would yield $\gamma \leq s < \frac{1}{2}$. On the other hand, if $I(s) > \log 2$, then by Chebyshev’s inequality $\mathbf{P}\{N(s, n) \geq 1\} \leq \mathbf{E}(N(s, n))$, $\sum_n \mathbf{P}\{N(s, n) \geq 1\} < \infty$, so a Borel–Cantelli argument tells that $\gamma \geq s$.

If the heuristics were true, then one would get $\gamma = \inf\{s < \frac{1}{2} : I(s) < \log 2\} = \sup\{s < \frac{1}{2} : I(s) > \log 2\}$ (admitting that the last identity holds).

In the next subsections, we will develop the heuristics into a rigorous argument. We will prove the Biggins martingale convergence theorem (which is a spatial version of the Kesten–Stigum theorem) by means of size-biased branching random walks; the martingale convergence theorem allows us to avoid discussing the technical point about whether $N(s, n) \geq 1$ when $\mathbf{E}(N(s, n))$ is very large.

2.2. Law of large numbers

The (discrete-time one-dimensional) branching random walk is a natural extension of the Galton–Watson tree in the spatial sense; its distribution is governed by a point process which we denote by Θ .

¹¹Throughout, $|x|$ denote the generation of the vertex x .

¹²The Cramér–Chernoff theorem says that if X_1, X_2, \dots are i.i.d. real-valued random variables, then under some mild general assumptions, $\mathbf{P}\{\sum_{i=1}^n X_i \leq sn\}$ (for $s < \mathbf{E}(X_1)$) and $\mathbf{P}\{\sum_{i=1}^n X_i \geq s'n\}$ (for $s' > \mathbf{E}(X_1)$) decay exponentially when $n \rightarrow \infty$. This can be formulated in a more general setting, known as the *large deviation* theory; see Dembo and Zeitouni [20].

An initial ancestor is born at the origin of the real line. Its children, who form the first generation, are positioned according to the point process Θ . Each of the individuals in the first generation produces children who are thus in the second generation and are positioned (with respect to their born places) according to the same point process Θ . And so on. We assume that each individual reproduces independently of each other and of everything else. The resulting system is called a branching random walk.

It is clear that if we only count the number of individuals in each generation, we get a Galton–Watson process, with $\#\Theta$ being its reproduction distribution.

The example in Subsection 2.1 corresponds to the special case that the point process Θ consists of two independent uniform-[0, 1] random variables.

Let $(V(x), |x| = n)$ denote the positions of the individuals in the n -th generation. We are interested in the asymptotic behaviour of $\inf_{|x|=n} V(x)$.

Let us introduce the (log-)Laplace transform of the point process

$$\psi(t) := \log \mathbf{E} \left(\sum_{|x|=1} e^{-tV(x)} \right) \in (-\infty, \infty], \quad t \geq 0. \quad (2.7)$$

Throughout this section, we assume:

- $\psi(t) < \infty$ for some $t > 0$;
- $\psi(0) > 0$.

The assumption $\psi(0) > 0$ is equivalent to $\mathbf{E}(\#\Theta) > 1$, i.e., the associated Galton–Watson tree is supercritical.

The main result of this section is as follows.

Theorem 2.1. *If $\psi(0) > 0$ and $\psi(t) < \infty$ for some $t > 0$, then almost surely on the set of non-extinction,*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \inf_{|x|=n} V(x) = \gamma,$$

where

$$\gamma := \inf \{ a \in \mathbb{R} : J(a) > 0 \}, \quad J(a) := \inf_{t \geq 0} [ta + \psi(t)], \quad a \in \mathbb{R}. \quad (2.8)$$

If instead we want to know about $\sup_{|x|=n} V(x)$, we only need to replace the point process Θ by $-\Theta$.

Theorem 2.1 is proved in Subsection 2.3 for the lower bound, and in Subsection 2.5 for the upper bound.

We close this paragraph with a few elementary properties of the functions $J(\cdot)$ and $\psi(\cdot)$. We assume that $\psi(t) < \infty$ for some $t > 0$.

Clearly, J is concave, being the infimum of concave (linear) function; in particular, it is continuous on the interior of $\{a \in \mathbb{R} : J(a) > -\infty\}$. Also, it is obvious that J is non-decreasing.

We recall that a function f is said to be lower semi-continuous at point t if for any sequence $t_n \rightarrow t$, $\liminf_{n \rightarrow \infty} f(t_n) \geq f(t)$, and that f is lower semi-continuous if it is lower semi-continuous at all points. It

is well-known and easily checked by definition that f is lower semi-continuous if and only if for any $a \in \mathbb{R}$, $\{t : f(t) \leq a\}$ is closed.

Exercise 2.2. Prove that ψ is convex and lower semi-continuous on $[0, \infty)$.

Exercise 2.3. Assume that $\psi(t) < \infty$ for some $t > 0$. Then for any $t \geq 0$,

$$\psi(t) = \sup_{a \in \mathbb{R}} [J(a) - at].$$

Hint: A property of the Legendre transformation $f^*(x) := \sup_{a \in \mathbb{R}} (ax - f(a))$: a function $f : \mathbb{R} \rightarrow (-\infty, \infty]$ is convex and lower semi-continuous if and only if¹³ $(f^*)^* = f$.

Exercise 2.4. Assume that $\psi(t) < \infty$ for some $t > 0$. Then

$$\gamma = \sup\{a \in \mathbb{R} : J(a) < 0\}$$

Furthermore, γ is the unique solution of $J(\gamma) = 0$.

2.3. Proof of the law of large numbers: lower bound

Fix an $a \in \mathbb{R}$ such that $J(a) < 0$. Let

$$L_n = L_n(a) := \sum_{|x|=n} \mathbf{1}_{\{V(x) \leq na\}}.$$

By Chebyshev's inequality, $\mathbf{P}(L_n > 0) = \mathbf{P}(L_n \geq 1) \leq \mathbf{E}(L_n)$. Let $t \geq 0$. Since $\mathbf{1}_{\{V(x) \leq na\}} \leq e^{nat-tV(x)}$, we have

$$\mathbf{P}(L_n > 0) \leq e^{nat} \mathbf{E}\left(\sum_{|x|=n} e^{-tV(x)}\right).$$

To compute the expectation expression on the right-hand side, let \mathcal{F}_k (for any k) be the sigma-algebra generated by the first k generations of the branching random walk; then

$$\mathbf{E}\left(\sum_{|x|=n} e^{-tV(x)} \mid \mathcal{F}_{n-1}\right) = \sum_{|y|=n-1} e^{-tV(y)} e^{\psi(t)}.$$

Taking expectation on both sides gives $\mathbf{E}(\sum_{|x|=n} e^{-tV(x)}) = e^{\psi(t)} \mathbf{E}(\sum_{|x|=n-1} e^{-tV(x)})$, which is $e^{n\psi(t)}$ by induction. As a consequence, $\mathbf{P}(L_n > 0) \leq e^{nat+n\psi(t)}$, for any $t \geq 0$. Taking infimum over all $t \geq 0$, this leads to:

$$\mathbf{P}(L_n > 0) \leq \exp\left(n \inf_{t \geq 0} (at + \psi(t))\right) = e^{nJ(a)}.$$

By assumption, $J(a) < 0$, so that $\sum_n \mathbf{P}(L_n > 0) < \infty$. By the Borel–Cantelli lemma, with probability one, for all sufficiently large n , we have $L_n = 0$, i.e., $\inf_{|x|=n} V(x) > na$. The lower bound in Theorem 2.1 follows.

□

¹³The “if” part is trivial, whereas the “only if” part is proved by the fact that $f(x) = \sup\{g(x) : g \text{ affine, } g \leq f\}$.

2.4. Size-biased branching random walk and martingale convergence

Let $\beta \in \mathbb{R}$ be such that $\psi(\beta) := \log \mathbf{E}\{\sum_{|x|=1} e^{-\beta V(x)}\} \in \mathbb{R}$. Let

$$W_n(\beta) := \frac{1}{e^{n\psi(\beta)}} \sum_{|x|=n} e^{-\beta V(x)} = \sum_{|x|=n} e^{-\beta V(x) - n\psi(\beta)}, \quad n \geq 1.$$

[When $\beta = 0$, $W_n(0)$ is the martingale we studied in Section 1.] Using the branching structure, we immediately see that $(W_n(\beta), n \geq 1)$ is a martingale with respect to (\mathcal{F}_n) , where \mathcal{F}_n is the sigma-field induced by the first n generations of the branching random walk. Therefore, $W_n(\beta) \rightarrow W(\beta)$ a.s., for some non-negative random variable $W(\beta)$. Fatou's lemma says that $\mathbf{E}[W(\beta)] \leq 1$.

Here is Biggins's martingale convergence theorem, which is the analogue of the Kesten–Stigum theorem for the branching random walk.

Theorem 2.5. (Biggins [8]) *If $\psi(\beta) < \infty$ and $\psi'(\beta) := -e^{-\psi(\beta)} \mathbf{E}\{\sum_{|x|=1} V(x)e^{-\beta V(x)}\}$ exists and is finite, then*

$$\begin{aligned} \mathbf{E}[W(\beta)] = 1 &\Leftrightarrow \mathbf{P}\{W(\beta) > 0 \mid \text{non-extinction}\} = 1 \\ &\Leftrightarrow \mathbf{E}[W_1(\beta) \log^+ W_1(\beta)] < \infty \text{ and } \beta\psi'(\beta) < \psi(\beta). \end{aligned}$$

A similar remark as in the Kesten–Stigum theorem applies here: the condition

$$\mathbf{P}\{W(\beta) > 0 \mid \text{non-extinction}\} = 1,$$

which means $\mathbf{P}[W(\beta) = 0] = q$, is easily seen to be equivalent to $\mathbf{P}[W(\beta) = 0] < 1$ via a similar argument as in Proposition 1.2.

The proof of the theorem relies the notion of size-biased branching random walks, which is an extension in the spatial sense of size-biased Galton–Watson processes. We only describe the main idea; for more details, we refer to Lyons [40].

Some basic notation is in order (for more details, see Neveu [44]). Let $\mathcal{U} := \{\emptyset\} \cup \bigcup_{k=1}^{\infty} (\mathbb{N}^*)^k$ as in Subsection 1.2. Let $\overline{\mathcal{U}} := \{(u, V(u)) : u \in \mathcal{U}, V : \mathcal{U} \rightarrow \mathbb{R}\}$. Let Ω be Neveu's space of *marked* trees, which consists of all the subsets ω of $\overline{\mathcal{U}}$ such that the first component of ω is a tree. [Attention: Ω is different from the Ω in Subsection 1.2.] Let $\mathbb{T} : \Omega \rightarrow \Omega$ be the identity application. According to Neveu [44], there exists a probability \mathbf{P} on Ω such that the law of \mathbb{T} under \mathbf{P} is the law of the branching random walk described in the previous section.

According to Kolmogorov's extension theorem, there exists a unique probability $\widehat{\mathbf{P}} = \widehat{\mathbf{P}}(\beta)$ on Ω such that for any $n \geq 1$,

$$\widehat{\mathbf{P}}_{|\mathcal{F}_n} = W_n(\beta) \bullet \mathbf{P}_{|\mathcal{F}_n}.$$

The law of \mathbb{T} under $\widehat{\mathbf{P}}$ is called the law of a size-biased branching random walk.

Here is a simple description of the size-biased branching random walk (i.e., the distribution of the branching random walk under $\widehat{\mathbf{P}}$). The offspring of the root $\emptyset =: w_0$ is generated according to the biased distribution¹⁴ $\widehat{\Theta} = \widehat{\Theta}(\beta)$. Pick one of these offspring w_1 at random; the probability that a given vertex x is picked up as w_1 is proportional to $e^{-\beta V(x)}$. The children other than w_1 give rise to independent ordinary branching random walks, whereas the offspring of w_1 is generated according to the biased distribution $\widehat{\Theta}$ (and independently of others). Again, pick one of the children of w_1 at random with probability which is inversely exponentially proportional to $(\beta \text{ times})$ the displacement, call it w_2 , with the others giving rise to independent ordinary branching random walks while w_2 produce offspring according to the biased distribution $\widehat{\Theta}$, and so on.

Proof of Theorem 2.5. If $\beta = 0$ or if the point process Θ is deterministic, then Theorem 2.5 is reduced to the Kesten–Stigum theorem (Theorem 1.5) proved in the previous section. So let us assume that $\beta \neq 0$ and that Θ is not deterministic.

(i) Assume the condition on the right-hand side fails. We claim that in this case, $\limsup_{n \rightarrow \infty} W_n(\beta) = \infty$ $\widehat{\mathbf{P}}$ -a.s.; thus by (1.6), $\mathbf{E}[W(\beta)] = 0$; a fortiori, $\mathbf{P}\{W(\beta) > 0 \mid \text{non-extinction}\} = 0$.

To see why $\limsup_{n \rightarrow \infty} W_n(\beta) = \infty$ $\widehat{\mathbf{P}}$ -a.s., we distinguish two possibilities.

First possibility: $\beta\psi'(\beta) \geq \psi(\beta)$. Then

$$\limsup_{n \rightarrow \infty} [-\beta V(w_n) - n\psi(\beta)] = \infty, \quad \widehat{\mathbf{P}}\text{-a.s.}$$

[This is obvious if $\beta\psi'(\beta) > \psi(\beta)$, because by the law of large numbers,

$$\frac{V(w_n)}{n} \rightarrow \mathbf{E}_{\widehat{\mathbf{P}}}[V(w_1)] = \mathbf{E}\left[\sum_{|x|=1} V(x)e^{-\beta V(x)-\psi(\beta)}\right] = -\psi'(\beta), \quad \widehat{\mathbf{P}}\text{-a.s.}$$

If $\beta\psi'(\beta) = \psi(\beta)$, then the law of large numbers says $\frac{V(w_n)}{n} \rightarrow -\frac{\psi(\beta)}{\beta}$, $\widehat{\mathbf{P}}$ -a.s., and we still have¹⁵

$$\liminf_{n \rightarrow \infty} [\beta V(w_n) + n\psi(\beta)] = -\infty, \quad \widehat{\mathbf{P}}\text{-a.s.}]$$

Since $W_n(\beta) \geq e^{-\beta V(w_n) - n\psi(\beta)}$, we immediately get $\limsup_{n \rightarrow \infty} W_n(\beta) = \infty$ $\widehat{\mathbf{P}}$ -a.s., as desired.

