Weierstraß-Institut
für Angewandte Analysis und Stochastik
Leibniz-Institut im Forschungsverbund Berlin e. V.

Preprint ISSN 0946 – 8633

The many-to-few lemma and multiple spines

Simon C. Harris[1], Matthew I. Roberts[2]

submitted: June 29, 2011

[1] University of Bath
Department of Mathematical Sciences
Bath, BA2 7AY
United Kingdom
E-Mail: S.C.Harris@bath.ac.uk

[2] Weierstrass Institute
Mohrenstr. 39
10117 Berlin
Germany
E-Mail: mattiroberts@gmail.com

No. 1624
Berlin 2011

2010 *Mathematics Subject Classification.* 60J80.

Key words and phrases. Branching processes, many-to-one, many-to-few, spine, change of measure.

The research of Matthew I. Roberts was supported by a Weierstrass Postdoctoral Fellowship.

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Leibniz-Institut im Forschungsverbund Berlin e. V.
Mohrenstraße 39
10117 Berlin
Germany

Fax: +49 30 2044975
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

Abstract: We develop an extension to the spine theory of branching processes, and use it to give a simple and intuitive identity for calculating additive functionals of such processes, generalizing the well-known many-to-one lemma.

1 Introduction

1.1 The many-to-two lemma

Consider a branching Brownian motion (BBM): one particle starts at 0 and moves like a Brownian motion until a random exponentially distributed time with mean 1. It then dies and leaves in its place two new particles, which independently follow, relative to their initial position, the same random behaviour as their parent. Let $N(t)$ be the set of particles alive at time t, and for a particle $u \in N(t)$ let $X_u(t)$ be the position of particle u. Let B_t, $t \geq 0$ be a standard Brownian motion, and $f : \mathbb{R} \to \mathbb{R}$ be some measurable function. The following result is well-known:

Lemma 1 (Simple many-to-one lemma)**.**

$$\mathbb{E}\left[\sum_{u \in N(t)} f(X_u(t))\right] = e^t \mathbb{E}[f(B_t)]. \tag{1}$$

The most useful aspect of this lemma is that it turns questions about a system of many dependent particles into questions about a single Brownian motion. For example, let $A(x,t) = \#\{u \in N(t) : X_u(t) > x\}$, the number of particles that are above x at time t. For which x and t is $A(x,t)$ non-zero? (This question is related to solutions of the FKPP equation.) Markov's inequality and the many-to-one lemma give us an easy upper bound:

$$\mathbb{P}(A(x,t) \geq 1) \leq \mathbb{E}[A(x,t)] = \mathbb{E}\left[\sum_{u \in N(t)} \mathbb{1}_{\{X_u(t) > x\}}\right] = \frac{e^t}{\sqrt{2\pi t}} \int_x^\infty e^{-y^2/2} dy.$$

For a lower bound, one would like to use a second-moment method, applying

$$\mathbb{P}(A(x,t) \geq 1) \geq \frac{\mathbb{E}[A(x,t)]^2}{\mathbb{E}[A(x,t)^2]},$$

but the many-to-one lemma does not tell us how to calculate $\mathbb{E}[A(x,t)^2]$. Instead we should use a *many-to-two* lemma. Lemma 2 gives an example of a many-to-two lemma for BBM.

Lemma 2 (Simple many-to-two lemma)**.** *For measurable f and g,*

$$\mathbb{E}\left[\sum_{u,v \in N(t)} f(X_u(t))\, g(X_v(t))\right] = e^{2t} \mathbb{E}[e^{T \wedge t} f(B_t)\, g(B'_t)] \tag{2}$$

where

$$B'_t = \begin{cases} B_t & \text{if } t < T \\ B_T + W_{t-T} & \text{if } t \geq T \end{cases}$$

with T exponentially distributed with parameter 2 and W_t, $t \geq 0$ a standard Brownian motion independent of B_t.

The main result of this article will be the many-to-few lemma, Lemma 3, which is a much more general version of Lemma 2. In fact we will be able to calculate additive functionals not just of two particles, but of arbitrarily many particles. We also incorporate the possibility of using a change of measure for the motion of the particles to allow for easier calculation of the right-hand side of the identity.

Results similar to Lemma 3 have existed for some time in various forms[1], usually proved by arguments specific to the particular model or problem. Our article provides several advantages over these previous results. Firstly, we state Lemma 3 for a rather general model, and our methods are robust and may be adapted for use with other branching processes. In addition the multiple spine setup outlined in Section 2 gives an intuitive backdrop for understanding many-to-few results. Thus we hope that this article will provide a general framework that will allow the reader to quickly understand and construct a many-to-few lemma for whichever branching process they wish to consider. Finally, to our knowledge there is no existing work — for any model — that allows one to change measure as part of the result. This technique can be extremely useful: we give an example in Section 4.2.

There are already several applications of this theory underway. Aïdékon and Harris [1] use the k-particle (for general k) version to 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. Döring and Roberts [6] calculate moments of numbers of particles in a catalytic branching model, for which the multiple spine theory gives an intuitive combinatorial derivation for a collection of constants which otherwise appear abstractly from the analysis. Ortgiese and Roberts (work in progress) also apply the k-particle version to the parabolic Anderson model to show that the large-time behaviour of the underlying branching process is rather different from that anticipated by its moments. Roberts [15] uses the full power of our general many-to-two lemma, with a particular choice of measure change, to give simple proofs of large-time asymptotics for the position of the extremal particle in a branching Brownian motion.

