A new formulation of the spine approach to branching diffusions

Robert Hardy and Simon C. Harris

March 8, 2004

Abstract

We present a new formalization of the spine change of measure approach for branching diffusions that improves on the scheme laid out for branching Brownian motion in Kyprianou [20], which itself made use of earlier works of Lyons *et al* [23, 22, 19]. We use our new formulation to interpret certain 'Gibbs-Boltzmann' weightings of particles and use this to give a new, intuitive and more general proof of a 'Many-to-One' result which enables expectations of sums over particles in the branching diffusion to be calculated purely in terms of an expectation of one particle. Significantly, our formalization has already facilitated a variety of new, greatly simplified and more intuitive proofs in branching diffusions: see, for example, the $\mathcal{L}^p$ convergence of additive martingales in Hardy and Harris [12], the path large deviation results for branching Brownian motion in Hardy and Harris [10] and the large deviations for a continuous-typed branching diffusion in Git, J.W.Harris and S.C.Harris [8] and Hardy and Harris [11].

1 Overview of spine ideas

One of the central elements of the spine approach is to interpret the behaviour of a branching process under a certain change of measure. Such an interpretation was first laid out by Chauvin and Rouault [5] in the case of branching Brownian motion, and we first briefly review the main ideas on a heuristic level.

Consider a branching Brownian motion (BBM) with constant breeding rate r, that is, a branching process whereby particles diffuse independently according to a Brownian motion and at any moment undergo fission at a rate r to produce two particles that each evolve independently from their birth position, and so on. We suppose that the probabilities of this process are given by $\{P^x : x \in \mathbb{R}\}$, where P^x is a measure defined on the natural filtration $(\mathcal{F}_t)_{t\geq 0}$ such that it is the law of the process initiated from a single particle positioned at x. Suppose that the configuration of this branching Brownian motion at time t is given by the $\mathbb{R}$-valued point process $\mathbb{X}_t := \{X_u(t) : u \in N_t\}$ where N_t is the set of individuals alive at time t. It is well known that for any $\lambda \in \mathbb{R}$,

$$Z_\lambda(t) := \sum_{u\in N_t} e^{-rt} e^{\lambda X_u(t) - \frac{1}{2}\lambda^2 t} \tag{1}$$

defines a *strictly-positive* P-martingale, so $Z_\lambda(\infty) := \lim_{t\to\infty} Z_\lambda(t)$ is almost surely finite under P^x. The important contribution of Chauvian and Rouault [5] was to determine a pathwise construction of the measure $\mathbb{Q}^x_\lambda$ where

$$\left.\frac{\mathrm{d}\mathbb{Q}^x_\lambda}{\mathrm{d}P^x}\right|_{\mathcal{F}_t} = \frac{Z_\lambda(t)}{Z_\lambda(0)}, \tag{2}$$

where the term $Z_\lambda(0)$ acts as a normalizing factor.

Definition 1.1 *The measure $\mathbb{Q}_\lambda^x$ defined at (2) is equivalent to the following pathwise construction:*

- *starting from position x, the original ancestor diffuses according to a Brownian motion on $\mathbb{R}$ with drift λ;*
- *at rate $2r$ the particle undergoes fission producing two particles;*
- *with equal probability, one of these two particles is selected;*
- *this chosen particle repeats stochastically the behaviour of the parent;*
- *the other particle initiates, from its birth position, an independent copy of a $P^{\cdot}$ branching Brownian motion with branching rate r.*