Second possibility: $\mathbf{E}[W_1(\beta) \log^+ W_1(\beta)] = \infty$. In this case, we argue that

$$\begin{aligned} W_{n+1}(\beta) &= \sum_{|x|=n} e^{-\beta V(x) - (n+1)\psi(\beta)} \sum_{|y|=n+1, y>x} e^{-\beta[V(y)-V(x)]} \\ &\geq e^{-\beta V(w_n) - (n+1)\psi(\beta)} \sum_{|y|=n+1, y>w_n} e^{-\beta[V(y)-V(w_n)]} =: I_n \times II_n. \end{aligned}$$

¹⁴That is, a point process whose distribution (under \mathbf{P}) is the distribution of Θ under $\widehat{\mathbf{P}}$.

¹⁵It is an easy consequence of the central limit theorem that if X_1, X_2, \dots are i.i.d. random variables with $\mathbf{E}(X_1) = 0$ and $\mathbf{E}(X_1^2) < \infty$, then $\limsup_{n \rightarrow \infty} \sum_{i=1}^n X_i = \infty$ and $\liminf_{n \rightarrow \infty} \sum_{i=1}^n X_i = -\infty$, a.s., as long as $\mathbf{P}\{X_1 = 0\} < 1$.

Since I_n are i.i.d. (under $\widehat{\mathbf{P}}$) with $\mathbf{E}_{\widehat{\mathbf{P}}}(\log^+ I_0) = \mathbf{E}[W_1(\beta) \log^+ W_1(\beta)] = \infty$, it follows from Exercise 1.7 (ii) that

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log^+ I_n = \infty, \quad \widehat{\mathbf{P}}\text{-a.s.}$$

On the other hand, $\frac{V(w_n)}{n} \rightarrow -\psi'(\beta)$, $\widehat{\mathbf{P}}$ -a.s. (law of large numbers), this yields $\limsup_{n \rightarrow \infty} I_n \times II_n = \infty$, $\widehat{\mathbf{P}}$ -a.s., which, again, leads to $\limsup_{n \rightarrow \infty} W_n(\beta) = \infty$ $\widehat{\mathbf{P}}$ -a.s.

(ii) We now assume that the condition on the right-hand side of the Biggins martingale convergence theorem is satisfied, i.e., $\beta\psi'(\beta) < \psi(\beta)$ and $\mathbf{E}[W_1(\beta) \log^+ W_1(\beta)] < \infty$. Let \mathcal{G} be the sigma-algebra generated by w_n and $V(w_n)$ as well as offspring of w_n , for all $n \geq 0$. Then

$$\begin{aligned} \mathbf{E}_{\widehat{\mathbf{P}}}[W_n(\beta) | \mathcal{G}] &= \sum_{k=0}^{n-1} e^{-\beta V(w_k) - (n+1)\psi(\beta)} \sum_{\substack{|x|=k+1: \\ x > w_k, x \neq w_{k+1}}} e^{-\beta[V(x) - V(w_k)]} e^{[n-(k+1)]\psi(\beta)} \\ &\quad + e^{-\beta V(w_n) - (n+1)\psi(\beta)} \\ &= \sum_{k=0}^{n-1} e^{-\beta V(w_k) - (k+1)\psi(\beta)} \sum_{|x|=k+1: x > w_k} e^{-\beta[V(x) - V(w_k)]} - \sum_{i=1}^{n-1} e^{-\beta V(w_i) - i\psi(\beta)} \\ &= \sum_{k=0}^{n-1} I_k \times II_k - e^{\psi(\beta)} \sum_{i=1}^{n-1} I_i. \end{aligned}$$

Since I_n are i.i.d. (under $\widehat{\mathbf{P}}$) with $\mathbf{E}_{\widehat{\mathbf{P}}}(\log^+ I_0) = \mathbf{E}[W_1(\beta) \log^+ W_1(\beta)] < \infty$, it follows from Exercise 1.7 (i) that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log^+ I_n = 0, \quad \widehat{\mathbf{P}}\text{-a.s.}$$

On the other hand, I_n decays exponentially fast (because $\beta\psi'(\beta) < \psi(\beta)$). It follows that $\sum_{k=0}^{n-1} I_k \times II_k - e^{\psi(\beta)} \sum_{i=1}^{n-1} I_i$ converges $\widehat{\mathbf{P}}$ -a.s. By (the conditional version of) Fatou's lemma, $\mathbf{E}_{\widehat{\mathbf{P}}}\{\liminf_{n \rightarrow \infty} W_n(\beta) | \mathcal{G}\} < \infty$, $\widehat{\mathbf{P}}$ -a.s.

Recall that $\mathbf{P}|_{\mathcal{F}_n} = \frac{1}{W_n(\beta)} \bullet \widehat{\mathbf{P}}|_{\mathcal{F}_n}$. We claim that $\frac{1}{W_n(\beta)}$ is a $\widehat{\mathbf{P}}$ -supermartingale¹⁶: indeed, for any $n \geq j$ and $A \in \mathcal{F}_j$, we have $\mathbf{E}_{\widehat{\mathbf{P}}}\left[\frac{1}{W_n(\beta)} \mathbf{1}_A\right] = \mathbf{P}\{W_n(\beta) > 0, A\} \leq \mathbf{P}\{W_j(\beta) > 0, A\} = \mathbf{E}_{\widehat{\mathbf{P}}}\left[\frac{1}{W_j(\beta)} \mathbf{1}_A\right]$, thus $\mathbf{E}_{\widehat{\mathbf{P}}}\left[\frac{1}{W_n(\beta)} | \mathcal{F}_j\right] \leq \frac{1}{W_j(\beta)}$ as claimed. Since $\frac{1}{W_n(\beta)}$ is a positive $\widehat{\mathbf{P}}$ -supermartingale, it converges $\widehat{\mathbf{P}}$ -a.s., to a limit which, according to what we have proved in the last paragraph, is $\widehat{\mathbf{P}}$ -almost surely (strictly) positive. We can write as follow: $\limsup_{n \rightarrow \infty} W_n(\beta) < \infty$, $\widehat{\mathbf{P}}$ -a.s. According to (1.5), this yields $\mathbf{E}[W(\beta)] = 1$, which obviously implies $\mathbf{P}[W(\beta) = 0] < 1$, and which is equivalent to $\mathbf{P}\{W(\beta) > 0 | \text{non-extinction}\} = 1$ as pointed out in the remark after Theorem 2.5. This completes the proof of Theorem 2.5. \square

We end this paragraph with the following re-statement of the size-biased branching random walk: for any $n \geq 1$, $\widehat{\mathbf{P}}|_{\mathcal{F}_n} := W_n(\beta) \bullet \mathbf{P}|_{\mathcal{F}_n}$, $\widehat{\mathbf{P}}(w_n = x | \mathcal{F}_n) = \frac{e^{-\beta V(x)}}{\sum_{|y|=n} e^{-\beta V(y)}} = \frac{e^{-\beta V(x) - n\psi(\beta)}}{W_n(\beta)}$ for any $|x| = n$; under $\widehat{\mathbf{P}}$, we have an (infinite) spine and i.i.d. copies of the usual branching random walk. Along the

¹⁶In general, $\frac{1}{W_n(\beta)}$ is not a $\widehat{\mathbf{P}}$ -martingale; see Harris and Roberts [29]. I am grateful to an anonymous referee for having pointed out an erroneous claim in a first draft.

spine, $V(w_i) - V(w_{i-1})$, $i \geq 1$, are i.i.d. under $\widehat{\mathbf{P}}$, and the distribution of $V(w_1)$ under $\widehat{\mathbf{P}}$ is given by $\mathbf{E}_{\widehat{\mathbf{P}}}[F(V(w_1))] = \mathbf{E}\{\sum_{|x|=1} F(V(x))e^{-\beta V(x)-\psi(\beta)}\}$, for any measurable function $F : \mathbb{R} \rightarrow \mathbb{R}_+$.

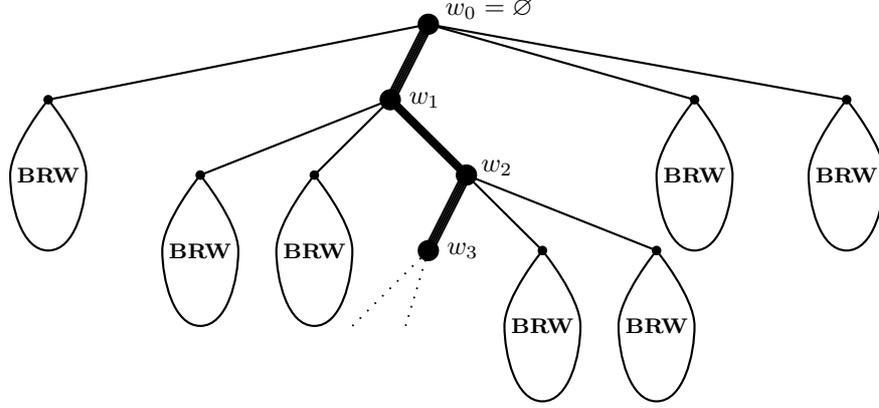


Figure 6: A size-biased branching random walk

Here is a consequence of the decomposition of the size-biased branching random walk, which will be of frequent use.

Corollary 2.6. *Assume $\psi(\beta) < \infty$. Then for any $n \geq 1$ and any measurable function $g : \mathbb{R} \rightarrow \mathbb{R}_+$,*

$$\mathbf{E}\left\{\sum_{|x|=n} e^{-\beta V(x)-n\psi(\beta)} g(V(x))\right\} = \mathbf{E}[g(S_n)], \quad (2.9)$$

where $S_n := \sum_{i=1}^n X_i$, and (X_i) is a sequence of i.i.d. random variables such that

$$\mathbf{E}[F(X_1)] = \mathbf{E}\left\{\sum_{|x|=1} F(V(x))e^{-\beta V(x)-\psi(\beta)}\right\},$$

for any measurable function $F : \mathbb{R} \rightarrow \mathbb{R}_+$.

Proof of Corollary 2.6. Let $\text{LHS}_{(2.9)}$ denote the expression on the left hand-side of (2.9). By definition, $\text{LHS}_{(2.9)} = \mathbf{E}_{\widehat{\mathbf{P}}}\left\{\sum_{|x|=n} \frac{e^{-\beta V(x)-n\psi(\beta)}}{W_n(\beta)} g(V(x))\right\}$. Recall that $\widehat{\mathbf{P}}(w_n = x | \mathcal{F}_n) = \frac{e^{-\beta V(x)-n\psi(\beta)}}{W_n(\beta)}$, this yields $\text{LHS}_{(2.9)} = \mathbf{E}_{\widehat{\mathbf{P}}}\left\{\sum_{|x|=n} \mathbf{1}_{\{w_n=x\}} g(V(x))\right\}$, which is $\mathbf{E}_{\widehat{\mathbf{P}}}[g(V(w_n))]$. It remains to recall that $V(w_i) - V(w_{i-1})$, $i \geq 1$, are i.i.d. under $\widehat{\mathbf{P}}$ having the distribution of X_1 . \square

2.5. Proof of the law of large numbers: upper bound

We prove the upper bound in the law of large numbers under the additional assumptions that $\psi(t) < \infty$, $\forall t \geq 0$ and that $\mathbf{E}[W_1(\beta) \log^+ W_1(\beta)] < \infty$, $\forall \beta > 0$.

Exercise 2.7. Assume that $\psi(t) < \infty, \forall t \geq 0$, and that $\psi(0) > 0$. Let $a \in \mathbb{R}$ be such that $0 < J(a) < \psi(0)$. Then there exists $\beta > 0$ such that $\psi'(\beta) = -a$.

Let $a \in \mathbb{R}$ be such that $0 < J(a) < \psi(0)$. According to Exercise 2.7, $\psi'(\beta) = -a$ for some $\beta > 0$. In particular, $0 < J(a) = a\beta + \psi(\beta) = -\beta\psi'(\beta) + \psi(\beta)$.

Let $W_n(\beta) := \sum_{|x|=n} e^{-\beta V(x) - n\psi(\beta)}$ as before. Let $\varepsilon > 0$ and let

$$\Delta_n := \sum_{|x|=n: |V(x) - na| > \varepsilon n} e^{-\beta V(x) - n\psi(\beta)}.$$

By Corollary 2.6 and in its notation, $\mathbf{E}(\Delta_n) = \mathbf{P}(|S_n - na| > \varepsilon n)$. Recall that $S_n = \sum_{i=1}^n X_i$, with (X_i) i.i.d. such that $\mathbf{E}[F(X_1)] = \mathbf{E}\{\sum_{|x|=1} F(V(x))e^{-\beta V(x) - \psi(\beta)}\}$, for any measurable function $F : \mathbb{R} \rightarrow \mathbb{R}_+$. In particular, $\mathbf{E}(X_1) = a$. By the Cramér–Chernoff large deviation theorem, $\sum_n \mathbf{P}(|S_n - na| > \varepsilon n) < \infty$. Therefore, $\sum_n \Delta_n < \infty$ a.s. In particular, $\Delta_n \rightarrow 0$, a.s.

