Branching Random Walks with Selection

Patrik Gerber

Supervised by Prof. Julien Berestycki

Dissertation on a topic in Statistics
presented for MMath in Mathematics and Statistics

April 10, 2019

UNIVERSITY OF OXFORD

DEPARTMENT OF STATISTICS

Contents

1 INTRODUCTION

1.1 Branching-Selection Systems

In this essay our main objects of interest are models called Branching Random Walks with selection (N-BRWs), which the reader can think of as a cloud of particles on $\mathbb{R}$ indexed by discrete time. Loosely speaking these 'branching-selection' systems evolve according to two mechanisms:

1 **branching** Each particle gives birth to offspring whose position is scattered on the real line.

2 **selection** Out of all children, the rightmost N are selected to form the next generation.

Regular BRWs (without selection) essentially correspond to the case $N = \infty$. For the models that we care about the branching step happens independently for each particle in the population: the position of the offspring of a given particle at say $x \in \mathbb{R}$ is conditionally independent of the past of the process and the other particles' offspring's position given x. Of course one can also consider continuous time branching(-selection) systems, for example Branching Brownian Motion (BBM) which is one of the first and most natural such models to be studied. In standard/dyadic BBM the particles follow independent Brownian motions while branching independently at exponential rate 1. At the time of branching the particle in question is replaced by two particles at the same position, who continue independently on their Brownian path. To this description it is then straightforward to add the selection step, which results in the N-BBM model.

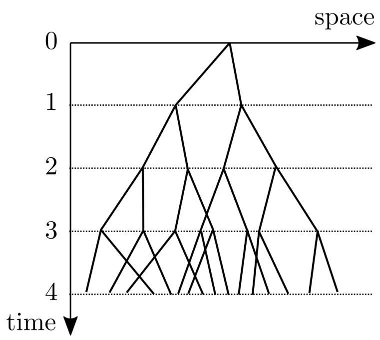

Figure 1: Sketch of a binary BRW.

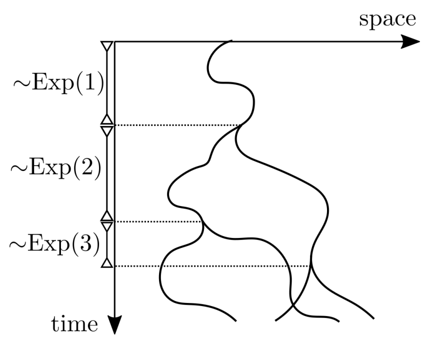

Figure 2: Sketch of a standard BBM.

Remark 1.1 (Related models). BRWs without selection and BBM have been studied extensively in the past. The book [28] gives a good overview of the results and methods employed in the study of BRWs, while Professor Berestycki's lecture notes[1] give a good introduction to BBM. The introduction of selection makes the models significantly more difficult to study. However, many related models have been introduced. One such model is the BRW/BBM killed under a linear boundary. The idea is that instead of selecting the N rightmost particles, the next generation is formed by the children whose position is greater than $v \times g$; where v is some fixed velocity and g is the current generation. Questions such as when the system survives, what speed the front propagates with and what the genealogical tree looks like are natural and have been studied for example in [18, 19, 5]. The notion of an L-BRW was introduced in [8] where only particles within distance L to the rightmost particle survive. In [11] the authors studied a model in which exactly

[1] Said lecture notes can be found at www.stats.ox.ac.uk/~berestyc/Articles/EBP18_v2.pdf

N particles are selected, but in a random manner depending on the position of the particles. What we have listed here is just a fraction of the work on problems like these, however we hope it shows the reader the general level of interest of the community in this subject.

1.2 Traveling waves and the FKPP-equation

This section relies mainly on [22] and partly on [10, Section 2] to give an introduction to the FKPP-equation and its traveling wave solutions. Let $u : [0, \infty) \times \mathbb{R} \to [0, 1]$ be sufficiently differentiable. The general form of the FKPP-equation as presented in [22] is

$$\frac{\partial}{\partial t} u(t, x) = \kappa \frac{\partial^2}{\partial x^2} u(t, x) + F(u) \tag{1.1}$$

with $\kappa > 0$ and forcing term F that satisfies

$$F(0) = F(1) = 0 \tag{1.2}$$

$$F(u) > 0, \ \forall\, u \in (0, 1) \tag{1.3}$$

$$F'(0) > 0 \text{ and } F'(u) < F'(0), \ \forall\, u \in (0, 1]. \tag{1.4}$$

Fisher [16] was the first to study a special case of (1.1): he used (1.1) with $F(u) = u(1 - u)$ to describe the spread of a favorable gene over time in a population along one dimension where the function u denoted the fraction of individuals posessing the gene at position x at time t. It turns out that (1.1) is intimately related to problems of front propagation in many problems of physics, chemistry and biology. Kolmogorov, Petrovskii and Piskunov [22] were the first to study the equation analytically and prove properties of the traveling wave solutions of the FKPP-equation. A function $w_v \in C^2$ is called a traveling wave solution if there exists $v \in \mathbb{R}$ such that $(t, x) \mapsto w(x + vt)$ solves the equation. The name 'traveling wave' is descriptive: the solution $w_v(x + vt)$ to (1.1) is just a fixed shape $w_v : \mathbb{R} \to \mathbb{R}$ that travels to the left at speed v. Remarkably, the FKPP-equation is one of the simplest equations that admits traveling wave solutions.

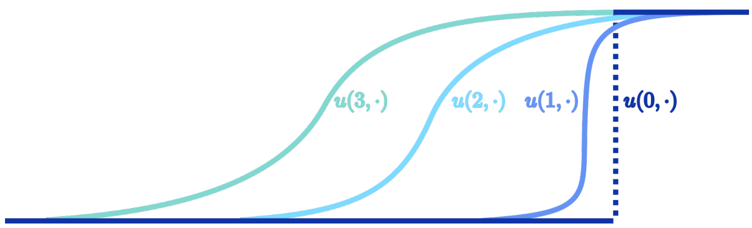

Figure 3: A typical solution to the FKPP equation with initial data $\mathbb{1}_{\{x>0\}}$ (inspired by [22]).

Suppose that we're interested in the case where the initial data is given by $u(0, x) := \mathbb{1}_{\{x>0\}}$. A sketch of a typical solution can be found in Figure 3. We see that as t grows, the shape seems to converge, while also moving to the left. The natural question to ask is then: what shape does it converge to, and what is limiting rate at which it propagates to the left. Instead of answering this straight away, let us look for a traveling wave solution w_v to the FKPP equation which satisfies $w_v(-\infty) = 0$ and $w_v(\infty) = 1$. Substituting into (1.1) we see that w_v must satisfy

$$v \frac{d}{dx} w_v(x) = \kappa \frac{d^2}{dx^2} w_v(x) + F(w_v). \tag{1.5}$$

With heuristic arguments following **??** (which can be made rigorous), we can deduce the possible values of v: for small values of $w_v(x)$ i.e. for large negative values of x we can linearise (1.5) to get

$$v\frac{d}{dx}w_v(x) = \kappa\frac{d^2}{dx^2}w_v(x) + F'(0)w_v.$$

It is easy to see that the above has a solution whose tail stays inside $[0,1]$ if and only if $v \geq 2\sqrt{\kappa F'(0)}$. Further, the left tail of w_v behaves like $e^{\gamma x}$ as $x \to -\infty$ where γ is the smaller root of $v = \gamma\kappa + F'(0)/\gamma$. The smallest speed which admits a traveling wave solution, sometimes called the critical velocity, is denoted $v_c = 2\sqrt{\kappa F'(0)}$ and the corresponding value of γ by $\gamma_c = v_c/(2\kappa)$.

Back to the problem of finding the limiting shape in Figure 3: we want to find conditions under which the solution 'selects a speed' and what that speed is. More precisely, find conditions on $u(0,x)$ under which there exists $m : [0,\infty) \to [0,\infty)$ and $v \in \mathbb{R}$ such that

$$u(t, x + m(t)) \to w_v \qquad \text{and} \qquad \frac{m(t)}{t} \to v$$

as $t \to \infty$, preferably uniformly in $x \in \mathbb{R}$. In the particular case when $u(0,x) = \mathbb{1}_{\{x>0\}}$ it turns out that

$$m(t) = v_c\, t - \frac{3}{2\gamma_c}\log t + const.$$

works and the limit shape is w_{v_c}, as shown by Bramson [7]. The above in fact holds for more general initial conditions, in particular for ones with support bounded from below.

As mentioned before, the FKPP-equation has ties to many physical problems of interest. Branching Brownian motion is a well-studied probabilistic model that have a very precise connection to the FKPP-equation, with Bramson's works such as [6, 7] being fundamental in this topic. Recall the informal description of standard Branching Brownian Motion from the previous section. For each time $t \geq 0$ let $u(t,\cdot)$ denote the distribution function of the right-most particle's position at time t. We now present a short and informal proof of the fact that $1 - u(t,-x)$ satisfies the FKPP-equation with $F(u) = u(1-u)$ and $\kappa = 1/2$, based on Professor Berestycki's lecture notes. Indeed, let $M(t)$ denote the position of the right-most particle at time t so that $u(t,x) = \mathbb{P}(M(t) \leq x)$. For $dt > 0$ consider the right-most particle over the time interval $[t, t+dt]$:

- With probability $1 - dt + o(dt)$ it does not branch
- With probability $dt + o(dt)$ it branches exactly once
- With probability $o(dt)$ it branches more than once

By the law of total probability and ignoring terms of order $o(dt)$ we find that

$$\begin{aligned}\mathbb{P}(M(t+dt) < x) &= (1-dt)\,\mathbb{P}(M(t) \leq x - B_{dt}) + dt\,\mathbb{P}(M(t) \leq x)^2 + o(dt)\\ &= \mathbb{E}\left[u(t, x - B_{dt})\right] + dt\left(u(t,x)^2 - u(t,x)\right) + o(dt).\end{aligned} \tag{1.6}$$

where $(B_t)_{t\geq 0}$ is an independent standard Brownian motion. Let $f(z) = u(t,z)$ so that $\mathbb{E}\left[u(t,x-B_{dt})\right] = \mathbb{E}_x f(B_{dt})$. Applying the Kolmogorov's backwards equation to f (assuming $u(t,\cdot)$ is twice differentiable), we have

$$\lim_{dt\downarrow 0}\frac{\mathbb{E}\left[u(t,x-B_{dt})\right] - u(t,x)}{dt} = \frac{1}{2}\frac{\partial^2}{\partial x^2}u(t,x).$$

Combining with (1.6) we get

$$\frac{\partial}{\partial t}u(t,x) = \frac{1}{2}\frac{\partial^2}{\partial x^2}u(t,x) + u(t,x)(u(t,x)-1) \tag{1.7}$$

with initial condition

$$u(0,x) = \mathbb{1}_{x\leq 0}.$$

After the transformation $\tilde{u}(t,x) = 1 - u(t,-x)$, (1.7) becomes the FKPP-equation with forcing term $F(\tilde{u}) = \tilde{u}(1-\tilde{u})$ and initial condition $\tilde{u}(0,x) = \mathbb{1}_{\{x>0\}}$ so that $\tilde{u}$ selects the critical speed $v_c = \sqrt{2}$ so that the limit shape of u propagates to the right at speed $\sqrt{2}$.

1.3 Brunet-Derrida behaviour

Assume for simplicity that $\kappa = 1$ in 1.1. In their seminal paper [9] Brunet and Derrida studied the FKPP-equation where the forcing term is multiplied by a cutoff $\mathbb{1}_{\{u\geq\epsilon\}}$ and asked the question how the critical velocity v_ϵ behaves as $\epsilon \downarrow 0$. Using non-rigorous arguments they show that v_ϵ converges to v_c, the critical velocity of the problem without a cutoff. The speed at which this convergence occurs is found to be unusually slow:

$$v_\epsilon = v_c - \frac{\pi^2\gamma_c^2 v''(\gamma_c)}{2(\log\epsilon)^2} + o(1/(\log\epsilon)^2) \tag{1.8}$$

where $v(\gamma) = \gamma + F'(0)/\gamma$. They also introduce a discrete (in both space and time) front equation (and its cutoff version) which the authors then go on to relate to a probabilistic finite particle model whose limit is governed by said equation. The model, which appears in the study of directed polymers can be described as follows. Given $N \geq 1$ and $p \in (0,1)$, at any time $n \in \mathbb{N}$ there are N particles alive at integer positions $x_1(n), ..., x_N(n)$, possibly multiple particles in the same position. For each $i \in [N]$ we choose $j_1^i, j_2^i \in [N]$ uniformly at random and set $x_i(n+1) = x_{j_1^i}(n) + b_1^i \wedge x_{j_2^i}(n) + b_2^i$ where b_1^i, b_2^i are i.i.d. $\mathrm{Ber}\,(p)$. They define $h(n,m)$ to be the fraction of particles at time n that fall above m and show in a loose sense that $h(n,m)$ is governed by the discrete front equation in the limit $N \uparrow \infty$. The remarkable observation they make is that the asymptotic velocity v_N of the stochastic model's rightmost particle converges to the critical velocity of the discrete front equation at rate $1/(\log N)^2$. Furthermore, based on the authors' large scale simulations, the constants seem to agree with their predictions of the critical speed of the discrete front equation with cutoff $\epsilon = 1/N$.

