Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2017, Vol. 53, No. 1, 226–242
DOI: 10.1214/15-AIHP714

ANNALES DE L'INSTITUT HENRI POINCARÉ
PROBABILITÉS ET STATISTIQUES
www.imstat.org/aihp

The many-to-few lemma and multiple spines

Simon C. Harris and Matthew I. Roberts

Department of Mathematical Sciences, University of Bath, Bath BA2 7AY, UK. E-mail: S.C.Harris@bath.ac.uk; mattiroberts@gmail.com

Received 11 September 2014; revised 8 September 2015; accepted 11 September 2015

Abstract. We develop a simple and intuitive identity for calculating expectations of weighted k-fold sums over particles in branching processes, generalising the well-known many-to-one lemma.

Résumé. On développe une identité simple et intuitive pour calculer l'espérance des sommations k-plier sur particules dans les processus de branchement, la généralisation du lemme bien connu 'many-to-one'.

MSC: 60J80

Keywords: Branching processes; Many-to-one; Many-to-few; Spine

1. Introduction

Consider the following simple branching random walk on $\mathbb{Z}$. We begin with one particle at 0, which has two children, whose positions are independent copies of some random variable X. Each of these two new particles has two children of its own, whose positions relative to their parent are independent copies of X, and so on. If the initial particle is the 0th generation, and its children are the first generation, then in the nth generation we have 2^n particles. This is a very basic stochastic model and a classical question asks for the position M_n of the maximal particle in the nth generation when n is large. If we let $Y_n(x)$ be the number of particles in generation n whose position is at least x, then we anticipate that

$$M_n \approx \sup\bigl\{x : \mathbb{E}\bigl[Y_n(x)\bigr] \geq 1\bigr\}.$$

Indeed, for an upper bound, $\mathbb{P}(M_n \geq x) = \mathbb{P}(Y_n(x) \geq 1) \leq \mathbb{E}[Y_n(x)]$. We would therefore like to calculate $\mathbb{E}[Y_n(x)]$, and of course by linearity of expectation we have

$$\mathbb{E}\bigl[Y_n(x)\bigr] = 2^n \mathbb{P}(S_n \geq x), \tag{1}$$

where S_i, $i \geq 0$ is a random walk with step distribution X. Thus a question about the 2^n particles in the nth generation becomes one about a single random walk, and we call (1) a many-to-one formula. There are ways of extending this concept to far more complicated branching processes.

For a lower bound on M_n, we note that by Cauchy–Schwarz,

$$\mathbb{P}(M_n \geq x) = \mathbb{P}\bigl(Y_n(x) \geq 1\bigr) \geq \frac{\mathbb{E}[Y_n(x)]^2}{\mathbb{E}[Y_n(x)^2]},$$

and hence we want to calculate the second moment $\mathbb{E}[Y_n(x)^2]$. By counting the number of pairs of particles whose last common ancestor was alive at time j for each $j = 0, \ldots, n-1$, we see that

$$\mathbb{E}\big[Y_n(x)^2\big] = \mathbb{E}\big[Y_n(x)\big] + \sum_{j=0}^{n-1} 2^{2n-j-1}\mathbb{P}\big(S_{j,n} \geq x, S'_{j,n} \geq x\big),$$

where for each j, $(S_{j,i}, i \geq 0)$ and $(S'_{j,i}, i \geq 0)$ are random walks with step distribution X such that

- $S_{j,i} = S'_{j,i}$ for all $i \leq j$, and
- $(S_{j,j+i} - S_{j,j}, i \geq 0)$ and $(S'_{j,j+i} - S'_{j,j}, i \geq 0)$ are independent.

Thus a question about the second moment of a branching random walk becomes one about two dependent random walks: a many-to-two formula.

It turns out that this formula can also be greatly generalised, and in fact extends to higher moments. Questions about kth moments of branching processes turn into questions about k dependent random walks.

Several results of this type are already known. A simple version for branching Brownian motion was given by Sawyer [18]. Kallenberg [12] proved a version for discrete trees, which he calls a "backward tree formula." Gorostiza and Wakolbinger [10] extend Kallenberg's formula to a class of continuous-time processes. Dawson and Perkins generate what they call "extended Palm formulas" for historical processes (superprocesses enriched with information on genealogy) in [8]. For the parabolic Anderson model with Weibull upper tails, Albeverio et al. [3] gave a similar result by considering existence and uniqueness of solutions to a Cauchy problem. Bansaye et al. [4] develop a quite general many-to-two lemma for Markov branching processes, allowing particles to be born away from their parent. This list is unlikely to be exhaustive, but reflects the fact that many-to-few results exist in various specialised forms with little in the way of a consistent underlying theory.

The theory in the many-to-one case is much more complete. The single random walk on the right-hand side of the formula can be interpreted as a special particle or *spine* present in the original branching process, and this additional structure can be used to construct and understand changes of measure on the branching system, which turns out to be a powerful tool: see for example [2,13].

The aim of this article is to state a quite general kth moment formula which we call the many-to-few lemma, but also to develop a corresponding theory involving multiple spines. This underlying structure will allow us to incorporate similar changes of measure to those that have proved so useful for first moment calculations. It should also allow the reader to transfer the many-to-few lemma to processes not covered by our setup.

There are already several applications of the many-to-few formula either published or underway. To name a few, Aïdékon and Harris [1] compute moments in order to show that the number of particles hitting a certain level in a branching Brownian motion with killing at the origin converges in distribution in the limit approaching criticality. Both Carmona and Hu [5] and Döring and Roberts [9] investigate a catalytic branching model. Gün, König and Sekulović [11] apply our result to a branching random walk in random environment. Maillard [15] uses many-to-two to bound numbers and positions of particles in a branching Brownian motion with selection. Chen [7] also uses many-to-two to investigate the time taken for particles in a branching Brownian motion to become the rightmost. Roberts [17] uses the many-to-two formula at two different times to look at the consistent maximal displacement of branching Brownian motion.

The article is arranged as follows. Mostly we work in continuous time, since this is slightly trickier to handle than discrete time. In Section 2 we give a summary of the multi-spine setup, and then state our main result – the many-to-few lemma – in Section 3. Since the resulting formula can be difficult to handle, we follow this with a discussion of some special cases and fully worked examples in Section 4. In Section 5 we give full constructions of the measures and filtrations used in the theory, and then prove the many-to-few lemma in Section 6. We then give an extension in Section 7 that allows us to take sums over particles at two different times. Finally, in Section 8 we give a discrete time version of the many-to-few lemma.

2. Multiple spines

In this section we detail the general continuous-time branching process that we will consider for most of the article, and introduce the multi-spine setup that will be needed to state our main result.

We consider a branching process starting with one particle at x under a probability measure $\mathbb{P}_x$. This particle moves within a measurable space $(J, \mathcal{B})$ according to a Markov process with generator $\mathcal{M}$. When at position y, the particle branches at rate $R(y)$ (more precisely, the probability that the particle has not branched by time t is $e^{-\int_0^t R(X(s))\,ds}$ where $X(s)$ is the position of the particle at time s), dying and giving birth to a random number of new particles with distribution $\mu^{(y)}$, supported on $\{0, 1, 2, \ldots\}$. Each of these particles then independently repeats the stochastic behaviour of its parent from its starting point.

We denote by $N(t)$ the set of all particles alive at time t. For a particle $v \in N(t)$ we let σ_v be the time of its birth and τ_v the time of its death, and define $\sigma_v(t) = \sigma_v \wedge t$ and $\tau_v(t) = \tau_v \wedge t$. If $v \in N(t)$ then for $s \leq t$ we write $X_v(s)$ for the position of the unique ancestor of v alive at time s. For technical reasons we keep track of particles that die without giving birth to any children by introducing a graveyard state $\Delta \notin J$; if v has 0 children then we write $X_v(s) = \Delta$ for all $t \geq \tau_v$.

2.1. *The k-spine measures $\mathbb{P}_x^k$ and $\mathbb{Q}_x^k$*

We define new measures $\mathbb{P}_x^k$ and $\mathbb{Q}_x^k$ under which there are k distinguished lines of descent, which we call spines. Briefly, $\mathbb{P}_x^k$ is simply an extension of $\mathbb{P}_x$ in that all particles behave as in the original branching process; the only difference is that some particles carry marks showing that they are part of a spine. Under $\mathbb{Q}_x^k$ the marked particles will behave differently from under $\mathbb{P}_x^k$, but non-marked particles will be unchanged. We will eventually see the relationship between $\mathbb{Q}_x^k$ and $\mathbb{P}_x^k$ in terms of a Radon–Nikodym derivative, but for now it is enough to state their properties.

Under $\mathbb{P}_x^k$ particles behave as follows:

- We begin with one particle at position x which (as well as its position) carries k marks $1, 2, \ldots, k$.
- All particles move as Markov processes with generator $\mathcal{M}$, independently of each other given their birth times and positions, just as under $\mathbb{P}_x$.
- We think of each of the marks $1, \ldots, k$ as distinguishing a particular line of descent or "spine," and define ξ_t^i to be the position of whichever particle carries mark i at time t.
- A particle at position y carrying j marks $b_1 < b_2 < \cdots < b_j$ at time t branches at rate $R(y)$, dying and being replaced by a random number of particles with law $\mu^{(y)}$ independently of the rest of the system, just as under $\mathbb{P}_x$.
- Given that a particles $v_1, \ldots, v_a$ are born at a branching event as above, the j marks each choose a particle to follow independently and uniformly at random from amongst the a available. Thus for each $1 \leq l \leq a$ and $1 \leq i \leq j$ the probability that v_l carries mark b_i just after the branching event is $1/a$, independently of all other marks.
- If a particle carrying $j > 0$ marks $b_1 < b_2 < \cdots < b_j$ dies and is replaced by 0 particles, then its marks remain with it as it moves to the graveyard state Δ.