By definition,

$$\begin{aligned} W_n(\beta) &= \sum_{|x|=n: |V(x) - na| \leq \varepsilon n} e^{-\beta V(x) - n\psi(\beta)} + \Delta_n \\ &\leq e^{-\beta n(a - \varepsilon) - n\psi(\beta)} \#\{|x| = n : V(x) \leq (a + \varepsilon)n\} + \Delta_n. \end{aligned}$$

We know $\Delta_n \rightarrow 0$ a.s., so that

$$\liminf_{n \rightarrow \infty} e^{-\beta n(a - \varepsilon) - n\psi(\beta)} \#\{|x| = n : V(x) \leq (a + \varepsilon)n\} \geq W(\beta), \quad \text{a.s.}$$

Since $\beta\psi'(\beta) < \psi(\beta)$ and $\mathbf{E}[W_1(\beta) \log^+ W_1(\beta)] < \infty$, it follows from the Biggins martingale convergence theorem (Theorem 2.5) that $W_n(\beta) \rightarrow W(\beta) > 0$ almost surely on non-extinction. As a consequence, on the set of non-extinction,

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \inf_{|x|=n} V(x) \leq a + \varepsilon, \quad \text{a.s.}$$

This completes the proof of the upper bound in Theorem 2.1. □

2.6. Notes

The law of large numbers (Theorem 2.1) was first proved by Hammersley [26] for the Bellman–Harris process, by Kingman [36] for the (strictly) positive branching random walk, and by Biggins [7] for the branching random walk.

The proof of the upper bound in the law of large numbers, presented in Subsection 2.5 based on martingale convergence, is not the original proof given by Biggins [7]. Biggins’ proof relies on constructing an auxiliary branching process, as suggested by Kingman [36].

The short proof of the Biggins martingale convergence theorem in Subsection 2.4, via size-biased branching random walks, is borrowed from Lyons [40].

3. BRANCHING RANDOM WALKS AND THE CENTRAL LIMIT THEOREM

We continue our study of the branching random walk. In Subsection 3.1, we state the main result, a central limit theorem for the minimal position in the branching random walk. We do not directly study the minimal position, but rather investigate some martingales involving *all* the living particles in the same generation, but to which only the minimal positions make significant contributions. The study of these martingales relies, again, on the idea of size-biased branching random walks, and is also connected to problems for directed polymers on a tree. Unfortunately, our central limit theorem does not apply to all branching random walks; a particular pathological case is analyzed in Subsection 3.9. At the end of the section, we mention a few related models of branching random walks.

3.1. Central limit theorem

Let $(V(x), |x| = n)$ denote the positions of the branching random walk. The law of large numbers proved in the previous section states that conditional on non-extinction,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \inf_{|x|=n} V(x) = \gamma, \quad \text{a.s.},$$

where γ is the constant in (2.8). It is natural to ask about the rate of convergence in this theorem.

Let us look at the special case of i.i.d. random variables assigned to the edges of a rooted regular tree. It turns out that $\inf_{|x|=n} V(x)$ has few fluctuations with respect to, say, its median $m_V(n)$. In fact, the law of $\inf_{|x|=n} V(x) - m_V(n)$ is tight! This was first proved by Bachmann [5] for the branching random walk under the technical condition that the common distribution of the i.i.d. random variables assigned on the edges of the regular tree admits a density function which is log-concave (a most important example being the Gaussian law). This technical condition was recently removed by Bramson and Zeitouni [15]. See also Subsection 5 of the survey paper by Aldous and Bandyopadhyay [2] for other discussions. Finally, let us mention the recent paper of Lifshits [38], where an example of branching random walk is constructed such that the law of $\inf_{|x|=n} V(x) - m_V(n)$ is tight but does not converge weakly.

Throughout the section, we assume that for some $\delta > 0$, $\delta_+ > 0$ and $\delta_- > 0$,

$$\mathbf{E}\left\{\left(\sum_{|x|=1} 1\right)^{1+\delta}\right\} < \infty, \quad (3.10)$$

$$\mathbf{E}\left\{\sum_{|x|=1} e^{-(1+\delta_+)V(x)}\right\} + \mathbf{E}\left\{\sum_{|x|=1} e^{\delta_- V(x)}\right\} < \infty, \quad (3.11)$$

We recall the log-Laplace transform

$$\psi(t) := \log \mathbf{E}\left\{\sum_{|x|=1} e^{-tV(x)}\right\} \in (-\infty, \infty], \quad t \geq 0.$$

By (3.11), $\psi(t) < \infty$ for $t \in [-\delta_-, 1 + \delta_+]$. Following Biggins and Kyprianou [11], we assume¹⁷

$$\psi(0) > 0, \quad \psi(1) = \psi'(1) = 0. \tag{3.12}$$

For comments on this assumption, see Remark (ii) below after Theorem 3.1.

Under (3.12), the value of the constant γ defined in (2.8) is $\gamma = 0$, so that Theorem 2.1 reads: almost surely on the set of non-extinction,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \inf_{|x|=n} V(x) = 0. \tag{3.13}$$

On the other hand, under (3.12), Theorem 2.5 tells us that $\sum_{|x|=n} e^{-V(x)} \rightarrow 0$ a.s., which yields that, almost surely,

$$\inf_{|x|=n} V(x) \rightarrow +\infty.$$

In other words, on the set of non-extinction, the system is transient to the right.

A refinement of (3.13) is obtained by McDiarmid [43]. Under the additional assumption $\mathbf{E}\{(\sum_{|x|=1} 1)^2\} < \infty$, it is proved in [43] that for some constant $c_1 < \infty$ and conditionally on the system's non-extinction,

$$\limsup_{n \rightarrow \infty} \frac{1}{\log n} \inf_{|x|=n} V(x) \leq c_1, \quad \text{a.s.}$$

We now state a central limit theorem.

Theorem 3.1. *Assume (3.10), (3.11) and (3.12). On the set of non-extinction, we have*

$$\limsup_{n \rightarrow \infty} \frac{1}{\log n} \inf_{|x|=n} V(x) = \frac{3}{2}, \quad \text{a.s.} \tag{3.14}$$

$$\liminf_{n \rightarrow \infty} \frac{1}{\log n} \inf_{|x|=n} V(x) = \frac{1}{2}, \quad \text{a.s.} \tag{3.15}$$

$$\lim_{n \rightarrow \infty} \frac{1}{\log n} \inf_{|x|=n} V(x) = \frac{3}{2}, \quad \text{in probability.} \tag{3.16}$$

Remark. (i) The most interesting part of Theorem 3.1 is possibly (3.14)–(3.15). It reveals the presence of fluctuations of $\inf_{|x|=n} V(x)$ on the logarithmic level, which is in contrast with a known result of Bramson [14] stating that for a class of branching random walks, $\frac{1}{\log \log n} \inf_{|x|=n} V(x)$ converges almost surely to a finite and positive constant.

(ii) Some brief comments on (3.12) are in order. In general (i.e., without assuming $\psi(1) = \psi'(1) = 0$), if

$$t^* \psi'(t^*) = \psi(t^*) \tag{3.17}$$

for some $t^* \in (0, \infty)$, then the branching random walk associated with the point process $\widehat{V}(x) := t^* V(x) + \psi(t^*)|x|$ satisfies (3.12). That is, as long as (3.17) has a solution (which is the case for example if $\psi(1) = 0$ and $\psi'(1) > 0$), the study will boil down to the case (3.12).

¹⁷We recall that the assumption $\psi(0) > 0$ in (3.12) means that the associated Galton–Watson tree is supercritical.

It is, however, possible that (3.17) has no solution. In such a situation, Theorem 3.1 does not apply. For example, we have already mentioned a class of branching random walks exhibited in Bramson [14], for which $\inf_{|x|=n} V(x)$ has an exotic $\log \log n$ behaviour. See Subsection 3.9 for more details.

(iii) Under (3.12) and suitable integrability assumptions, Addario-Berry and Reed [1] obtain a very precise asymptotic estimate of $\mathbf{E}[\inf_{|x|=n} V(x)]$, as well as an exponential upper bound for the deviation probability for $\inf_{|x|=n} V(x) - \mathbf{E}[\inf_{|x|=n} V(x)]$, which, in particular, implies (3.16).

(iv) In the case of (continuous-time) branching Brownian motion, a refined version of the analogue of (3.16) was proved by Bramson [13], by means of some powerful explicit analysis.

3.2. Directed polymers on a tree

The following model is borrowed from the well-known paper of Derrida and Spohn [22]: Let \mathbb{T} be a rooted Cayley tree; we study all self-avoiding walks (= directed polymers) of n steps on \mathbb{T} starting from the root. To each edge of the tree, is attached a random variable (= potential). We assume that these random variables are independent and identically distributed. For each walk ω , its energy $E(\omega)$ is the sum of the potentials of the edges visited by the walk. So the partition function is

$$Z_n(\beta) := \sum_{\omega} e^{-\beta E(\omega)},$$

where the sum is over all self-avoiding walks of n steps on \mathbb{T} , and $\beta > 0$ is the inverse temperature.¹⁸

More generally, we take \mathbb{T} to be a Galton–Watson tree, and observe that the energy $E(\omega)$ corresponds to (the partial sum of) the branching random walk described in the previous paragraphs. The associated partition function becomes

$$Z_{n,\beta} := \sum_{|x|=n} e^{-\beta V(x)}, \quad \beta > 0. \quad (3.18)$$

If $0 < \beta < 1$, the study of $Z_{n,\beta}$ boils down to the case $\psi'(1) < 0$ which is covered by the Biggins martingale convergence theorem (Theorem 2.5). In particular, on the set of non-extinction, $\frac{Z_{n,\beta}}{\mathbf{E}\{Z_{n,\beta}\}}$ converges almost surely to a (strictly) positive random variable.

We now study the case $\beta \geq 1$. If $\beta = 1$, we write simply Z_n instead of $Z_{n,1}$.

Theorem 3.2. *Assume (3.10), (3.11) and (3.12). On the set of non-extinction, we have*

$$Z_n = n^{-1/2+o(1)}, \quad \text{a.s.} \quad (3.19)$$

¹⁸There is hopefully no confusion possible between Z_n here and the number of individuals in a Galton–Watson process studied in Section 1.

Theorem 3.3. *Assume (3.10), (3.11) and (3.12), and let $\beta > 1$. On the set of non-extinction, we have*

$$\limsup_{n \rightarrow \infty} \frac{\log Z_{n,\beta}}{\log n} = -\frac{\beta}{2}, \quad \text{a.s.} \quad (3.20)$$

$$\liminf_{n \rightarrow \infty} \frac{\log Z_{n,\beta}}{\log n} = -\frac{3\beta}{2}, \quad \text{a.s.} \quad (3.21)$$

$$Z_{n,\beta} = n^{-3\beta/2+o(1)}, \quad \text{in probability.} \quad (3.22)$$

Again, the most interesting part in Theorem 3.3 is probably (3.20)–(3.21), which describes a new fluctuation phenomenon. Also, there is no phase transition any more for $Z_{n,\beta}$ at $\beta = 1$ from the point of view of upper almost sure limits.

An important step in the proof of Theorems 3.2 and 3.3 is to estimate all small moments of Z_n and $Z_{n,\beta}$, respectively. This is done in the next theorems.

Theorem 3.4. *Assume (3.10), (3.11) and (3.12). For any $a \in [0, 1)$, we have*

$$0 < \liminf_{n \rightarrow \infty} \mathbf{E} \left\{ (n^{1/2} Z_n)^a \right\} \leq \limsup_{n \rightarrow \infty} \mathbf{E} \left\{ (n^{1/2} Z_n)^a \right\} < \infty.$$

Theorem 3.5. *Assume (3.10), (3.11) and (3.12), and let $\beta > 1$. For any $0 < r < \frac{1}{\beta}$, we have*

$$\mathbf{E} \left\{ Z_{n,\beta}^r \right\} = n^{-3r\beta/2+o(1)}, \quad n \rightarrow \infty. \quad (3.23)$$

We prove the theorems of this section under the additional assumption that for some constant C , $\sup_{|x|=1} |V(x)| + \#\{x : |x| = 1\} \leq C$ a.s. This assumption is not necessary, but allows us to avoid some technical discussions.

3.3. Small moments of partition function: upper bound in Theorem 3.5

This subsection is devoted to (a sketch of) the proof of the upper bound in Theorem 3.5; the upper bound in Theorem 3.4 can be proved in a similar spirit, but needs more care.

We assume (3.10), (3.11) and (3.12), and fix $\beta > 1$.