Consequently, the slow convergence rate $1/(\log N)^2$ has become known as the 'Brunet-Derrida behaviour'. Along with it came several questions and conjectures

(i) Can we understand the relationship between branching(-selection) systems and the FKPP-equation (with cutoff)?

(ii) What equations exhibit Brunet-Derrida behaviour?

(iii) What branching-selection systems exhibit Brunet-Derrida behaviour?

There has been progress towards answering (ii) in [13] where the authors prove the results of [9] rigorously, in the case $f(u) = (u - g(u)) \times \phi(u)$ for a specified class of cutoffs ϕ including

$u \mapsto \mathbb{1}_{\{u \geq \epsilon\}}$ and functions g including $u \mapsto u^2$. Another related question that emerged was analysing the noisy FKPP-equation. An analogue of Brunet-Derrida behaviour was confirmed in [25] for the equation

$$\frac{\partial u}{\partial t} = \frac{\partial u}{\partial x^2} + u(1-u) + \epsilon\sqrt{u(1-u)}\dot{W}$$

where $\dot{W}$ is 2-d white noise.

1.4 Outline

The remainder of this essay will essentially be devoted to question (iii). In [3] Bérard and Gouéré provided the first rigorous example of a branching-selection system that exhibits the Brunet-Derrida behaviour. They studied binary branching random walks where each particle gives birth to two independent offspring positioned according to some well-behaved distribution on $\mathbb{R}$. In Section 2 we give a brief introduction to branching random walks and the necessary background to discuss [3]. In Section 3 we present a full proof of the Brunet-Derrida behaviour of a generalisation of Bérard and Gouér's example. In Section 4 we discuss [4] in which the authors consider binary branching random walks, but with less well-behaved (heavy-tailed) offspring distributions. In Section 5 we give some concrete examples and visualisations of the models that we discussed thus far. Finally, in the appendix we record some original lemmas, known results and omitted proofs that we rely on in Section 3.

2 BRANCHING RANDOM WALKS

In this section we present some important results for BRWs that satisfy an exponential moment assumption. At the heart of the result is a probability change which enables us to say things about the BRW via its spine, a one-dimensional random walk. The idea was first introduced for BRWs by Lyons in [23]. In our exposition we rely on [24] and [28, Section 4.7].

2.1 Definition and notation

It will be convenient to think of BRWs and N-BRWs as stochastic processes taking values in the set $\mathfrak{M}$ of counting measures μ on $\mathbb{R}$ which put non-negative integer mass on every real number and further satisfy $\mu([x,\infty)) < \infty$ for all $x \in \mathbb{R}$. The latter condition is needed for the phrase 'rightmost particles' to be meaningful. We will write $\mathfrak{M}_N \subset \mathfrak{M}$ for measures which have total mass N and $\delta_{x_0} \in \mathfrak{M}_1$ for the unit mass at x_0. The interpretation is that if μ is the value of the (N-)BRW at some time n, then there are exactly $\mu(\{x\})$ particles at position x at time n. There is a natural partial order on $\mathfrak{M}$: we say that $\mu \preceq \nu$ if $\mu([x,\infty)) \leq \nu([x,\infty))$ for all $x \in \mathbb{R}$. For random elements (also referred to as point processes) $\mathscr{L}, \mathscr{G}$ of $\mathfrak{M}$ we say that $\mathscr{L} \preceq \mathscr{G}$ if there exists a coupling $(\mathscr{L}, \mathscr{G})$ such that $\mathscr{L} \preceq \mathscr{G}$ almost surely. In summations involving $\mathscr{L}$ we'll write $\sum_{l\in\mathscr{L}}[\cdots]$ for the sum over positions l of the particles in $\mathscr{L}$. We will write $\#\mathscr{L}$ for the total mass of $\mathscr{L}$. We'll further write $\mathscr{L}(k)$ for the random variable defined as follows: if $\mathscr{L}$ has at least k particles then set $\mathscr{L}(k)$ to be the position of the k-th particle from the right; otherwise set it equal to $\min \mathscr{L}$. We will always assume

$$1 \leq \#\mathscr{L} \qquad \text{almost surely} \qquad \text{and} \qquad 1 < \mathbb{E}\left[\#\mathscr{L}\right] < \infty. \tag{2.1}$$

The former is necessary and sufficient for the N-BRW to survive with positive probability, while the latter ensures that we avoid the trivial case of $\#\mathscr{L} \equiv 1$ where the model reduces to N independent random walks.

Definition 2.1 (N-BRW) — We call an $\mathfrak{M}$-valued Markov chain $(X_n)_{n\geq 0}$ an N-BRW driven by the point process $\mathscr{L}$ if $X_0 \in \mathfrak{M}_N$ is deterministic and given X_n, X_{n+1} is defined as follows. Each particle in X_n gives birth to children according to i.i.d. copies of $\mathscr{L}$ centred around the parent. Then, out of all children the N rightmost are chosen to form the next generation.

Not surprisingly, regular BRWs are defined similarly except no selection happens. Alternatively one can take $N = \infty$ in Definition 2.1 to get the definition of BRWs. In more mathematical terms we can construct (N-)BRWs using the notation we have introduced so far: Suppose that $\mathscr{L}$ is the point process driving the (N-)branching random walk $X := (X_n)_{n\geq 0}$ and that X is given up to time $n \geq 0$. Define

$$\widetilde{X}_{n+1} := \sum_{j=1}^{\#X_n} \sum_{l\in\mathscr{L}_{n,j}} \delta_{X_n(j)+l}. \tag{2.2}$$

If X is a regular BRW just set $X_{n+1} := \widetilde{X}_{n+1}$, in the N-BRW case set X_{n+1} to be the N rightmost particles of $\widetilde{X}_{n+1}$. This construction allows for a natural and important coupling between (N-)BRWs. This coupling was first described in [3], the way we present it here is more general and similar to [24] Lemma 4.1.

LEMMA 2.1 — *Let* $1 \le N_1 \le N_2$, $\mu_i \in \mathfrak{M}_{N_i}$ *for* $i = 1, 2$. *Take two random elements* $\mathscr{L}_i \mathfrak{M}_{N_i}$ *for* $i = 1, 2$. *If* $\{(X_n^{(i)})_{n \ge 0}\}_{i=1,2}$ *are two (N-)BRWs driven by the point processes* $(\mathscr{L}_i)_{i=1,2}$ *and are started from* $(\mu_i)_{i=1,2}$ *respectively, then there exists a coupling such that* $X_n^{(1)} \preceq X_n^{(2)}$ *almost surely for all* $n \ge 0$.

Sketch of proof. We construct the coupling inductively. The base case clearly holds. Given $X_n^{(1)} \preceq X_n^{(2)}$, independently take N_2 i.i.d. copies $\{(\mathscr{L}_i^{(1)}, \mathscr{L}_i^{(2)})\}_{i=1}^{N_2}$ of the coupling of $\mathscr{L}_1$ and $\mathscr{L}_2$ that witnesses the partial order. Using these, construct $\widetilde{X}_{n+1}^{(1)}$ and $\widetilde{X}_{n+1}^{(2)}$ as in (2.2). If the $X^{(i)}$ are regular BRWs just set $X_{n+1}^{(i)} = \widetilde{X}_{n+1}^{(i)}$, if they are N-BRWs take the rightmost N-particles like before. Either way, we have $X_{n+1}^{(1)} \preceq X_{n+1}^{(2)}$ as desired. □

Let us introduce more notation that's commonly used in the literature of BRWs. For a BRW started from a single particle at zero, denote by $\mathbb{T}$ the genealogical tree of the system and write $(V(x))_{x \in \mathbb{T}}$ for the positions of the particles on the real line. Further let $|x|$ be the generation of x and write x_i for the ancestor of x in generation i so that $x_0 = \varnothing$ where $\varnothing$ denotes the root of $\mathbb{T}$. Let $(\mathbb{T}, V)$ be a BRW started from 0 and $\mathscr{L} \in \mathfrak{M}$ be the point process that governs its evolution. It is easy to see then that $\mathbb{T}$ is a Galton-Watson tree with $\#\mathscr{L}$ as its reproduction law. Consider now the logarithmic moment generating function

$$\psi(t) := \log \mathbb{E} \int_{\mathbb{R}} e^{tx} \mathscr{L}(dx) = \log \mathbb{E} \sum_{x \in \mathbb{T}: |x| = 1} e^{tV(x)}$$

for $t \in \mathbb{R}$ where it is defined. The crucial assumption in this section is the following:

$$0 < \zeta := \sup\{t > 0 \mid \psi(t) < \infty\}. \tag{2.3}$$

This is the case that is usually studied in the classical BRW literature. The positivity of ζ ensures that the rightmost particle has an asymptotic velocity by Theorem 1.3 [28]. Note that $t \mapsto \exp \psi(t)$ is convex and $\psi(0) = \mathbb{E}[\#\mathscr{L}] \in (1, \infty)$ by (2.1) which gives $\psi(t) < \infty$ for $t \in [0, \zeta)$. Convexity also implies continuity and differentiability almost everywhere on this interval.

2.2 The Many-to-One Lemma

Suppose now that V satisfies

$$e^{\psi(1)} = \mathbb{E} \sum_{|x|=1} e^{V(x)} = 1. \tag{2.4}$$

Then we can define a random variable X by giving it's distribution function:

$$\mathbb{P}(X \le x) = \mathbb{E} \sum_{|u|=1} \mathbb{1}_{\{V(u) \le x\}} e^{V(u)}.$$

If $(S_n)_{n \ge 0}$ is a random walk with step distribution X started from $S_0 = 0$ then we have the following result:

Lemma 2.2 (Many-to-one) — *Let $(\mathbb{T}, V)$ be a BRW governed by the point process $\mathscr{L}$ which satisfies (2.4) and (2.1). Take g measurable and $n \geq 1$. Provided the integrals exist,*

$$\mathbb{E} \sum_{|x|=n} g(V(x_1), ..., V(x_n)) = \mathbb{E}\left[e^{-S_n} g(S_1, ..., S_n)\right].$$

Lemma 2.2 is a consequence of the Spinal Decomposition Theorem, one of the most important tools in the study of BRWs. For more details we refer the reader to Chapter 4 of [28] and Section 4.7 in particular. $(S_n)_{n\geq 0}$ is sometimes called the random walk associated with the BRW $(\mathbb{T}, V)$. An application of the many-to-one formula with $g(x) = xe^{-x}$ yields $\mathbb{E}X = \mathbb{E}\sum_{|x|=1} V(x)\exp(V(x))$ while for the variance we get $\mathbb{V}(X) = \mathbb{E}\sum_{|x|=1} V(x)^2 \exp(V(x)) - (\mathbb{E}X)^2$ provided that the necessary integrability conditions hold.

One might ask themselves what use the Many-to-One lemma is if it relies on assumption (2.4). However, if $\mathscr{L}$ is a point process which satisfies (2.3) and $t > 0$ is such that $\psi(t) < \infty$, then we can simply consider the point process $t\mathscr{L}$ that satisfies (2.4).