Again we emphasise that under $\mathbb{P}_x^k$, the system behaves exactly as under $\mathbb{P}_x$ except that some particles carry extra marks showing the lines of descent of k spines. We call the collection of particles that have carried at least one spine up to time t the *skeleton* at time t, and write $\mathrm{skel}(t)$; see Figure 1. Of course $\mathbb{P}_x^k$ is not defined on the same σ-algebra as $\mathbb{P}_x$. We let $\mathcal{F}_t^k$ be the filtration containing all information about the system (including the k spines) up to time t; then $\mathbb{P}_x^k$ is defined on $\mathcal{F}_\infty^k$. This will be clarified in Section 5.

Now, for each $n \geq 0$ and $y \in \mathbb{R}$ let

$$m_n(y) = \sum_{a \in \mathbb{Z}_+} a^n \mu^{(y)}(a),$$

the nth moment of the offspring distribution. Define

$$\mu_n^{(y)}(a) = \frac{a^n \mu^{(y)}(a)}{m_n(y)}, \quad a \in \mathbb{Z}_+;$$

$\mu_n^{(y)}$ is called the nth *size-biased* distribution with respect to $\mu^{(y)}$. Let

$$\alpha_n(y) = \big(m_n(y) - 1\big) R(y).$$

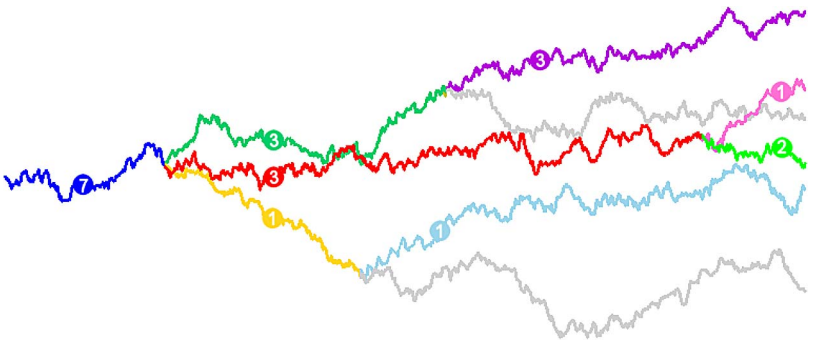

Fig. 1. Each particle in the skeleton is a different colour, and particles not in the skeleton are drawn in grey. The numbers show how many spines are carried by each particle.

For $1 \leq i, j \leq k$ define $T(i, j)$ to be the first split time of the ith and jth spines, i.e. the first time at which marks i and j are carried by different particles. Let $D(v)$ be the total number of marks carried by particle v.

Suppose that $\zeta(X, t)$ is a functional of a process $(X_t, t \geq 0)$ such that if $(X_t, t \geq 0)$ is a Markov process with generator $\mathcal{M}$ then $\zeta(X, t)$ is a non-negative martingale with respect to the natural filtration of $(X_t, t \geq 0)$, with $\zeta(X, 0) = 1$ almost surely. For example if X is a Brownian motion on $\mathbb{R}$ then we might take $\zeta(X, t) = e^{X_t - t/2}$. We will sometimes slightly abuse notation by writing $\zeta(X_v, t)$, or even $\zeta(v, t)$, where $v \in N(t)$. Since $\zeta(X, t)$ must be measurable with respect to $\sigma(X_s, s \leq t)$, it does not matter that $X_v(u)$ is not defined for $u > t$.

Under $\mathbb{Q}_x^k$ particles behave as follows:

- We begin with one particle at position x which (as well as its position) carries k marks $1, 2, \ldots, k$.
- Just as under $\mathbb{P}_x^k$, we think of each of the marks $1, \ldots, k$ as a spine, and write ψ_t^i for whichever particle carries mark i at time t, and ξ_t^i for its position (i.e. $\xi_t^i = X_{\psi_t^i}(t)$).
- A particle with mark i at time t moves as if under the changed measure $Q_x^i|_{\sigma(\xi_s^i, s \leq t)} := \zeta(\xi^i, t)\mathbb{P}_x^k|_{\sigma(\xi_s^i, s \leq t)}$.
- A particle at position y carrying j marks at time t branches at rate $m_j(y)R(y)$, dying and being replaced by a random number of particles with law $\mu_j^{(y)}$ independently of the rest of the system.
- Given that a particles $v_1, \ldots, v_a$ are born at such a branching event, the j marks each choose a particle to follow independently and uniformly at random.
- Particles not in the skeleton (those carrying no marks) behave just as under $\mathbb{P}$, branching at rate $R(y)$ and giving birth to numbers of particles with law $\mu^{(y)}$ when at y.

In other words, under $\mathbb{Q}_x^k$ spine particles move as if weighted by the martingale ζ, they breed at an modified rate, and they give birth to size-biased numbers of children. The birth rate and number of children depend on how many marks the spine particle is carrying, whereas the motion does not.

3. The many-to-few lemma

If Y is measurable with respect to $\mathcal{F}_t^k$, then it can be expressed as the sum

$$Y = \sum_{v_1, \ldots, v_k \in N(t) \cup \{\Delta\}} Y(v_1, \ldots, v_k) \mathbb{1}_{\{\psi_t^1 = v_1, \ldots, \psi_t^k = v_k\}}, \tag{2}$$

where for any $v_1, \ldots, v_k \in N(t) \cup \{\Delta\}$, the random variable $Y(v_1, \ldots, v_k)$ is $\mathcal{F}_t$-measurable. We sometimes write $Y(\psi_t^1, \ldots, \psi_t^k)$ for Y, but emphasise that Y need not depend only on the k spines and can depend on the entire process up to time t.

For example, if $k = 2$ we might take $Y = \mathbb{1}_{\{\xi_t^1 \geq x, \xi_t^2 \geq x\}}$ and then $Y(v_1, v_2) = \mathbb{1}_{\{X_{v_1}(t) \geq x, X_{v_2}(t) \geq x\}}$. This choice of Y would allow us to calculate

$$\mathbb{E}\big[\#\big\{v \in N(t) : X_v(t) \geq x\big\}^2\big].$$

To prove that Y can be written in the form (2), one can generalize the argument on pages 24–25 of [16]. Since this is a purely measure-theoretic argument and will be clear for most Y of interest, we leave it as an exercise for the reader. We now state our main result (in continuous time) in full. A similar statement will be given in discrete time in Section 8.

Lemma 1 (Many-to-few). *For any $k \geq 1$ and $\mathcal{F}_t^k$-measurable Y as above,*

$$\mathbb{P}_x\Bigg[\sum_{v_1,\dots,v_k\in N(t)} Y(v_1,\dots,v_k)\mathbb{1}_{\{\zeta(v_i,t)>0\forall i=1,\dots,k\}}\Bigg] = \mathbb{Q}_x^k\Bigg[Y\prod_{v\in\mathrm{skel}(t)}\frac{\zeta(X_v,\sigma_v(t))}{\zeta(X_v,\tau_v(t))}\exp\Bigg(\int_{\sigma_v(t)}^{\tau_v(t)}\alpha_{D(v)}\big(X_v(s)\big)\,ds\Bigg)\Bigg].$$

We will see in the next section that although the quantity on the left-hand side depends on all the particles in the branching Brownian motion, for many natural choices of Y the right-hand side will depend only the particles in the skeleton, of which there are at most k at any time. Hence the name "many-to-few."

Note that the many-to-few lemma is only interested in particles $v \in N(t)$ such that $\zeta(v,t) > 0$. This can be useful in applications: if we wish to introduce a model incorporating killing of particles in some subset of J, we can choose ζ to be zero on J so that we only count those particles still alive at time t.

4. Examples

Lemma 1 contains a large amount of information within a single identity. Here we expand some of that information by working out the details of some simple cases.

4.1. *The many-to-one formula*

If $k = 1$, then the skeleton simply consists of a single spine particle ξ. We obtain

$$\mathbb{P}_x\Bigg[\sum_{v\in N(t)} Y(v)\mathbb{1}_{\{\zeta(v,t)>0\}}\Bigg] = \mathbb{Q}_x^1\Bigg[Y\frac{1}{\zeta(\xi,t)}e^{\int_0^t\alpha_1(\xi_s)\,ds}\Bigg].$$

For example, if $x = 0$ and our branching process is branching Brownian motion (i.e. $\mathcal{M} = \frac{1}{2}\Delta$) with $\mu^{(y)}(2) \equiv 1$ (binary branching) and $R \equiv \beta > 0$, we might choose $Y = f(\xi_t)$ for some function $f : \mathbb{R} \to \mathbb{R}$, and $\zeta(X,t) = e^{\lambda X_t - \lambda^2 t/2}$. We then obtain

$$\mathbb{P}_0\Bigg[\sum_{v\in N(t)} f\big(X_v(t)\big)\Bigg] = \mathbb{Q}_0^1\big[f(\xi_t)e^{-\lambda\xi_t+\lambda^2 t/2+\beta t}\big].$$

To carry out an actual calculation, take $f(z) = \mathbb{1}_{\{z\geq\lambda t\}}$ and $\lambda > 0$ to get

$$\mathbb{P}_0\big[\#\big\{v\in N(t) : X_v(t)\geq\lambda t\big\}\big] = \mathbb{Q}_0^1\big[\mathbb{1}_{\{\xi_t\geq\lambda t\}}e^{-\lambda\xi_t+\lambda^2 t/2+\beta t}\big] \leq e^{(\beta-\lambda^2/2)t}.$$

We will use the many-to-two formula to get a similar lower bound in the next section.