The chosen line of descent in such pathwise constructions of the measure, here $\mathbb{Q}_\lambda$, has come to be known as the *spine* as it can be thought of as the backbone of the branching process $\mathbb{X}_t$ from which all particles are born. Although Chauvin and Rouault's work on the measure change continued in a paper co-authored with Wakolbinger [6], where the new measure is interpreted as the result of building a conditioned tree using the concepts of Palm measures, it wasn't until the so-called 'conceptual proofs' of Lyons, Kurtz, Peres and Pemantle published around 1995 ([23, 22, 19]) that the spine approach really began to crystalize. These papers laid out a formal basis for spines using a series of new measures on two underlying spaces of sample trees with and without distinguished lines of descent (the spine). Of particular interest is the paper by Lyons [22] which gave a spine-based proof of the $\mathcal{L}^1$-convergence of the well-known martingale for the Galton-Watson process. Here we first saw the *spine decomposition* of the martingale as the key to using the intuition provided by Chauvin and Rouault's pathwise construction of the new measure – Lyons used this together with a previously known measure-theoretic result on Radon-Nikodym derivatives that allows us to deduce the behaviour of the change-of-measure martingale under the original measure by investigating its behaviour under the second measure. Similar ideas have more recently been used by Kyprianou [20] to investigate the $\mathcal{L}^1$-convergence of the BBM martingale (1) as well as Olofsson [25], to mention a just a couple of more recent spine-based studies.

In this article we present a new formalization of the spine approach that improves the schemes originally laid out by Lyons *et al* [23, 22, 19] and later for BBM in Kyprianou [20]. Although the set-up costs of our spine formalization are quite large, at least in terms of definitions and notation, the underlying ideas are all extremely simple and intuitive. The power of this approach should not be underestimated. The further techniques developed subsequently in the papers by Hardy and Harris [12, 10, 11] and in Git *et. al.* [8] manage to completely bypass many previous technical problems and difficult non-linear calculations, with spine calculations facilitating the reduction to relatively straightforward classical one-particle situations.

In the first instance our improvements correct a perceived weakness in the Lyons *et al* scheme where one of the measures they defined had a time-dependent mass and could therefore not be normalized to be a probability measure and lacked a clear interpretation in terms of any direct process construction; an immediate consequence of this improvement is that here *all* measure changes are carried out by *martingales* and we regain a clear intuitive construction.

Another difference is in our use of filtrations and sub-filtrations, where Lyons *et al* instead used marginalizing. As we shall show, this brings substantial benefits since it allows us to relate the spine and the branching diffusion through the conditional-expectation operation, and in this way gives us a proper methodology for building *new* martingales for the branching diffusion based on known martingales for the spine. In other work we have used this powerful construction to obtain martingales for large-deviations problems, including a neat proof of a large deviations principle for branching Brownian motion – see Hardy and Harris [10, 11].

The conditional-expectation approach also leads directly to new and simple proofs of some key results for branching diffusions. The first of these concerns the relation that becomes clear between the spine and the 'Gibbs-Boltzmann' weightings for the branching particles. Such weightings are well known in the theory of branching process, and Harris [14] contains some analysis for a model of a typed branching diffusion that is similar in spirit to the models we shall be considering here. In our formulation these important weightings can be interpreted as a conditional expectation of a spine event, and we can use them to immediately obtain a new and very useful interpretation of the additive operations previously seen only within the context of the Kesten-Stigum theorem and related problems. As a second application of our approach we give a substantially easier proof of a far more general *Many-to-One* theorem than that used, for example, in Champneys *et al* [1] or Harris and Williams [13] as a key tool in their more classical approaches to branching diffusions.

2 Branching Markov models

Before we present the underlying constructions for spines, it will be useful to give the reader an idea of the branching-diffusion models that we have in mind for applications. We first discuss a finite-type branching diffusion, and then present a more general model that shall be used as the basis of our spine constructions in the following sections.