2.3 Centering the spine

Suppose that we have a BRW $(\widehat{\mathbb{T}}, \widehat{V})$ governed by $\widehat{\mathscr{L}}$ with logarithmic moment generating function $\widehat{\psi}$. Assume that it satisfies (2.3). We have seen that there is a very powerful connection between the BRW and its associated random walk via the Many-to-One lemma, which we would like to exploit. To this end it will be fruitful to have X to be centred. It follows by a simple calculation that the BRW can be transformed to have a centred spine if and only if $\exists t^* \in (0, \zeta)$ such that

$$\widehat{\psi}(t^*) = t^* \widehat{\psi}'(t^*). \tag{2.5}$$

Let $x_{max} = \sup \text{ess}\{\max \widehat{\mathscr{L}}\}$. Jaffuel gives the following characterization of when t^* exists:

Proposition 2.1 ([21, Proposition A.2]) — *Suppose $\zeta = \infty$. There exists $t^* > 0$ such that (2.5) holds if and only if*

- $x_{max} = \infty$ *or*
- $\mathbb{E}\sum_{|x|=1} \mathbb{1}_{\{x=x_{max}\}} < 1$

For a complete discussion (including the case $\zeta < \infty$) see the appendix of [21]. Let $\gamma_s : \mathscr{B}(\mathbb{R}) \to \mathscr{B}(\mathbb{R})$ to be the right shift operator on the space $\mathscr{B}(\mathbb{R})$ of Borel-measurable subsets of $\mathbb{R}$ and define

$$\mathscr{L} := (t^* \widehat{\mathscr{L}}) \circ \gamma_{-\widehat{\psi}(t^*)}. \tag{2.6}$$

If $(\mathbb{T}, V)$ is the BRW that corresponds to $\mathscr{L}$ then it is easy to see that V and its logarithmic moment generator function ψ are given by

$$V(x) = t^* \widehat{V}(x) - |x|\widehat{\psi}(t^*) \qquad \text{and} \qquad \psi(t) = \widehat{\psi}(tt^*) - t\widehat{\psi}(t^*) \tag{2.7}$$

for $x \in \mathbb{T}$. The transformation is chosen specifically so that two conditions are satisfied. First, $\psi(1) = 0$ i.e. V satisfies (2.4) and second, $\mathbb{E}X = 0$ i.e. the associated random walk is centred.

2.4 Killed below a linear boundary

Let $1 \leq m \leq \infty$ and call a sequence of vertices $(u_n)_{0\leq n\leq m}$ a path if u_i is the parent of u_{i+1} for each $0 \leq i \leq m-1$. For $v \in \mathbb{R}$ we say that the vertex $u \in \mathbb{T}$ is (m, v)-good if there exists a path $(u_n)_{0\leq n\leq m}$ with $u_0 := u$ such that $V(u_i) - V(u) \geq vi$ for all $i \in [m]$. This is essentially saying that there exists a path starting at u that stays to the right of the space-time line through $(u, V(u))$ with slope v, for at least m steps.

Let $(\mathbb{T}, V)$ be the BRW governed by $\mathscr{L}$ with logarithmic moment generating function ψ which is obtained after the transformation (2.6). We now state two results lifted from [18] that we will use in Section 3. Suppose that

$$1 < \mathbb{E}\left[(\#\mathscr{L})^{1+\delta}\right] < \infty \qquad \text{for some } \delta > 0, \tag{2.8}$$

and that there exist $\delta_- < 0 < \delta_+$ such that

$$\psi(\delta_-) < \infty \qquad \text{and} \qquad \psi(\delta_+) < \infty. \tag{2.9}$$

Let (ζ_-, ζ_+) be the interior of the domain of ψ noting that it includes both 0 and 1. Because ψ is convex and satisfies (2.9), by standard results it is C^∞ on its domain and satisfies $d^n\psi(t)/dt^n = \mathbb{E}\sum_{|x|=1} V^n(x)\exp(tV(x))$. To see this we can apply the Many-to-One lemma:

$$\psi(t) = \log \mathbb{E}\sum_{|x|=1} e^{tV(x)} = \log \mathbb{E}\left[e^{(t-1)X}\right].$$

Since 1 is in the interior of the domain of ψ we see that X must have finite moment generating function in a neighbourhood of 0. Differentiating and using basic properties of moment generating functions we obtain the required conclusion. With existence out of the way, we see that

$$\mathbb{E}X = \psi'(1) = 0 \qquad \text{and} \qquad \mathbb{V}(X) = \psi''(1) = (t^*)^2\widehat{\psi}''(t^*).$$

Under assumptions (2.8) and (2.9) the following holds.

Theorem 2.3 ([18, Theorem 1.2]) — *Let $\rho(\infty, -\epsilon)$ denote the probability that the root of $(\mathbb{T}, V)$ is $(\infty, -\epsilon) -$ good. Then, as $\epsilon > 0$ goes to zero,*

$$\log \rho(\infty, -\epsilon) = -\pi\sqrt{\frac{\mathbb{V}(X) + o(1)}{2\epsilon}}.$$

A similar result can be obtained for the probability of observing a $(m, -\epsilon)$-good root with m finite:

Theorem 2.4 ([18, Consequence of proof of Theorem 1.2]) — *Let $\rho(m, -\epsilon)$ denote the probability that the root of $(\mathbb{T}, V)$ is $(m, -\epsilon)$-good. For any $0 < \beta < \mathbb{V}(X)$, there exists $\theta > 0$ such that for all large m,*

$$\log \rho(m, -\epsilon) \leq -\pi\sqrt{\frac{\mathbb{V}(X) - \beta}{2\epsilon}}, \qquad \textit{with } \epsilon := \theta m^{-2/3}.$$

3 N-BRW WITH LIGHT TAILS

In [3] Bérard and Gouéré studied the N-BRW whose point process is given by two independent draws from a distribution p that has finite moment generating function in some neighbourhood of zero and whose spine can be centred (see 2.5). In this section we adapt their results to the more general case of N-Branching Random Walks with finite logarithmic moment generating function in a neighbourhood of zero whose spine can be centered. Loosely speaking we extend their results to the case where the number of offspring are random and their position not necessarily independent of each other. We use the methods of [3] for most proofs, adding details and adapting the arguments as necessary.

3.1 The model

Consider the BRW $(\widehat{\mathbb{T}}, \widehat{V})$ governed by the point process $\widehat{\mathscr{L}}$ with logarithmic moment generating function $\widehat{\psi}$. Assume that (2.1), (2.5) and (2.9) holds. Let $(\widehat{X}_n)_{n\geq 0}$ be the corresponding N-BRW and write $\max \widehat{X}_n$ and $\min \widehat{X}_n$ for the position of the right- and leftmost particle of $\widehat{X}_n$. We will show that

$$\hat{v}_N := \lim_{n\to\infty} n^{-1} \max \widehat{X}_n = \lim_{n\to\infty} n^{-1} \min \widehat{X}_n$$

exists in L^1 and almost surely. The main result in this section is the analogue of Theorem 1 of [3]:

Theorem 3.1 — *As N goes to infinity,*

$$\widehat{v}_N = \widehat{\psi}'(t^*) - \frac{\pi^2 t^* \widehat{\psi}''(t^*)}{2(\log N)^2} + o((\log N)^{-2}).$$

It will be more convenient to work with the BRW that is obtained after the transformation (2.6). The resulting BRW will be denoted by $(\mathbb{T}, V)$ with point process $\mathscr{L}$ and logarithmic moment generating function ψ that satisfies (2.4). Observe that the associated random walk $(S_n)_{n\geq 0}$ with step distribution X then satisfies (**??**), i.e. $\mathbb{E}X = 0$. Let $X := (X_n)_{n\geq 0}$ denote the N-BRW with point process $\mathscr{L}$ and take $X_0 \in \mathfrak{M}_N$ deterministic. Assuming they exists, $v_N := \lim_{n\to\infty} n^{-1} \max X_n$ is related to $\hat{v}_N$ by

$$v_N + \widehat{\psi}(t^*) = \hat{v}_N t^*, \tag{3.1}$$

as can be seen from (2.7). Therefore, what we will prove the following:

Theorem 3.2 (Brunet-Derrida behaviour, centred form) — *As N goes to infinity,*

$$v_N = -\frac{\pi^2 (t^*)^2 \widehat{\psi}''(t^*)}{2(\log N)^2} + o((\log N)^{-2}).$$

3.2 Properties of the model

Clearly for all $t \in \mathbb{R}$ and $k \geq 1$ we have

$$0 \leq e^{t\,\mathscr{L}_{n,j}(k)} \leq \sum_{|x|=1} e^{t\,V(x)}$$

so by assumption (2.9) the $\mathscr{L}_{n,j}(k)$ have finite moment generating function in a neighbourhood of zero which implies exponentially decaying tails. It follows by Lemma 6.1 that $\min X_n$ and $\max X_n$ are integrable and hence finite almost surely. Let $d(X_n) := \max X_n - \min X_n$ be the diameter of X_n. The following result is analogous to Corollary 1 of [3] and says that the diameter doesn't get too large:

Proposition 3.1 — *For any $N \geq 1$ and initial population $X_0 \in \mathfrak{M}_N$, we have*

$$\frac{d(X_n)}{n} \xrightarrow[n\to\infty]{a.s.,\, L^1} 0.$$

Remark 3.1. In [24] Mallein shows something stronger: he shows that there exists $C > 0$ such that for all N,

$$\mathbb{P}\left(d(X_n) > y\right) \leq C \left(\frac{N(\log N + \log y)}{y}\right)^2$$

for all sufficiently large n. Using this we can get an upper bound on $\mathbb{E}\, d(X_n)$:

$$\mathbb{E}\, d(X_n) = \int_0^\infty \mathbb{P}\left(d(X_n) > x\right) dx \leq 1 + CN^3 \int_N^\infty \left[\frac{\log x}{x}\right]^2 dx = 1 + CN^2[\log^2 N + 2\log N + 2].$$

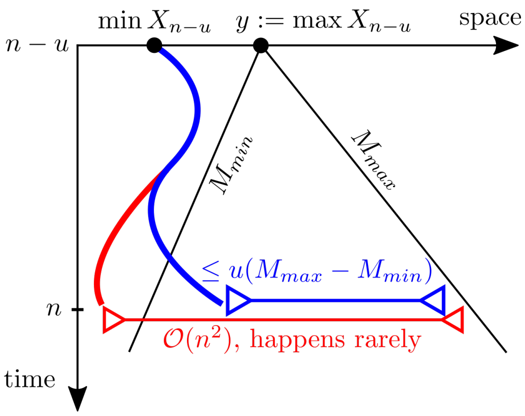

Figure 4: Proof of Proposition 3.1.

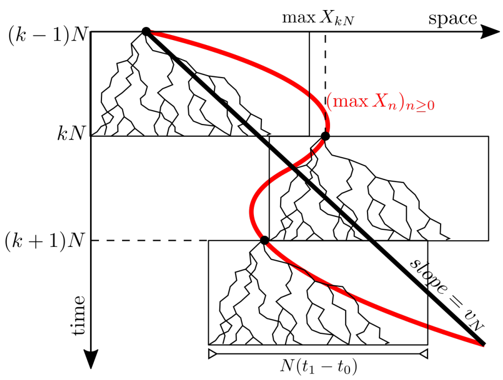

Figure 5: Alternative proof of Proposition 3.1.

Proof of Proposition 3.1. Let $u \in \mathbb{N}_+$ and for $n \geq u$ consider the process X in the timeframe $[\![n-u, n]\!]$. Define

$$M_{min} := \min\{\mathscr{L}_{i,j}(N) \mid i \in [\![n-u, n-1]\!],\, j \in [N]\}$$
$$M_{max} := \max\{\mathscr{L}_{i,j}(1) = \max \mathscr{L}_{i,j} \mid i \in [\![n-u, n-1]\!],\, j \in [N]\}$$

to be the smallest respectively largest random walk steps made between times $n-u$ and n. By Lemma 6.1 both M_{min} and M_{max} are integrable. Write $y := \max X_{n-u}$ for the rightmost particle's position at time $n-u$.