4.2. *The many-to-two formula*

If $k = 2$, then under $\mathbb{Q}_x^2$ the first particle in the skeleton branches at rate $m_2(y)R(y)$ when at position y, into a particles with probability $a^2\mu^{(y)}(a)/m_2(y)$. At such a branching event, the two marks follow different particles with probability $1 - 1/a$. Thus $T(1,2)$ – the time at which the two spines split – satisfies

$$\mathbb{Q}_x^2\big(T(1,2)\geq t|\xi_s^1, s\in[0,t]\big) = e^{-\int_0^t(m_2(\xi_s^1)-m_1(\xi_s^1))R(\xi_s^1)\,ds}.$$

Writing out the many-to-two formula and simplifying, we obtain

$$\begin{aligned}
&\mathbb{P}_x\Big[\sum_{v_1,v_2\in N(t)} Y(v_1,v_2)\mathbb{1}_{\{\zeta(v_i,t)>0\,\forall i=1,2\}}\Big]\\
&\quad=\mathbb{Q}_x^2\Big[Y\frac{\zeta(\xi^1,T(1,2)\wedge t)}{\zeta(\xi^1,t)\zeta(\xi^2,t)}e^{\int_0^{T(1,2)\wedge t}(m_2(\xi_s^1)-2m_1(\xi_s^1)+1)R(\xi_s^1)\,ds+\int_0^t\alpha_1(\xi_s^1)\,ds+\int_0^t\alpha_1(\xi_s^2)\,ds}\Big]\\
&\quad=\mathbb{Q}_x^2\Big[Y\frac{1}{\zeta(\xi^1,t)}e^{\int_0^t\alpha_1(\xi_u^1)\,du}\Big|T(1,2)\geq t\Big]\\
&\qquad+\int_0^t\mathbb{Q}_x^2\Big[Y\frac{\zeta(\xi^1,s)}{\zeta(\xi^1,t)\zeta(\xi^2,t)}\big(\alpha_2(\xi_s^1)-\alpha_1(\xi_s^1)\big)e^{\int_0^t\alpha_1(\xi_u^1)\,du+\int_s^t\alpha_1(\xi_u^2)\,du}\Big|T(1,2)=s\Big]ds.
\end{aligned}$$

Simplification: *Many-to-two with binary branching at constant rate with* $\zeta\equiv 1$
For a very simple version of the many-to-two formula, suppose that $\mu^{(y)}(2)=1$ for all y and $R(y)=r$ for all y, so under $\mathbb{P}_x$ we have binary branching at constant rate r. Suppose also that $\zeta\equiv 1$ and that Y depends only on the positions of the two spines, $Y=f(\xi_t^1,\xi_t^2)$. Then

$$\mathbb{P}_x\Big[\sum_{v_1,v_2\in N(t)} f\big(X_{v_1}(t),X_{v_2}(t)\big)\Big]=\mathbb{Q}_x^2\big[f\big(\xi_t^1,\xi_t^2\big)e^{2rt+r(T(1,2)\wedge t)}\big].$$

But $T(1,2)$ is exponentially distributed with parameter $2r$ (the particle carrying the two spines 1 and 2 breeds at rate $2^2r=4r$, and at each of these events the two spines follow different children with probability $1/2$), and the motion of the spines is the same under $\mathbb{Q}_x^2$ as under $\mathbb{P}_x^2$, so

$$\begin{aligned}
&\mathbb{P}_x\Big[\sum_{v_1,v_2\in N(t)} f\big(X_{v_1}(t),X_{v_2}(t)\big)\Big]\\
&\quad=e^{rt}\mathbb{P}_x^2\big[f\big(\xi_t^1,\xi_t^1\big)|T(1,2)>t\big]+\int_0^t 2re^{r(2t-s)}\mathbb{P}_x^2\big[f\big(\xi_t^1,\xi_t^2\big)|T(1,2)=s\big]ds.
\end{aligned}$$

Simplification: *Many-to-two with homogeneous breeding*
To include slightly more generality than the simplification above, suppose that $m_2(y)\equiv m_2$, $m_1(y)\equiv m_1$ and $R(y)\equiv\beta>0$ do not depend on y. Then we may write

$$\begin{aligned}
&\mathbb{P}_x\Big[\sum_{v_1,v_2\in N(t)} Y(v_1,v_2)\mathbb{1}_{\{\zeta(v_i,t)>0\,\forall i=1,2\}}\Big]\\
&\quad=e^{(m_1-1)\beta t}\mathbb{Q}_x^2\Big[Y\frac{1}{\zeta(\xi^1,t)}\Big|T(1,2)\geq t\Big]\\
&\qquad+\int_0^t(m_2-m_1)\beta e^{m_1\beta(2t-s)}\mathbb{Q}_x^2\Big[Y\frac{\zeta(\xi^1,s)}{\zeta(\xi^1,t)\zeta(\xi^2,t)}\Big|T(1,2)=s\Big]ds.
\end{aligned}$$

Example: *Large deviations for branching Brownian motion*
Fix $\lambda,\beta>0$. In the previous section we saw that for branching Brownian motion with binary branching at rate β, we have

$$\mathbb{P}_0\big(\exists v\in N(t):X_v(t)\geq\lambda t\big)\leq\mathbb{P}_0\big[\#\big\{v\in N(t):X_v(t)\geq\lambda t\big\}\big]\leq e^{(\beta-\lambda^2/2)t}.$$

We will now use the many-to-two lemma to give a lower bound on the same probability when $\beta-\lambda^2/2<0$. We use the random variable

$$Y=\mathbb{1}_{\{\xi_s^1\leq\lambda s+1\,\forall s\leq t,\xi_s^2\leq\lambda s+1\,\forall s\leq t,\xi_t^1\geq\lambda t,\xi_t^2\geq\lambda t\}}$$

and the same martingale as before, $\zeta(X,t) = e^{\lambda X_t - \lambda^2 t/2}$. These choices give

$$\begin{aligned}
&\mathbb{P}_0\big[\#\big\{v \in N(t) : X_v(s) \le \lambda s + 1 \,\forall s \le t, X_v(t) \ge \lambda t\big\}^2\big] \\
&= e^{\beta t}\mathbb{Q}_0^2\bigg[\mathbb{1}_{\{\xi_s^1 \le \lambda s+1\,\forall s\le t, \xi_t^1 \ge \lambda t\}} \frac{1}{e^{\lambda\xi_t^1 - \lambda^2 t/2}}\Big| T(1,2) \ge t\bigg] \\
&\quad + \int_0^t 2\beta e^{\beta(2t-s)}\mathbb{Q}_x^2\bigg[\mathbb{1}_{\{\xi_s^1 \le \lambda s+1\,\forall s\le t, \xi_s^2 \le \lambda s+1\,\forall s\le t, \xi_t^1 \ge \lambda t, \xi_t^2 \ge \lambda t\}} \frac{e^{\lambda\xi_s^1 - \lambda^2 s/2}}{e^{\lambda\xi_t^1 + \lambda\xi_t^2 - \lambda^2 t}}\Big| T(1,2) = s\bigg] ds \\
&\le e^{\beta t - \lambda^2 t/2} + 2\beta\int_0^t e^{\beta(2t-s)}\frac{e^{\lambda^2 s/2 + \lambda}}{e^{\lambda^2 t}}\, ds.
\end{aligned}$$

When $\beta - \lambda^2/2 < 0$, this is at most $(1 + \frac{2\beta e^{\lambda}}{\lambda^2/2 - \beta})e^{(\beta - \lambda^2/2)t}$.

Using the many-to-one lemma with the same martingale and $Y = \mathbb{1}_{\{\xi_s \le \lambda s+1\,\forall s\le t, \xi_t \ge \lambda t\}}$, we get

$$\begin{aligned}
&\mathbb{P}_0\big[\#\big\{v \in N(t) : X_v(s) \le \lambda s + 1 \,\forall s \le t, X_v(t) \ge \lambda t\big\}\big] \\
&= \mathbb{Q}_0^1\bigg[\mathbb{1}_{\{\xi_s \le \lambda s+1\,\forall s\le t, \xi_t \ge \lambda t\}} \frac{1}{e^{\lambda\xi_t - \lambda^2 t/2}} e^{\beta t}\bigg] \\
&\ge e^{-\lambda} e^{(\beta - \lambda^2/2)t}\mathbb{Q}_0^1(\xi_s \le \lambda s + 1\,\forall s \le t, \xi_t \ge \lambda t).
\end{aligned}$$

Now, under $\mathbb{Q}_0^1$, the process $(\xi_s, s \ge 0)$ moves as if under the changed measure $Q_0|_{\sigma(\xi_s, s\le t)} := e^{\lambda\xi_t - \lambda^2 t/2}\mathbb{P}_0^1|_{\sigma(\xi_s, s\le t)}$. By Girsanov's theorem, $(\xi_s - \lambda s, s \ge 0)$ is therefore a standard Brownian motion, and we have

$$\mathbb{P}_0\big[\#\big\{v \in N(t) : X_v(s) \le \lambda s + 1 \,\forall s \le t, X_v(t) \ge \lambda t\big\}\big] \ge e^{-\lambda} e^{(\beta - \lambda^2/2)t}\mathbb{P}_0^1(\xi_s \le 1\,\forall s \le t, \xi_t \ge 0).$$

It is an easy exercise using the reflection principle to prove that this is at least a positive constant times $e^{(\beta - \lambda^2/2)t} t^{-3/2}$.