Let $\widehat{\mathbf{P}}$ be such that $\widehat{\mathbf{P}}|_{\mathcal{F}_n} = Z_n \bullet \mathbf{P}|_{\mathcal{F}_n}$, $\forall n$. For any $Y \geq 0$ which is \mathcal{F}_n -measurable, we have $\mathbf{E}\{Z_{n,\beta}Y\} = \mathbf{E}_{\widehat{\mathbf{P}}}\left\{\sum_{|x|=n} \frac{e^{-\beta V(x)}}{Z_n} Y\right\} = \mathbf{E}_{\widehat{\mathbf{P}}}\left\{\sum_{|x|=n} \mathbf{1}_{\{w_n=x\}} e^{-(\beta-1)V(x)} Y\right\}$, and thus

$$\mathbf{E}\{Z_{n,\beta}Y\} = \mathbf{E}_{\widehat{\mathbf{P}}}\left\{e^{-(\beta-1)V(w_n)} Y\right\}. \quad (3.24)$$

Let $s \in (\frac{\beta-1}{\beta}, 1)$, and $\lambda > 0$. (We will choose $\lambda = \frac{3}{2}$.) Then

$$\begin{aligned} \mathbf{E} \left\{ Z_{n,\beta}^{1-s} \right\} &\leq n^{-(1-s)\beta\lambda} + \mathbf{E} \left\{ Z_{n,\beta}^{1-s} \mathbf{1}_{\{Z_{n,\beta} > n^{-\beta\lambda}\}} \right\} \\ &= n^{-(1-s)\beta\lambda} + \mathbf{E}_{\widehat{\mathbf{P}}} \left\{ \frac{e^{-(\beta-1)V(w_n)}}{Z_{n,\beta}^s} \mathbf{1}_{\{Z_{n,\beta} > n^{-\beta\lambda}\}} \right\}. \end{aligned} \quad (3.25)$$

We now estimate the expectation expression $\mathbf{E}_{\widehat{\mathbf{P}}}\{\cdots\}$ on the right-hand side. Let $a > 0$ and $\varrho > b > 0$ be constants such that $(\beta - 1)a > s\beta\lambda + \frac{3}{2}$ and $[\beta s - (\beta - 1)]b > \frac{3}{2}$. (The choice of ϱ will be made precise later on.) Let $\underline{w}_n \in \llbracket \emptyset, w_n \rrbracket$ be such that $V(\underline{w}_n) = \min_{x \in \llbracket \emptyset, w_n \rrbracket} V(x)$, and consider the following events:

$$\begin{aligned} E_{1,n} &:= \{V(w_n) > a \log n\} \cup \{V(w_n) \leq -b \log n\}, \\ E_{2,n} &:= \{V(\underline{w}_n) < -\varrho \log n, V(w_n) > -b \log n\}, \\ E_{3,n} &:= \{V(\underline{w}_n) \geq -\varrho \log n, -b \log n < V(w_n) \leq a \log n\}. \end{aligned}$$

Clearly, $\widehat{\mathbf{P}}(\cup_{i=1}^3 E_{i,n}) = 1$.

On the event $E_{1,n} \cap \{Z_{n,\beta} > n^{-\beta\lambda}\}$, we have either $V(w_n) > a \log n$, in which case $\frac{e^{-(\beta-1)V(w_n)}}{Z_{n,\beta}^s} \leq n^{s\beta\lambda - (\beta-1)a}$, or $V(w_n) \leq -b \log n$, in which case we use the trivial inequality $Z_{n,\beta} \geq e^{-\beta V(w_n)}$ to see that $\frac{e^{-(\beta-1)V(w_n)}}{Z_{n,\beta}^s} \leq e^{[\beta s - (\beta-1)]V(w_n)} \leq n^{-[\beta s - (\beta-1)]b}$ (recalling that $\beta s > \beta - 1$). Since $s\beta\lambda - (\beta - 1)a < -\frac{3}{2}$ and $[\beta s - (\beta - 1)]b > \frac{3}{2}$, we obtain:

$$\mathbf{E}_{\widehat{\mathbf{P}}}\left\{\frac{e^{-(\beta-1)V(w_n)}}{Z_{n,\beta}^s} \mathbf{1}_{E_{1,n} \cap \{Z_{n,\beta} > n^{-\beta\lambda}\}}\right\} \leq n^{-3/2}. \quad (3.26)$$

We now study the integral on $E_{2,n} \cap \{Z_{n,\beta} > n^{-\beta\lambda}\}$. Since $s > 0$, we can choose $s_1 > 0$ and $s_2 > 0$ (with s_2 sufficiently small) such that $s = s_1 + s_2$. We have, on $E_{2,n} \cap \{Z_{n,\beta} > n^{-\beta\lambda}\}$,

$$\frac{e^{-(\beta-1)V(w_n)}}{Z_{n,\beta}^s} = \frac{e^{\beta s_2 V(\underline{w}_n) - (\beta-1)V(w_n)}}{Z_{n,\beta}^{s_1}} \frac{e^{-\beta s_2 V(\underline{w}_n)}}{Z_{n,\beta}^{s_2}} \leq n^{-\beta s_2 \varrho + (\beta-1)b + \beta \lambda s_1} \frac{e^{-\beta s_2 V(\underline{w}_n)}}{Z_{n,\beta}^{s_2}}.$$

We admit that for small $s_2 > 0$, $\mathbf{E}_{\widehat{\mathbf{P}}}\left[\frac{e^{-\beta s_2 V(\underline{w}_n)}}{Z_{n,\beta}^{s_2}}\right] \leq n^K$, for some $K > 0$. [This actually is true for any $s_2 > 0$.] We choose (and fix) the constant ϱ so large that $-\beta s_2 \varrho + (\beta - 1)b + \beta \lambda s_1 + K < -\frac{3}{2}$. Therefore, for all large n ,

$$\mathbf{E}_{\widehat{\mathbf{P}}}\left\{\frac{e^{-(\beta-1)V(w_n)}}{Z_{n,\beta}^s} \mathbf{1}_{E_{2,n} \cap \{Z_{n,\beta} > n^{-\beta\lambda}\}}\right\} \leq n^{-3/2}. \quad (3.27)$$

We make a partition of $E_{3,n}$: let $M \geq 2$ be an integer, and let $a_i := -b + \frac{i(a+b)}{M}$, $0 \leq i \leq M$. By definition,

$$E_{3,n} = \bigcup_{i=0}^{M-1} \{V(\underline{w}_n) \geq -\varrho \log n, a_i \log n < V(w_n) \leq a_{i+1} \log n\} =: \bigcup_{i=0}^{M-1} E_{3,n,i}.$$

Let $0 \leq i \leq M - 1$. There are two possible situations. First situation: $a_i \leq \lambda$. In this case, we argue that on the event $E_{3,n,i}$, we have $Z_{n,\beta} \geq e^{-\beta V(w_n)} \geq n^{-\beta a_{i+1}}$ and $e^{-(\beta-1)V(w_n)} \leq n^{-(\beta-1)a_i}$, thus $\frac{e^{-(\beta-1)V(w_n)}}{Z_{n,\beta}^s} \leq n^{\beta s a_{i+1} - (\beta-1)a_i} = n^{\beta s a_i - (\beta-1)a_i + \beta s(a+b)/M} \leq n^{[\beta s - (\beta-1)]\lambda + \beta s(a+b)/M}$. Accordingly, in this situation,

$$\mathbf{E}_{\widehat{\mathbf{P}}}\left\{\frac{e^{-(\beta-1)V(w_n)}}{Z_{n,\beta}^s} \mathbf{1}_{E_{3,n,i}}\right\} \leq n^{[\beta s - (\beta-1)]\lambda + \beta s(a+b)/M} \widehat{\mathbf{P}}(E_{3,n,i}).$$

Second (and last) situation: $a_i > \lambda$. We have, on $E_{3,n,i} \cap \{Z_{n,\beta} > n^{-\beta\lambda}\}$, $\frac{e^{-(\beta-1)V(w_n)}}{Z_{n,\beta}^s} \leq n^{\beta\lambda s - (\beta-1)a_i} \leq n^{[\beta s - (\beta-1)]\lambda}$; thus, in this situation,

$$\mathbf{E}_{\widehat{\mathbf{P}}}\left\{\frac{e^{-(\beta-1)V(w_n)}}{Z_{n,\beta}^s} \mathbf{1}_{E_{3,n,i} \cap \{Z_{n,\beta} > n^{-\beta\lambda}\}}\right\} \leq n^{[\beta s - (\beta-1)]\lambda} \widehat{\mathbf{P}}(E_{3,n,i}).$$

We have therefore proved that

$$\begin{aligned} \mathbf{E}_{\widehat{\mathbf{P}}}\left\{\frac{e^{-(\beta-1)V(w_n)}}{Z_{n,\beta}^s} \mathbf{1}_{E_{3,n} \cap \{Z_{n,\beta} > n^{-\beta\lambda}\}}\right\} &= \sum_{i=0}^{M-1} \mathbf{E}_{\widehat{\mathbf{P}}}\left\{\frac{e^{-(\beta-1)V(w_n)}}{Z_{n,\beta}^s} \mathbf{1}_{E_{3,n,i} \cap \{Z_{n,\beta} > n^{-\beta\lambda}\}}\right\} \\ &\leq n^{[\beta s - (\beta-1)]\lambda + \beta s(a+b)/M} \widehat{\mathbf{P}}(E_{3,n}). \end{aligned}$$

We need to estimate $\widehat{\mathbf{P}}(E_{3,n})$. Recall from Subsection 2.4 that $V(w_i) - V(w_{i-1})$, $i \geq 1$, are i.i.d. under $\widehat{\mathbf{P}}$, and the distribution of $V(w_1)$ under $\widehat{\mathbf{P}}$ is that of X_1 given in Corollary 2.6. Therefore,

$$\widehat{\mathbf{P}}(E_{3,n}) = \mathbf{P}\left\{\min_{0 \leq k \leq n} S_k \geq -\varrho \log n, -b \log n \leq S_n \leq a \log n\right\},$$

with $S_k := \sum_{i=1}^k X_i$ as before. Since $\mathbf{E}(X_1) = 0$ (which is a consequence of the assumption $\psi'(1) = 0$), the random walk (S_k) is centered; so the probability above is $n^{-(3/2)+o(1)}$. Combining this with (3.25), (3.26) and (3.27) yields

$$\mathbf{E}\left\{Z_{n,\beta}^{1-s}\right\} \leq 2n^{-(1-s)\beta\lambda} + 2n^{-3/2} + n^{[\beta s - (\beta-1)]\lambda + \beta s(a+b)/M - (3/2) + o(1)}.$$

We choose $\lambda := \frac{3}{2}$. Since M can be as large as possible, this yields the upper bound in Theorem 3.5 by posing $r := 1 - s$. \square

3.4. Small moments of partition function: lower bound in Theorem 3.5

We now turn to (a sketch of) the proof of the lower bound in Theorem 3.5. [Again, the lower bound in Theorem 3.4 can be proved in a similar spirit, but with more care.]

Assume (3.10), (3.11) and (3.12). Let $\beta > 1$ and $s \in (1 - \frac{1}{\beta}, 1)$.

By means of the elementary inequality $(a+b)^{1-s} \leq a^{1-s} + b^{1-s}$ (for $a \geq 0$ and $b \geq 0$), we have, for some constant K ,

$$Z_{n,\beta}^{1-s} \leq K \sum_{j=1}^n e^{-(1-s)\beta V(w_{j-1})} \sum_{u \in \mathcal{I}_j} \left(\sum_{|x|=n, x>u} e^{-\beta[V(x)-V(u)]} \right)^{1-s} + e^{-(1-s)\beta V(w_n)},$$

where \mathcal{I}_j is the set of brothers¹⁹ of w_j .

Let \mathcal{G}_n be the sigma-field generated by everything on the spine in the first n generations. Since we are dealing with a regular tree, $\#\mathcal{I}_j$ is bounded by a constant. So the decomposition of size-biased branching

¹⁹That is, the set of children of w_{j-1} but excluding w_j .

random walks described in Subsection 2.4 tells us that for some constant K_1 ,

$$\mathbf{E}_{\hat{\mathbf{P}}}\left\{Z_{n,\beta}^{1-s} \mid \mathcal{G}_n\right\} \leq K_1 \sum_{j=1}^n e^{-(1-s)\beta V(w_{j-1})} \mathbf{E}\{Z_{n-j,\beta}^{1-s}\} + e^{-(1-s)\beta V(w_n)}.$$

Let $\varepsilon > 0$ be small, and let $r := \frac{3}{2}(1-s)\beta - \varepsilon$. By means of the already proved upper bound for $\mathbf{E}(Z_{n,\beta}^{1-s})$, this leads to:

$$\begin{aligned} \mathbf{E}_{\hat{\mathbf{P}}}\left\{Z_{n,\beta}^{1-s} \mid \mathcal{G}_n\right\} &\leq K_2 \sum_{j=1}^n e^{-(1-s)\beta V(w_{j-1})} (n-j+1)^{-r} + e^{-(1-s)\beta V(w_n)}. \\ &\leq K_3 \sum_{j=0}^n e^{-(1-s)\beta V(w_j)} (n-j+1)^{-r}. \end{aligned} \quad (3.28)$$

Since $\mathbf{E}(Z_{n,\beta}^{1-s}) = \mathbf{E}_{\hat{\mathbf{P}}}\left\{\frac{e^{-(\beta-1)V(w_n)}}{Z_{n,\beta}^s}\right\}$ (see (3.24)), we have, by Jensen's inequality (noticing that $V(w_n)$ is \mathcal{G}_n -measurable),

$$\mathbf{E}\left(Z_{n,\beta}^{1-s}\right) \geq \mathbf{E}_{\hat{\mathbf{P}}}\left\{\frac{e^{-(\beta-1)V(w_n)}}{\{\mathbf{E}_{\hat{\mathbf{P}}}(Z_{n,\beta}^{1-s} \mid \mathcal{G}_n)\}^{s/(1-s)}}\right\},$$

which, in view of (3.28), yields

$$\mathbf{E}\left(Z_{n,\beta}^{1-s}\right) \geq \frac{1}{K_3^{s/(1-s)}} \mathbf{E}_{\hat{\mathbf{P}}}\left\{\frac{e^{-(\beta-1)V(w_n)}}{\{\sum_{j=0}^n e^{-(1-s)\beta V(w_j)} (n-j+1)^{-r}\}^{s/(1-s)}}\right\}.$$