Suppose that for each $k \in [u]$ we have $\min X_{n-u+k} < y + km$. As all steps during branching are $\geq M_{min}$, this implies in particular that the descendants of the particle at space-time point

$(y, n-u)$ survive all selection steps until time n. Consider a Galton-Watson process $G := (G_n)_{n\geq 0}$ with offspring distribution $\mathscr{L}$ coupled with the descendants of $(y, n-u)$ and consider the event $A_u := \{G_u > N\}$. Since $\mathbb{E}[\#\mathscr{L}] > 1$ and $\#\mathscr{L} \geq 1$ almost surely, $\mathbb{P}(A_u) \to 1$ as $u \uparrow \infty$ where $\mathbb{P}(A_u)$ is independent of n. Since at most N descendants of y can be alive at any time, on A_u we must have $\min X_{n-u+k} \geq y + k_0 M_{min}$ for some k_0. By the definition of M_{min} this must also hold for all $k \in [\![k_0, u]\!]$, in particular for $k = u$. Noting that $\max X_n \leq y + uM_{max}$, it follows that

$$d(X_n)\mathbb{1}_{A_u} \leq u(M_{max} - M_{min}), \tag{3.2}$$

with probability one. Fix $\epsilon > 0$ and take u large enough so that $\mathbb{P}(A_u^c) < \epsilon^2$ and consider the decomposition

$$\frac{d(X_n)}{n} = \frac{d(X_n)}{n}\mathbb{1}_{A_u} + \frac{d(X_n)}{n}\mathbb{1}_{A_u^c}. \tag{3.3}$$

The reader is referred to Figure 4 to help understanding. Taking expectations and then taking n to infinity, the first term vanishes by (3.2). The second term is upper bounded by $\left(\mathbb{P}(A_u^c)\,\mathbb{E}\left[d(X_n)^2/n^2\right]\right)^{1/2}$ using Hölder's inequality. A rough bound on $d(X_n)$ suffices: $d(X_n)$ is stochastically dominated by the sum of n i.i.d. copies of $\max_{j\in[N]} \mathscr{L}_{0,j}(1) - \min_{j\in[N]} \mathscr{L}_{0,j}(N)$. Since the $\mathscr{L}_{n,j}(k)$ have exponentially decaying tails, by Lemma 6.1 this yields $\mathbb{E}\left[d(X_n)^2\right] = \mathcal{O}(n^2)$ which implies that the second term in 3.3 is $\mathcal{O}(\epsilon)$. Taking ϵ to zero concludes the proof of L^1 convergence. Almost sure convergence is a consequence of the proof of the next Proposition. □

The original proof of Bérard and Gouéré is identical to ours up to the point where we consider the Galton Watson process G. Since they only considered binary BRWs, for $u > \log_2 N$ it is clear that $\mathbb{P}(A_u) = 1$ so that there is no need to decompose, they just conclude $d(X_n) \leq u(M_{max} - M_{min}$ almost surely.

Proposition 3.2 ([3, Proposition 2]) — *There exists $v_N \in \mathbb{R}$ such that for any initial population $X_0 \in \mathfrak{M}_N$ the following holds almost surely and in L^1:*

$$\lim_{n\to\infty} \frac{\min X_n}{n} = \lim_{n\to\infty} \frac{\max X_n}{n} = v_N.$$

Proof. First we treat the case $X_0 = N\delta_0$. Recall the definition of $(\mathscr{L}_{n,j})_{n\geq 0, j\in[N]}$ from the construction of X. For each $l \geq 0$ we define the process $(X_n^r)_{n\geq 0}$ by shifting the origin of time by r. More precisely, $X_0^r = N\delta_0$ for each $r \geq 0$ and given the process up to time $n \geq 0$, X_{n+1}^r is be given by the N rightmost particles of

$$\sum_{j=1}^{N} \sum_{l\in\mathscr{L}_{r+n+1,j}} \delta_{X_n^r(j)+l}.$$

It is clear that given their initial state, $((X_n^r)_{n\geq 0})_{r\geq 0}$ are identically distributed. In particular, $(X_n^0)_{n\geq 0} = (X_n)_{n\geq 0}$ almost surely. From Lemma 2.1 it follows easily that

$$\max X_{n+m}^0 \leq \max X_n^0 + \max X_m^n \qquad \forall\, n, m \geq 0. \tag{3.4}$$

For clarity define $Z_{i,j} = \max X_{j-i}^i$ for $0 \leq i \leq j$. Then (3.4) reads $Z_{0,j} \leq Z_{0,i} + Z_{i,j}$ for all $0 \leq i \leq j$, which is familiar territory for Kingman's Subadditive Ergodic Theorem. We postpone showing that the conditions of the theorem hold to Lemma 6.5. Applying Kingman's Subadditive Theorem yields

$$\lim_{n\to\infty} n^{-1}\max X_n = \lim_{n\to\infty} \mathbb{E}\left[n^{-1}\max X_n\right] = \inf_n \mathbb{E}\left[n^{-1}\max X_n\right] = v_N \in \mathbb{R}$$

where the first limit is almost sure. Noting that the process $(-X_n)_{n\geq 0}$ also satisfies the hypothesis of the theorem by our assumptions, we can deduce from the identity $\min X_n = -\max(-X_n)$ that

$$\lim_{n\to\infty} n^{-1}\min X_n = \lim_{n\to\infty} \mathbb{E}\left[n^{-1}\min X_n\right] = \inf_n \mathbb{E}\left[n^{-1}\min X_n\right] = \tilde{v}_N \in \mathbb{R}$$

exists too, where the first limit is almost sure. From the proof of Proposition 3.1 we immediately get $\tilde{v}_N = v_N$ by uniqueness of L^1 limits, which gives $\lim_{n\to\infty} n^{-1} d(X_n) = v_N - \tilde{v}_N = 0$ almost surely as claimed in the previous proposition. The proof is complete in the case $X_0 = N\delta_0$. By translation invariance of the dynamics of the system the result also follows for initial states of the form $N\delta_{x_0}$ for any $x_0 \in \mathbb{R}$. Finally, for arbitrary $X_0 \in \mathfrak{M}_N$ note that the result is a consequence of Lemma 2.1 and a sandwiching argument between the initial states $N\delta_{\min X_0}$ and $N\delta_{\max X_0}$. □

Inspecting the previous proof, we see that the existence of v_N and $\tilde{v}_N$ (the almost sure and L^1 limits of the left- and rightmost particles) when started from $X_0 = N\delta_0$ was shown without relying on Proposition 3.1. We can in fact deduce Proposition 3.1 by an argument inspired by one of Prof. Berestycki's suggestions:

Alternative proof of Proposition 3.1. Let $Y = (Y_n)_{n\geq 0}$ be a branching random walk (without selection) with point process $\mathscr{L}$. Start Y from δ_0 noting that initially there is only one particle. Since $\mathbb{E}[\#\mathscr{L}] > 1$ by assumption, the probability ρ_1 that the number of particles in Y has reached N by time N is strictly positive. Define M_{min} and M_{max} for Y analogously to their definition in the proof of Proposition 3.1. By Lemma 6.1 one can find $t > 0$ large enough such that $\rho_2 := \mathbb{P}\left(-t \leq m_Y \leq M_Y \leq t \middle| \#Y_N \geq N\right) > 0$. We can now write

$$\begin{aligned}\mathbb{P}\left(\{\#Y_N \geq N\} \cap \{-t \leq m_Y \leq M_Y \leq t\}\right) &= \mathbb{P}\left(\#Y_N \geq N\right)\mathbb{P}\left(-t \leq m_Y \leq M_Y \leq t \middle| \#Y_N \geq N\right) \\ &\geq \rho_1\rho_2 > 0.\end{aligned}$$

Suppose that we couple X with independent copies of Y placed at the space-time points $(\max X_{kN}, kN)_{k\geq 0}$ denoted by $((Y_n^{(k)})_{0\leq n\leq N})_{k\geq 0}$. By the second Borel-Cantelli lemma it follows that almost surely infinitely many of the $(Y_n^{(k)})_{0\leq n\leq N}$ must have N particles by time N and have $-t \leq m_Y \leq M_Y \leq t$. This in turn implies that for infinitely many $k \geq 0$ the diameter is less than $2Nt$ for some time $n \in [Nk, N(k+1))$. This immediately yields $\tilde{v}_N = v_N$. □

Proposition 3.3 ([3, analogue of Proposition 3]) — *The sequence $(v_N)_{N\geq 1}$ is non-decreasing.*

Proof. This is again a consequence of Lemma 2.1. □

3.3 Brunet-Derrida behaviour

First let us describe the coupling between the N-branching random walk and N independent branching random walks which allows us apply Theorems 2.3 and 2.4. Let $(\mathrm{BRW}_j)_{j\in[N]} =$

$((\mathbb{T}_j, V_j))_{j\in[N]}$ be N independent copies of the BRW with point process $\mathscr{L}$. Define $\mathscr{T}_n := \bigsqcup_{j=1}^N \{x \in \mathbb{T}_j : |x| = n\}$ to be the disjoint union of vertices at depth n. We now inductively define a sequence $(G_n)_{n\geq 0}$ of random subsets of $\mathscr{T}_n$, each with exactly N elements. These random subsets will correspond to the particles alive in the coupled N-braching random walk at time n. Define $G_0 = \mathscr{T}_0$ and given G_n, define H_n to be the vertices in $\mathscr{T}_{n+1}$ that descend from vertices in G_n. Finally, set G_{n+1} to be the set of N vertices in H_n with the gratest value. If we now define (with some abuse of notation) $\mathfrak{X}_n = \sum_{u,j:u\in G_n\cap\mathbb{T}_j} \delta_{V_j(u)}$ then $(\mathfrak{X}_n)_{n\geq 0}$ has the same distribution as X started from $N\delta_0$.

3.3.1 Lower bound

Proof of upper bound in Theorem 3.2. As before, we first treat the case $X_0 = N\delta_0$. For notational simplicity define $\chi = \pi^2(t^*)^2\widehat{\psi}''(t^*)/2$. Our aim is to show $v_N := \lim_{n\to\infty} \mathbb{E}\left[n^{-1}\max X_n\right] \leq -\chi/(\log N)^2 + o((\log N)^{-2})$. Recalling the proof of Proposition 3.2, we know by subadditivity that

$$v_N \leq \frac{\mathbb{E}\left[\max X_n\right]}{n} \qquad \forall n \in \mathbb{N}. \tag{3.5}$$

In light of the desired correction term, we decompose (3.5) into

$$\mathbb{E}\left[\max X_n\right] \leq -\frac{\chi}{(\log N)^2} + \mathbb{E}\left[\max X_n \mathbb{1}_{\{\max X_n \geq -n\chi(\log N)^{-2}\}}\right].$$

However, this is not the exact form of the RHS that we will work with. Let $\gamma \in (0,1)$ and $\epsilon = \epsilon(N)$ whose precise form we will choose later, but will be approximately $\chi(\log N)^{-2}$ as $N \to \infty$. We go on to show that in

$$\mathbb{E}\left[\frac{\max X_n}{n}\right] \leq -(1-\gamma)\epsilon + \mathbb{E}\left[\frac{\max X_n}{n}\mathbb{1}_{\{\max X_n \geq -n(1-\gamma)\epsilon\}}\right]$$

the last term is $o((\log N)^{-2})$. This will yield the desired upper bound on v_N if we take $\gamma \to 0$. We further decompose the problem: for $\delta > 0$ we have

$$\mathbb{E}\left[\frac{\max X_n}{n}\mathbb{1}_{\{\max X_n \geq -n(1-\gamma)\epsilon\}}\right] \leq \delta\, \underbrace{\mathbb{P}\left(\frac{\max X_n}{n} \geq -(1-\gamma)\epsilon\right)}_{1} + \underbrace{\mathbb{E}\left[\frac{\max X_n}{n}\mathbb{1}_{\{\delta n \leq \max X_n\}}\right]}_{2}.$$