Putting these two calculations together, and using the inequality (from Cauchy–Schwarz) $\mathbb{P}(X > 0) \ge \mathbb{P}[X]^2/\mathbb{P}[X^2]$, we have that for $\beta - \lambda^2/2 < 0$,

$$\begin{aligned}
\mathbb{P}_0\big(\exists v \in N(t) : X_v(t) \ge \lambda t\big) &\ge \mathbb{P}_0\big(\exists v \in N(t) : X_v(s) \le \lambda s + 1\,\forall s \le t, X_v(t) \ge \lambda t\big) \\
&\ge \frac{\mathbb{P}_0[\#\{v \in N(t) : X_v(s) \le \lambda s + 1\,\forall s \le t, X_v(t) \ge \lambda t\}]^2}{\mathbb{P}_0[\#\{v \in N(t) : X_v(s) \le \lambda s + 1\,\forall s \le t, X_v(t) \ge \lambda t\}^2]} \\
&\ge \frac{ce^{2(\beta - \lambda^2/2)t} t^{-3}}{e^{(\beta - \lambda^2/2)t}} = ce^{(\beta - \lambda^2/2)t} t^{-3/2}
\end{aligned}$$

for some constant $c > 0$. Thus, together with the upper bound from the previous section, we have that when $\lambda > 0$ and $\beta - \lambda^2/2 < 0$,

$$\lim_{t\to\infty} \frac{1}{t}\log\mathbb{P}\big(\exists v \in N(t) : X_v(t) \ge \lambda t\big) = \beta - \lambda^2/2.$$

With only slightly more work, it is possible to show that the number of particles above λt at time t when $\beta - \lambda^2/2 > 0$ is approximately $e^{\beta - \lambda^2/2}$. Similar techniques can be used to prove far more delicate estimates.

4.3. *The many-to-few formula*

In this section we will apply the many-to-few lemma to a very simple model. There are many other ways of doing the same calculations, but we hope this will allow the reader to see how the many-to-few lemma is – despite appearances – relatively intuitive even for higher moments. We then mention a further extension of the theory, which we will not detail in this article.

Example: *Yule tree*
We take the simplest possible choices: $Y \equiv 1$, each $\zeta^j \equiv 1$, $\mu(2) \equiv 1$ (purely binary branching, so $m_k \equiv 2^k$) and $R \equiv 1$. This completely ignores the spatial movement of the particles, so we shall simply be calculating the moments of the number of particles in a Yule tree (a continuous-time Galton–Watson process with 2 children at every branch point). Let $T = \inf_{1\le i,j\le k} T(i,j)$ be the first time at which any two spines split, and let S_j be the event that at time T, j of the spines follow the first child and $k-j$ follow the second child. Recall that under $\mathbb{Q}^k$, the first birth time is exponentially distributed with parameter 2^k, and then each spine independently chooses to follow either of the two children with probability $1/2$. In particular T is exponentially distributed with parameter $2^k(1-2^{-k+1}) = 2^k - 2$, corresponding to a birth where not all of the spines follow the same child, and $\mathbb{Q}^k(S_j | T \in ds) = \binom{k}{j}\frac{2^{-k}}{1-2^{-k+1}}$. Therefore

$$\begin{aligned}\mathbb{E}\big[|N(t)|^k\big] &= \mathbb{Q}^k\Bigg[\prod_{v\in\mathrm{skel}(t)} e^{(2^{D(v)}-1)(\tau_v(t)-\sigma_v(t))}\Bigg]\\ &= \mathbb{Q}^k\big[e^{(2^k-1)t}\mathbb{1}_{\{T>t\}}\big] + \sum_{j=1}^{k-1}\int_0^t \mathbb{Q}^k\Bigg[\prod_{v\in\mathrm{skel}(t)} e^{(2^{D(v)}-1)(\tau_v(t)-\sigma_v(t))}\mathbb{1}_{\{T\in ds\}}\mathbb{1}_{S_j}\Bigg]\\ &= e^t + \sum_{j=1}^{k-1}\binom{k}{j}\int_0^t e^s \mathbb{E}\big[|N(t-s)|^j\big]\mathbb{E}\big[|N(t-s)|^{k-j}\big]\,ds.\end{aligned}$$

Thus $\mathbb{E}[|N(t)|^2] = 2e^{2t} - e^t$, $\mathbb{E}[|N(t)|^3] = 6e^{3t} - 6e^{2t} + e^t$, $\mathbb{E}[|N(t)|^4] = 24e^{4t} - 36e^{3t} + 14e^{2t} + 3e^t$, and so on.

Extension: *Stopping lines*
Rather than looking at a fixed time t, we might like, for example, to count the number of particles that hit some subset of J at the time they hit. The theory of *stopping lines* allows us to extend the many-to-few lemma to cover this kind of calculation. However, the concept of a stopping line involves a large amount of notation in itself, and combining this with the many-to-few lemma would make this article longer than we would like. We therefore leave it to the reader to extend our methods in this way. A detailed discussion can be found in [14].

5. Multiple spines and changes of measure

Our main aim in this section is to give full details of the setup introduced in Section 2.

5.1. *Trees*

We use the *Ulam–Harris labelling system*: define a set of labels

$$\Omega := \{\varnothing\} \cup \bigcup_{n\in\mathbb{N}} \mathbb{N}^n.$$

We often call the elements of Ω *particles*. We think of $\varnothing$ as our inital ancestor, and $(3,2,7)$ for example as representing the seventh child of the second child of the third child of the initial ancestor. For a particle $u \in \Omega$ we define $|u|$, the generation of u, to be the length of u (so if $u \in \mathbb{N}^n$ then $|u| = n$, and $|\varnothing| = 0$). For two labels $u, v \in \Omega$ we write uv for the concatenation of u and v, taking $\varnothing u = u\varnothing = u$. We write $u \le v$ and say that u is an *ancestor* of v if there exists $w \in \Omega$ such that $uw = v$.

We define $\mathbb{T}$ to be the set of all *trees*: subsets $\tau \subseteq \Omega$ such that

- $\varnothing \in \tau$: the initial ancestor is part of τ;
- for all $u, v \in \Omega$, $uv \in \tau \Rightarrow u \in \tau$: if τ contains a particle then it contains all the ancestors of that particle;
- for each $u \in \tau$, there exists $A_u \in \{0, 1, 2, \ldots\}$ such that for $j \in \mathbb{N}$, $uj \in \tau$ if and only if $1 \le j \le A_u$: each particle in τ has a finite number of children.

5.2. *Marked trees*

Since we wish to have a particular view of trees, as systems evolving in time and space, we define a *marked tree* to be a set T of triples of the form (u, l_u, X_u) such that $u \in \Omega$, the set

$$\text{tree}(T) := \left\{u : \exists l_u, X_u \text{ such that } (u, l_u, X_u) \in T\right\}$$

forms a tree, $l_u \in [0, \infty)$ is the *lifetime* of u, and, setting $\sigma_u := \sum_{v<u} l_v$ and $\tau_u := \sum_{v \leq u} l_u$,

$$X_u : [\sigma_u, \tau_u) \to J$$

is the *position function* of u. We think of the inital ancestor $\varnothing$ moving around in space according to its position function $X_\varnothing$ until time $l_\varnothing$. It then disappears and a number $A_\varnothing$ of new particles appear; each moves according to its position function for a period of time equal to its lifetime, before being replaced by a number of new particles; and so on.

We let $\mathcal{T}$ be the set of all marked trees, and for $T \in \mathcal{T}$ we define

$$N(t) := \left\{u \in \text{tree}(T) : \sigma_u \leq t < \tau_u\right\},$$

the set of particles alive at time t. For convenience, we extend the position path of a particle v to all times $t \in [0, \tau_v)$, to include the paths of all its ancestors:

$$X_v(t) := \begin{cases} X_v(t) & \text{if } \sigma_v \leq t < \tau_v, \\ X_u(t) & \text{if } u < v \text{ and } \sigma_u \leq t < \tau_u \end{cases}$$

and if $A_v = 0$ then we write $X_v(t) = \Delta \ \forall t \geq \tau_v$.

5.3. *Marked trees with spines*

We now enlarge our state space further to include the notion of *spines*. A spine ψ on a marked tree τ is a subset of $\text{tree}(\tau)$ such that

- $\varnothing \in \psi$;
- $\psi \cap (N(t) \cup \{\Delta\})$ contains exactly one particle for each t;
- if $v \in \psi$ and $u < v$ then $u \in \psi$;
- if $v \in \psi$ and $A_v > 0$, then $\exists j \in \{1, \ldots, A_v\}$ such that $vj \in \psi$; otherwise $\psi \cap N(t) = \varnothing \ \forall t \geq \tau_v$.

If $v \in \psi \cap N(t)$ then we write $\psi_t := v$, and write $\xi_t := X_v(t)$ for the position of the spine at time t. We say that a marked tree with spines is a sequence $(\tau, \psi^1, \psi^2, \psi^3, \ldots)$ where $\tau \in \mathcal{T}$ is a marked tree and each ψ^j, $j \geq 1$ is a spine on τ. We let $\tilde{\mathcal{T}}$ be the set of all marked trees with spines.

5.4. *Filtrations*

We now work exclusively on the space $\tilde{\mathcal{T}}$ of marked trees with spines, and use different filtrations on this space to encapsulate different amounts of information. We give descriptions of these filtrations below; formal definitions are similar to those in [16] and are left to the reader.

The filtration $(\mathcal{F}_t, t \geq 0)$**:** We define $(\mathcal{F}_t, t \geq 0)$ to be the natural filtration of the branching process – it does not know anything about the spines.

The filtrations $(\mathcal{F}_t^k, t \geq 0)$**:** For each $k \geq 1$ we let $(\mathcal{F}_t^k, t \geq 0)$ be the natural filtration for the branching process and the first k spines. It does not know anything about spines $\psi^{k+1}, \psi^{k+2}, \ldots$.