1.2 The spine approach

Three articles [11, 13, 14] by Kurtz, Lyons, Pemantle and Peres — building on work of Chauvin and Rouault [4] among others — gave the subject of branching processes a new set of tools, known as *spine* methods. These techniques have since been used by many authors to prove new results and to give intuitive new proofs of old results.

Just like the many-to-one lemma, the spine methods retain one essential theme: at large times the branching structure may be very complicated and we may have very many particles, but one can understand much of this complicated

[1] An even simpler form of Lemma 2 was given by Sawyer [17]. Kallenberg [10] proved a version for discrete trees, which he calls a "backward tree formula". Gorostiza and Wakolbinger [7] 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 [5]. For the parabolic Anderson model with Weibull upper tails, Albeverio *et al.* [2] gave a similar result by considering existence and uniqueness of solutions to a Cauchy problem. Bansaye *et al.* [3] develop quite general many-to-two lemmas for Markov branching processes, allowing particles to be born away from their parent.

behaviour to first order by carefully studying just one special particle. It is no great surprise, then, that spine techniques allowed simple proofs of much more general versions of the many-to-one lemma that would not have been accessible otherwise.

We develop a theory of multiple spines in order to gain further information about the system. This approach leads naturally to a quite simple proof of our main result. However, just as general many-to-one theorems are far from the only application of single-spine techniques, the detailed multiple-spine theory that we develop in proving our results may also be useful in other ways.

This article is arranged as follows. In Section 2 we give a summary of the multi-spine setup, and then state our main result in Section 3. Section 4 provides some examples of how this result can be applied. Then in Section 5 we give full constructions of the measures and filtrations used in the theory. Section 5 is rather technical and may be ignored by readers wishing only to apply our methods. We prove the many-to-few lemma in Section 6. Finally, in Section 7 we state a discrete-time version of the many-to-few lemma.

2 Multiple spines

We state here the general continuous-time branching setup that we will study in this paper.

We consider a branching process starting with one particle at x under a probability measure $\mathbb{P}_x$. This particle moves withing a measurable space $(J, \mathcal{B})$ according to a Markov process with generator $\mathcal{C}$. When at position y, a particle branches at rate $R(y)$ (informally, in a period of time dt the particle branches with probability $R(y)dt$), dying and giving birth to a random number of new particles with distribution μ_y (where for each y, μ_y has support on $\{0, 1, 2, \ldots\}$). Each of these particles then independently repeats the stochastic behaviour of its parent from its starting point.

We label our particles using the Ulam-Harris scheme: the first particle is $\emptyset$, its l children are labelled 1, 2, ..., l, the m children of particle 1 are labelled 11, 12, ..., $1m$, and so on. We denote by $N(t)$ the set of all particles alive at time t. For a particle $u \in N(t)$ we let σ_u be the time of its birth and τ_u the time of its death, and define $\sigma_u(t) = \sigma_u \wedge t$ and $\tau_u(t) = \tau_u \wedge t$. If $u \in N(t)$ then for all $s \leq t$ we write $X_u(s)$ for the position of the unique ancestor of u alive at time s. If u has 0 children then we write $X_u(s) = \Delta$ for all $t \geq \tau_u$, where $\Delta \notin J$ is a graveyard state.

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

We define new measures $\mathbb{P}^k_x$ and $\mathbb{Q}^k_x$ under which there are k distinguished lines of descent which we call spines. The actual construction of $\mathbb{P}^k_x$ is slightly technical, and the construction of $\mathbb{Q}^k_x$ relies on a carefully chosen change of measure (see Section 5), but we do not necessarily have to understand these constructions. It is most important simply to understand the dynamics of the system under these new measures.

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

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

In other words, under $\mathbb{P}_x^k$ the system behaves exactly as under $\mathbb{P}_x$; the only difference is 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.

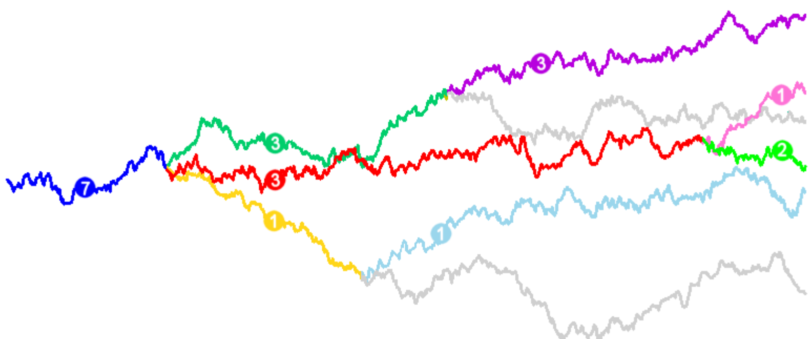

Figure 1: A realisation of the start of the process. Each particle in the skeleton is a different colour, and particles not in the skeleton are drawn in grey. The numbers show how many spines are carried by each particle in the skeleton.

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

$$m^n(y) = \sum_a a^n \mu_y(a),$$

the nth moment of the offspring distribution. Let

$$\mu_y^n(a) = \frac{a^n \mu_y(a)}{m^n(y)}\ ;$$

μ_y^n is called the nth *size-biased* distribution with respect to μ_y. 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 $(Y_t, t \geq 0)$ is a Markov process with generator $\mathcal{B}$ then $\zeta(Y,t)$ is a unit-mean martingale with respect to the natural filtration of $(Y_t, t \geq 0)$. For example if Y is a Brownian motion on $\mathbb{R}$ then we might take

$$\zeta(X,t) = e^{X_t - t/2}.$$

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, with ξ_t^i the position of whichever particle carries mark i at time t.
- A particle with mark i at time t moves as if under the changed measure $Q_x^i|_{\mathcal{G}_t^{\{i\}}} := \zeta(\xi^i, t)\mathbb{P}_x^k|_{\mathcal{G}_t^{\{i\}}}$.
- 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 μ_y^j 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 spines each choose a particle to follow independently and uniformly at random, just as under $\mathbb{P}_x^k$.
- 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 μ_y when at y.