First we show that for an appropriate scaling of $\epsilon = \epsilon(N)$ and $n = n(N)$ we have $1 = o((\log N)^{-2})$. Set $n = \lceil N^\xi \rceil$ for some $0 < \xi < \gamma$ and $m = m(N) \leq n$ whose exact form we'll specify soon. Let B be the number of vertices in $\sqcup_{i=0}^n G_i$ that are $(m, -\epsilon)$-good with respect to their corresponding BRWs. Let $u_0, u_1, ..., u_n$ be a path in $\mathbb{T}_{i_0}$ for some $i_0 \in [N]$ such that $u_0 = root_{i_0}$ and $u_n \in G_n$ with $V_{i_0}(u_n) = \max X_n$. In other words, let BRW_{i_0} be the random walk that the rightmost particle of the coupled N-branching random walk lives in at time n, and let $u_0, ..., u_n$ be the path connecting it to the corresponding root. Define the event $E := \{\max X_n \geq -n(1-\gamma)\epsilon\}$, and apply Lemma 6.6 to the sequence of real numbers $(V_{i_0}(u_i))_{i\in[n]}$. It follows that on E, for any $K > 0$, either one of the random walk steps along the path is $> K$ or $B \geq \frac{\gamma\epsilon}{K+\epsilon}\frac{n}{m} - \frac{K}{K+\epsilon}$. This yields

$$\textcircled{1} = \mathbb{P}(E) \le \mathbb{P}(M \ge K) + \mathbb{P}\left(B \ge \frac{\gamma\epsilon}{K+\epsilon}\frac{n}{m} - \frac{K}{K+\epsilon}\right), \tag{3.6}$$

where $M := \max\{\max \mathscr{L}_{l,j} \mid l \in [\![0, n-1]\!],\ j \in [N]\}$. Set $\beta \in (0, \mathbb{V}(X))$ and let $\theta > 0$ be as in Theorem 2.4. Let $\lambda > 0$, and define

$$m := \left\lceil \left[\frac{2\theta}{\pi^2}\right]^{3/2} \left(\frac{(1+\lambda)\log N}{(\mathbb{V}(X)-\beta)^{1/2}}\right)^3 \right\rceil,$$

and $\epsilon := \theta\, m^{-2/3}$. m is carefully chosen so that by Theorem 2.4,

$$\rho(m, \epsilon) \le N^{-(1+\lambda)} \tag{3.7}$$

for all large N. Also let $K = \kappa \log N$ for $\kappa > 0$ and observe that $\frac{\gamma\epsilon}{K+\epsilon}\frac{n}{m} - \frac{K}{K+\epsilon} > 0$ for large enough N independently of κ. Thus (3.6) turns into

$$\textcircled{1} = \mathbb{P}(E) \le \underbrace{\mathbb{P}(M \ge K)}_{\textcircled{1a}} + \underbrace{\mathbb{P}(B \ge 1)}_{\textcircled{1b}}.$$

1a: As noted before, $(\max \mathscr{L}_{i,j})_{i\ge 0, j\in[N]} = (\mathscr{L}_{i,j}(1))_{i\ge 0, j\in[N]}$ are i.i.d. with common distribution $\max \mathscr{L}$ and have exponentially decaying tails so that there exists $C, \phi, K_0 > 0$ such that for all $K > K_0$ it holds that $\mathbb{P}(\max \mathscr{L} > K) \le C\exp(-\phi K)$. By a calculation similar to the proof of Lemma 6.1, for all large enough κ we get

$$\mathbb{P}(M \ge K) = 1 - (1 - \mathbb{P}(\max \mathscr{L} > K))^{Nn} \le 1 - (1 - C\exp(-\phi K))^{Nn} \tag{3.8}$$

$$= 1 - (1 - CN^{-\phi\kappa})^{Nn} \le CN^{-\phi\kappa}Nn = CN^{1+\xi-\phi\kappa}. \tag{3.9}$$

Thus, $\mathbb{P}(M \ge K) = o((\log N)^{-2})$ for $\kappa > 2/\phi$.

1b: Consider a vertex $u \in \mathbb{T}_{j_0}$ for some $j_0 \in [N]$ with $|u| \le n$. The event $\{u \in G_d\}$ is measurable with respect to the sigma algebra generated by the random variables $\{V_j(v) \mid j \in [N],\ v \in \mathbb{T}_j,\ |v| \le |u|\}$. On the other hand, the event $\{u \text{ is } (m, -\epsilon)\text{-good}\}$ is determined by the variables $\{V_{j_0}(v) - V_{j_0}(u) \mid v \in \mathbb{T}_{j_0},\ |u| < |v|\}$, so that the two events are independent. We can write B as

$$B = \sum_{i=1}^{N} \sum_{u \in \mathbb{T}_i} \mathbb{1}_{\{u \text{ is } (m,-\epsilon)\text{-good}\}} \mathbb{1}_{\{u \in G_d \text{ for some d} \le n\}}.$$

Taking expectations gives

$$\mathbb{E}[B] \le N(n+1)\rho(m,\epsilon) = \mathcal{O}(N^{\xi-\lambda}) = o((\log N)^{-2}) \qquad \text{as } N \text{ goes to infinity,}$$

where we used that G_n has N elements for all n. Applying Markov's inequality to B and combining with our estimate of 1a gives $\textcircled{1} = o((\log N)^{-2})$ as desired.

We now turn to showing $\textcircled{2} = o((\log N)^{-2})$. Consider the obvious inequality $\exp(\max X_n) \le$

$\sum_{j\in[N]}\sum_{x\in\mathbb{T}_j:|x|=n}\exp(V_j(x))$. It follows from the Many-to-One lemma that

$$\mathbb{E}\left[\exp(\max X_n)\right] \le N\,\mathbb{E}\sum_{|x|=n}\exp(V(x)) = N.$$

Now apply Lemma 6.2 with $x = \max X_n$ and $a = \delta n$ and take expectations to get

$$\mathbb{E}\left[\max X_n \mathbb{1}_{\{\max X_n \ge \delta n\}}\right] \le \mathbb{E}\left[\exp(X_n - \delta n/2)\right] \le N\,e^{-\delta n/2} = o((\log N)^{-2}).$$

This concludes the proof of the upper bound: we have shown that for any choice of $\gamma \in (0,1)$, $\beta \in (0, \mathbb{V}(X))$ and $\lambda > \xi > 0$ we have

$$v_N \le \mathbb{E}\left[\lceil N^{\xi}\rceil^{-1}\max X_{\lceil N^{\xi}\rceil}\right] \le -(1-\gamma)\frac{\pi^2(\mathbb{V}(X)-\beta)}{2(1+\lambda)^2(\log N)^2} + o((\log N)^{-2}). \tag{3.10}$$

Taking γ, β, λ and ξ to zero gives the desired result. $\square$

Here is a summary of the variables used in the previous proof, collected in a table to ease understanding

γ	$\in (0,1)$	$\mathbb{V}(X)$	$2\chi/\pi^2$
δ	$\in (0,\infty)$	β	$\in (0, \mathbb{V}(X))$
ξ	$\in (0,\gamma)$	θ	as in Theorem 2.4
κ	$\in (0,\infty)$	K	$\kappa \log N$
χ	$\pi^2(t^*)^2\widehat{\psi}''(t^*)/2$	m	$\left\lceil \left[\frac{2\theta}{\pi^2}\right]^{3/2}\left(\frac{(1+\lambda)\log N}{(\mathbb{V}(X)-\beta)^{1/2}}\right)^3\right\rceil$
λ	$\in (0,\infty)$	ϵ	$\theta\, m^{-2/3}$
n	$\lceil N^{\xi}\rceil$		

3.3.2 Upper bound

Proof of lower bound in Theorem 3.2. Let $\lambda, \eta \in (0,1)$ and let

$$\epsilon := \frac{\chi}{(1-\lambda)^2(\log N)^2}.$$

With this choice Theorem 2.3 gives $\rho(\infty, -\epsilon) = N^{\lambda-1+o(1)}$ as $N \to \infty$. Recall now the construction of $(X_n^r)_{n,l\ge 0}$ using $(\mathscr{L}_{n,j})_{n\ge 0, j\in[N]}$ from the proof of Proposition 3.2. Define the random variables $\Gamma_0 := 0$ and $J_0 := 0$ and set $i = 0$. Given Γ_i and J_i inductively define $L_{i+1} := n \wedge \inf\{k \mid \min X_k^{\Gamma_i} \ge -(1+\eta)\epsilon k\}$. Finally, let $\Gamma_{i+1} := \Gamma_i + L_{i+1}$ and $J_{i+1} := J_i + \min X^{\Gamma_i}_{L_{i+1}}$. By Lemma 2.1 it follows that

$$\min X_{\Gamma_i} \ge J_i \qquad \forall\, i \ge 0. \tag{3.11}$$

Observe now that $\Gamma_{i+1} - \Gamma_i$ is an i.i.d. sequence with common distribution $L := n \wedge \inf\{k \mid \min X_k \ge -(1+\eta)\epsilon k\}$. Similarly, the sequence $J_{i+1} - J_i$ is also i.i.d. with common distribution $\min X_L$. The strong law of large numbers gives

$$\lim_{i\to\infty}\frac{\min X_{\Gamma_i}}{i} = \lim_{i\to\infty}\frac{\min X_{\Gamma_i}}{\Gamma_i}\frac{\Gamma_i}{i} = v_N \mathbb{E}L \tag{3.12}$$

almost surely. However, by (3.11) we also have

$$\liminf_{i\to\infty}\frac{\min X_{\Gamma_i}}{i}\geq\liminf_{i\to\infty}\frac{J_i}{i}=\mathbb{E}\left[\min X_L\right]. \tag{3.13}$$

Denote $B:=\{\min X_k<-(1+\eta)\epsilon \text{ for all } k\in[n]\}$ and define the random variable Θ_n to have the law of $\min\{\min\mathscr{L}_{i,j}\mid j\in[N],\,0\leq i\leq n-1\}$. Combining the obvious inequalities $\min X_L\geq -(1+\eta)\epsilon L\mathbb{1}_{B^c}+n\Theta_n\mathbb{1}_B$ and $1\leq L\leq n$ with the above bounds we get

$$v_N\geq\frac{\mathbb{E}\left[\min X_L\right]}{\mathbb{E}L}\geq-(1+\eta)\epsilon(1+n\mathbb{P}\left(B\right))-n\mathbb{E}\left[|\Theta_n|\mathbb{1}_B\right]. \tag{3.14}$$

By Hölder's inequality $\mathbb{E}\left[|\Theta_n|\mathbb{1}_B\right]\leq\sqrt{\mathbb{E}\left[\Theta_n^2\right]\mathbb{P}\left(B\right)}$. By Lemma 6.1 we know that Θ_n has exponentially decaying tails, so in particular is square-integrable. To prove the lower bound we only need a suitably strong upper bound on $\mathbb{P}\left(B\right)$. The remainder of the proof is dedicated to this.

Since $1<\mathbb{E}\left[\#\mathscr{L}\right]=\mathbb{E}\sum_{|x|=1}\mathbb{1}$, by the monotone convergence theorem we can take $R\in\mathbb{R}$ such that $1<\mathbb{E}\sum_{|x|=1}\mathbb{1}_{\{V(x)\geq R\}}$. If we let $(M_n)_{n\geq0}$ be the Galton Watson process started from $M_0=1$ with offspring law $\sum_{|x|=1}\mathbb{1}_{\{V(x)\geq -R\}}$ then by Lemma 6.3 there exist $\phi>1$ and $r>0$ such that $\mathbb{P}\left(M_n\geq\phi^n\right)\geq r$ for all $n\geq0$. Define now $s:=\lceil\frac{\log N}{\log\phi}\rceil+1$ and $m:=\lceil\frac{-Rs}{\eta\,\epsilon}\rceil$ and finally set $n=s+m$. Consider a vertex u at depth m in one of the BRW_i. The probability that there are at least ϕ^s distinct paths $u=:u_m,...,u_n$ with $V_i(u_{k+1})-V_i(u_k)\geq R$ for all $k\in[\![m,n-1]\!]$ is greater than r by our previous discussion. Recall that the probability of the root being $(m,-\epsilon)$-good is $\rho(m,-\epsilon)$. We see that then the probability of observing a path $root=w_0,...,w_n$ in BRW_i such that $V_i(w_k)\geq-k\epsilon$ for $k\in[m]$ and $V_i(w_{k+1})-V_i(w_k)\geq R$ for $k\in[\![m,n-1]\!]$ is at least $\rho(m,-\epsilon)r$. By the choice of s and m, such a path must also be $(n,-(1+\eta)\epsilon)$-good. For $i\in[N]$ define A_j to be the event that BRW_i contains no more than ϕ^s distinct $(n,-(1+\eta)\epsilon)$-good paths starting at the root. By independence we get

$$\mathbb{P}\left(\bigcap_{i=1}^{N}A_i\right)\leq(1-\rho(m,\epsilon)r)^N. \tag{3.15}$$

On the event $B\cap[\cap_{i=1}^N A_i]^c$ one of the BRW_is has $>\phi^s>N$ particles at time n that have stayed to the right of the space time line with slope $-(1+\eta)\epsilon$ until time n. By definition, on B this implies that there are $>N$ particles alive in the N-branching random walk which is impossible. Therefore we must have $B\subset\cap_{i=1}^N A_i$. Using the fact that $\rho(m,-\epsilon)\leq\rho(\infty,-\epsilon)=N^{\lambda-1+o(1)}$ and the inequality $1-x\leq\exp(-x)$ for all $x\in\mathbb{R}$, we get

$$\mathbb{P}\left(B\right)\leq\mathbb{P}\left(\bigcap_{i=1}^{N}A_i\right)\leq\exp(-N^{\lambda+o(1)}) \tag{3.16}$$

Combining the above estimate of $\mathbb{P}\left(B\right)$ with (3.14) results in

$$v_N\geq-\frac{\chi(1+\eta)}{(1-\lambda)^2(\log N)^2}+o((\log N)^{-2}). \tag{3.17}$$

As λ and η can be taken arbitrarily small, the desired result follows. □

3.4 Discussion

This section is based in part on [3] Section 8. Write v_N for the asymptotic speed of the N-BRW and $v := \lim_{N\to\infty} v$ for its limit, to which it converges at rate $(\log N)^{-2}$. The proof of the Brunet-Deridda behaviour hinged on the idea that the following two are comparable:

(a) $(BRW_j)_{j\in[N]}$ do not survive killing below the speed $v - \epsilon$
(b) $v_N < v - \epsilon$.
It is intuitively clear why this might be true: if (a) holds then we certainly can't expect the N-BRW to propagate faster than $v - \epsilon$ since it is dominated by $(BRW_j)_{j\in[N]}$ through the coupling introduced in section 3.3. Conversely, suppose (a) doesn't hold. Consider the particle in the N-BRW at time zero that corresponds to a BRW_i that survives killing. With probability $\sim 1/N$ the descendants of this particle will survive and form the entirety of the N-BRW. If this doesn't happen we can 'restart' the argument. After a rougly geometric number of attempts we will succed in which case (b) cannot hold.