The filtrations $(\mathcal{G}_t^j, t \geq 0)$**:** For each j we define $\mathcal{G}_t^j := \sigma(\xi_s^j, s \in [0, t])$, where ξ_s^j represents the position of the jth spine at time s. $\mathcal{G}_t^j$ contains just the spatial information about the jth spine up to time t (and whether or not it has died), but does not know which *nodes* of the tree actually make up that spine.

The filtrations $(\tilde{\mathcal{G}}_t^{\{i_1,\dots,i_j\}}, t \geq 0)$**:** For each j-tuple $i_1, \dots, i_j$ we define

$$\tilde{\mathcal{G}}_t^{\{i_1,\dots,i_j\}} := \sigma\big(\mathcal{G}_t^k \cup \mathcal{A}_t^k \cup \mathcal{C}_t^k, k \in \{i_1, \dots, i_j\}\big),$$

where

$$\mathcal{A}_t^k = \big\{\big\{v = \psi_s^k\big\} : v \in \Omega, s \in [0,t]\big\}$$

and

$$\mathcal{C}_t^k = \big\{\big\{v < \psi_t^k, A_v = a, \sigma_v \leq \sigma\big\} : v \in \Omega, a \geq 2, \sigma \in [0,\infty)\big\}.$$

In words, $\tilde{\mathcal{G}}_t^{\{i_1,\dots,i_j\}}$ contains all the information about spines $\psi^{i_1}, \dots, \psi^{i_j}$ up to time t: which nodes make up the spines, their positions, and for all spine nodes not in $N(t)$ (so all the strict ancestors of the spines at time t) their lifetimes and number of children.

The filtration $(\tilde{\mathcal{G}}_t^k, t \geq 0)$**:** We use the shorthand $\tilde{\mathcal{G}}_t^k = \tilde{\mathcal{G}}_t^{\{1,\dots,k\}}$, so that $\tilde{\mathcal{G}}_t^k$ knows everything about the first k spines up to time t. (Note in particular that $\tilde{\mathcal{G}}_t^k$ is different from $\tilde{\mathcal{G}}_t^{\{k\}}$, which only knows about the kth spine.)

5.5. *Probability measures*

We may now take a probability measure $\mathbb{P}_x$ on $\tilde{\mathcal{T}}$ such that under $\mathbb{P}_x$, the system evolves as a branching process starting with one particle at x, each particle moves as a Markov process with generator $\mathcal{M}$ independently of all others given its birth time and position, and a particle at position y branches at rate $R(y)$ into a random number of particles with distribution $\mu^{(y)}$. This is the system described in Section 2. This measure, however, has no knowledge of the spines (since it sees only the filtration $\mathcal{F}_t$). We would like to extend this to a measure on each of the finer filtrations $\tilde{\mathcal{F}}_t^k$. To do this, we imagine each spine, at each fission event, choosing uniformly from the available children. Then it is easy to see that, for any particle u in a marked tree T and any $j \geq 1$, we would like the probability that u is in the jth spine, conditional on the family sizes $(A_v, v < u)$, to be $\prod_{v<u} \frac{1}{A_v}$. We recall from Section 2 that if Y is an $\tilde{\mathcal{F}}_t^k$-measurable random variable then we can write:

$$Y = \sum_{v_1,\dots,v_k \in N(t) \cup \{\Delta\}} Y(v_1, \dots, v_k) \mathbb{1}_{\{\psi_t^1 = v_1, \dots, \psi_t^k = v_k\}}, \tag{3}$$

where each $Y(v_1, \dots, v_k)$ is $\mathcal{F}_t$-measurable.

Definition 2. *We define the probability measure $\mathbb{P}_x^k$ on $(\tilde{\mathcal{T}}, \tilde{\mathcal{F}}_\infty)$, by setting*

$$\mathbb{P}_x^k[Y] = \mathbb{P}_x\left[\sum_{v_1,\dots,v_k \in N(t) \cup \{\Delta\}} Y(v_1, \dots, v_k) \prod_{j=1}^{k} \prod_{u<v_j} \frac{1}{A_u}\right] \tag{4}$$

for each $\mathcal{F}_t^k$-measurable Y with representation (3). Note that $\mathbb{P}_x = \mathbb{P}_x^k|_{\mathcal{F}_\infty}$.

In summary, particles carrying spines behave just as they would under $\mathbb{P}_x$, and when such a particle branches, each spine makes an independent choice uniformly from amongst the available children.

5.6. *Martingales and a change of measure*

As in Section 2 define $T(i,j) := \inf\{t \geq 0 : \psi_t^i \neq \psi_t^j\}$, and suppose that we are given a functional $\zeta(\cdot, t)$, $t \geq 0$, such that $\zeta(Y,t)$ is a non-negative unit-mean martingale with respect to the natural filtration of the Markov process

$(Y_t, t \geq 0)$ with generator $\mathcal{M}$. We call ζ the single-particle martingale. We recall that we sometimes slightly abuse notation by writing $\zeta(X_v, t)$, or even $\zeta(v, t)$, where $v \in N(t)$. Since $\zeta(Y, t)$ must be measurable with respect to $\sigma(Y_s, s \leq t)$, it does not matter that $X_v(u)$ is not defined for $u > t$.

Recall that we defined $\text{skel}^k(t)$, the skeleton, to be the subtree up to time t generated by those particles carrying at least one of the k spines,

$$\text{skel}^k(t) = \big\{u \in \Omega : \exists s \leq t,\, j \leq k \text{ such that } \psi_s^j = u\big\}.$$

We also set

$$D^k(v) = \#\big\{j \leq k : \exists t \text{ with } v = \psi_t^j\big\}$$

to be the number of spines following particle v, and define

$$E^k(v, t) = \exp\bigg(-\int_{\sigma_v(t)}^{\tau_v(t)} \alpha_{D(v)}\big(X_v(s)\big)\, ds\bigg),$$

where we recall that $\alpha_n(y) = (m_n(y) - 1)R(y)$. Since we will not always know which particles are the spines (when we are working on $\mathcal{F}_t$ for example), it will sometimes be helpful to have the above concepts defined for a general skeleton of k particles $u_1, \ldots, u_k$ instead of the spines. For this reason we define

$$\text{skel}_{u_1,\ldots,u_k}(t) = \{v \in \Omega : \sigma_v \leq t, \exists j \text{ with } v \leq u_j\},$$
$$D_{u_1,\ldots,u_k}(v) = \#\{j : v \leq u_j\},$$

and

$$E_{u_1,\ldots,u_k}(v, t) = \exp\bigg(-\int_{\sigma_v(t)}^{\tau_v(t)} \alpha_{D_{u_1,\ldots,u_k}(v)}\big(X_v(s)\big)\, ds\bigg)$$

so that

$$\text{skel}^k(t) = \text{skel}_{\psi_t^1,\ldots,\psi_t^k}(t), \qquad D^k(v) = D_{\psi^1_{\sigma_v},\ldots,\psi^k_{\sigma_v}}(v) \quad \text{and} \quad E^k(v, t) := E_{\psi^1_{\sigma_v},\ldots,\psi^k_{\sigma_v}}(v, t).$$

Remark. *We note that, with the notation given above,*

$$\mathbb{P}_x^k\big(\psi_t^1 = u_1, \ldots, \psi_t^k = u_k | \mathcal{F}_t\big) = \prod_{v \in \text{skel}_{u_1,\ldots,u_k}(t) \setminus N(t)} A_v^{-D_{u_1,\ldots,u_k}(v)}.$$

Definition 3. *We define an $\tilde{\mathcal{F}}_t^k$-adapted (and, in fact, $\tilde{\mathcal{G}}_t^k$-adapted) process $\tilde{\zeta}^k(t)$, $t \geq 0$ by*

$$\tilde{\zeta}^k(t) = \mathbb{1}_{\{\zeta(\xi^i, t) > 0 \,\forall i = 1,\ldots,k\}} \prod_{v \in \text{skel}^k(t)} \bigg(\frac{\zeta(X_v, \tau_v(t))}{\zeta(X_v, \sigma_v(t))} E^k(v, t)\bigg) \prod_{v \in \text{skel}^k(t) \setminus N(t)} A_v^{D^k(v)}$$

(if $A_v = 0$ then we define $\zeta(X_v, \tau_v(t)) = 0$) and an $\mathcal{F}_t$-adapted process $Z^k(t)$, $t \geq 0$ by

$$Z^k(t) = \sum_{u_1,\ldots,u_k \in N(t)} \mathbb{1}_{\{\zeta(u_i, t) > 0 \,\forall i = 1,\ldots,k\}} \prod_{v \in \text{skel}_{u_1,\ldots,u_k}(t)} \frac{\zeta(X_v, \tau_v(t))}{\zeta(X_v, \sigma_v(t))} E_{u_1,\ldots,u_k}(v, t).$$

We remark here that Z^k and $\zeta(\xi^j, \cdot)$ are, in fact, simply the projections of $\tilde{\zeta}^k$ onto the relevant filtrations:

$$Z^k(t) = \mathbb{P}_x^k\big[\tilde{\zeta}^k(t) | \mathcal{F}_t\big] \quad \text{and} \quad \zeta\big(\xi^j, t\big) = \mathbb{P}_x^k\big[\tilde{\zeta}^k(t) | \mathcal{G}_t^{\{j\}}\big].$$

Lemma 4. *The process $\tilde{\zeta}^k(t)$, $t \geq 0$ is a martingale with respect to the filtrations $\tilde{\mathcal{G}}_t^k$ and $\tilde{\mathcal{F}}_t^k$.*

Proof. Let $\chi = (v_1, v_2, \ldots)$ be a single line of descent (so in particular $v_1 < v_2 < \cdots$), with χ_t representing the position of the unique v_i that is alive at time t. The births along χ form a Cox process driven by χ_t with rate function R. Thus for any $j \geq 0$,

$$\mathbb{P}_x\left[\prod_{v<\chi_t} A_v^j \Big| \chi_s, s \in [0,t]\right] = \exp\left(\int_0^t \alpha_j(\chi_s)\, ds\right).$$

We work by induction on k. The case $k = 1$ is just the single spine case, and is proved by conditioning first on $\mathcal{G}_t^1$, since the births along the spine form a Cox process driven by ξ_t^1 with rate function R. Then, by induction, it is enough to consider the process up to the first split time of the skeleton, since after this time no particle carries more than $k-1$ spines. But up to the first split we have a single particle carrying k spines, so the same argument holds as for the single spine case: the births again form a Cox process driven by ξ_t^1 with rate function R. □

Definition 5. *We define the measure $\mathbb{Q}_x^k$ by setting*

$$\left.\frac{d\mathbb{Q}_x^k}{d\mathbb{P}_x^k}\right|_{\mathcal{F}_t^k} = \tilde{\zeta}^k(t).$$

The proof that $\mathbb{Q}_x^k$ behaves as claimed in Section 2.1 is identical to the proof for one spine given by Chauvin and Rouault [6], applied to each branch of the skeleton independently.