2.1 A finite-type branching diffusion

Suppose that for a fixed $n \in \mathbb{N}$ we are given two sets of positive constants $a(1), \dots, a(n)$ and $r(1), \dots, r(n)$. Consider a typed branching diffusion in which the type of each particle moves as a finite, irreducible and time-reversible Markov chain on the set $I := \{1, \dots, n\}$ with Q-matrix θQ (θ is a strictly positive constant that could be considered as the *temperature* of the system) and invariant measure $\pi = \{\pi(1), \dots, \pi(n)\}$. The spatial movement of a particle of type y is a driftless Brownian motion with instantaneous variance $a(y)$, so that if $\big(X_u(t), Y_u(t)\big) \in \mathbb{R} \times I$ is the space-type location of individual u at time t then we have

$$\mathrm{d}X_u(t) = a(Y_u(t))\, \mathrm{d}B_t$$

for a Brownian motion B_t. Fission of a particle of type y occurs at a rate $r(y)$ to produce two particles at the same space-type location as the parent.

We define $J := \mathbb{R} \times I$, and suppose that the configuration of this whole branching diffusion at time t is given by the J-valued point process $\mathbb{X}_t = \big\{\big(X_u(t), Y_u(t)\big) : u \in N_t\big\}$ where N_t is the set of individuals alive at time t. Let the measures $\big\{P^{x,y} : (x, y) \in \mathbb{R}^2\big\}$ on the filtered space $(\Omega, \mathcal{F}_\infty, (\mathcal{F}_t)_{t\geq 0})$ be such that under $P^{x,y}$ a single initial ancestor starting at (x, y) evolves as the branching diffusion, $\mathbb{X}_t$, as described above and where $(\mathcal{F}_t)_{t\geq 0}$ the natural filtration generated by the point process $\mathbb{X}_t$.

In this branching diffusion, each particle moves in a stochastically similar manner: let a process (ξ_t, η_t) on J under a measure $\mathbb{P}$ behaves stochastically like a single particle in the branching-diffusion $\mathbb{X}_t$ with no branching occurring. Thus, η_t is an irreducible, time-reversible Markov chain on I with Q-matrix θQ and invariant measure $\pi = \{\pi_1, \dots, \pi_n\}$, whilst ξ_t moves as a Brownian motion with zero drift and diffusion coefficient $a(y) > 0$ whenever η_t is in state y:

$$\mathrm{d}\xi_t = a(\eta_t)^{\frac{1}{2}}\, \mathrm{d}B_t,$$

for a $\mathbb{P}$-Brownian motion B_t. We note that the formal generator of this process (ξ_t, η_t) is:

$$\mathcal{H}F(x, y) = \frac{1}{2}a(y)\frac{\partial^2 F}{\partial x^2} + \theta \sum_{j\in I} Q(y, j)F(x, j), \qquad (F : J \to \mathbb{R}). \tag{3}$$

We shall often refer to the typed diffusion $(\xi_t, \eta_t) \in \mathbb{R} \times I$ as the *single particle model*, after the work carried out by Harris and Williams [13].

In Hardy [9] and Hardy and Harris [12] this finite-type branching diffusion has been investigated, and we briefly mention that the proofs have been based on the following two martingales, the first based on the whole branching diffusion and the second based only on the single-particle model:

$$Z_\lambda(t) := \sum_{u \in N_t} v_\lambda(Y_u(t)) e^{\lambda X_u(t) - E_\lambda t}, \tag{4}$$

$$\zeta_\lambda(t) := e^{\int_0^t R(\eta_s)\,\mathrm{d}s} v_\lambda(\eta_s) e^{\lambda \xi_t - E_\lambda t}, \tag{5}$$

where v_λ and E_λ satisfy:

$$\Big(\frac{1}{2}\lambda^2 A + \theta Q + R\Big) v_\lambda = E_\lambda v_\lambda,$$

which is to say that v_λ is an eigenvector of the matrix $\frac{1}{2}\lambda^2 A + \theta Q + R$, with eigenvalue E_λ. These two martingales should be compared with the corresponding martingales (1) and $e^{\lambda B_t - \frac{1}{2}\lambda^2 t}$ for BBM and a single Brownian motion respectively.

2.2 A general branching Markov process

The spine constructions in our formulation can be applied to a much more general branching Markov model, and we shall base the presentation on the following model, where particles move independently in a general space J as a stochastic copy of some given Markov process Ξ_t, and at a location-dependent rate undergo fission to produce a location-dependent random number of offspring that each carry on this branching behaviour independently.