In other words, under $\mathbb{Q}^k$ spine particles move as if weighted by the martingale ζ; they breed at an accelerated rate; and they give birth to size-biased numbers of children.

3 The many-to-few lemma

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

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

where each $Y(v_1, \ldots, v_k)$ is $\mathcal{F}_t$-measurable. To see this one can generalize the argument in [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 full.

Lemma 3 (Many-to-few). *For any $k \geq 1$ and $\mathcal{F}_t^k$-measurable Y as above,*

$$\mathbb{P}\left[\sum_{v_1,\dots,v_k \in N(t)} Y(v_1,\dots,v_k)\right]$$
$$= \mathbb{Q}^k\left[Y \prod_{v\in \mathrm{skel}(t)} \frac{\zeta(X_v,\sigma_v(t))}{\zeta(X_v,\tau_v(t))} \exp\left(\int_{\sigma_v(t)}^{\tau_v(t)} \left(m^{D(v)}(X_v(s)) - 1\right) R(X_v(s))ds\right)\right].$$

Note that this is much more general than the simple version stated in Lemma 2. As well as using the more general branching setup and allowing us to calculate additive functionals of arbitrarily many particles rather than just two, we are also able to use the martingales $\zeta(\xi^i, t)$ to change the motion of the spines, which in many situations will make calculation of the right-hand side easier. We also state a discrete-time version of Lemma 3 in Section 7.

4 Examples

4.1 Simple applications of Lemma 3

The section above states the many-to-few lemma in some generality. It may be enlightening to look instead at some particular simple examples of branching processes and see how the result can easily be used to calculate moments of population numbers. We do this below.

Example 1. The simplest possibility is to take $Y \equiv 1$, each $\zeta^j \equiv 1$, $A \equiv 2$ (purely binary branching, so $m^k \equiv 2^k$) and $R \equiv 1$. This completely ignores the spatial movement of the particles: 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). Because of the simplicity of this model there are many other ways of getting the same result.

$$\begin{aligned}
\mathbb{E}[|N(t)|^2] &= e^{2t}\mathbb{Q}^2[e^{T(1,2)\wedge t}] \\
&= e^{2t}\int_0^t e^s \mathbb{Q}(T(1,2)\in ds) + e^{2t}\mathbb{Q}(T(1,2) > t) \\
&= e^{2t}\int_0^t 2e^{-s} ds + e^t \\
&= 2e^{2t} - e^t.
\end{aligned}$$

In order to calculate the kth moment let $T = \inf_{1\leq i,j\leq k} T(i,j)$ 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.

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

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

Example 2. A more interesting example is to take the same setup as in Example 1 above but with each particle moving as a Brownian motion (so that we have a standard branching Brownian motion), and to attempt to calculate the probability that a particle has position above λt at time t. The first moment method, with the many-to-one lemma, gives us an upper bound: setting

$$W = |\{u \in N(t) : X_u(t) \geq \lambda t\}|$$

we have

$$\mathbb{P}(\exists u \in N(t) : X_u(t) \geq \lambda t) \leq \mathbb{E}[W] = e^t\mathbb{P}(\xi_t \geq \lambda t) \sim e^{t-\lambda^2 t/2}$$

where we use $\sim$ to indicate that we are ignoring terms of at most polynomial order.

For the lower bound we use the second moment method with the many-to-two lemma. Let $W = \#\{u \in N(t) : X_u(t) \geq \lambda t\}$; then

$$\mathbb{P}(\exists u \in N(t) : X_u(t) \geq \lambda t) \geq \frac{\mathbb{E}[W]^2}{\mathbb{E}[W^2]}$$

so to get asymptotic agreement with the upper bound, we require

$$\mathbb{E}[W^2] \lesssim e^{t-\lambda^2 t/2}.$$

Now, from Lemma 3, taking $Y = \mathbb{1}_{\{\xi_t^1\geq\lambda t,\xi_t^2\geq\lambda t\}}$,

$$
\begin{aligned}
\mathbb{E}[W^2] &= e^{2t}\mathbb{Q}^2[e^{T(1,2)\wedge t}\mathbb{1}_{\{\xi_t^1\geq\lambda t,\xi_t^2\geq\lambda t\}}]\\
&= e^t\mathbb{P}(\xi_t \geq \lambda t) + e^{2t}\int_0^t e^s\cdot 2e^{-2s}\mathbb{Q}^2\left(\xi_t^1 \geq \lambda t, \xi_t^2 \geq \lambda t\,\middle|\, T(1,2) = s\right)ds\\
&\sim e^{t-\lambda^2 t/2} + 2e^{2t}\int_0^t e^{-2s}\int_{-\infty}^\infty \frac{1}{\sqrt{2\pi s}}e^{-x^2/2s}\\
&\qquad\qquad \cdot\mathbb{Q}^2\left(\xi_t^1 \geq \lambda t, \xi_t^2 \geq \lambda t\,\middle|\, T(1,2) = s, \xi_s = x\right)dx\,ds\\
&\sim e^{t-\lambda^2 t/2} + 2e^{2t}\int_0^t e^{-2s}\int_{-\infty}^\infty \frac{1}{\sqrt{2\pi s}}e^{-x^2/2s-(\lambda t-x)^2/(t-s)}dxds\\
&= e^{t-\lambda^2 t/2} + 2e^{2t}\int_0^t e^{-2s}\sqrt{\frac{2\pi(t-s)}{t+s}}e^{-\lambda^2t^2/(t+s)}ds.
\end{aligned}
$$