6. Proof of the many-to-few lemma

We first calculate the probability that particles $(u_1, \ldots, u_k)$ make up the skeleton at time t.

Lemma 6 (Gibbs–Boltzmann weights for $\mathbb{Q}^k$). *For any $u_1, \ldots u_k \in N(t) \cup \{\Delta\}$,*

$$\mathbb{Q}_x^k\left(\psi_t^1 = u_1, \ldots, \psi_t^k = u_k | \mathcal{F}_t\right) = \frac{1}{Z(t)} \prod_{v \in \mathrm{skel}_{u_1,\ldots,u_k}(t)} \frac{\zeta(X_v, \tau_v(t))}{\zeta(X_v, \sigma_v(t))} E_{u_1,\ldots,u_k}(v,t).$$

Proof. By the fact that $\mathbb{P}_x^k[\tilde{\zeta}(t)|\mathcal{F}_t] = Z(t)$ and standard properties of conditional expectation,

$$\begin{aligned}
\mathbb{Q}_x^k\left(\psi_t^1 = u_1, \ldots, \psi_t^k = u_k | \mathcal{F}_t\right) &= \frac{\mathbb{P}_x^k[\tilde{\zeta}(t)\mathbb{1}_{\{\psi_t^1 = u_1, \ldots, \psi_t^k = u_k\}}|\mathcal{F}_t]}{\mathbb{P}_x^k[\tilde{\zeta}(t)|\mathcal{F}_t]} \\
&= \frac{1}{Z(t)}\left(\prod_{v \in \mathrm{skel}_{u_1,\ldots,u_k}(t)} \frac{\zeta(X_v, \tau_v(t))}{\zeta(X_v, \sigma_v(t))} E_{u_1,\ldots,u_k}(v,t)\right) \\
&\quad \cdot \left(\prod_{v \in \mathrm{skel}_{u_1,\ldots,u_k}(t) \setminus N(t)} A_v^{D_{u_1,\ldots,u_k}(v)}\right) \mathbb{P}_x^k\left(\psi_t^1 = u_1, \ldots, \psi_t^k = u_k | \mathcal{F}_t\right) \\
&= \frac{1}{Z(t)} \prod_{v \in \mathrm{skel}_{u_1,\ldots,u_k}(t)} \frac{\zeta(X_v, \tau_v(t))}{\zeta(X_v, \sigma_v(t))} E_{u_1,\ldots,u_k}(v,t).
\end{aligned}$$

□

The proof of the many-to-few lemma is now straightforward.

Proof of Lemma 1. We begin with the right-hand side.

$$\begin{aligned}
&\mathbb{Q}_x^k\Big[Y \prod_{v\in\mathrm{skel}(t)} \frac{\zeta(X_v,\sigma_v(t))}{\zeta(X_v,\tau_v(t))}\frac{1}{E(v,t)}\Big]\\
&= \mathbb{Q}_x^k\Big[\sum_{u_1,\dots,u_k\in N(t)\cup\{\Delta\}} Y(u_1,\dots,u_k) \prod_{v\in\mathrm{skel}_{u_1,\dots,u_k}(t)} \frac{\zeta(X_v,\sigma_v(t))}{\zeta(X_v,\tau_v(t))}\frac{1}{E_{u_1,\dots,u_k}(v,t)}\mathbb{1}_{\{\psi_t^1=u_1,\dots,\psi_t^k=u_k\}}\Big]\\
&= \mathbb{Q}_x^k\Big[\sum_{u_1,\dots,u_k\in N(t)\cup\{\Delta\}} Y(u_1,\dots,u_k) \prod_{v\in\mathrm{skel}_{u_1,\dots,u_k}(t)} \frac{\zeta(X_v,\sigma_v(t))}{\zeta(X_v,\tau_v(t))}\frac{\mathbb{Q}_x^k(\psi_t^1=u_1,\dots,\psi_t^k=u_k|\mathcal{F}_t)}{E_{u_1,\dots,u_k}(v,t)}\Big]\\
&= \mathbb{Q}_x^k\Big[\frac{1}{Z^k(t)}\sum_{u_1,\dots,u_k\in N(t)} Y(u_1,\dots,u_k)\Big]\\
&= \mathbb{Q}_x^k\Big[\frac{1}{Z^k(t)}\sum_{u_1,\dots,u_k\in N(t)} Y(u_1,\dots,u_k)\mathbb{1}_{\{\zeta(u_i,t)>0\,\forall i=1,\dots,k\}}\Big]\\
&= \mathbb{P}_x^k\Big[\sum_{u_1,\dots,u_k\in N(t)} Y(u_1,\dots,u_k)\mathbb{1}_{\{\zeta(u_i,t)>0\,\forall i=1,\dots,k\}}\Big],
\end{aligned}$$

where for the last step we used the fact that $\frac{d\mathbb{Q}_x^k}{d\mathbb{P}_x^k}|_{\mathcal{F}_t} = Z^k(t)$. □

7. Many-to-two at two different times

We can also calculate things like

$$\mathbb{P}_x\big[\#\big\{v\in N(s), w\in N(t) : X_v(s)\geq x_s, X_w(t)\geq x_t\big\}\big],$$

where $s<t$ and $x, x_s, x_t\in\mathbb{R}$. This is useful for example in [17] to estimate the number of particles that stay near the maximum in a branching Brownian motion. In this case we might expect an expression involving one spine at time s and the other at time t. A calculation using the many-to-few formula confirms this, and indeed similar statements for k particles at k different times. To save ourselves from having to carry around too much notation, we restrict to the case $k=2$.

One complication is as follows. When working with one time t, we asked that our random variable be $\mathcal{F}_t^2$-measurable. Now that we are handling two different times $s<t$, we need something more subtle: Y should depend only on some part of the tree after time s. The following definition makes this precise.

Fix $s<t$. Suppose that we have an $\mathcal{F}_t^2$-measurable random variable Y. We say that Y *respects the tree at time s* if Y can be written in the form

$$Y=\sum_{v\in N(s)}\sum_{w\in N(t)} Y(v,w)\mathbb{1}_{\{\psi_s^1=v,\psi_t^2=w\}},$$

where for each $v\in N(s)$ and $w\in N(t)$, $Y(v,w)$ is $\mathcal{F}_t$-measurable and, given $\mathcal{F}_s^2$, on the event $\{v\not\leq w\}$, $Y(v,w)$ is independent of the subtree generated by v (that is, the labels, positions, lifetimes, and number of children of v and its descendants).

For example, for any measurable functions $f,g:\mathbb{R}\to\mathbb{R}$, the random variable $f(\xi_t^1)g(\xi_s^2)$ respects the tree at time s.

Lemma 7 (Many-to-two at two different times). *Fix $s < t$. Suppose that we have an $\mathcal{F}_t^2$-measurable random variable Y that respects the tree at time s. Then*

$$\mathbb{P}_x\Bigg[\sum_{v\in N(s)}\sum_{w\in N(t)} Y(v,w)\mathbb{1}_{\{\zeta(v,s)>0,\zeta(w,t)>0\}}\Bigg] = \mathbb{Q}_x^2\Bigg[Y\frac{\zeta(\xi^1,T(1,2)\wedge s)}{\zeta(\xi^1,s)\zeta(\xi^2,t)}e^{\int_0^{T(1,2)\wedge s}\alpha_2(\xi_u^1)\,du+\int_{T(1,2)\wedge s}^{s}\alpha_1(\xi_u^1)\,du+\int_{T(1,2)\wedge s}^{t}\alpha_1(\xi_u^2)\,du}\Bigg]. \tag{5}$$

Proof. For a particle $v \in N(t)$, let v_s be the ancestor of v that was alive at time s. Write T as shorthand for $T(1,2)$, the split time for the two spines. For $v, w \in N(t)$ let $S(v,w)$ be the death time of the most recent common ancestor of v and w (in particular if $v = w$ then $S(v,w) = \tau_v > t$). Also set

$$\begin{aligned}\tilde{Y}(v,w) = {}& \mathbb{1}_{\{S(v,w)\le s\}}\mathbb{1}_{\{\zeta(v,s)>0,\zeta(w,t)>0\}}Y(v_s,w)\mathbb{1}_{\{v=v_s\}}e^{\int_s^t R(X_v(u))\,du}\\ &+\mathbb{1}_{\{S(v,w)>t\}}\mathbb{1}_{\{\zeta(w,t)>0\}}Y(v_s,w).\end{aligned} \tag{6}$$

We will prove the result by showing that both sides of (5) are equal to $\mathbb{P}_x[\sum_{v,w\in N(t)}\tilde{Y}(v,w)]$.