Definition 2.1 (A general branching Markov process) *We suppose that three initial elements are given to us:*

- *a Markov process Ξ_t in a measurable set $(J, \mathcal{B})$,*
- *a measurable function $R : J \to [0, \infty)$,*
- *for each $x \in J$ we are given a random variable $A(x)$ whose probability distribution on the natural numbers $\{0, 1, \dots\}$ is $P\big(A(x) = k\big) = p_k(x)$, and whose mean is $m(x) := \sum_{k=0}^{\infty} k p_k(x)$.*

From these ingredients we can build a branching process in J according to the following recipe:

- *Each particle of the branching process will live, move and die in this space $(J, \mathcal{B})$, and if an individual u is alive at time t we refer to its location in J as $X_u(t)$. Therefore the time-t configuration of the branching process is a J-valued point process $\mathbb{X}_t := \{X_u(t) : u \in N_t\}$ where N_t denotes the collection of all particles alive at time t.*
- *For each individual u, the stochastic behaviour of its motion in J is an independent copy of the given process Ξ_t.*
- *The function $R : J \to [0, \infty)$ determines the rate at which each particle dies: given that u is alive at time t, its probability of dying in the interval $[t, t + \mathrm{d}t)$ is $R(X_u(t))\mathrm{d}t + o(\mathrm{d}t)$.*
- *If a particle u dies at location $x \in J$ it is replaced by $1 + A_u$ particles all positioned at x, where A_u is an independent copy of the random variable $A(x)$. All particles, once born, progress independently of each other.*

We suppose that the probabilities of this branching process are $\{P^x : x \in J\}$ where under P^x one initial ancestor starts out at x.

We shall first give a formal construction of the underlying probability space, made up of the sample trees of the branching process $\mathbb{X}_t$ in which the spines are the distinguished lines of descent. Once built, this space will be filtered in a natural way by the underlying family relationships of each sample tree, the diffusing branching particles and the diffusing spine, and then in section 4 we shall explain how we can define new probability measures $\tilde{P}^x$ that extend each P^x up to the finest filtration that contains all information about the spine and the branching particles.

For greater clarity, we often use the finite-type branching diffusion to introduce the ideas, before following up with the general formulation. For example, to deal with this finite-type model we would take the process Ξ_t to be the single-particle process (ξ_t, η_t) described in the above section, which lives in the space $J = \mathbb{R} \times I$. The birth rate of that model was given by the function $R(y)$ for all $y \in I$, and since we only dealt with binary branching we would suppose that $P(A(y) = 1) = 1$ for all $y \in I$.

Finally, we remark that much of the *notation* that we use for the underlying space of trees, the filtrations and the measures is closely related to that used in Kyprianou [20].

3 The underlying space for spines

3.1 Marked Galton-Watson trees with spines

The set of Ulam-Harris labels is to be equated with the set Ω of finite sequences of strictly-positive integers:

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

where we take $\mathbb{N} = \{1, 2, \ldots\}$. For two words $u, v \in \Omega$, uv denotes the concatenated word ($u\emptyset = \emptyset u = u$), and therefore Ω contains elements like '213' (or '$\emptyset$213'), which we read as 'the individual being the 3rd child of the 1st child of the 2nd child of the initial ancestor $\emptyset$'. For two labels $v, u \in \Omega$ the notation $v < u$ means that v is an *ancestor* of u, and $|u|$ denotes the length of u. The set of all ancestors of u is equally given by

$$\{v : v < u\} = \{v : \exists w \in \Omega \text{ such that } vw = u\}.$$

Collections of labels, ie. subsets of Ω, will therefore be groups of individuals. In particular, a subset $\tau \subset \Omega$ will be called a *Galton-Watson tree* if:

1. $\emptyset \in \tau$,
2. if $u, v \in \Omega$, then $uv \in \tau$ implies $u \in \tau$,
3. for all $u \in \tau$, there exists $A_u \in 0, 1, 2, \ldots$ such that $uj \in \tau$ if and only if $1 \le j \le 1 + A_u$, (where $j \in \mathbb{N}$).