To make such an argument rigorous we needed to get a handle on the probability of a BRW exhibiting paths of various slopes and lengths near the critical slope 0; the expectation of the BRW's associated random walk. In [18] they did just that using the Many-to-One lemma and one of Mogulskii's results which accurately describes the probabilitiy of a centred random walk's path to stay in a general space-time region. Loosely speaking they show that on the scale $m \propto \epsilon^{-u}$ for $u \in (0, 3/2]$ we have

$$\log \rho(m, -\epsilon) \propto -\epsilon^{-u/3} \qquad \text{as } \epsilon \downarrow 0. \tag{3.18}$$

Similarly, they show $\log \rho(\infty, -\epsilon) \propto -\epsilon^{1/2}$ as $\epsilon \downarrow 0$. Using this the convergence rate $(\log N)^{-2}$ can be heuristically deduced: setting $\epsilon_N = v - v_N$ we would expect

$$\rho(\infty, -\epsilon_N) \propto \frac{1}{N}. \tag{3.19}$$

To see why, suppose for contradiction that $\rho(\infty, -\epsilon_N) \ll 1_N$. Then with probability close to one none of the BRW_i survive killing which would imply that v_N needs to be slower. Conversely, if we had $\rho(\infty, -\epsilon_N) \gg 1/N$ then with probability close to one at least one of the BRW_i would survive killing, implying that v_N needs to be faster. From 3.19 and the asymptotic form of $\rho(\infty, -\epsilon)$ for small ϵ it follows that for all large N, ϵ_N should satisfy

$$\epsilon_N \propto (\log N)^{-2}. \tag{3.20}$$

3.5 α-stable spine

In [24, Lemma 4.2] Mallein studies a generalisation of the model we considered in this section. Recall that if $(\mathbb{T}, V)$ is a BRW with point process $\mathscr{L}$ that satisfies (2.4) i.e.

$$e^{\psi(1)} = \mathbb{E} \sum_{|x|=1} e^{V(x)} = 1, \tag{3.21}$$

then the step distribution of the spine X is given by

$$\mathbb{P}(X \leq x) = \mathbb{E} \sum_{|u|=1} \mathbb{1}_{\{V(u) \leq x\}} e^{V(u)}. \tag{3.22}$$

In Section 3 we considered the case when X had finite mean and variance so that by the central limit theorem it belonged to the domain of attraction of the Gaussian law. Mallein puts less stringent conditions on X: he requires that X be in the domain of attraction of some stable random variable Y that veryfies $\mathbb{P}(Y \geq 0) \in (0, 1)$. The family of stable probability distributions are often referred to as α-stable distributions, signifying the dependence on the parameter $\alpha \in (0, 2]$. The case $\alpha = 2$ corresponds to the Gaussian while for $\alpha \in (0, 2)$ we get α-regularly varying distributions (for a definition see Section 4). He defines the truncated second moment function

$$L^*(x) := x^{\alpha-2} \mathbb{E}\left[Y^2 \mathbb{1}_{|Y| \leq x}\right], \tag{3.23}$$

and goes on to show that under some further technical conditions on $\mathscr{L}$ that we omit, it holds that

$$v_N \sim -C_* \frac{L^*(\log N)}{(\log N)^\alpha} \qquad \text{as } N \to \infty \tag{3.24}$$

for $C_* > 0$ depending on Y whose form he also specifies.

4 POLYNOMIAL TAILS

Placeholder text.

4.1 Overview of heavy tailed distributions

In this section we rely on sections 2 and 3 of [17]. For a random variable X with cumulative distribution function F, let $\overline{F}(x) := \mathbb{P}\,(X > x) = 1 - F(x)$ be its right tail-function. F is said to be heavy-tailed if

$$\int_{\mathbb{R}} e^{\lambda x} dF = \infty \qquad \forall\, \lambda > 0. \tag{4.1}$$

F is said to be light-tailed if it is not heavy-tailed. It is clear that if F is positive and light-tailed then it has finite moments of all orders. A positive function $f : \mathbb{R} \to \mathbb{R}_+$ is said to be heavy-tailed if

$$\limsup_{x\to\infty} e^{\lambda x} f(x) = \infty \qquad \forall\, \lambda > 0. \tag{4.2}$$

We have the following characterisation of heavy tailed distributions:

Theorem 4.1 ([17, Theorem 2.6, adapted]) — *Let F be a distribution function. The following are equivalent:*

- *F is heavy-tailed in the sense (4.1)*
- *$\overline{F}$ is heavy-tailed in the sense (4.2).*

If any of the above holds and F has density f with respect to the Lebesgue measure, then f is also heavy-tailed. The converse however is not true in general.

Let f be a positive function as before. For a positive, non-decreasing function $h : \mathbb{R} \to \mathbb{R}_+$ we say that f is h-insensitive if

$$\lim_{x\to\infty} \sup_{|y|\le h(x)} \left| \frac{f(x+y)}{f(x)} - 1 \right| = 0. \tag{4.3}$$

If f is h-insensitive for all constant functions h we say that f is long-tailed. For f to be long-tailed, h-insensitivity for some h with $h(\infty) = \infty$ is obviously sufficient (and also necessary by Lemma 2.19 [17]). Furthermore f being long-tailed implies that f is heavy-tailed by Lemma 2.17 [17]. By varying the choice of h, we can classify long-tailed functions according to the heaviness of their tails. An important subset of long-tailed functions are those of regular variation: A positive function $f : \mathbb{R} \to \mathbb{R}_+$ is said to be regularly varying at ∞ with index $\alpha \in \mathbb{R}$ if for all $c > 0$,

$$f(cx) \sim c^{\alpha} f(x) \qquad \text{as } x \to \infty. \tag{4.4}$$

Functions that are 0-regularly varying are called slowly varying. It is a well known result that if f is α-regularly varying then it can be written as $f(x) = x^{\alpha} f_0(x)$ for some slowly varying f_0. The definitions of long-tailedness and regular variation naturally extend to probability distributions F: we say F is regularly varying with index $\alpha > 0$ if $\overline{F}$ is regularly varying with index $-\alpha$. A probability distribution having long-tails has a simple probabilistic interpretation: the distribution function F is long-tailed by definition if

$$\mathbb{P}\,(\,X > x + y|\,X > x) \to 1 \qquad \text{as } x \to \infty\,, \forall\, y > 0, \tag{4.5}$$

where X has distribution F. Notice that this doesn't give uniform convergence in y on closed intervals as required by definition (4.3), however it is a standard result that the two are equivalent. Interpreting (4.5) in words, given that F exceeds some high level the probability of it exceeding an even higher level is large. Finally, we introduce the class of subexponential functions: a distribution F is called subexponential if it is long-tailed and

$$\overline{F * F}(x) \sim 2\overline{F}(x) \qquad \text{as } x \to \infty, \tag{4.6}$$

where $*$ denotes convolution. (4.6) in fact implies $\overline{F^{*(n)}} \sim n\overline{F}$ for $n \geq 1$ so that we have the following probabilistic interpretation: if $(X_i)_{1 \leq i \leq n}$ is an i.i.d. sequence with common distribution F then

$$\mathbb{P}\left(X_1 + ... + X_n > x\right) \sim \mathbb{P}\left(\max_{1 \leq i \leq n} X_i > x\right) \qquad \text{as } x \to \infty. \tag{4.7}$$

Let $\mathcal{H}, \mathcal{L}, \mathcal{S}, \mathcal{R}$ denote the class of heavy-tailed, long-tailed, subexponential and regularly varying probability distributions respectively. We have the following inclusions:

$$\mathcal{R} \subset \mathcal{S} \subset \mathcal{L} \subset \mathcal{H}, \tag{4.8}$$

where the leftmost inclusion is a consequence of Theorem 3.29 [17]. Probability distributions arising in practice virtually always fall in the class of light-tailed distributions or that of the subexponential distributions.

4.2 The result

Here we present the model as studied in [4]. Let $(X_n)_{n \geq 0}$ be the N-BRW evolving as follows: At each step each particle gives birth to two children whose position is distributed independently and identically around the position of the parent, according to the law of a random variable X. Then, out of all children the N rightmost are selected tp form the next generation where $N \geq 2$ is fixed. The authors further impose the condition $X \geq 0$ almost surely and that the random variable X be regularly varying with index $\alpha > 0$. The authors remark that the methods of their analysis could be extended to more general reproduction laws and to X taking values in $\mathbb{R}$, however since no new phenomena can be observed the don't do so. For $x \geq 0$ define h to be given by

$$\mathbb{P}\left(X > x\right) =: \overline{F}(x) =: \frac{1}{h(x)}. \tag{4.9}$$

Then our assumptions imply that h is regularly varying at infinity with index α. Define the sequence $(c_N)_{N \geq 1}$ by $c_N := h^{-1}(2N \log_2 N)$ where $h^{-1}(x) := \inf\{y : h(y) > x\}$ is the generalised inverse of h. Let $\mathcal{L}(Y)$ denote the law of the random variable Y and let $d(\cdot, \cdot)$ be the Prokhorov metric on the space of probability distributions on $\mathbb{R}$, which induces the topology of weak convergence. We now present Theorem 1.2 of [4] as it is found there.

Theorem 4.2 ([4, Theorem 1.2]) — *We distinguish the following cases:*

(a) $\alpha > 1$*: The limits* $v_N := \lim_{n\to\infty} n^{-1} \max X_n = \lim_{n\to\infty} n^{-1} \min X_n$ *exist almost surely and in* L^1 *and satisfies* $v_N \sim \rho_\alpha c_N / \log_2 N$ *as* $N \to \infty$ *where* $\rho_\alpha > 0$.