From the definition of $\tilde{Y}(v,w)$,

$$\begin{aligned}\mathbb{P}_x\Bigg[\sum_{v,w\in N(t)}\tilde{Y}(v,w)\Bigg] = {}& \mathbb{P}_x\Bigg[\sum_{v\in N(s)}\sum_{\substack{w\in N(t):\\ v\not\le w}}\mathbb{1}_{\{\zeta(v,s)>0,\zeta(w,t)>0\}}Y(v,w)\mathbb{1}_{\{\tau_v>t\}}e^{\int_s^t R(X_v(u))\,du}\Bigg]\\ &+\mathbb{P}_x\Bigg[\sum_{v\in N(s)}\sum_{\substack{w\in N(t):\\ v\le w}}\mathbb{1}_{\{\zeta(w,t)>0\}}Y(v,w)\Bigg].\end{aligned} \tag{7}$$

By the fact that Y respects the tree at time s, given $\mathcal{F}_s^2$, if $v \in N(s)$, $w \in N(t)$ and $v \not\le w$, then $Y(v,w)$ is independent of the subtree generated by v. Also $\mathbb{P}(\tau_v > t|\mathcal{F}_s^2) = e^{-\int_s^t R(X_v(u))\,du}$, so

$$\begin{aligned}\mathbb{P}_x\Bigg[\sum_{v,w\in N(t)}\tilde{Y}(v,w)\Bigg] &= \mathbb{P}_x\Bigg[\sum_{v\in N(s)}\sum_{\substack{w\in N(t):\\ v\not\le w}}\mathbb{1}_{\{\zeta(v,s)>0,\zeta(w,t)>0\}}Y(v,w)\Bigg]\\ &\quad+\mathbb{P}_x\Bigg[\sum_{v\in N(s)}\sum_{\substack{w\in N(t):\\ v\le w}}\mathbb{1}_{\{\zeta(w,t)>0\}}Y(v,w)\Bigg]\\ &= \mathbb{P}_x\Bigg[\sum_{v\in N(s)}\sum_{w\in N(t)}\mathbb{1}_{\{\zeta(v,s)>0,\zeta(w,t)>0\}}Y(v,w)\Bigg].\end{aligned}$$

We have shown that the left-hand side of (5) is equal to $\mathbb{P}_x[\sum_{v,w\in N(t)}\tilde{Y}(v,w)]$. We essentially want to apply the standard many-to-two lemma to this quantity, but it turns out that this does not quite give us the required expression and thus we need to rework the proof to adapt it to $\tilde{Y}(v,w)$.

We return to the definition (6) of $\tilde{Y}(v,w)$. We have that

$$\begin{aligned}\mathbb{P}_x\Bigg[\sum_{v,w\in N(t)}\tilde{Y}(v,w)\Bigg] = {}& \mathbb{P}_x\Bigg[\sum_{v,w\in N(t)}\mathbb{1}_{\{S(v,w)\le s\}}\mathbb{1}_{\{\zeta(v,s)>0,\zeta(w,t)>0\}}Y(v_s,w)\mathbb{1}_{\{v=v_s\}}e^{\int_s^t R(X_v(u))\,du}\Bigg]\\ &+\mathbb{P}_x\Bigg[\sum_{v,w\in N(t)}\mathbb{1}_{\{S(v,w)>t\}}\mathbb{1}_{\{\zeta(v,t)>0\}}Y(v_s,w)\Bigg].\end{aligned}$$

First note that

$$\begin{aligned}\mathbb{P}_x\Big[\sum_{v,w\in N(t)}\mathbb{1}_{\{S(v,w)>t\}}\mathbb{1}_{\{\zeta(v,t)>0\}}Y(v_s,w)\Big] &= \mathbb{Q}^2_x\Big[Y\mathbb{1}_{\{T>t\}}\frac{1}{\zeta(\xi^2,t)}e^{\int_0^t\alpha_2(\xi^2_u)\,du}\Big]\\ &= \mathbb{Q}^2_x\Big[Y\frac{1}{\zeta(\xi^2,t)}e^{\int_0^t\alpha_2(\xi^2_u)\,du}\mathbb{1}_{\{T>s\}}\mathbb{Q}^2_x\big(T>t|\mathcal{G}^2_t,\mathcal{F}^2_s\big)\Big]\\ &= \mathbb{Q}^2_x\Big[Y\frac{1}{\zeta(\xi^2,t)}e^{\int_0^s\alpha_2(\xi^2_u)\,du+\int_s^t\alpha_1(\xi^2_u)\,du}\mathbb{1}_{\{T>s\}}\Big].\end{aligned}$$

We will also show that

$$\begin{aligned}&\mathbb{P}_x\Big[\sum_{v,w\in N(t)}\mathbb{1}_{\{S(v,w)\le s\}}\mathbb{1}_{\{\zeta(v,s)>0,\zeta(w,t)>0\}}Y(v_s,w)\mathbb{1}_{\{v=v_s\}}e^{\int_s^t R(X_v(u))\,du}\Big]\\ &= \mathbb{Q}^2_x\Big[Y\mathbb{1}_{\{T\le s\}}\frac{\zeta(\xi^1,T)}{\zeta(\xi^1,s)\zeta(\xi^2,t)}e^{\int_0^T\alpha_2(\xi^1_u)\,du+\int_T^s\alpha_1(\xi^1_u)\,du+\int_T^t\alpha_1(\xi^2_u)\,du}\Big].\end{aligned}\tag{8}$$

Combining these two equalities, we get that

$$\mathbb{P}_x\Big[\sum_{v,w\in N(t)}\tilde{Y}(v,w)\Big] = \mathbb{Q}^2_x\Big[Y\frac{\zeta(\xi^1,T\wedge s)}{\zeta(\xi^1,s)\zeta(\xi^2,t)}e^{\int_0^{T\wedge s}\alpha_2(\xi^1_u)\,du+\int_{T\wedge s}^s\alpha_1(\xi^1_u)\,du+\int_{T\wedge s}^t\alpha_1(\xi^2_u)\,du}\Big]$$

which will be enough to complete the proof.

It remains to show (8). By the definition (4) of $\mathbb{P}^2_x$,

$$\begin{aligned}&\mathbb{P}_x\Big[\sum_{v,w\in N(t)}\mathbb{1}_{\{S(v,w)\le s\}}\mathbb{1}_{\{\zeta(v,s)>0,\zeta(w,t)>0\}}Y(v_s,w)\mathbb{1}_{\{v=v_s\}}e^{\int_s^t R(X_v(u))\,du}\Big]\\ &= \mathbb{P}^2_x\Big[\mathbb{1}_{\{T\le s\}}\mathbb{1}_{\{\zeta(\xi^1,s)>0,\zeta(\xi^2,t)>0\}}Y\mathbb{1}_{\{\psi^1_t=\psi^1_s\}}e^{\int_s^t R(\xi^1_u)\,du}\Big(\prod_{v\le\psi^1_T}A^2_v\Big)\Big(\prod_{\psi^1_T<v<\psi^1_t}A_v\Big)\Big(\prod_{\psi^2_T<v<\psi^2_t}A_v\Big)\Big].\end{aligned}$$

On the event $\{\psi^1_t=\psi^1_s\}$, the second product above can be restricted to $v<\psi^1_s$ without changing anything. Then, using the fact that Y respects the tree at time s, the above is

$$\mathbb{P}^2_x\Big[\mathbb{1}_{\{T\le s\}}\mathbb{1}_{\{\zeta(\xi^1,s)>0,\zeta(\xi^2,t)>0\}}Y\Big(\prod_{v\le\psi^1_T}A^2_v\Big)\Big(\prod_{\psi^1_T<v<\psi^1_s}A_v\Big)\Big(\prod_{\psi^2_T<v<\psi^2_t}A_v\Big)\Big].$$

Using again the fact that Y respects the tree at time s, we see that given $\mathcal{F}^2_s$, on the event $\{T\le s\}$,

$$\frac{\zeta(\xi^1,r)}{\zeta(\xi^1,s)}e^{-\int_s^r\alpha_1(\xi^1_u)\,du}\mathbb{1}_{\{\zeta(\xi^1,r)>0\}}\prod_{\psi^1_s<v<\psi^1_r}A_v,\quad r\ge s$$

is a martingale that is independent of $Y\prod_{\psi^2_s<v<\psi^2_t}A_v$. Putting this together with the above, we have that

$$\begin{aligned}&\mathbb{P}_x\Big[\sum_{v,w\in N(t)}\mathbb{1}_{\{S(v,w)\le s\}}\mathbb{1}_{\{\zeta(v,s)>0,\zeta(w,t)>0\}}Y(v_s,w)\mathbb{1}_{\{v=v_s\}}e^{\int_s^t R(X_v(u))\,du}\Big]\\ &= \mathbb{P}^2_x\Big[\mathbb{1}_{\{T\le s,\zeta(\xi^1,t)>0,\zeta(\xi^2,t)>0\}}Y\frac{\zeta(\xi^1,t)}{\zeta(\xi^1,s)}e^{-\int_s^t\alpha_1(\xi^1_u)\,du}\Big(\prod_{v\le\psi^1_T}A^2_v\Big)\Big(\prod_{\psi^1_T<v<\psi^1_t}A_v\Big)\Big(\prod_{\psi^2_T<v<\psi^2_t}A_v\Big)\Big].\end{aligned}\tag{9}$$