By the decomposition of size-biased branching random walks, the $\mathbf{E}_{\hat{\mathbf{P}}}\{\dots\}$ expression on the right-hand side is

$$\begin{aligned} &= \mathbf{E}\left\{\frac{e^{-(\beta-1)S_n}}{\{\sum_{j=0}^n (n-j+1)^{-r} e^{-(1-s)\beta S_j}\}^{s/(1-s)}}\right\} \\ &= \mathbf{E}\left\{\frac{e^{[\beta s - (\beta-1)]\tilde{S}_n}}{\{\sum_{k=0}^n (k+1)^{-r} e^{(1-s)\beta \tilde{S}_k}\}^{s/(1-s)}}\right\}, \end{aligned}$$

where

$$\tilde{S}_\ell := S_n - S_{n-\ell}, \quad 0 \leq \ell \leq n.$$

Consequently,

$$\mathbf{E}\left(Z_{n,\beta}^{1-s}\right) \geq \frac{1}{K_3^{s/(1-s)}} \mathbf{E}\left\{\frac{e^{[\beta s - (\beta-1)]\tilde{S}_n}}{\{\sum_{k=0}^n (k+1)^{-r} e^{(1-s)\beta \tilde{S}_k}\}^{s/(1-s)}}\right\}.$$

Let $K_4 > 0$ be a constant, and define

$$\begin{aligned} E_{n,1} &:= \bigcap_{k=1}^{\lfloor n^\varepsilon \rfloor - 1} \left\{ \tilde{S}_k \leq -K_4 k^{1/3} \right\} \cap \left\{ -2n^{\varepsilon/2} \leq \tilde{S}_{\lfloor n^\varepsilon \rfloor} \leq -n^{\varepsilon/2} \right\}, \\ E_{n,2} &:= \bigcap_{k=\lfloor n^\varepsilon \rfloor + 1}^{n - \lfloor n^\varepsilon \rfloor - 1} \left\{ \tilde{S}_k \leq -[k^{1/3} \wedge (n-k)^{1/3}] \right\} \cap \left\{ -2n^{\varepsilon/2} \leq \tilde{S}_{n - \lfloor n^\varepsilon \rfloor} \leq -n^{\varepsilon/2} \right\}, \\ E_{n,3} &:= \bigcap_{k=n - \lfloor n^\varepsilon \rfloor + 1}^{n-1} \left\{ \tilde{S}_k \leq \frac{3}{2} \log n \right\} \cap \left\{ \frac{3-\varepsilon}{2} \log n \leq \tilde{S}_n \leq \frac{3}{2} \log n \right\}. \end{aligned}$$

On $\cap_{i=1}^3 E_{n,i}$, we have $\sum_{k=0}^n (k+1)^{-r} e^{(1-s)\beta\tilde{S}_k} \leq K_5 n^{2\varepsilon}$, while $e^{[\beta s - (\beta-1)]\tilde{S}_n} \geq n^{(3-\varepsilon)[\beta s - (\beta-1)]/2}$ (recalling that $\beta s > \beta - 1$). Therefore,

$$\mathbf{E} \left(Z_{n,\beta}^{1-s} \right) \geq (K_3 K_5)^{-s/(1-s)} n^{-2\varepsilon s/(1-s)} n^{(3-\varepsilon)[\beta s - (\beta-1)]/2} \mathbf{P}\{\cap_{i=1}^3 E_{n,i}\}. \quad (3.29)$$

We need to bound $\mathbf{P}(\cap_{i=1}^3 E_{n,i})$ from below. Note that $\tilde{S}_\ell - \tilde{S}_{\ell-1}$, $1 \leq \ell \leq n$, are i.i.d., distributed as $S_1 = X_1$. For $j \leq n$, let $\tilde{\mathcal{G}}_j$ be the sigma-field generated by \tilde{S}_k , $1 \leq k \leq j$. Then $E_{n,1} \in \tilde{\mathcal{G}}_{n-\lfloor n^\varepsilon \rfloor}$ and $E_{n,2} \in \tilde{\mathcal{G}}_{n-\lfloor n^\varepsilon \rfloor}$, whereas writing $N := \lfloor n^\varepsilon \rfloor$, we see by the Markov property that $\mathbf{P}(E_{n,3} | \tilde{\mathcal{G}}_{n-\lfloor n^\varepsilon \rfloor})$ is greater than or equal to, on the event $\{\tilde{S}_{n-\lfloor n^\varepsilon \rfloor} \in I_n := [-2n^{\varepsilon/2}, -n^{\varepsilon/2}]\}$,

$$\inf_{z \in I_n} \mathbf{P} \left\{ S_i \leq \frac{3}{2} \log n - z, \forall 1 \leq i \leq N-1, \frac{3-\varepsilon}{2} \log n - z \leq S_N \leq \frac{3}{2} \log n - z \right\},$$

which is greater than $N^{-(1/2)+o(1)}$. As a consequence,

$$\mathbf{P}\{\cap_{i=1}^3 E_{n,i}\} \geq n^{-(\varepsilon/2)+o(1)} \mathbf{P}(E_{n,1} \cap E_{n,2}).$$

We now condition on $\tilde{\mathcal{G}}_{\lfloor n^\varepsilon \rfloor}$. Since $\mathbf{P}(E_{n,2} | \tilde{\mathcal{G}}_{\lfloor n^\varepsilon \rfloor}) \geq n^{-(3-\varepsilon)/2+o(1)}$, this yields

$$\mathbf{P}\{\cap_{i=1}^3 E_{n,i}\} \geq n^{-(\varepsilon/2)+o(1)} n^{-(3-\varepsilon)/2+o(1)} \mathbf{P}(E_{n,1}).$$

We choose the constant $K_4 > 0$ sufficiently small so that $\mathbf{P}(E_{n,1}) \geq n^{-(\varepsilon/2)+o(1)}$. Accordingly,

$$\mathbf{P}\{\cap_{i=1}^3 E_{n,i}\} \geq n^{-(3+\varepsilon)/2+o(1)}, \quad n \rightarrow \infty.$$

Substituting this into (3.29) yields

$$\mathbf{E} \left(Z_{n,\beta}^{1-s} \right) \geq n^{-2\varepsilon s/(1-s)} n^{(3-\varepsilon)[\beta s - (\beta-1)]/2} n^{-(3+\varepsilon)/2+o(1)}.$$

Since ε can be as small as possible, this implies the lower bound in Theorem 3.5. \square

3.5. Partition function: all you need to know about exponents $-\frac{3\beta}{2}$ and $-\frac{1}{2}$

In this subsection, we prove Theorem 3.2, as well as parts (3.21)–(3.22) of Theorem 3.3. We assume (3.10), (3.11) and (3.12) throughout the subsection.

Proof of Theorem 3.2 and (3.21)–(3.22) of Theorem 3.3: upper bounds. Let $\varepsilon > 0$. By Theorem 3.5 and Chebyshev's inequality, $\mathbf{P}\{Z_{n,\beta} > n^{-(3\beta/2)+\varepsilon}\} \rightarrow 0$. Therefore, $Z_{n,\beta} \leq n^{-(3\beta/2)+o(1)}$ in probability, yielding the upper bound in (3.22).

The upper bound in (3.21) follows trivially²⁰ from the upper bound in (3.22).

²⁰Convergence in probability is equivalent to saying that for any subsequence, there exists a sub-subsequence along which there is a.s. convergence.

It remains to prove the upper bound in Theorem 3.2. Fix $a \in (0, 1)$. Since Z_n^a is a non-negative supermartingale, the maximal inequality tells that for any $n \leq m$ and any $\lambda > 0$,

$$\mathbf{P}\left\{\max_{n \leq j \leq m} Z_j^a \geq \lambda\right\} \leq \frac{\mathbf{E}(Z_n^a)}{\lambda} \leq \frac{c_{82}}{\lambda n^{a/2}},$$

the last inequality being a consequence of Theorem 3.4. Let $\varepsilon > 0$ and let $n_k := \lfloor k^{2/\varepsilon} \rfloor$. Then

$$\sum_k \mathbf{P}\left\{\max_{n_k \leq j \leq n_{k+1}} Z_j^a \geq n_k^{-(a/2)+\varepsilon}\right\} < \infty.$$

By the Borel–Cantelli lemma, almost surely for all large k , $\max_{n_k \leq j \leq n_{k+1}} Z_j < n_k^{-(1/2)+(\varepsilon/a)}$. Since $\frac{\varepsilon}{a}$ can be arbitrarily small, this yields the desired upper bound: $Z_n \leq n^{-(1/2)+o(1)}$ a.s. \square

Proof of Theorem 3.2 and (3.21)–(3.22) of Theorem 3.3: lower bounds. To prove the lower bound in (3.21)–(3.22), we use the Paley–Zygmund inequality and Theorem 3.5, to see that

$$\mathbf{P}\{Z_{n,\beta} > n^{-(3\beta/2)+o(1)}\} \geq n^{o(1)}, \quad n \rightarrow \infty. \quad (3.30)$$

Let $\varepsilon > 0$ and let $\tau_n := \inf\{k \geq 1 : \#\{x : |x| = k\} \geq n^{2\varepsilon}\}$. Then

$$\begin{aligned} & \mathbf{P}\left\{\tau_n < \infty, \min_{k \in [\frac{n}{2}, n]} Z_{k+\tau_n, \beta} \leq n^{-(3\beta/2)-\varepsilon} \exp[-\beta \max_{|x|=\tau_n} V(x)]\right\} \\ & \leq \sum_{k \in [\frac{n}{2}, n]} \mathbf{P}\left\{\tau_n < \infty, Z_{k+\tau_n, \beta} \leq n^{-(3\beta/2)-\varepsilon} \exp[-\beta \max_{|x|=\tau_n} V(x)]\right\} \\ & \leq \sum_{k \in [\frac{n}{2}, n]} \left(\mathbf{P}\left\{Z_{k,\beta} \leq n^{-(3\beta/2)-\varepsilon}\right\}\right)^{\lfloor n^{2\varepsilon} \rfloor}, \end{aligned}$$

which, according to (3.30), is bounded by $n \exp(-n^{-\varepsilon} \lfloor n^{2\varepsilon} \rfloor)$ (for all sufficiently large n), thus summable in n . By the Borel–Cantelli lemma, almost surely for all sufficiently large n , we have either $\tau_n = \infty$, or $\min_{k \in [\frac{n}{2}, n]} Z_{k+\tau_n, \beta} > n^{-(3\beta/2)-\varepsilon} \exp[-\beta \max_{|x|=\tau_n} V(x)]$. On the set of non-extinction, we have $\tau_n \sim \frac{2\varepsilon \log n}{\log m}$ a.s., $n \rightarrow \infty$ (a consequence of the Kesten–Stigum theorem in Section 1), therefore $Z_{n,\beta} \geq \min_{k \in [\frac{n}{2}, n]} Z_{k+\tau_n, \beta}$ for all sufficiently large n . Since $\frac{1}{\ell} \max_{|x|=\ell} V(x)$ converges a.s. to a constant on the set of non-extinction when $\ell \rightarrow \infty$ (law of large numbers in Section 2), this readily yields lower bound in (3.21) and (3.22): on the set of non-extinction, $Z_{n,\beta} \geq n^{-(3\beta/2)+o(1)}$ almost surely (and a fortiori, in probability).

The proof of the lower bound in Theorem 3.2 is along exactly the same lines, but using Theorem 3.4 instead of Theorem 3.5. \square

3.6. Central limit theorem: the $\frac{3}{2}$ limit

We prove parts (3.14) and (3.16) of Theorem 3.1.

Assume (3.10), (3.11) and (3.12). Let $\beta > 1$. We trivially have $Z_{n,\beta} \leq Z_n \exp\{-(\beta - 1) \inf_{|x|=n} V(x)\}$ and $Z_{n,\beta} \geq \exp\{-\beta \inf_{|x|=n} V(x)\}$. Therefore, $\frac{1}{\beta} \log \frac{1}{Z_{n,\beta}} \leq \inf_{|x|=n} V(x) \leq \frac{1}{\beta-1} \log \frac{Z_n}{Z_{n,\beta}}$ on the set of non-extinction. Since β can be as large as possible, by means of Theorem 3.2 and of parts (3.21)–(3.22) of Theorem 3.3, we immediately get (3.14) and (3.16).

3.7. Central limit theorem: the $\frac{1}{2}$ limit

This subsection is devoted to the proof of (3.15) of Theorem 3.1.

Since $\inf_{|x|=n} V(x) \geq \log \frac{1}{Z_n}$ (on the set of non-extinction), it follows from (the upper bound in) Theorem 3.2 that, on the set of non-extinction,

$$\liminf_{n \rightarrow \infty} \frac{1}{\log n} \inf_{|x|=n} V(x) \geq \frac{1}{2}, \quad \text{a.s.}$$

It remains to prove the upper bound. We fix $-\infty < a < b < \infty$ and $\varepsilon > 0$. Let $\ell \geq 1$ be an integer. Consider $n \in [\ell, 2\ell] \cap \mathbb{Z}$. Fix a small constant $c > 0$, and let A_n be the set of $|x| = n$ such that

$$\begin{aligned} a \log \ell &\leq V(x) \leq b \log \ell, \\ V(x_k) &\geq \min\{ck^{1/3}, c(n-k)^{1/3} + a \log \ell\}, \quad \forall 0 \leq k \leq n, \end{aligned}$$

where $x_0 := \emptyset, x_1, \dots, x_n := x$ are the vertices on the shortest path relating the root \emptyset and the vertex x , with $|x_k| = k$ for any $0 \leq k \leq n$. We consider the measurable event

$$E_\ell := \bigcup_{n=\ell}^{2\ell} \bigcup_{|x|=n} \{x \in A_n\}.$$

After some elementary and tedious computations using size-biased branching random walks, we arrive at: for any $\varepsilon > 0$ and all sufficiently large ℓ ,

$$\frac{\mathbf{E}[(\#E_\ell)^2]}{[\mathbf{E}(\#E_\ell)]^2} \leq \ell^{b-a+\varepsilon} + \ell^{b-2a+(1/2)+\varepsilon}.$$

For any $b > \frac{1}{2}$, we can choose $a > \frac{1}{2}$ as close to b as possible. By the Cauchy–Schwarz inequality, $\mathbf{P}\{E_\ell \neq \emptyset\} \geq \frac{[\mathbf{E}(\#E_\ell)]^2}{\mathbf{E}[(\#E_\ell)^2]}$; thus for any $b > \frac{1}{2}$, any $\varepsilon > 0$ and all sufficiently large ℓ ,

$$\mathbf{P}\left\{ \min_{\ell \leq |x| \leq 2\ell} V(x) \leq b \log \ell \right\} \geq \ell^{-\varepsilon}.$$

This is the analogue of (3.30). From here, we can use the same argument as in Subsection 3.5 to see that, on the set of non-extinction,

$$\liminf_{n \rightarrow \infty} \frac{1}{\log n} \inf_{|x|=n} V(x) \leq \frac{1}{2}, \quad \text{a.s.}$$

This completes the proof of (3.15). □

3.8. Partition function: exponent $-\frac{\beta}{2}$

We prove part (3.20) in Theorem 3.3.