It is not difficult to see that if $\lambda > \sqrt{2}$ then

$$2s + \frac{\lambda^2 t^2}{t+s} \geq t + \frac{1}{2}\lambda^2 t \quad \text{for} \quad s \in [0,t]$$

(expand out to get a quadratic in s; if $\lambda \in (\sqrt{2}, \sqrt{18})$ then there are no roots, and if $\lambda \geq \sqrt{18}$ then both roots are larger than t — the easiest way to check this latter fact is to note that the equation is satisfied for $s = 0$ and $s = t$, and has negative derivative for $s \in [0,t]$). Thus

$$\mathbb{E}[W^2] \sim e^{t - \lambda^2 t/2}$$

and we have proved that if $\lambda > \sqrt{2}$ then

$$\lim_{t\to\infty} \frac{1}{t} \log \mathbb{P}(\exists u \in N(t) : X_u(t) \geq \lambda t) = 1 - \frac{1}{2}\lambda^2.$$

Of course we could have taken more care in the approximations above to gain a more detailed result, but we prefer to demonstrate a simple use of the many-to-two lemma without getting bogged down in carefully approximating integrals. For a more detailed application to a similar problem see Roberts [15].

4.2 Large deviations for BBM

A large deviations result for branching Brownian motion was first proved by Lee [12]. Later a probabilistic proof was given by Hardy and Harris [8]. In this section we give an outline of a proof using the many-to-two lemma, showing how a careful choice of single-particle martingale can ease the required calculations.

For $A \subseteq C[0,1]$, let

$$M(A,T) = \{u \in N(T) : X(sT)/T = g(s) \;\; \forall s \in [0,1] \text{ for some } g \in A\}$$

and define

$$H_1 = \left\{ g \in C[0,1] : g(0) = 0, \; \exists h \in L^2[0,1] \text{ with } g(s) = \int_0^t h(s)ds \; \forall t \in [0,1] \right\}.$$

Theorem 4. *For any closed set $F \subseteq C[0,1]$,*

$$\limsup_{T\to\infty} \frac{1}{T} \log \mathbb{P}(M(F,T) \neq \emptyset) \leq -\inf_{g\in F} J(g)$$

and for any open set $U \subseteq C[0,1]$,

$$\liminf_{T\to\infty} \frac{1}{T} \log \mathbb{P}(M(U,T) \neq \emptyset) \geq -\inf_{g\in U} J(g)$$

where

$$J(g) := \begin{cases} \sup_{\theta\in[0,1]} \left(\int_0^\theta g'(s)^2 ds - \theta \right) & \text{if } g \in H_1 \\ \infty & \text{otherwise.} \end{cases}$$

Proof. For a C^2 function $f : [0,T] \to \mathbb{R}$ such that $f(0) = 0$ and $t \in [0,T]$ we define

$$\hat{N}(t) = \#\{u \in N(t) : |X_u(s) - f(s)| < \varepsilon T \ \ \forall s \in [0,t]\}$$

where $\varepsilon > 0$ and $T > 0$ are fixed constants (sometimes we shall write $\hat{N}_\varepsilon(t)$ to indicate the dependence on ε). Itô's formula shows that if $(B_t, t \geq 0)$ is a standard Brownian motion, then

$$V(B,t) := e^{\int_0^t f'(s)dB_s - \frac{1}{2}\int_0^t f'(s)^2 ds + \frac{\pi^2 t}{8\varepsilon^2 T^2}} \cos\Big(\frac{\pi}{2\varepsilon T}(B_t - f(t))\Big)$$

is a local martingale. The optional stopping theorem then tells us that

$$\zeta(B,t) := V(B,t)\mathbb{1}_{\{|B_s - f(s)| < \varepsilon T \ \forall s \leq t\}}$$

is a martingale. Applying the many-to-one lemma,

$$\mathbb{E}[\hat{N}(t)] = e^t \mathbb{Q}^1\left[\frac{1}{\zeta(\xi^1, t)}\right] \geq e^{\frac{\pi^2 t}{8\varepsilon^2 T} + t}\mathbb{Q}^1[e^{-\int_0^t f'(s)d\xi^1_s + \frac{1}{2}\int_0^t f'(s)^2 ds}].$$

Integration by parts tells us that

$$\begin{aligned}
&\int_0^t f'(s)d\xi^1_s - \int_0^t f'(s)^2 ds \\
&= f'(t)\xi^1_t - \int_0^t f''(s)\xi^1_s ds - f'(t)f(t) + \int_0^t f(s)f''(s)ds \\
&= f'(t)(\xi^1_t - f(t)) - \int_0^t f''(s)(\xi^1_s - f(s))ds
\end{aligned}$$

so that under $\mathbb{Q}^1$,

$$\left|\int_0^t f'(s)d\xi^1_s - \int_0^t f'(s)^2 ds\right| \leq \varepsilon T|f'(t)| + \varepsilon T\int_0^t |f''(s)|ds.$$

Thus

$$\mathbb{E}[\hat{N}(t)] \geq e^{t - \frac{1}{2}\int_0^t f'(s)^2 ds - \varepsilon T|f'(t)| - \varepsilon T\int_0^t |f''(s)|ds}.$$