That is just to say that a Galton-Watson tree:

1. has a single initial ancestor $\emptyset$,
2. contains all ancestors of any of its individuals v,
3. has the $1 + A_u$ children of an individual u labelled in a consecutive way,

and is therefore just what we imagine by the picture of a family tree descending from a single ancestor. Note that the '$1 \le j \le 1 + A_u$' condition in 3 means that each individual has *at least* one child, so that in our model we are insisting that Galton-Watson trees *never die out*.

The set of all Galton-Watson trees will be called $\mathbb{T}$. Typically we use the name τ for a particular tree, and whenever possible we will use the letters u or v or w to refer to the labels in τ, which we may also refer to as *nodes of* τ or *individuals in* τ or just as *particles*.

Each individual should have a *location* in J at each moment of its *lifetime*. Since a Galton-Watson tree $\tau \in \mathbb{T}$ in itself can express only the *family* structure of the individuals in our branching random walk, in order to give them these extra features we suppose that each individual $u \in \tau$ has a mark (X_u, σ_u) associated with it which we read as:

- $\sigma_u \in \mathbb{R}^+$ is the *lifetime* of u, which determines the *fission time* of particle u as $S_u := \sum_{v \leq u} \sigma_v$ (with $S_\emptyset := \sigma_\emptyset$). The times S_u may also be referred to as the *death* times;
- $X_u : [S_u - \sigma_u, S_u) \to J$ gives the *location* of u at time $t \in [S_u - \sigma_u, S_u)$.

To avoid ambiguity, it is always necessary to decide whether a particle is in existence or not at its death time.

Remark 3.1 *Our convention throughout will be that a particle u dies 'just before' its death time S_u (which explains why we have defined $X_u : [S_u - \sigma_u, S_u) \to \cdot$ for example). Thus at the time S_u the particle u has* disappeared*, replaced by its $1 + A_u$ children which are all alive and ready to go.*

We denote a single marked tree by (τ, X, σ) or (τ, M) for shorthand, and the set of all marked Galton-Watson trees by $\mathcal{T}$:

- $\mathcal{T} := \Big\{ (\tau, X, \sigma) : \tau \in \mathbb{T} \text{ and for each } u \in \tau, \sigma_u \in \mathbb{R}^+, X_u : [S_u - \sigma_u, S_u) \to J \Big\}$.
- For each $(\tau, X, \sigma) \in \mathcal{T}$, the set of particles that are alive at time t is defined as $N_t := \big\{ u \in \tau : S_u - \sigma_u \leq t < S_u \big\}$.

Where we want to highlight the fact that these values depend on the underlying marked tree we write e.g. $N_t((\tau, X, \sigma))$ or $S_u((\tau, M))$.

Any particle $u \in \tau$ that comes into existence creates a *subtree* made up from the collection of particles (and all their marks) that have u as an ancestor – and u is the original ancestor of this subtree.

- $(\tau, X, \sigma)^u_j$, or $(\tau, M)^u_j$ for shorthand, is defined as the *subtree* growing from individual u's jth child uj, where $1 \leq j \leq 1 + A_u$.

This subtree is a marked tree itself, but when considered as a part of the original tree we have to remember that it comes into existence at the space-time location $(X_u(0), S_u - \sigma_u)$ – which is just the space-time location of the death of particle u (and therefore the space-time location of the birth of its child uj).