The upper bound follows from Theorem 3.2 and the elementary inequality $Z_{n,\beta} \leq Z_n^\beta$, the lower bound from (3.15) and the relation $Z_{n,\beta} \geq \exp\{-\beta \inf_{|x|=n} V(x)\}$. \square

3.9. A pathological case

In Subsection 3.1, the central limit theorem says that as long as there exists $t^* > 0$ such that $t^*\psi'(t^*) = \psi(t^*)$, $\inf_{|x|=n} V(x) - \gamma n$ is of order $\log n$, where $\gamma := \lim_{n \rightarrow \infty} \frac{1}{n} \inf_{|x|=n} V(x)$ a.s. on the set of non-extinction. We also mentioned that the equation $t^*\psi'(t^*) = \psi(t^*)$ does not always have a solution.

If the equation fails to have a solution, $\inf_{|x|=n} V(x) - \gamma n$ can indeed have pathological behaviours. Here is a simple example.

Recall that the distribution of a branching random walk is governed by a point process Θ . Let $p_i := \mathbf{P}(\#\Theta = i)$, $i \geq 0$. We consider the example that conditional on $\#\Theta = i$, Θ consists of i independent Bernoulli(p) random variables, where $p \in (0, 1)$ is a fixed constant. In this case, the law of large numbers reads as follows: on the set of non-extinction,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \inf_{|x|=n} V(x) = \gamma, \quad \text{a.s.},$$

where²¹ $\gamma = 0$ if $p \leq \frac{1}{m}$, while for $p > \frac{1}{m}$, γ is the unique solution in $(0, 1)$ of

$$v \log\left(\frac{v}{1-v}\right) + (1-v) \log\left(\frac{1-v}{q}\right) - \log m = 0.$$

It is easily checked in this example that the equation $t^*\psi'(t^*) = \psi(t^*)$ has a solution if and only if $p > \frac{1}{m}$.

Bramson [14] goes much further than this. He proves that if $p = \frac{1}{m}$, then on the set of non-extinction,

$$\lim_{n \rightarrow \infty} \frac{1}{\log \log n} \inf_{|x|=n} V(x) = \frac{1}{\log m}, \quad \text{a.s.}$$

Therefore, the branching random walk is much slower in this case than when the equation $t^*\psi'(t^*) = \psi(t^*)$ has a solution.

If $p > \frac{1}{m}$, the branching random walk becomes the slowest possible: on the set of non-extinction,

$$\lim_{n \rightarrow \infty} \inf_{|x|=n} V(x) < \infty, \quad \text{a.s.}$$

This is proved by Bramson [14]; see also Theorem 11.2 of Révész [48].

3.10. The Seneta–Heyde norming for the branching random walk

In Subsection 2.4, the Biggins martingale convergence theorem says that the limit of the martingale $W_n(\beta)$ does not vanish (on the set of non-extinction) if and only if $\mathbf{E}[W_1(\beta) \log^+ W_1(\beta)] < \infty$ and $\beta\psi'(\beta) < \psi(\beta)$.

²¹As before, $m := \sum_{i=0}^{\infty} i p_i$.

A natural question is whether it is possible to find an appropriate normalisation (to get a non-degenerate limit) when the condition fails.

When $\beta = 0$, only a Galton–Watson process is involved, for which the Seneta–Heyde theorem (Theorem 1.12) gives an affirmative answer.

To treat the case $\beta \neq 0$, we can assume, without loss of generality, that²² $\beta = 1$, and that²³ $\psi(1) = 0$. We write $Z_n := W_n(1) = \sum_{|x|=1} e^{-V(x)}$.

When $\psi'(1) < 0$ (assuming it exists and is finite), a beautiful theorem of Biggins and Kyprianou [10] tells us that there exists a sequence of positive constants (c_n) such that $\frac{Z_n}{c_n}$ converges in probability to a random variable, which is (strictly) positive on the set of non-extinction.

A few seconds of thought convince us that the case $\psi'(1) > 0$ boils down²⁴ to the case $\psi'(1) = 0$. It was an open question of Biggins and Kyprianou [11] to study the Seneta–Heyde problem for the case $\psi(1) = \psi'(1) = 0$. We have the following (partial) answer.

Theorem 3.6. *Under the assumptions of Theorem 3.1, there exists a deterministic positive sequence (λ_n) with $0 < \liminf_{n \rightarrow \infty} \frac{\lambda_n}{n^{1/2}} \leq \limsup_{n \rightarrow \infty} \frac{\lambda_n}{n^{1/2}} < \infty$, such that conditional on non-extinction, $\lambda_n Z_n$ converges in distribution to \mathcal{L} , with $\mathcal{L} > 0$ a.s.*

We have not been able to work out whether convergence in distribution can be strengthened into convergence in probability. The (conditional) limit distribution \mathcal{L} , on the other hand, is explicitly known, and is connected to Mandelbrot’s multiplicative cascades.

The proof of this theorem, relying on Theorem 3.4, can be found in [32].

3.11. Branching random walks with selection, I

Historically, the model of the Galton–Watson process was introduced by Galton and Watson in 1874 to study the survival probability of distinguished or ordinary families. In the supercritical case, the number of living members in a family grows exponentially rapidly; they may saturate the environment and compete with one another. It looks therefore quite natural to impose a criterion of selection at each generation. Such criteria have, indeed, been introduced in problems in physics ([16], [17], [18], [19], [21], [52]), in probability ([12], [27], [28], [34]), or in computer science ([39], [46]).

To avoid trivial situations, we assume in this subsection that $p_0 = 0$; thus the system survives with probability one.

²²Otherwise, we only need to replace $V(\cdot)$ by $\frac{1}{\beta}V(\cdot)$.

²³Otherwise, replace $V(\cdot)$ by $V(\cdot) - \psi(1)$.

²⁴In which case, there exists $t^* \in (0, 1)$ such that $t^* \psi'(t^*) - \psi(t^*) = 0$, so that the discussions in Remark (ii) after Theorem 3.1 confirm that it boils down to the case $\psi'(1) = 0$.

Let $N \geq 1$ be an integer. Following Brunet and Derrida [16], we keep at each generation only the N left-most particles if there are more than N , and remove from the system all other particles as well as their offspring. The resulting new system is called a branching random walk with selection.

We refer to Brunet et al. [19] for a list of no less than 23 references, as well as for an explanation of the relation with noisy Fisher–Kolmogorov–Petrovskii–Piscounov travelling wave equations.

Our goal here is to compare the new system with the original branching random walk (without selection) when N goes to infinity. Let $(V(x), |x| = n)$ denote the positions of particles at generation n in the original branching random walk (without selection), and let $(V_N(x), |x| = n)$ denote those at generation n in the new system. By the law of large numbers in Section 2,

$$\frac{1}{n} \inf_{|x|=n} V(x) \rightarrow \gamma, \quad \text{a.s.}$$

For the new system with selection, assuming that

$$\frac{1}{n} \inf_{|x|=n} V_N(x) \rightarrow \gamma_N, \quad \text{a.s.}, \quad (3.31)$$

for some constant $\gamma_N \in \mathbb{R}$, then it is clear that $\gamma_N \geq \gamma$. The basic question is whether $\gamma_N \rightarrow \gamma$ when N goes to infinity, and if it is the case, what the rate of convergence is.

The last part of the question turns out to be very challenging.

For continuous-time branching Brownian motion, it is conjectured by Brunet and Derrida [16] that

$$\lim_{N \rightarrow \infty} (\log N)^2 (\gamma_N - \gamma) = c \in (0, \infty), \quad (3.32)$$

for some constant c . Numerical simulations presented in Brunet and Derrida [16]–[17] seem to support the conjecture.

The conjecture turns out to be true.

Theorem 3.7. (Bérard and Guéré [6]) *Under suitable general assumptions, the limit γ_N in (3.31) is well-defined, and the conjecture in (3.32) holds true.*

The “suitable general assumptions” in Theorem 3.7 are slightly stronger²⁵ than those in Theorem 3.8 (see next subsection), except that the branching random walk $(V(x))$ here plays the role of the branching random walk $(-U(x))$ in next subsection. Indeed, the proof of Theorem 3.7 relies on the proof of Theorem 3.8. The value of the constant c in (3.32) is also determined. For more details, see Bérard and Guéré [6].

²⁵More precisely, an additional one-sided uniform ellipticity condition is assumed here.

3.12. Branching random walks with selection, II

Let \mathbb{T}_{bs} be a binary tree (“bs” for binary search), rooted at \emptyset . We assign i.i.d. Bernoulli(p) random variables on each edge of \mathbb{T}_{bs} , where $p \in (0, \frac{1}{2})$ is a fixed parameter. Let $(U_{\text{bs}}(x), x \in \mathbb{T}_{\text{bs}})$ denote the associated branching random walk. Recall from the law of large numbers in Section 2 that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \max_{|x|=n} U_{\text{bs}}(x) = \gamma_{\text{bs}}, \quad \text{a.s.},$$

where the constant $\gamma_{\text{bs}} = \gamma_{\text{bs}}(p) \in (0, 1)$ is the unique solution of

$$\gamma_{\text{bs}} \log \frac{\gamma_{\text{bs}}}{p} + (1 - \gamma_{\text{bs}}) \log \frac{1 - \gamma_{\text{bs}}}{1 - p} - \log 2 = 0. \quad (3.33)$$

For any $\varepsilon > 0$, let $\varrho_{\text{bs}}(\varepsilon, p)$ denote the probability that there exists an infinite ray²⁶ $\{\emptyset =: x_0, x_1, x_2, \dots\}$ such that $U_{\text{bs}}(x_j) \geq (\gamma_{\text{bs}} - \varepsilon)j$ for any $j \geq 1$. It is conjectured by Pemantle [46] that there exists a constant $\beta_{\text{bs}}(p)$ such that²⁷

$$\log \varrho_{\text{bs}}(\varepsilon, p) \sim -\frac{\beta_{\text{bs}}(p)}{\varepsilon^{1/2}}, \quad \varepsilon \rightarrow 0. \quad (3.34)$$

We prove the conjecture, and give the value of $\beta_{\text{bs}}(p)$. Let $\psi_{\text{bs}}(t) := \log[2(pe^t + 1 - p)]$, $t > 0$. Let $t^* = t^*(p) > 0$ be the unique solution of $\psi_{\text{bs}}(t^*) = t^* \psi'_{\text{bs}}(t^*)$. [One can then check that the solution of equation (3.33) is $\gamma_{\text{bs}} = \frac{\psi_{\text{bs}}(t^*)}{t^*}$.] Our main result, Theorem 3.8 below, implies that conjecture (3.34) holds, with

$$\beta_{\text{bs}}(p) := \frac{\pi}{2^{1/2}} [t^* \psi''(t^*)]^{1/2}.$$

We consider a general branching random walk, and denote by $(U(x), |x| = n)$ the positions of the particles in the n -th generation, and by $Z_n := \sum_{|x|=n} 1$ the number of particles in the n -th generation.

We assume that for some $\delta > 0$,

$$\mathbf{E}(Z_1^{1+\delta}) < \infty, \quad \mathbf{E}(Z_1) > 1; \quad (3.35)$$

in particular, the Galton–Watson process $(Z_n, n \geq 0)$ is supercritical. We also assume that there exists $\delta_+ > 0$ such that

$$\mathbf{E}\left(\sum_{|x|=1} e^{\delta_+ U(x)}\right) < \infty. \quad (3.36)$$

An additional assumption is needed (which, in Pemantle’s problem, corresponds to the condition $p < \frac{1}{2}$). Let us define the logarithmic generating function for the branching walk:

$$\psi_U(t) := \log \mathbf{E}\left(\sum_{|x|=1} e^{tU(x)}\right), \quad t > 0. \quad (3.37)$$

²⁶By an infinite ray, we mean that each x_j is the parent of x_{j+1} .

²⁷By $a(\varepsilon) \sim b(\varepsilon)$, $\varepsilon \rightarrow 0$, we mean $\lim_{\varepsilon \rightarrow 0} \frac{a(\varepsilon)}{b(\varepsilon)} = 1$.

Let $\zeta := \sup\{t : \psi_U(t) < \infty\}$. Under condition (3.36), we have $0 < \zeta \leq \infty$, and ψ_U is C^∞ on $(0, \zeta)$. We assume that there exists $t^* \in (0, \zeta)$ such that

$$\psi_U(t^*) = t^* \psi'_U(t^*). \quad (3.38)$$

By the law of large numbers in Section 2, on the set of non-extinction,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \max_{|x|=n} U(x) = \gamma_U, \quad \text{a.s.}, \quad (3.39)$$

where $\gamma_U := \frac{\psi_U(t^*)}{t^*}$ is a constant, with t^* and $\psi_U(\cdot)$ defined in (3.38) and (3.37), respectively.

For $\varepsilon > 0$, let $\varrho_U(\varepsilon)$ denote the probability that there exists an infinite ray $\{\emptyset =: x_0, x_1, x_2, \dots\}$ such that $U(x_j) \geq (\gamma_U - \varepsilon)j$ for any $j \geq 1$. Our main result is as follows.