On the other hand, for $\delta < \varepsilon$,

$$\begin{aligned}
\mathbb{E}[\hat{N}_\delta(t)] &= e^t\mathbb{Q}^1\left[\frac{1}{\zeta(\xi^1,t)}\mathbb{1}_{\{|\xi^1_s - f(s)| < \delta T \ \forall s \leq t\}}\right] \\
&\leq \frac{e^{\frac{\pi^2 t}{8\varepsilon^2 T} + t}}{\cos\left(\frac{\pi\delta}{2\varepsilon}\right)}\mathbb{Q}^1[e^{-\int_0^t f'(s)d\xi^1_s + \frac{1}{2}\int_0^t f'(s)^2 ds}] \\
&\leq \frac{e^{t - \frac{1}{2}\int_0^t f'(s)^2 ds + \varepsilon T|f'(t)| + \varepsilon T\int_0^t |f''(s)|ds + \frac{\pi^2}{8\varepsilon^2 T}}}{\cos\left(\frac{\pi\delta}{2\varepsilon}\right)}
\end{aligned}$$

Similarly, setting

$$R(T) = \frac{e^{3\varepsilon T \sup_{u \leq T}|f'(u)| + 3\varepsilon T\int_0^T |f''(u)|du + \frac{\pi^2}{8\varepsilon^2 T}}}{\cos\left(\frac{\pi\delta}{2\varepsilon}\right)}$$

we have

$$\begin{aligned}\mathbb{E}[\hat{N}_\delta(t)^2] &= e^t\mathbb{Q}^2\bigg[\frac{1}{\zeta(\xi^1,t)}\mathbb{1}_{\{|\xi^1_s-f(s)|<\delta T\ \forall s\leq t\}}\bigg]+\int_0^t\mathbb{Q}\left[\frac{2e^{2t-s}\zeta(\xi^1,s)}{\zeta(\xi^1,t)\zeta(\xi^2,t)}\middle| T=s\right]\\ &\leq R(T)e^{t-\frac{1}{2}\int_0^t f'(s)^2ds}+R(T)\int_0^t e^{2t-s+\frac{1}{2}\int_0^s f'(u)^2du-\int_0^t f'(u)^2du}ds\\ &\leq R(T)e^{t-\frac{1}{2}\int_0^t f'(s)^2ds}+R(T)te^{2t-\int_0^t f'(s)^2ds}\sup_{r\in[0,t]}e^{\frac{1}{2}\int_0^r f'(s)^2ds-r}.\end{aligned}$$

Choosing τ such that

$$e^{\frac{1}{2}\int_0^\tau f'(s)^2ds-\tau}=\sup_{r\in[0,T]}e^{\frac{1}{2}\int_0^r f'(s)^2ds-r}$$

we see that

$$\mathbb{E}[\hat{N}_\delta(t)^2]\leq R(T)(T+1)e^{2t-\int_0^t f'(s)^2ds+\frac{1}{2}\int_0^\tau f'(s)^2ds-\tau}.$$

Putting our estimates for the first and second moments together,

$$\begin{aligned}\mathbb{P}(\hat{N}(T)\geq 1)&\leq\mathbb{P}(\hat{N}(\tau)\geq 1)\\ &\leq\mathbb{E}[\hat{N}(\tau)]\leq\frac{e^{\tau-\frac{1}{2}\int_0^\tau f'(s)^2ds+2\varepsilon T|f'(\tau)|+2\varepsilon T\int_0^\tau|f''(s)|ds+\frac{\pi^2}{32\varepsilon^2T}}}{\cos\left(\frac{\pi}{4}\right)}\end{aligned}$$

and

$$\mathbb{P}(\hat{N}_\varepsilon(T)\geq 1)\geq\mathbb{P}(\hat{N}_\delta(T)\geq 1)\geq\frac{\mathbb{E}[\hat{N}_\delta(T)]^2}{\mathbb{E}[\hat{N}_\delta(T)^2]}\geq\frac{e^{\tau-\frac{1}{2}\int_0^\tau f'(s)^2ds}}{R(T)(T+1)e^{2\varepsilon T|f'(T)|+2\varepsilon\int_0^T|f''(s)|ds}}.$$

Now setting $g(t)=f(tT)/T$ and $\theta=\tau/T$, we obtain

$$\frac{1}{T}\log\mathbb{P}(\hat{N}(T)\geq 1)\leq\theta-\frac{1}{2}\int_0^\theta g'(s)^2ds+2\varepsilon|g'(\theta)|+2\varepsilon\int_0^1|g''(s)|ds+o(1)$$

and

$$\frac{1}{T}\log\mathbb{P}(\hat{N}(T)\geq 1)\geq\theta-\frac{1}{2}\int_0^\theta g'(s)^2ds-5\varepsilon\sup_{s\in[0,1]}|g'(s)|-5\varepsilon\int_0^1|g''(s)|ds+o(1).$$

This establishes the required estimates for balls about smooth functions, to within an error which goes to zero with the radius of the ball. It remains to apply techniques from large deviations theory. For the lower bound it suffices to choose ε small. For the upper bound we must rule out the possibility of particles following extreme paths, so that we are left with a compact set; then use upper semicontinuity of the rate function to check that we may choose an appropriate ε. These details are carried out fully in [8], and are similar to those in the proof of Schilder's theorem for one Brownian motion (see [18] for example). □

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. We take, more or less, the route laid out by Hardy and Harris [9] for a single spine.

5.1 Trees

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

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

(as usual $\mathbb{N} = \{1, 2, 3, \ldots\}$).