Recall now that on the event $\{T \le s\}$, since $s \le t$, we have that $T \le t$ and hence by Definitions 3 and 5,

$$\left.\frac{d\mathbb{Q}_x^2}{d\mathbb{P}_x^2}\right|_{\mathcal{F}_t^2} = \frac{\zeta(\xi^1,t)\zeta(\xi^2,t)}{\zeta(\xi^1,T)}\mathbb{1}_{\{\zeta(\xi^1,t)>0,\zeta(\xi^2,t)>0\}}e^{-\int_0^T \alpha_2(\xi_u^1)\,du-\int_T^t \alpha_1(\xi_u^1)\,du-\int_T^t \alpha_1(\xi_u^2)\,du}$$
$$\cdot\Big(\prod_{v\le\psi_T^1} A_v^2\Big)\Big(\prod_{\psi_T^1<v<\psi_t^1} A_v\Big)\Big(\prod_{\psi_T^2<v<\psi_t^2} A_v\Big).$$

Applying this to (9), we get that

$$\mathbb{P}_x\Big[\sum_{v,w\in N(t)} \mathbb{1}_{\{S(v,w)\le s\}}\mathbb{1}_{\{\zeta(v,s)>0,\zeta(w,t)>0\}}Y(v_s,w)\mathbb{1}_{\{v=v_s\}}e^{\int_s^t R(X_v(u))\,du}\Big]$$
$$= \mathbb{Q}_x^2\Big[Y\mathbb{1}_{\{T\le s\}}\frac{\zeta(\xi^1,T)}{\zeta(\xi^1,s)\zeta(\xi^2,t)}e^{\int_0^T \alpha_2(\xi_u^1)\,du+\int_T^s \alpha_1(\xi_u^1)\,du+\int_T^t \alpha_1(\xi_u^2)\,du}\Big].$$

This establishes (8) and completes the proof. □

8. Many-to-few in discrete time

We state here a version of the many-to-few lemma for discrete-time processes. We shall not prove it, as it is very similar to the continuous-time version studied above.

We begin, under a probability measure $\mathbb{P}_x$, with one particle in generation 0 located at $x \in J$. Any particle at position y has children whose number and positions are decided according to a finite point process $\mathcal{D}_y$ on J. The children of particles in generation n make up generation $n+1$. We define $G(n)$ to be the set of all particles in generation n, $N(n) = \#G(n)$ to be the number of such particles, and X_v to be the position of particle v. We set $m_j(y) = \mathbb{P}_y[N(1)^j]$ to be the jth moment of the number of particles created by the point process $\mathcal{D}_y$. Write $|v|$ to be the generation of particle v. For a particle v in generation $n \ge 1$, let $p(v)$ be its parent in generation $n-1$.

8.1. *The measure $\mathbb{Q}_x^k$ and the main result in discrete time*

We define a new measure $\mathbb{P}_x^k$ which has k distinguished lines of descent $\psi^1,\dots,\psi^k$ just as in the continuous-time case, which we call spines. Under $\mathbb{P}_x^k$, if a particle carrying j marks (i.e. the particle is part of j spines) in generation n has l children in generation $n+1$, then each of its j marks chooses a particle to follow in generation $n+1$ uniformly at random from the l children. We let ξ_n^i be the position of the ith spine in generation n and define skel(n) to be the set of all particles of generation at most n which are part of at least one spine. Let D_v be the number of marks carried by particle v.

For any i, we note that $X_{\xi_0^i}, X_{\xi_1^i}, X_{\xi_2^i},\dots$ is a Markov chain with some generator $\mathcal{M}'$ not depending on i. Suppose that $\zeta(X,n), n \ge 0$ is a functional of a process $(X_n, n \ge 0)$ such that if $(X_n, n \ge 0)$ is a Markov process with generator $\mathcal{M}'$ then $\zeta(X,n), n \ge 0$ is a martingale with respect to the natural filtration of $(X_n, n \ge 0)$.

Under $\mathbb{Q}_x^k$ particles behave as follows:

- A particle at position y carrying j marks has children whose number and positions are decided by a point process such that:
  - for each j and $l \ge 0$, $\mathbb{Q}_y^j(N(1)=l) = l^j\mathbb{P}_y(N(1)=l)/\mathbb{P}_y[N(1)^j]$;
  - for each i, the sequence $X_{\xi_0^i}, X_{\xi_1^i}, X_{\xi_2^i},\dots$ is a Markov chain distributed as if under the changed measure $Q_x^i|_{\mathcal{G}_n^{\{i\}}} := \zeta(\xi^i,n)\mathbb{P}_x^k|_{\mathcal{G}_n^{\{i\}}}$.
- Given that a particles $v_1,\dots,v_a$ are born at such a branching event, the j spines each choose a particle to follow independently and uniformly at random.
- Particles not in the skeleton (those carrying no marks) have children according to the point process $\mathcal{D}_y$ when at position y, just as under $\mathbb{P}$.

In other words, under $\mathbb{Q}_x^k$ spine particles move as if weighted by the martingale ζ, they breed at a modified rate, and they give birth to size-biased numbers of children. The birth rate and number of children depend on how many marks the spine particle is carrying, whereas the motion does not.

Lemma 8 (Many-to-few in discrete time). *For any $k \geq 1$ and $\mathcal{F}_n^k$-measurable Y such that*

$$Y = \sum_{v_1,\dots,v_k \in G(n) \cup \{\Delta\}} Y(v_1,\dots,v_k) \mathbb{1}_{\{\psi_n^1 = v_1,\dots,\psi_n^k = v_k\}}$$

we have

$$\mathbb{P}_x\left[\sum_{v_1,\dots,v_k \in G(n)} Y(v_1,\dots,v_k) \mathbb{1}_{\{\zeta(v_i,n)>0 \,\forall i=1,\dots,k\}}\right] = \mathbb{Q}_x^k\left[Y \prod_{v \in \mathrm{skel}(n)\setminus\{\varnothing\}} \frac{\zeta(p(v),|v|-1)}{\zeta(v,|v|)} m_{D_{p(v)}}(X_{p(v)})\right].$$

Acknowledgements

The authors are indebted to both Elie Aïdékon and Julien Berestycki for some very helpful discussions, and also thank Pascal Maillard for checking an earlier draft.

References

[1] E. Aïdékon and S. C. Harris. Near-critical survival probability of branching Brownian motion with an absorbing barrier. Preprint.
[2] E. Aidekon. Convergence in law of the minimum of a branching random walk. *Ann. Probab.* **41** (3A) (2013) 1362–1426. MR3098680
[3] S. Albeverio, L. V. Bogachev, S. A. Molchanov and E. B. Yarovaya. Annealed moment Lyapunov exponents for a branching random walk in a homogeneous random branching environment. *Markov Process. Related Fields* **6** (4) (2000) 473–516. MR1805091
[4] V. Bansaye, J.-F. Delmas, L. Marsalle and V.C. Tran. Limit theorems for Markov processes indexed by continuous time Galton–Watson trees. *Ann. Appl. Probab.* **21** (6) (2011) 2263–2314. MR2895416
[5] P. Carmona and Y. Hu. The spread of a catalytic branching random walk. *Ann. Inst. Henri Poincaré Probab. Stat.* **50** (2) (2014) 327–351. MR3189074
[6] B. Chauvin and A. Rouault. KPP equation and supercritical branching Brownian motion in the subcritical speed area. Application to spatial trees. *Probab. Theory Related Fields* **80** (2) (1988) 299–314. MR0968823
[7] X. Chen. Waiting times for particles in a branching Brownian motion to reach the rightmost position. *Stochastic Process. Appl.* **123** (8) (2013) 3153–3182. MR3062441
[8] D. A. Dawson and E. A. Perkins. *Historical Processes*. Amer. Math. Soc., Providence, RI, 1991. MR1079034
[9] L. Döring and M. I. Roberts. Catalytic branching processes via spine techniques and renewal theory. In *Séminaire de Probabilités, XLV* 305–322. *Lecture Notes in Math.* **2078**. Springer, Berlin, 2013. MR3185920
[10] L. G. Gorostiza and A. Wakolbinger. Persistence criteria for a class of critical branching particle systems in continuous time. *Ann. Probab.* **19** (1) (1991) 266–288. MR1085336
[11] O. Gün, W. König and O. Sekulovic. Moment asymptotics for branching random walks in random environment. *Electron. J. Probab.* **18** (63) (2013) 1–18. MR3078022
[12] O. Kallenberg. Stability of critical cluster fields. *Math. Nachr.* **77** (1) (1977) 7–43. MR0443078
[13] R. Lyons, R. Pemantle and Y. Peres. Conceptual proofs of $L \log L$ criteria for mean behavior of branching processes. *Ann. Probab.* **23** (3) (1995) 1125–1138. MR1349164
[14] P. Maillard. Branching Brownian motion with selection of the N right-most particles: An approximate model. Ph.D. thesis, 2012.
[15] P. Maillard. Speed and fluctuations of N-particle branching Brownian motion with spatial selection. *Probab. Theory Related Fields* **166** (3) (2016) 1061–1173. MR3568046
[16] M. I. Roberts. Spine changes of measure and branching diffusions. Ph.D. thesis, University of Bath, 2010. Available at http://people.bath.ac.uk/mir20/thesis.pdf.
[17] M. I. Roberts. Fine asymptotics for the consistent maximal displacement of branching Brownian motion. *Electron. J. Probab.* **20** (2015) article 28. MR3325098
[18] S. Sawyer. Branching diffusion processes in population genetics. *Adv. in Appl. Probab.* **8** (4) (1976) 659–689. MR0432250