(b) $\alpha = 1$, $\mathbb{E}X < \infty$*:* v_N *exists as above and satisfies*

$$v_N \sim \frac{c_N h(c_N)}{\log_2 N} \int_1^\infty \frac{1}{h(xc_N)} dx \qquad \text{as } N \to \infty. \tag{4.10}$$

(c) $\alpha = 1$, $\mathbb{E}X = \infty$*: Let* $b_n = \int_1^{h^{-1}(n)} h(c_N)/h(c_N x)dx$*. For* $i = 1$ *and* $i = N$ *we have*

$$\lim_{N\to\infty} \limsup_{n\to\infty} d(\mathcal{L}\left(\frac{\log_2 N}{c_N} \frac{X_n(i)}{nb_n}\right), \delta_1) = 0.$$

(d) $\alpha \in (0,1)$*: Let* W_α *be the random variable whose moment generating function is given by*

$$\mathbb{E}e^{-\lambda W_\alpha} = exp(-\alpha \int_0^\infty (1 - e^{-\lambda x})x^{-\alpha-1}dx).$$

Then for $i = 1$ *and* $i = N$*,*

$$\lim_{N\to\infty} \limsup_{n\to\infty} d(\mathcal{L}\left(\frac{1}{(2N)^{1/\alpha}} \frac{X_n(i)}{h^{-1}(n)}\right), \mathcal{L}(W_\alpha)) = 0.$$

The way Bérard and Maillard show this is by considering a stochastic process which they call the 'stairs process'. As opposed to the light tailed models considered in Section 3, when the tails are heavy the cloud of particles moves mainly through large occasional jumps. The stairs process is devised so that it mimics this behaviour. They prove certain properties of these stairs processes, most importantly their asymptotic speed of propagation when suitably scaling space and time, which is the subject of Theorem 2.5 [4]. Using delicate couplings between the N-BRW and the continuous and discretized stairs processes, they obtain that $c_N^{-1} \max X_n$ converges to the stairs process in the J_1 topology while the minimum converges in the M_1.

From Theorem 4.2 it is hard to tell what v_N looks like in practice because of expressions like c_N, $h^{-1}(c_N)$. In Section 5 we present a fully worked example as well as simulations which should give us an idea of what these processes look like. From our discussion in Section 4.1 we know that Bérard and Maillard didn't cover all classes of heavy-tailed distributions, in particular examining what happens for distributions in $\mathscr{S} \setminus \mathscr{L}$ might be the subject of future research.

5 EXAMPLES

As before, let $(X_n)_{n\geq 0}$ be the N-BRW governed by the point process $\mathscr{L}$ with logarithmic moment generating function ψ. In this section we provide some concrete examples and reflect upon how they fit into the previous sections' results.

5.1 Binary Gaussian N-BRW

Suppose that $\mathscr{L} = \delta_{Y_1} + \delta_{Y_2}$ where Y_1, Y_2 are i.i.d. normal with mean μ and variance $\sigma^2 > 0$. The logarithmic moment generating function takes the form

$$\psi(t) = \log \mathbb{E}\left[e^{Y_1} + e^{Y_2}\right] = \mu t + \frac{\sigma^2 t^2}{2} + \log 2.$$

Clearly $\psi(t)$ is finite for all $t \in \mathbb{R}$, and solving for $t^* > 0$ in $\psi'(t^*)t^* = \psi(t^*)$ we get $t^* = \sqrt{\frac{2\log 2}{\sigma^2}}$. Therefore $\mathscr{L}$ satisfies the hypothesis of Theorem 3.1, and so

$$\lim_{n\to\infty} \frac{\max X_n}{n} = v_N = \mu + \sqrt{\sigma^2 \log 4} - \frac{\pi^2\sqrt{\sigma^2 \log 4}}{2(\log N)^2} + o((\log N)^{-2}).$$

In fact, we could replace the Gaussian distribution with any distribution on $\mathbb{R}$ that has finite moment generating function on all of $\mathbb{R}$ and which puts less than $1/2$ mass on its essential supremum. The resulting point process would still satisfy the hypothesis of Theorem 3.1 by Proposition 2.1.

5.2 Binary Bernoulli N-BRW

Another natural choice might be Y_1, Y_2 i.i.d. Bernoulli with parameter $\alpha \in (0,1)$. However, it turns out that the hypothesis of Theorem 3.1 is satisfied if and only if $\alpha \in (0, 1/2)$. This is because for $\alpha \geq 1/2$ the Y_i put $\geq 1/2$ mass on their essential supremum (which in this case is equal to 1).

To see this consider the following. $\psi(t) = \log 2 + \log(\alpha e^t + 1 - \alpha)$ so if $f(t) := t\psi'(t) - \psi(t)$ then $f'(t) = t\psi''(t) > 0$ for all $t > 0$. Now, as $f(0) = -\log 2 < 0$, the equation $f(t^*) = 0$ has a solution $t^* > 0$ if and only if $\lim_{t\to\infty} f(t) > 0$. We have

$$\begin{aligned}
\lim_{t\to\infty} f(t) &= \lim_{t\to\infty} \left[\frac{\alpha t e^t}{\alpha e^t + 1 - \alpha} - \log(\alpha e^t + 1 - \alpha) - \log 2\right] \\
&= \lim_{t\to\infty} \left[t\big(1 - \frac{1-\alpha}{\alpha e^t + 1 - \alpha}\big) - t - \log(1 + \frac{1-\alpha}{\alpha} e^{-t}) - \log(2\alpha)\right] \\
&= -\log(2\alpha).
\end{aligned}$$

Above expression is positive if and only if $\alpha \in (0, 1/2)$ as claimed. In [3] Bérard and Gouéré note this as well and go on to show the following

THEOREM 5.1 ([3, Theorems 4, 5]) — *For $\alpha = 1/2$ it holds that*

$$1 - v_N = \Theta(N^{-1})$$

while for $\alpha \in (1/2, 1]$ *we have*

$$-\log(1 - v_N) = \Theta(N).$$

Remark 5.1. A slightly more general case of Bernoulli N-BRWs was studied by the authors [12] using elementary methods, making it an accessible and enjoyable read.

5.3 N-BBM

Let $(B_t)_{t\geq 0}$ be a standard N-BBM (as described in Section **??**) started from $N\,\delta_0$. Clearly $(B_n)_{n\in\mathbb{N}}$ is an N-BRW, however explicitly describing its point process is not trivial. Nevertheless, its logarithmic moment generating function is easy to obtain. It is clear that $\#\mathscr{L}$ is distributed as the simple birth process $(M_t)_{t\geq 0}$ after one unit of time that is started from $M_0 = 1$ with escape rate equal to the number of particles alive. It is an elementary fact that $\mathbb{E}M_t = e^t$ so we get

$$\psi(t) = 1 + \frac{t^2}{2}.$$

Solving $\psi'(t^*)t^* = \psi(t^*)$ we obtain $t^* = \sqrt{2}$. By Theorem 3.1 we get

$$v_N := \lim_{n\to\infty} \frac{\max B_n}{n} = \sqrt{2} - \frac{\pi^2}{\sqrt{2}(\log N)^2} + o((\log N)^{-2})$$

where the limit is both in L^1 and almost surely. We can use a simple sandwiching argument to get the corresponding result for continuous time. We have

$$\frac{\max B_t}{t} = \left(\frac{\max B_t - \max B_{\lfloor t\rfloor}}{\lfloor t\rfloor} + \underbrace{\frac{\max B_{\lfloor t\rfloor}}{\lfloor t\rfloor}}_{\to v_N \ a.s.\ \&\ L^1} \right) \underbrace{\frac{\lfloor t\rfloor}{t}}_{\to 1}. \tag{5.1}$$

Fix $t \geq 0$ and let $\mathcal{P} := (P_j)_{j\in[N]}$ be an i.i.d. collection of Poisson (1) random variables. Further, let $\mathcal{Z} := (Z_t^{j,k})_{j\in\mathbb{N},\, k\geq 1,\, t\geq 0}$ be a collection of i.i.d. standard Brownian motions independent of $\mathcal{N}$. There is an obvious coupling of $(B_s)_{\lfloor t\rfloor\leq s<t}$ and the collections $\mathcal{P}$ and $\mathcal{Z}$. This coupling immediately yields the following upper bound:

$$\max B_t - \max B_{\lfloor t\rfloor} \overset{st.}{\leq} \sum_{j=1}^{N} \sum_{k=1}^{P_j} \sup_{s\in[0,1)} Z_s^{j,k}. \tag{5.2}$$

It is a well known fact that $\sup_{s\in[0,1)} Z_s$ and $|Z_1| \in L^1$ have the same distribution for a standard Brownian motion $(Z_t)_{t\geq 0}$. In the same fashion we can obtain a lower bound and we find that

$$|\max B_t - \max B_{\lfloor t\rfloor}| \overset{st.}{\leq} \mathcal{B} \tag{5.3}$$

where the distribution of $\mathcal{B} \in L^1$ is independent of t. This gives $t^{-1}\max B_t \to v_N$ in L^1 while almost sure convergence follows by the Borel-Cantelli lemma. Indeed, let $(\mathcal{B}_j)_{j\geq 0}$ be an i.i.d. collection distributed like $\mathcal{B}$. Then

$$\mathbb{P}\left(t^{-1}\max B_t \nrightarrow v_N\right) \leq \sum_{k=1}^{\infty} \mathbb{P}\left(\mathcal{B}_n > n/k \text{ for infinitely many } n \in \mathbb{N}\right) = 0,$$

since $\int_0^\infty \mathbb{P}(\mathcal{B} > \lambda)\, d\lambda = \mathbb{E}\mathcal{B} < \infty$.

5.4 N-BRW with Cauchy spine

Consider the model where $\mathscr{L}$ is given by a Poisson Point Process with intensity measure $e^{-x}\nu_\alpha(dx)$ where ν_α is some α-stable distribution with $\nu_\alpha(0,\infty) \in (0,1)$. By The Slivnyak-Mecke Theorem [2, Theorem 1.13] we have

$$\mathbb{E}\sum_{l\in\mathscr{L}} e^l = \int_{\mathbb{R}} e^x e^{-x}\nu_\alpha(dx) = 1$$

so $\mathscr{L}$ satisfies (2.4). We claim that the spine is in the domain of attraction of ν_α. Indeed, we just have to observe that

$$\mathbb{P}(X \le x) := \mathbb{E}\sum_{l\in\mathscr{L}} \mathbb{1}_{\{l\le x\}} e^l = \int_{\mathbb{R}}^{x} \nu_\alpha(d\lambda), \tag{5.4}$$

by the Slivnyak-Mecke theorem. Since stable distributions are in their own domains of attraction ([15, p. 576]) we are done. Thus, by [24, Theorem 1.1] we have

$$v_N \sim -C_* \frac{L^*(\log N)}{(\log N)^\alpha} \qquad \text{as } N \to \infty \tag{5.5}$$

for some $C_* > 0$ and $L^*(x) = x^{\alpha-2}\int_x^x \lambda^2 \nu_\alpha(d\lambda)$. Using [15, (5.18) and (5.22)] we can write (5.5) as

$$v_N \sim -C_* \frac{L^*(\log N)}{(\log N)^\alpha} \propto -\nu_\alpha(\log N, \infty). \tag{5.6}$$

Now we can deduce the asymptotic behaviour of v_N using [26, Theorem 1.12] which states that $\nu_\alpha(x) \sim c\, x^{-\alpha}$ for some $c > 0$ as $x \to \infty$. This gives

$$v_N \sim -\frac{K}{(\log N)^\alpha} \qquad \text{as } N \to \infty \tag{5.7}$$

for some $K > 0$.