We often call the elements of Ω *particles*. We think of $\emptyset$ as our "inital ancestor", and a label $(3, 2, 7)$ (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 $|\emptyset| = 0$). For two labels $u, v \in \Omega$ we write uv for the concatenation of u and v, so for example $(3, 2, 7)(1, 5, 4) := (3, 2, 7, 1, 5, 4)$ (and we take $\emptyset u = u\emptyset = u$). We write $u \leq v$ and say that u is an *ancestor* of v if there exists $w \in \Omega$ such that $uw = v$.

We define a *tree* to be a subset $\tau \subseteq \Omega$ such that

- $\emptyset \in \tau$: 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 $1 \leq j \leq A_u$: each particle in τ has a finite number of children.

We let $\mathbb{T}$ be the set of all such trees.

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

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

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 $X_u(t)$ as describing the spatial position of the particle u at time t. To paint the picture more clearly, we think of the inital ancestor $\emptyset$ moving around in space according to its position function $X_\emptyset$ until just before time $l_\emptyset$. At this time it disappears and a number $A_\emptyset$ of new particles appear; each of these then moves around in space 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 the set of particles alive at time t to be

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

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:

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

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

5.3 Marked trees with spines

We now enlarge our state space further to include the notion of *spines*. We define a spine to be a single maximal distinguished line of descent. That is, a spine ξ on a marked tree τ is a subset of tree(τ) such that

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

If $v \in \xi \cap N(t)$ then we define $\xi_t := X_v(t)$, the position of the spine at time t. At certain points we shall also use the notation ξ_t to mean the particle v itself — beyond this introduction it should always be clear from the context which meaning is intended, and so this should not lead to any ambiguity. For clarity within this section we will use the less concise notation node(ξ_t) to denote the particle v itself — that is, the unique $v \in N(t) \cap \xi$. We say that a marked tree with spines is a sequence $(\tau, \xi^1, \xi^2, \xi^3, \ldots)$ where $\tau \in \mathcal{T}$ is a marked tree and ξ^1, ξ^2, ... are spines 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 $(\mathcal{F}_t, t \ge 0)$
We define $(\mathcal{F}_t, t \ge 0)$ to be the natural filtration of the branching process - it does not know anything about the spines.

The filtrations $(\mathcal{F}_t^k, t \ge 0)$
For each $k \ge 1$ we define $(\mathcal{F}_t^k, t \ge 0)$ to be the natural filtration for the branching process and the first k spines. It does not know anything about spines ξ^{k+1}, ξ^{k+2}, ..., but knows everything about the branching process and spines ξ^1, ..., ξ^k.

The filtrations $(\mathcal{G}_t^j, t \ge 0)$
For each j we define

$$\mathcal{G}_t^j := \sigma\left(\xi_s^j, s \in [0, t]\right)$$

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

$$\tilde{\mathcal{G}}_t^{\{i_1,\ldots,i_j\}} := \sigma\left(\mathcal{G}_t^k \cup \mathcal{A}_t^k \cup \mathcal{C}_t^k, k \in \{i_1, \ldots, i_j\}\right).$$

where

$$\mathcal{A}_t^k = \{\{u = \text{node}(\xi_s^k)\} : u \in \Omega, s \in [0,t]\}$$

and

$$\mathcal{C}_t^k = \{\{u < \text{node}(\xi_t^k), A_u = a, \sigma_u \leq \sigma\} : u \in \Omega, a \geq 2, \sigma \in [0,\infty)\}.$$

$\tilde{\mathcal{G}}_t^{\{i_1,\ldots,i_j\}}$ contains all the information about the relevant collection of spines 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

$$\tilde{\mathcal{G}}_t^k = \tilde{\mathcal{G}}_t^{\{1,\ldots,k\}}$$

so that $\tilde{\mathcal{G}}_t^k$ knows everything about the first k spines up to time t. Thus $\tilde{\mathcal{G}}_t^k$ is different from $\tilde{\mathcal{G}}_t^{\{k\}}$.

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 $\mathcal{C}$ 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 μ_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

$$\text{Prob}(u \in \xi^j) = \prod_{v<u} \frac{1}{A_v}.$$

We recall from Section 2 that if Y is an $\tilde{\mathcal{F}}_t^k$-measurable random variable then we can write:

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

where each $Y(v_1,\ldots,v_k)$ is $\mathcal{F}_t$-measurable. (Here when we write ξ_t^j we are talking really about the particle node(ξ_t^j) rather than its position.)

Definition 5. We define the probability measure $\mathbb{P}_x^k$ on $(\tilde{\mathcal{T}}, \tilde{\mathcal{F}}_\infty)$, by setting

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

for each $\mathcal{F}_t^k$-measurable Y with representation (3).

Remark. The measure $\tilde{\mathbb{P}}_x$ is an extension of $\mathbb{P}_x$ in that $\mathbb{P}_x = \tilde{\mathbb{P}}_x|_{\mathcal{F}_\infty}$, since

$$\sum_{v_1,\ldots,v_k \in N(t)\cup\Delta} \prod_{j=1}^k \prod_{u<v_j} \frac{1}{A_u} = 1.$$

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 $T(i,j) := \inf\{t \geq 0 : \xi_t^i \neq \xi_t^j\}$, and suppose that we are given a functional $\zeta(\cdot,t)$, $t \geq 0$, such that $\zeta(Y,t)$ is a unit-mean martingale with respect to the natural filtration of the Markov process $(Y_t, t \geq 0)$ with generator $\mathcal{C}$. We call ζ the single-particle martingale.