Before moving on there is a final extension of the notation to be made: for any particle u we extend the definition of X_u from the time interval $[S_u - \sigma_u, S_u)$ to allow all earlier times $t \in [0, S_u)$:

Definition 3.2 *Each particle u is alive in the time interval $[S_u - \sigma_u, S_u)$, but we extend the concept of its path in J to all earlier times $t < S_u$:*

$$X_u(t) := \begin{cases} X_u(t) & \text{if } S_u - \sigma_u \leq t < S_u \\ X_v(t) & \text{if } v < u \text{ and } J_v \leq t < S_v \end{cases}$$

Thus particle u inherits the path of its unique line of ancestors, and this simple extension will allow us to later write expressions like $e^{\int_0^t f(s)\,\mathrm{d}X_u(s)}$ whenever $u \in N_t$, without worrying about the birth time of u.

For any given marked tree $(\tau, M) \in \mathcal{T}$ we can identify distinguished lines of descent from the initial ancestor: $\emptyset, u_1, u_2, u_3, \ldots \in \tau$, in which u_3 is a child of u_2, which itself is a child of u_1 which is a child of the original ancestor $\emptyset$. We'll call such a subset of τ a *spine*, and will refer to it as ξ:

- a spine ξ is a subset of nodes $\{\emptyset, u_1, u_2, u_3, \ldots\}$ in the tree τ that make up a unique line of descent. We use ξ_t to refer to the unique node in ξ that that is alive at time t.

In a more formal definition, which can for example be found in the paper by Rouault and Liu [21], a spine is thought of as a point on $\partial\tau$ the boundary of the tree – in fact the boundary is *defined* as the set of all infinite lines of descent. This explains the notation $\xi \in \partial\tau$ in the following definition: we augment the space $\mathcal{T}$ of marked trees to become

- $\tilde{\mathcal{T}} := \Big\{(\tau, M, \xi) : (\tau, M) \in \mathcal{T} \text{ and } \xi \in \partial\tau\Big\}$ is the set of *marked trees with distinguished spines.*

It is natural to speak of the *position of the spine at time t* which think of just as the position of the unique node that is in the spine and alive at time t:

- we define the time-t position of the spine as $\xi_t := X_u(t)$, where $u \in \xi \cap N_t$.

By using the notation ξ_t to refer to both the node in the tree and that node's spatial position we are introducing potential ambiguity, but in practice the context will make clear which we intend. However, in case of needing to emphasize, we shall give the node a longer name:

- $\mathrm{node}_t((\tau, M, \xi)) := u$ if $u \in \xi$ is the node in the spine alive at time t,

which may also be written as $\mathrm{node}_t(\xi)$.

Finally, it will later be important to know how many fission times there have been in the spine, or what is the same, to know which generation of the family tree the node ξ_t is in (where the original ancestor $\emptyset$ is considered to be the 0th generation)

Definition 3.3 *We define the counting function*

$$n_t = \big|node_t(\xi)\big|,$$

which tells us which generation the spine node is in, or equivalently how many fission times there have been on the spine. For example, if $\xi_t = \big(\emptyset, u_1, u_2\big)$ *then both* $\emptyset$ *and* u_1 *have died and so* $n_t = 2$.

3.2 Filtrations

The reader who is already familiar with the Lyons *et al* [19, 22, 23] papers will recall that they used two separate underlying spaces of marked trees *with* and *without* the spines, then marginalized out the spine when wanting to deal only with the branching particles as a whole. Instead, we are going to use just the single underlying space $\tilde{\mathcal{T}}$, but define *four* filtrations of it that encapsulate different knowledge.

3.2.1 Filtration $(\mathcal{F}_t)_{t \geq 0}$

We define a filtration of $\tilde{\mathcal{T}}$ made up of the σ-algebras:

$$\mathcal{F}_t := \sigma\Big((u, X_u, \sigma_u) : S_u \leq t \;;\; (u, X_u(s) : s \in [S_u - \sigma_u, t]) : t \in [S_u - \sigma_u, S_u)\Big),$$

which in words means that $\mathcal{F}_t$ is generated by all the information regarding the branching particles that have *lived and died* before time t (this is the condition $S_u \leq t$), along with just the information up to time t of those particles *still alive* at time t (this is the $t \in [S_u - \sigma_u, S_u)$ condition). Each of these σ-algebras will be a subset of the limit defined as