5.5 Binary N-BRW with polynomial tails

As in Sections 5.1 and 5.2, let $\mathscr{L} = \delta_{Y_1} + \delta_{Y_2}$ where this time Y_1, Y_2 are i.i.d. with distribution given by

$$\mathbb{P}(Y_1 > x) = \frac{1}{x^\alpha}, \qquad \forall\, x \ge 1 \tag{5.8}$$

for some $\alpha > 0$. If $\alpha > 1$ we can use the notation of Section 4.2 to get $c_N = (2N\log_2 N)^{1/\alpha)}$ so that

$$v_N \sim \frac{\rho_\alpha c_N}{\log_2 N} \propto \left(\frac{N}{(\log N)^{\alpha-1}}\right)^{1/\alpha}$$

by Theorem 4.2. For $\alpha = 1$ we are in case (c) of Theorem 4.2 and we have $c_N = 2N\log_2 N$ and $b_N = \int_1^n dx/x = \log n + \mathcal{O}(1)$. Thus loosely speaking we get

$$\frac{\max X_n}{n\log n} \approx \delta_{2N}. \tag{5.9}$$

6 Appendix

6.1 Miscellaneous

Lemma 6.1 — *Let $\tau \in L^1$ be an $\mathbb{N}$-valued random variable and let $(\epsilon_n)_{n\geq 1}$ be an i.i.d. sequence of random variables with exponentially decaying tails, independent of τ. Then $M := \max_{1\leq n\leq\tau} \epsilon_n$ has exponentially decaying tails.*

Proof. Let $C, \gamma, t_0 > 0$ be such that $\mathbb{P}(|\epsilon_1| \leq t) \geq 1 - Ce^{-\gamma t}$ for all $t > t_0$. Then for $t > t_0$ large enough, Bernoulli's inequality gives

$$\begin{aligned} \mathbb{P}(M > t) &\leq 1 - \mathbb{E}\left[\mathbb{P}(|\epsilon_1| \leq t)^\tau\right] \leq 1 - \mathbb{E}\left[(1 - Ce^{-\gamma t})^\tau\right] \\ &\leq 1 - \mathbb{E}\left[1 - Ce^{-\gamma t}\tau\right] = \underbrace{C\,\mathbb{E}[\tau]}_{<\infty} e^{-\gamma t}. \end{aligned}$$

Similarly, looking at the lower tail we get

$$\mathbb{P}(M < -t) \leq 1 - \mathbb{E}\left[\mathbb{P}(|\epsilon_1| \leq t)^\tau\right] \leq C\,\mathbb{E}[\tau]\, e^{-\gamma t}.$$

□

Lemma 6.2 — *For all large enough a,*

$$x\mathbb{1}_{\{x\geq a\}} \leq e^{x-a/2}, \qquad \forall x \in \mathbb{R}. \tag{6.1}$$

Proof. Differentiate the map $f : x \mapsto \exp(x-a/2)-x$ to find that for large enough a, f is increasing on $[a, \infty)$. Noting that $f(a) \geq 0$ for all large a concludes the proof. □

Lemma 6.3 — *Let $(M_n)_{n\geq 0}$ be a supercritical Galton-Watson process with offspring distribution X and let $\mu := \mathbb{E}X > 1$. If $M_0 = 1$ then for all $\mu > \phi > 0$ we have $0 < \liminf_n \mathbb{P}(M_n > \phi^n)$.*

Proof. By the monotone convergence theorem we can take R large enough so that $\widetilde{\mu} := \mathbb{E}[R \wedge X] > 1$. Let $(\widetilde{M}_n)_{n\geq 0}$ be the Galton-Watson process with offspring distribution $\widetilde{X} := R \wedge X$ which by assumption is also supercritical. As $\widetilde{X}$ is bounded, Theorem 1 in Section 6, Chapter 1 of [1] gives $\widetilde{\mu}^{-n}\widetilde{M}_n \to M$ almost surely for some $M \geq 0$ with $\mathbb{P}(M > 0) > 0$ by Theorem 2 of the same section. By the obvious coupling

$$\mathbb{P}(M_n > \phi^n) \geq \mathbb{P}\left(\widetilde{M}_n\mu^{-n} > \phi^n\mu^{-n}\right) \tag{6.2}$$

so that $\liminf_{n\to\infty} \mathbb{P}(M_n > \phi^n) \geq \mathbb{P}(M > 0) > 0$. □

Theorem 6.4 ([14, Theorem 7.4.1.]) — *Let $\{X_{m,n} \mid 0 \leq m < n\}$ be a family of random variables satisfying*

(i) $X_{0,n} \leq X_{0,m} + X_{m,n}$ for all $m < n$.

(ii) $(X_{nk,(n+1)k})_{n\geq 0}$ is a stationary and ergodic sequence for each $k \geq 1$.

(iii) The distribution of $\{X_{m+k,n+k} \mid 0 \leq m < n\}$ does not depend on $k \in \mathbb{N}$.

(iv) $\mathbb{E}\left[X_{0,1}^{+}\right] < \infty$ *and there exists* $\gamma_0 > -\infty$ *such that* $\mathbb{E}\left[X_{0,n}\right] > \gamma_0 n$ *for all* $n \in \mathbb{N}$.

Then there exists $\gamma \in \mathbb{R}$ *such that*

$$\lim_{n\to\infty} \frac{\mathbb{E}X_{0,n}}{n} = \inf_n \frac{\mathbb{E}X_{0,n}}{n} = \gamma, \tag{6.3}$$

where the last limit is almost sure and in L^1.

Lemma 6.5 — *The random variables* $Z_{i,j}$ *as defined in the proof of Proposition 3.2 satisfy the hypothesis of Kingman's Subadditive Theorem.*

Proof. For each $k \geq 1$ the sequence $\{Z_{k,2k}, Z_{2k,3k}, ...\} = \{\max X_k^k, \max X_k^{2k}, ...\}$ is i.i.d. so stationary and ergodic. Clearly the distribution of $(Z_{i,i+k})_{k\geq 0} = (\max X_k^i)_{k\geq 0}$ is independent of i. $\mathbb{E}Z_{0,1}^{+} = \mathbb{E}(\max X_1)^{+} < \infty$ because $\max X_1 \in L^1$ by (**??**). Finally, $\mathbb{E}Z_{0,n} = \mathbb{E}\max X_n \geq n\,\mathbb{E}\mathscr{L}_{0,1}(1)$. □

Lemma 6.6 ([27, Adapted by Bérard and Gouéré from Lemma 5.2]) — *Let* $v_1 < v_2 \in \mathbb{R}$ *and* $1 \leq m \leq n \in \mathbb{N}$. *Suppose* $0 =: x_0, ..., x_n$ *is a sequence of real numbers such that* $\max_{i\in[\![0,n-1]\!]}(x_{i+1} - x_i) \leq K$ *for some* $K > 0$, *and define* $I := \{i \in [\![0, n-m]\!] \mid x_{i+j} - x_i \geq jv_1, \quad \forall j \in [\![0, m]\!]\}$. *If* $x_n \geq v_2 n$, *then* $\#I \geq \frac{v_2-v_1}{K-v_1}\frac{n}{m} - \frac{K}{K-v_1}$.

6.2 Skorokhod's topologies

This section is based on material from Sections 3, 11 and 12 of [20]. As usual, for $t > 0$ let $D([0,t]) := D([0,t],\mathbb{R})$ be the set of real-valued cádlág functions with domain $[0,t]$. Define Λ_t to be the space of continuous bijections from $[0,t]$ to itself. Skorokhod's J_1 topology (sometimes written SJ_1) on $D([0,t],\mathbb{R})$ is then defined by the metric

$$d_{J_1}(f,g) = \inf_{\lambda\in\Lambda_t}\{\|f\circ\lambda - g\|_\infty \vee \|\lambda - Id\|_\infty\}, \tag{6.4}$$

where $Id : x \mapsto x$. The intuition behind this definition is the following: Take the graph of f which is a (possibly discontinuous) curve in $\mathbb{R}^2$ and let $f\circ\lambda$ be some reparametrisation of it. The J_1 distance between f and g is small if there exists λ close to the identity such that the supremum distance between the graph of g and the graph of $f\circ\lambda$ is small too. J_1 convergence allows for a sequence of functions $(f_n)_{n\geq 0} \subset D([0,t])$ to convergence to a limit $f \in D([0,t])$ without the set of discontinuities of any of the f_n coinciding with that of f. Skorokhod's M_1 topology (sometimes written SM_1) is defined by a similar metric, the only difference being that the same idea is applied to completed graphs, which allows continuous functions to converge to a discontinuous one. Note that the topology J_1 is stronger. These definitions extend to functions with domain $[0,\infty)$ by saying that $f_n \in D([0,\infty)) =: D$ converges to $f \in D$ if convergence happens in $D([0,t])$ for all continuity points t of f. The topology this defines is in fact metrisable (see page 83 of [20]).

6.3 Stable distributions

The material covered here is mainly taken from [15, Section 17] and we will refer to distributions and their characteristic function itnerchangeably. Let ω be the characteristic function of some distribution on $\mathbb{R}$. ω is called *infinitely divisible* if for all $n \in \mathbb{N}$ there exists a characteristic function ω_n such that

$$\omega_n^n = \omega.$$

The distribution with characteristic function ϕ is in the *domain of attraction* of ω if there exist $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ with $a_n > 0$, $\forall\, n$ such that

$$\left(\phi(\frac{t}{a_n})e^{-itb_n}\right)^n \to \omega(t) \qquad \forall\, t \in \mathbb{R}.$$

The characteristic function ω is called *stable* if ϕ too can be replaced by ω in the above display. We now

References

[1] Krishna B Athreya, Peter E Ney, and PE Ney. *Branching processes*. Courier Corporation, 2004.

[2] François Baccelli and Bartłomiej Błaszczyszyn. Stochastic geometry and wireless networks, volume i—theory, volume 3, no 3–4 of foundations and trends in networking, 2009.

[3] Jean Bérard and Jean-Baptiste Gouéré. Brunet-derrida behavior of branching-selection particle systems on the line. *Communications in Mathematical Physics*, 298(2):323–342, 2010.

[4] Jean Bérard, Pascal Maillard, et al. The limiting process of n-particle branching random walk with polynomial tails. *Electronic Journal of Probability*, 19, 2014.

[5] Julien Berestycki, Nathanaël Berestycki, Jason Schweinsberg, et al. The genealogy of branching brownian motion with absorption. *The Annals of Probability*, 41(2):527–618, 2013.

[6] Maury Bramson. *Convergence of solutions of the Kolmogorov equation to travelling waves*, volume 285. American Mathematical Soc., 1983.

[7] Maury D Bramson. Maximal displacement of branching brownian motion. *Communications on Pure and Applied Mathematics*, 31(5):531–581, 1978.

[8] E Brunet, Bernard Derrida, AH Mueller, and S Munier. Phenomenological theory giving the full statistics of the position of fluctuating pulled fronts. *Physical Review E*, 73(5):056126, 2006.

[9] Eric Brunet and Bernard Derrida. Shift in the velocity of a front due to a cutoff. *Physical Review E*, 56(3):2597, 1997.

[10] Éric Brunet and Bernard Derrida. An exactly solvable travelling wave equation in the fisher–kpp class. *Journal of Statistical Physics*, 161(4):801–820, 2015.

[11] Aser Cortines and Bastien Mallein. A n-branching random walk with random selection. *arXiv preprint arXiv:1605.03401*, 2016.

[12] Olivier Couronné, Lucas Gerin, et al. A branching-selection process related to censored galton–walton processes. In *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, volume 50, pages 84–94. Institut Henri Poincaré, 2014.

[13] Freddy Dumortier, Nikola Popović, and Tasso J Kaper. The critical wave speed for the fisher–kolmogorov–petrowskii–piscounov equation with cut-off. *Nonlinearity*, 20(4):855, 2007.

[14] Rick Durrett. *Probability: theory and examples*. Cambridge university press, 2010.

[15] William Feller. An introduction to probability theory and its applications. *1957*.

[16] Ronald Aylmer Fisher. The wave of advance of advantageous genes. *Annals of eugenics*, 7(4):355–369, 1937.

[17] Sergey Foss, Dmitry Korshunov, Stan Zachary, et al. *An introduction to heavy-tailed and subexponential distributions*, volume 6. Springer, 2011.

[18] Nina Gantert, Yueyun Hu, and Zhan Shi. Asymptotics for the survival probability in a killed branching random walk. In *Annales de l'IHP Probabilités et statistiques*, volume 47, pages 111–129, 2011.

[19] John Harris, Simon Harris, et al. Survival probabilities for branching brownian motion with absorption. *Electronic Communications in Probability*, 12:81–92, 2007.

[20] Jean Jacod and Albert Shiryaev. *Limit theorems for stochastic processes*, volume 288. Springer Science & Business Media, 2013.

[21] B Jaffuel. The critical barrier for the survival of the branching random walk with absorption, 2009. *arXiv preprint math.PR/0911.2227*.

[22] Petrovskii I. Kolmogorov, A. and Piskunov. Study of a diffusion equation that is related to the growth of a quality of matter and its application to a biological problem. *Moscow University Mathematics Bulletin*, 1:1–26, 1937.

[23] Russell Lyons. A simple path to biggins' martingale convergence for branching random walk. In *Classical and modern branching processes*, pages 217–221. Springer, 1997.

[24] Bastien Mallein. N-branching random walk with α-stable spine. *Theory of Probability & Its Applications*, 62(2):295–318, 2018.

[25] Carl Mueller, Leonid Mytnik, and Jeremy Quastel. Effect of noise on front propagation in reaction-diffusion equations of kpp type. *Inventiones mathematicae*, 184(2):405–453, 2011.

[26] John Nolan. *Stable distributions: models for heavy-tailed data*. Birkhauser New York, 2003.

[27] Robin Pemantle. Search cost for a nearly optimal path in a binary tree. *The Annals of Applied Probability*, pages 1273–1291, 2009.

[28] Zhan Shi. *Branching random walks*. Springer, 2015.