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

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

We also set

$$D(v) = \#\{j : \exists t \text{ with } v = \xi_t^j\}$$

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

$$E(v,t) = \exp\left(-\int_{\sigma_v(t)}^{\tau_v(t)} \left(m^{D(v)}(X_v(s)) - 1\right) R(X_v(s))ds\right).$$

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

$$E_{u_1,\ldots,u_k}(v,t) = \exp\left(-\int_{\sigma_v(t)}^{\tau_v(t)} \left(m^{D_{u_1,\ldots,u_k}(v)}(X_v(s)) - 1\right) R(X_v(s))ds\right)$$

so that

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

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

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

Definition 6. 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

$$\tilde{\zeta}^k(t) = \prod_{v \in \mathrm{skel}(t)} \left(\frac{\zeta(X_v, \tau_v(t))}{\zeta(X_v, \sigma_v(t))} E(v,t) \right) \prod_{v \in \mathrm{skel}(t) \setminus N(t)} A_v^{D_v}$$

(if $A_v = 0$, that is to say that v has no children, then we may arbitrarily define $\zeta(X_v, \tau_v(t)) = 0$) and an $\mathcal{F}_t$-adapted process $Z^k(t)$, $t \geq 0$ by

$$Z^k(t) = \sum_{u_1,\ldots,u_k \in N(t)} \prod_{j=1}^{k} \prod_{v \leq u_j} \frac{\zeta(X_v, \tau_v(t))}{\zeta(X_v, \sigma_v(t))} E_{u_1,\ldots,u_k}(v,t).$$

Again we will often supress the dependence on k.

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

- $Z(t) = \tilde{\mathbb{P}}[\tilde{\zeta}(t)|\mathcal{F}_t]$
- $\zeta(\xi^j, t) = \tilde{\mathbb{P}}[\tilde{\zeta}(t)|\mathcal{G}_t^{\{j\}}]$.

Lemma 7. *The process $\tilde{\zeta}(t)$, $t \geq 0$ 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 $v_1 < v_2 < \ldots$), 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 $j \geq 0$,

$$\mathbb{P}\left[\prod_{v < \chi_t} A_v^j \middle| \chi_s, s \in [0,t] \right] = \exp\left(\int_0^t (m^j(\chi_s) - 1) R(\chi_s) ds \right).$$

Decomposing the process $\tilde{\zeta}(t)$ according to the splitting times of the k spines and repeatedly applying the above fact together with the optional stopping theorem and the Markov branching property (which ensures that different branches of the skeleton are independent given the information up to their split) gives the result. □

Definition 8. We define the measure $\mathbb{Q}_x^k$ by

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

The proof that $\mathbb{Q}_x^k$ behaves as claimed in Section 2.1 is just the same as the original proof (for one spine) given by Chauvin and Rouault [4], applied to each branch of the skeleton independently.

6 Proof of the many-to-few lemma

We first need to calculate the probability that a k-tuple of particles $(u_1, \dots, u_k)$ makes up the skeleton at time t.

Lemma 9 (Gibbs-Boltzmann weights for $\mathbb{Q}^k$)**.** *For any* $u_1, \dots u_k \in N(t) \cup \{\Delta\}$,

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

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

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

as required. □

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

Proof of Lemma 3. We begin with the right-hand side.

$$\mathbb{Q}^k\left[Y\prod_{v\in\text{skel}(t)}\frac{\zeta(X_v,\sigma_v(t))}{\zeta(X_v,\tau_v(t))}\frac{1}{E(v,t)}\right]$$

$$=\mathbb{Q}^k\left[\sum_{u_1,\dots,u_k\in N(t)\cup\{\Delta\}}Y(u_1,\dots,u_k)\right.$$

$$\left.\cdot\prod_{v\in\text{skel}_{u_1,\dots,u_k}(t)}\frac{\zeta(X_v,\sigma_v(t))}{\zeta(X_v,\tau_v(t))}\frac{1}{E_{u_1,\dots,u_k}(v,t)}\mathbb{1}_{\{\xi_t^1=u_1,\dots,\xi_t^k=u_k\}}\right]$$

$$=\mathbb{Q}^k\left[\sum_{u_1,\dots,u_k\in N(t)\cup\{\Delta\}}Y(u_1,\dots,u_k)\right.$$

$$\left.\cdot\prod_{v\in\text{skel}_{u_1,\dots,u_k}(t)}\frac{\zeta(X_v,\sigma_v(t))}{\zeta(X_v,\tau_v(t))}\frac{1}{E_{u_1,\dots,u_k}(v,t)}\mathbb{Q}^k(\xi_t^1=u_1,\dots,\xi_t^k=u_k|\mathcal{F}_t)\right]$$

$$=\mathbb{Q}^k\left[\frac{1}{Z(t)}\sum_{u_1,\dots,u_k\in N(t)}Y(u_1,\dots,u_k)\right]$$

$$=\mathbb{P}^k\left[\sum_{u_1,\dots,u_k\in N(t)}Y(u_1,\dots,u_k)\right]$$

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

7 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 this result, as it is very similar to the continuous-time version studied above.

7.1 A discrete-time branching process

We begin 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 $N(n)$ to be the total number of particles in generation n, 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 $n\geq 1$, let $p(v)$ be its parent in generation $n-1$. For any line of descent $v_0,v_1,v_2,\dots$ such that $|v_n|=n$ and $p(v_{n+1})=v_n$ for each $n\geq 0$, we note that $X_{v_0},X_{v_1},X_{v_2},\dots$ is a Markov chain with some generator $\mathcal{C}'$ not depending on the choice of $v_0,v_1,\dots$. Suppose that $\zeta(X,n)$, $n\geq 0$ is a functional of a process $(X_n,n\geq 0)$ such that if $(X_n,n\geq 0)$ is a Markov process with generator $\mathcal{C}'$ then $\zeta(X,n)$, $n\geq 0$ is a martingale with respect to the natural filtration of $(X_n,n\geq 0)$.