Theorem 3.8. *Assume (3.35) and (3.36). If (3.38) holds, then*

$$\log \varrho_U(\varepsilon) \sim -\frac{\pi}{(2\varepsilon)^{1/2}} [t^* \psi''_U(t^*)]^{1/2}, \quad \varepsilon \rightarrow 0, \quad (3.40)$$

where t^* and ψ_U are as in (3.38) and (3.37), respectively.

Since $(U(x), |x| = 1)$ is not a deterministic set (excluded by the combination of (3.38) and (3.35)), the function ψ_U is strictly convex on $(0, \zeta)$. In particular, we have $0 < \psi''_U(t^*) < \infty$.

We define

$$V(x) := -t^*U(x) + \psi_U(t^*). \quad (3.41)$$

Then

$$\mathbf{E} \left(\sum_{|x|=1} e^{-V(x)} \right) = 1, \quad \mathbf{E} \left(\sum_{|x|=1} V(x) e^{-V(x)} \right) = 0. \quad (3.42)$$

The new branching random walk $(V(x))$ satisfies $\lim_{n \rightarrow \infty} \frac{1}{n} \inf_{|x|=n} V(x) = 0$ a.s. conditional on non-extinction. Define $\varrho(\varepsilon) = \varrho(V, \varepsilon)$ by

$$\varrho(\varepsilon) := \mathbf{P} \left\{ \exists \text{ infinite ray } \{\emptyset =: x_0, x_1, x_2, \dots\} : V(x_j) \leq \varepsilon j, \forall j \geq 1 \right\}.$$

Theorem 3.8 is equivalent to the following: assuming (3.42), then

$$\log \varrho(\varepsilon) \sim -\frac{\pi\sigma}{(2\varepsilon)^{1/2}}, \quad \varepsilon \rightarrow 0, \quad (3.43)$$

where the constant σ is defined by

$$\sigma^2 := \mathbf{E} \left[\sum_{|x|=1} V(x)^2 e^{-V(x)} \right] = (t^*)^2 \psi''_U(t^*).$$

The proof of (3.43) relies on size-biased branching random walks described in Subsection 2.4 and on some ideas of Kesten [34] concerning branching Brownian motion with an absorbing wall. Details of the proof are in [24].

3.13. Notes

Convergence in probability for the partition function in Theorems 3.2 and 3.3 was first proved by Derrida and Spohn [22] for continuous-time branching Brownian motion.

The pathological Bernoulli example studied in Subsections 3.9 and 3.11 is a particular example borrowed from Bramson [14], who investigates a large class of pathological cases such that the displacements of the particles are i.i.d. and non-negative, with an atom at 0 which equals $\frac{1}{m}$.

Apart from the conjecture mentioned in Subsection 3.11, Brunet and Derrida have presented in [16] many other interesting conjectures for branching random walks with selection.

4. SOLUTION TO THE EXERCISES

Exercise 1.3. From the identity $f_n(s) = f(f_{n-1}(s))$, we obtain $f'_n(s) = f'(f_{n-1}(s))f'_{n-1}(s)$. Thus $f''_n(1) = f''(f_{n-1}(1))f'_{n-1}(1)^2 + f'(f_{n-1}(1))f''_{n-1}(1) = f''(1)f'_{n-1}(1)^2 + f'(1)f''_{n-1}(1)$. Since $f'(1) = m$ and $f'_{n-1}(1) = \mathbf{E}(Z_{n-1}) = m^{n-1}$, this leads to: $f''_n(1) = m^{2n-1} + mf''_{n-1}(1)$.

Solving this by induction, we obtain: $f''_n(1) = \sum_{i=n+1}^{2n-1} m^i + m^{n-1}f''(1)$. Since $f''(1) = \sum_{i=0}^{\infty} i(i-1)p_i = \mathbf{E}(Z_1^2) - \mathbf{E}(Z_1) = \sigma^2 + m^2 - m$ (where $\sigma^2 := \text{Var}(Z_1) < \infty$), we arrive at: $\mathbf{E}(Z_n^2) = f''_n(1) + \mathbf{E}(Z_n) = \sum_{i=n+1}^{2n-1} m^i + m^{n-1}(\sigma^2 + m^2 - m) + m^n$. In particular, $\sup_n \mathbf{E}(W_n^2) < \infty$. \square

Exercise 1.4. By Exercise 1.3, $\sup_n \mathbf{E}(W_n^2) < \infty$, which means (W_n) is a martingale bounded in L^2 (and a fortiori uniformly integrable). As a consequence, W_n converges in L^2 to a limit, which is W : $\mathbf{E}(W^2) = \lim \uparrow \mathbf{E}(W_n^2) > 0$, and $\mathbf{E}(W) = 1$.

It remains to check that $\mathbf{P}(W = 0) = q$. Let $q' := \mathbf{P}(W = 0)$; so $q' < 1$ because $\mathbf{E}(W) = 1$. By conditioning on the value of Z_1 , we obtain: $q' = \sum_{i=0}^{\infty} p_i \mathbf{P}(W = 0 | Z_1 = i) = \sum_{i=0}^{\infty} p_i (q')^i = f(q')$. Since q is the unique solution of $f(s) = s$ for $s \in [0, 1)$ in the supercritical case, we have $q' = q$. \square

Exercise 1.6. Without loss of generality²⁸, we may assume that A_1, A_2, \dots, A_k are elements of \mathcal{F}_n , for some n . Then

$$\widehat{\mathbf{P}}(N = k, \mathbb{T}_1 \in A_1, \dots, \mathbb{T}_k \in A_k) = \mathbf{E} \left\{ \frac{Z_n}{m^n} \mathbf{1}_{\{N=k, \mathbb{T}_1 \in A_1, \dots, \mathbb{T}_k \in A_k\}} \right\}.$$

²⁸Assume we can prove the desired identity for all n and all $A_1, A_2, \dots, A_k \in \mathcal{F}_n$. By the monotone class theorem, the identity holds true for any $A_1 \in \mathcal{F}$ and all n and all $A_2, A_3, \dots, A_k \in \mathcal{F}_n$. By the monotone class theorem again, it holds for any $A_1 \in \mathcal{F}, A_2 \in \mathcal{F}$, and all n and all $A_3, A_4, \dots, A_k \in \mathcal{F}_n$. Iterating the procedure n times completes the argument.

On the event $\{N = k\}$, we can write $Z_n = \sum_{i=1}^k Z_{n-1}^{(i)}$, where $Z_{n-1}^{(i)}$ denotes the number of individuals in the $(n-1)$ -th generation of the subtree rooted at the i -th individual in the first generation. Accordingly,

$$\widehat{\mathbf{P}}(N = k, \mathbb{T}_1 \in A_1, \dots, \mathbb{T}_k \in A_k) = \frac{1}{m^n} \mathbf{P}(N = k) \sum_{i=1}^k \mathbf{E} \left\{ Z_{n-1}^{(i)} \mathbf{1}_{\{\mathbb{T}_1 \in A_1, \dots, \mathbb{T}_k \in A_k\}} \right\}.$$

Since $\mathbf{P}(N = k) = p_k$, and

$$\mathbf{E} \left\{ Z_{n-1}^{(i)} \mathbf{1}_{\{\mathbb{T}_1 \in A_1, \dots, \mathbb{T}_k \in A_k\}} \right\} = \mathbf{E} [Z_{n-1} \mathbf{1}_{\{\mathbb{T} \in A_i\}}] \prod_{j \neq i} \mathbf{P}(A_j) = m^{n-1} \widehat{\mathbf{P}}(A_i) \prod_{j \neq i} \mathbf{P}(A_j)$$

, the desired identity follows. \square

Exercise 1.7. Part (i) follows from the law of large numbers, whereas part (ii) from the law of large numbers for i.i.d. non-negative random variables having infinite expectations (which can be easily checked by means of the Borel–Cantelli lemma). \square

Exercise 1.8. Write $\widehat{\mathbf{E}}$ for integration with respect to $\widehat{\mathbf{P}}$. Let $A \in \mathcal{F}_n$. Then $\mathbf{E}(\xi_{n+1} \mathbf{1}_{A_n}) = \widehat{\mathbf{P}}(A_n) = \mathbf{E}(\xi_n \mathbf{1}_{A_n})$. Therefore, $\mathbf{E}(\xi_{n+1} | \mathcal{F}_n) = \xi_n$, i.e., (ξ_n) is a \mathbf{P} -martingale. Since it is non-negative, it converges \mathbf{P} -almost surely to ξ , with $\xi < \infty$ \mathbf{P} -a.s. \square

Exercise 1.9. Assume first $\widehat{\mathbf{P}} \ll \mathbf{P}$ and let $\eta := \frac{d\widehat{\mathbf{P}}}{d\mathbf{P}}$. Exactly as we proved the martingale property of (ξ_n) , we see that $\xi_n = \mathbf{E}(\eta | \mathcal{F}_n)$ \mathbf{P} -a.s. Lévy’s martingale convergence theorem tells us that $\xi_n \rightarrow \eta$, \mathbf{P} -a.s. In particular, $\xi = \eta$ \mathbf{P} -a.s., so for any $A \in \mathcal{F}_\infty$, $\widehat{\mathbf{P}}(A, \xi < \infty) = \widehat{\mathbf{P}}(A, \eta < \infty) = \widehat{\mathbf{P}}(A) = \mathbf{E}(\xi \mathbf{1}_A)$, whereas $\widehat{\mathbf{P}}(A \cap \{\xi = \infty\}) = 0$: this implies (1.4).

In the general case (i.e., without assuming $\widehat{\mathbf{P}} \ll \mathbf{P}$), we use a usual trick. Let $\mathbf{Q} := \frac{1}{2}(\mathbf{P} + \widehat{\mathbf{P}})$; thus $\mathbf{P} \ll \mathbf{Q}$ and $\widehat{\mathbf{P}} \ll \mathbf{Q}$, so that we can apply what we have just proved to $r_n := \frac{d\mathbf{P}|_{\mathcal{F}_n}}{d\mathbf{Q}|_{\mathcal{F}_n}}$ and $s_n := \frac{d\widehat{\mathbf{P}}|_{\mathcal{F}_n}}{d\mathbf{Q}|_{\mathcal{F}_n}}$. Let $r := \limsup_{n \rightarrow \infty} r_n$ and $s := \limsup_{n \rightarrow \infty} s_n$. According to what we have just proved, $r_n \rightarrow r = \frac{d\mathbf{P}}{d\mathbf{Q}}$ and $s_n \rightarrow s = \frac{d\widehat{\mathbf{P}}}{d\mathbf{Q}}$, \mathbf{Q} -a.s. Since $r_n + s_n = 1$ \mathbf{Q} -a.s., we have $\mathbf{Q}(r = s = 0) = 0$. Therefore, \mathbf{Q} -almost surely,

$$\frac{s}{r} = \frac{\lim_{n \rightarrow \infty} s_n}{\lim_{n \rightarrow \infty} r_n} = \lim_{n \rightarrow \infty} \frac{s_n}{r_n} = \lim_{n \rightarrow \infty} \xi_n = \xi.$$

(In particular, $\{r = 0\} = \{\xi = \infty\}$, $\{r > 0\} = \{\xi < \infty\}$, \mathbf{Q} -a.s.)

Let $A \in \mathcal{F}_\infty$. We have

$$\widehat{\mathbf{P}}(A) = \int_A s \, d\mathbf{Q} = \int_A s \mathbf{1}_{\{r > 0\}} \, d\mathbf{Q} + \int_A s \mathbf{1}_{\{r = 0\}} \, d\mathbf{Q}.$$

Since $\int_A s \mathbf{1}_{\{r > 0\}} \, d\mathbf{Q} = \int_A r \xi \mathbf{1}_{\{\xi < \infty\}} \, d\mathbf{Q} = \mathbf{E}(\xi \mathbf{1}_{\{\xi < \infty\}} \mathbf{1}_A) = \mathbf{E}(\xi \mathbf{1}_A)$ (because we have already proved in Exercise 1.8 that $\xi < \infty$ \mathbf{P} -a.s.) and $\int_A s \mathbf{1}_{\{r = 0\}} \, d\mathbf{Q} = \int_A s \mathbf{1}_{\{\xi = \infty\}} \, d\mathbf{Q} = \widehat{\mathbf{P}}(A \cap \{\xi = \infty\})$, this yields (1.4). \square

Exercise 1.10. If $\widehat{\mathbf{P}} \ll \mathbf{P}$, then $\xi < \infty$ $\widehat{\mathbf{P}}$ -a.s. (since a \mathbf{P} -measure zero set is in this case a $\widehat{\mathbf{P}}$ -measure zero set). By (1.4), $\mathbf{E}(\xi) = 1$ (taking $A := \Omega$). Conversely, if $\mathbf{E}(\xi) = 1$, then by (1.4), $\widehat{\mathbf{P}}(\xi = \infty) = 0$ (taking $A := \Omega$ again), i.e., $\xi < \infty$, $\widehat{\mathbf{P}}$ -a.s. In particular, (1.4) becomes $\widehat{\mathbf{P}}(A) = \mathbf{E}(\xi \mathbf{1}_A)$, $\forall A \in \mathcal{F}_\infty$; thus $\widehat{\mathbf{P}} \ll \mathbf{P}$.