7.2 The skeleton and the measure $\mathbb{Q}^k$

We have k distinguished lines of descent just as in the continuous-time case, which we call spines. Under $\mathbb{P}$, 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^i_n 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. Set D_v to be the number of marks carried by particle v.

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

- We begin with one particle at position x which (as well as its position) carries k marks $1, 2, \dots, k$.
- Just as under $\mathbb{P}^k$, we think of each of the marks $1, \dots, k$ as a spine, with ξ^i_n the position of whichever particle carries mark i at time n.
- 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 $l \geq 0$, $\mathbb{Q}^j_y(N(1) = l) = l^j \mathbb{P}_y(N(1) = l)/\mathbb{P}_y[N(1)^j]$ (the number of children is j-size biased);
  - for each i, the sequence $X_{\xi^i_0}, X_{\xi^i_1}, X_{\xi^i_2}, \dots$ is a Markov chain distributed as if under the changed measure $Q^i_x|_{\mathcal{G}^{\{i\}}_n} := \zeta(\xi^i, n)\mathbb{P}^k_x|_{\mathcal{G}^{\{i\}}_n}$.
- 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, just as under $\mathbb{P}^k$.
- 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 $\mathbb{Q}^k$ spine particles move as if weighted by the martingale ζ, and they give birth to size-biased numbers of children.

7.3 The main result in discrete time

Lemma 10 (Many-to-few in discrete time)**.** *For any $k \geq 1$ and $\mathcal{F}^k_n$-measurable Y such that*

$$Y = \sum_{v_1, \dots, v_k \in N(n) \cup \{\Delta\}} Y(v_1, \dots, v_k) \mathbb{1}_{\{\xi^1_n = v_1, \dots, \xi^k_n = v_k\}}$$

we have

$$\mathbb{P}\left[\sum_{v_1, \dots, v_k \in N(n)} Y(v_1, \dots, v_k)\right] = \mathbb{Q}^k\left[Y \prod_{v \in \text{skel}(n)} \frac{\zeta(p(v), |v|-1)}{\zeta(v, |v|)} m^{D_{p(v)}}(X_{p(v)})\right].$$

The proof of this result is similar to that of Lemma 3.

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. MR would like to thank ANR MADCOF as well as WIAS for their support.

References

[1] E. Aïdékon and S. C. Harris. Near-critical survival probability of branching Brownian motion with an absorbing barrier. To appear, 2011.

[2] 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):473–516, 2000.

[3] V. Bansaye, J. F. Delmas, L. Marsalle and V. C. Tran. Limit theorems for Markov processes indexed by continuous time Galton-Watson trees To appear in *Ann. Appl. Probab.* Arxiv preprint: `http://arxiv.org/abs/0911.1973`, 2010.

[4] 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):299–314, 1988.

[5] D.A. Dawson and E.A. Perkins. *Historical processes.* AMS Bookstore, 1991.

[6] L. Döring and M. I. Roberts. Catalytic branching processes via spine techniques and renewal theory. Arxiv preprint: `http://arxiv.org/abs/1106.5428`, 2011.

[7] L.G. Gorostiza and A. Wakolbinger. Persistence criteria for a class of critical branching particle systems in continuous time. *Ann. Probab.*, 19(1):266–288, 1991.

[8] R. Hardy and S. C. Harris. A conceptual approach to a path result for branching Brownian motion. *Stochastic Process. Appl.*, 116(12):1992–2013, 2006.

[9] R. Hardy and S. C. Harris. A spine approach to branching diffusions with applications to L^p-convergence of martingales. In *Séminaire de Probabilités, XLII*, volume 1979 of *Lecture Notes in Math.* Springer, Berlin, 2009.

[10] O. Kallenberg. Stability of critical cluster fields. *Mathematische Nachrichten*, 77(1):7–43, 1977.

[11] T. Kurtz, R. Lyons, R. Pemantle, and Y. Peres. A conceptual proof of the Kesten-Stigum theorem for multi-type branching processes. In K. B. Athreya and P. Jagers, editors, *Classical and modern branching processes (Minneapolis, MN, 1994)*, volume 84 of *IMA Vol. Math. Appl.*, pages 181–185. Springer, New York, 1997.

[12] Tzong-Yow Lee. Some large-deviation theorems for branching diffusions. *Ann. Probab.*, 20(3):1288–1309, 1992.

[13] R. Lyons. A simple path to Biggins' martingale convergence for branching random walk. In K. B. Athreya and P. Jagers, editors, *Classical and modern branching processes (Minneapolis, MN, 1994)*, volume 84 of *IMA Vol. Math. Appl.*, pages 217–221. Springer, New York, 1997.

[14] R. Lyons, R. Pemantle, and Y. Peres. Conceptual proofs of $L \log L$ criteria for mean behavior of branching processes. *Ann. Probab.*, 23(3):1125–1138, 1995.

[15] M. I. Roberts. A simple path to asymptotics for the frontier of a branching Brownian motion. Arxiv preprint: `http://arxiv.org/abs/1106.4771`, 2011.

[16] M.I. Roberts. *Spine changes of measure and branching diffusions*. PhD thesis, University of Bath, 2010. Available online: `http://people.bath.ac.uk/mir20/thesis.pdf`.

[17] S. Sawyer. Branching diffusion processes in population genetics. *Advances in Appl. Probability*, 8(4):659–689, 1976.

[18] S. R. S. Varadhan. *Large deviations and applications*, volume 46 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1984.