It remains to check (1.6). If $\widehat{\mathbf{P}} \perp \mathbf{P}$, then $\widehat{\mathbf{P}}(A^c) = 0 = \mathbf{P}(A)$ for some $A \in \mathcal{F}_\infty$, in which case (1.4) becomes $\widehat{\mathbf{P}}(A) = \widehat{\mathbf{P}}(A \cap \{\xi = \infty\})$. Since $\widehat{\mathbf{P}}(A) = 1$, this yields $\widehat{\mathbf{P}}(\xi = \infty) = 1$, i.e., $\xi = \infty$ $\widehat{\mathbf{P}}$ -a.s. Thus (1.4) becomes $\widehat{\mathbf{P}}(A) = \mathbf{E}(\xi \mathbf{1}_A) + \widehat{\mathbf{P}}(A)$, $\forall A \in \mathcal{F}_\infty$. In particular, taking $A = \Omega$ yields $\mathbf{E}(\xi) = 0$. Conversely, if $\mathbf{E}(\xi) = 0$, then (1.4) becomes $\widehat{\mathbf{P}}(A) = \widehat{\mathbf{P}}(A \cap \{\xi = \infty\})$, $\forall A \in \mathcal{F}_\infty$; thus $\xi = \infty$ $\widehat{\mathbf{P}}$ -a.s. (taking $A := \Omega$). In particular, $\widehat{\mathbf{P}}(A^c) = 0 = \mathbf{P}(A)$ for $A := \{\xi = \infty\} \in \mathcal{F}_\infty$; thus $\widehat{\mathbf{P}} \perp \mathbf{P}$ as desired. \square

Exercise 1.11. Assume $\mathbf{E}(\log^+ Y_1) = \infty$. Then by Exercise 1.7, $\limsup_{n \rightarrow \infty} \frac{\log Y_n}{n} = \infty$ a.s. Since $Z_n \geq Y_n$, it follows that for any $c > 0$, $\limsup_{n \rightarrow \infty} \frac{Z_n}{c^n} = \infty$, a.s.

Assume now $\mathbf{E}(\log^+ Y_1) < \infty$. Exercise 1.7 tells us in this case that for any $c > 0$, $\sum_k \frac{Y_k}{c^k} < \infty$ a.s.

Let \mathcal{Y} be the sigma-algebra generated by (Y_n) . Clearly,

$$\mathbf{E}(Z_{n+1} | \mathcal{F}_n, \mathcal{Y}) = mZ_n + Y_{n+1} \geq mZ_n,$$

thus $(\frac{Z_n}{m^n})$ is a sub-martingale (conditionally on \mathcal{Y}), and $\mathbf{E}(\frac{Z_n}{m^n} | \mathcal{Y}) = \sum_{k=0}^n \frac{Y_k}{m^k}$. In particular,

$$\sup_n \mathbf{E}(\frac{Z_n}{m^n} | \mathcal{Y}) < \infty$$

on the set $A := \{\sum_{k=0}^\infty \frac{Y_k}{m^k} < \infty\}$. As a consequence, on the set A , $\lim_{n \rightarrow \infty} \frac{Z_n}{m^n}$ exists and is finite²⁹. Since $\mathbf{P}(A) = 1$, the result follows. \square

Exercise 2.2. The convexity is a straightforward consequence of Hölder's inequality. The lower semi-continuity follows from Fatou's lemma. \square

Exercise 2.3. Let $\psi(\cdot)$ be as in (2.7), and let $\psi(t) := \infty$ for $t < 0$. Then $J_1(a) := -J(-a) = \sup_{t \in \mathbb{R}} [ta - \psi(t)]$; in words, J_1 is the Legendre transform of the convex function ψ . It is well-known (Rockafellar [49], Theorem 12.2) that $\psi(t) = \sup_{a \in \mathbb{R}} [ta - J_1(a)] = \sup_{a \in \mathbb{R}} [J(a) - at]$. \square

Exercise 2.4. The identity $\inf\{a \in \mathbb{R} : J(a) = 0\} = \sup\{a \in \mathbb{R} : J(a) = 0\}$ is straightforward: the only difficulty would be if $J(a) = 0$ on an interval but this cannot happen as J is non-decreasing and concave (infimum of linear thus concave functions), hence J is (strictly) increasing up to its maximum which is $\psi(0)$ (Exercise 2.3 with $t = 0$) and which is (strictly) positive.

The argument also shows that γ is the unique solution of $J(\gamma) = 0$. \square

Exercise 2.7. Recall that $J(a) = \inf_{t \geq 0} [at + \psi(t)]$. So $\psi'(0) < -a$.

²⁹We again use the fact that a sub-martingale (X_n) converges a.s. to a finite limit if $\sup_n \mathbf{E}(X_n^+) < \infty$.

If $\psi'(\beta) \neq -a, \forall \beta > 0$, then $\sup_{t \geq 0} \psi'(t) \leq -a$, so for any $b < a$ sufficiently close to a such that $J(b) > 0$, we have $\sup_{t \geq 0} \psi'(t) < -b$, thus $0 < J(b) = \inf_{t \geq 0} [bt + \psi(t)] = \lim_{t \rightarrow \infty} [bt + \psi(t)] \leq J(a) < \psi(0)$, which yields $J(a) = \lim_{t \rightarrow \infty} [at + \psi(t)] = \lim_{t \rightarrow \infty} [bt + \psi(t) + (a - b)t] = \infty$: contradiction. \square

Acknowledgments. I am grateful to the organizers of the Symposium for the kind invitation, and to my co-authors for sharing the pleasure of random climbs.

REFERENCES

- [1] Addario-Berry, L.D. and Reed, B.A. (2009). Minima in branching random walks. *Ann. Probab.* **37**, 1044–1079.
- [2] Aldous, D.J. and Bandyopadhyay, A. (2005). A survey of max-type recursive distributional equations. *Ann. Appl. Probab.* **15**, 1047–1110.
- [3] Asmussen, S. and Hering, H. (1983). *Branching Processes*. Birkhäuser, Basel.
- [4] Athreya, K.B. and Ney, P.E. (1972). *Branching Processes*. Springer, New York.
- [5] Bachmann, M. (2000). Limit theorems for the minimal position in a branching random walk with independent logconcave displacements. *Adv. Appl. Probab.* **32**, 159–176.
- [6] Bérard, J. and Gouéré. (2008+). Brunet-Derrida behavior of branching-selection particle systems on the line. *Comm. Math. Phys.* (to appear) [ArXiv:0811.2782](https://arxiv.org/abs/0811.2782)
- [7] Biggins, J.D. (1976). The first- and last-birth problems for a multitype age-dependent branching process. *Adv. Appl. Probab.* **8**, 446–459.
- [8] Biggins, J.D. (1977). Martingale convergence in the branching random walk. *J. Appl. Probab.* **14**, 25–37.
- [9] Biggins, J.D. (1977). Chernoff’s theorem in the branching random walk. *J. Appl. Probab.* **14**, 630–636.
- [10] Biggins, J.D. and Kyprianou, A.E. (1997). Seneta-Heyde norming in the branching random walk. *Ann. Probab.* **25**, 337–360.
- [11] Biggins, J.D. and Kyprianou, A.E. (2005). Fixed points of the smoothing transform: the boundary case. *Electron. J. Probab.* **10**, Paper no. 17, 609–631.
- [12] Biggins, J.D., Lubachevsky, B.D., Shwartz, A. and Weiss, A. (1991). A branching random walk with barrier. *Ann. Appl. Probab.* **1**, 573–581.
- [13] Bramson, M.D. (1978). Maximal displacement of branching Brownian motion. *Comm. Pure Appl. Math.* **31**, 531–581.
- [14] Bramson, M.D. (1978). Maximal displacement of branching random walk. *Z. Wahrsch. Gebiete* **45**, 89–108.
- [15] Bramson, M.D. and Zeitouni, O. (2009). Tightness for a family of recursion equations. *Ann. Probab.* **37**, 615–653.
- [16] Brunet, E. and Derrida, B. (1997). Shift in the velocity of a front due to a cutoff. *Phys. Rev. E* **56**, 2597–2604.
- [17] Brunet, E. and Derrida, B. (2001). Effect of microscopic noise on front propagation. *J. Statist. Phys.* **103**, 269–282.
- [18] Brunet, E., Derrida, B., Mueller, A.H. and Munier, S. (2006). A phenomenological theory giving the full statistics of the position of fluctuating pulled fronts. *Phys. Rev. E* **73**, 056126.
- [19] Brunet, E., Derrida, B., Mueller, A.H. and Munier, S. (2007). Effect of selection on ancestry: an exactly soluble case and its phenomenological generalization. *Phys. Rev. E* **76**, 041104.
- [20] Dembo, A. and Zeitouni, O. (1998). *Large Deviations Techniques and Applications*. Second edition. Springer, New York.
- [21] Derrida, B. and Simon, D. (2007). The survival probability of a branching random walk in presence of an absorbing wall. *Europhys. Lett. EPL* **78**, Art. 60006, 6 pp.
- [22] Derrida, B. and Spohn, H. (1988). Polymers on disordered trees, spin glasses, and traveling waves. *J. Statist. Phys.* **51**, 817–840.
- [23] Duquesne, T. and Le Gall, J.-F. (2002). *Random Trees, Lévy Processes and Spatial Branching Processes*. Astérisque **281**. Société Mathématique de France, Paris.

- [24] Gantert, N., Hu, Y. and Shi, Z. (2008+). Asymptotics for the survival probability in a supercritical branching random walk. *Ann. Inst. H. Poincaré Probab. Statist.* (to appear) [ArXiv:0811.0262](https://arxiv.org/abs/0811.0262)
- [25] Grey, D.R. (1980). A new look at convergence of branching processes. *Ann. Probab.* **8**, 377–380.
- [26] Hammersley, J.M. (1974). Postulates for subadditive processes. *Ann. Probab.* **2**, 652–680.
- [27] Harris, J.W. and Harris, S.C. (2007). Survival probabilities for branching Brownian motion with absorption. *Electron. Comm. Probab.* **12**, 89–100.
- [28] Harris, J.W., Harris, S.C. and Kyprianou, A.E. (2006). Further probabilistic analysis of the Fisher-Kolmogorov-Petrovskii-Piscounov equation: one sided travelling waves. *Ann. Inst. H. Poincaré Probab. Statist.* **42**, 125–145.
- [29] Harris, S.C. and Roberts, M.I. (2009). Measure changes with extinction. *Statist. Probab. Letters* **79**, 1129–1133.
- [30] Harris, T.E. (1963). *The Theory of Branching Processes*. Springer, Berlin.
- [31] Heyde, C.C. (1970). Extension of a result of Seneta for the super-critical Galton–Watson process. *Ann. Math. Statist.* **41**, 739–742.
- [32] Hu, Y. and Shi, Z. (2009). Minimal position and critical martingale convergence in branching random walks, and directed polymers on disordered trees. *Ann. Probab.* **37**, 742–789.
- [33] Kahane, J.-P. and Peyrière, J. (1976). Sur certaines martingales de Mandelbrot. *Adv. Math.* **22**, 131–145.
- [34] Kesten, H. (1978). Branching Brownian motion with absorption. *Stoch. Proc. Appl.* **37**, 9–47.
- [35] Kesten, H. and Stigum, B.P. (1966). A limit theorem for multidimensional Galton–Watson processes. *Ann. Math. Statist.* **37**, 1211–1223.
- [36] Kingman, J.F.C. (1975). The first birth problem for an age-dependent branching process. *Ann. Probab.* **3**, 790–801.
- [37] Le Gall, J.-F. (2005). Random trees and applications. *Probab. Surveys* **2**, 245–311.
- [38] Lifshits, M.A. (2008+). Some limit theorems on binary trees. (in preparation)
- [39] Lubachevsky, B.D., Shwartz, A. and Weiss, A. (1991). An analysis of rollback-based simulation. *ACM Trans. Modeling Computer Simul.* **1**, 154–193.
- [40] Lyons, R. (1997). A simple path to Biggins’ martingale convergence for branching random walk. In: *Classical and Modern Branching Processes* (Eds.: K.B. Athreya and P. Jagers). *IMA Volumes in Mathematics and its Applications* **84**, 217–221. Springer, New York.
- [41] Lyons, R., Pemantle, R. and Peres, Y. (1995). Conceptual proofs of $L \log L$ criteria for mean behavior of branching processes. *Ann. Probab.* **23**, 1125–1138.
- [42] Lyons, R. with Peres, Y. (2008+). *Probability on Trees and Networks*. Cambridge University Press. In preparation. Current version available at <http://mypage.iu.edu/~rdlyons/prbtree/prbtree.html>
- [43] McDiarmid, C. (1995). Minimal positions in a branching random walk. *Ann. Appl. Probab.* **5**, 128–139.
- [44] Neveu, J. (1986). Arbres et processus de Galton-Watson. *Ann. Inst. H. Poincaré Probab. Statist.* **22**, 199–207.
- [45] Neveu, J. (1988). Multiplicative martingales for spatial branching processes. In: *Seminar on Stochastic Processes 1987* (Eds.: E. Çinlar et al.). *Progr. Probab. Statist.* **15**, 223–242. Birkhäuser, Boston.
- [46] Pemantle, R. (2009). Search cost for a nearly optimal path in a binary tree. *Ann. Appl. Probab.* **19**, 1273–1291.
- [47] Peres, Y. (1999). Probability on trees: an introductory climb. In: *École d’Été St-Flour 1997, Lecture Notes in Mathematics* **1717**, pp. 193–280. Springer, Berlin.
- [48] Révész, P. (1994). *Random Walks with Infinitely Many Particles*. World Scientific, Singapore.
- [49] Rockafellar, R.T. (1970). *Convex Analysis*. Princeton University Press, Princeton.
- [50] Seneta, E. (1968). On recent theorems concerning the supercritical Galton–Watson process. *Ann. Math. Statist.* **39**, 2098–2102.
- [51] Seneta, E. (1970). On the supercritical Galton-Watson process with immigration. *Math. Biosci.* **7**, 9–14.
- [52] Simon, D. and Derrida, B. (2008). Quasi-stationary regime of a branching random walk in presence of an absorbing wall. *J. Statist. Phys.* **131**, 203–233.