$$\mathcal{F}_\infty := \sigma\Big(\bigcup_{t \geq 0} \mathcal{F}_t\Big).$$

3.2.2 Filtration $(\tilde{\mathcal{F}}_t)_{t\geq 0}$

In order to know about the spine, we make this filtration finer, defining $\tilde{\mathcal{F}}_t$ by adding into $\mathcal{F}_t$ the knowledge of which node is the spine at time t:

$$\tilde{\mathcal{F}}_t := \sigma\big(\mathcal{F}_t, \mathrm{node}_t(\xi)\big), \qquad \tilde{\mathcal{F}}_\infty := \sigma\Big(\bigcup_{t\geq 0} \tilde{\mathcal{F}}_t\Big).$$

Consequently this filtration knows *everything* about the branching process and *everything* about the spine: it knows which nodes make up the spine, when they were born, when they died (ie. the fission times S_u), and their family sizes.

3.2.3 Filtration $(\mathcal{G}_t)_{t\geq 0}$

We define a filtration of $\tilde{\mathcal{T}}$, $\big\{\mathcal{G}_t\big\}_{t\geq 0}$, where the σ-algebras

$$\mathcal{G}_t := \sigma\big(\xi_s : 0 \leq s \leq t\big), \qquad \mathcal{G}_\infty := \sigma\Big(\bigcup_{t\geq 0} \mathcal{G}_t\Big),$$

are generated by *only* the spatial motion of the spine in the J. Note that the events $G \in \mathcal{G}_t$ do not know which *nodes* of the tree τ actually make up the spine.

3.2.4 Filtration $(\tilde{\mathcal{G}}_t)_{t\geq 0}$

We augment $\mathcal{G}_t$ by adding in information on the nodes that make up the spine (as we did from $\mathcal{F}_t$ to $\tilde{\mathcal{F}}_t$), as well as the knowledge of when the fission times occurred on the spine and how big the families were that were produced:

$$\tilde{\mathcal{G}}_t := \sigma\big(\mathcal{G}_t, (\mathrm{node}_s : s \leq t), (A_u : u < \xi_t)\big), \qquad \tilde{\mathcal{G}}_\infty := \sigma\Big(\bigcup_{t\geq 0} \tilde{\mathcal{G}}_t\Big).$$

3.2.5 Summary

In brief, the key filtrations we shall make key use of are:

- $\mathcal{F}_t$ knows everything that has happened to all the branching particles up to the time t, *but does not know which one is the spine*;
- $\tilde{\mathcal{F}}_t$ knows everything that $\mathcal{F}_t$ knows and also knows which line of descent is the spine (it is in fact the finest filtration);
- $\mathcal{G}_t$ knows only about the spine's motion in J up to time t, but does not actually know which line of descent in the family tree makes up the spine;
- $\tilde{\mathcal{G}}_t$ knows about the spine's motion and also knows which nodes it is composed of. Furthermore it knows about the fission times of these nodes and how many children were born at each time.

We note the following obvious relationships between the different filtrations of $\tilde{\mathcal{T}}$:

$$\mathcal{G}_t \subset \tilde{\mathcal{G}}_t \subset \tilde{\mathcal{F}}_t$$
$$\mathcal{F}_t \subset \tilde{\mathcal{F}}_t$$

Trivially, we also note that $\mathcal{G}_t \not\subseteq \mathcal{F}_t$, since the filtration $\mathcal{F}_t$ does not know *which* line of descent makes up the spine.

4 Probability measures

Having now carefully defined the underlying space for our probabilities, we remind ourselves of the probability measures:

Definition 4.1 *For each $x \in J$, let P^x be the measure on $(\tilde{\mathcal{T}}, \mathcal{F}_\infty)$ such that the filtered probability space $(\tilde{\mathcal{T}}, \mathcal{F}_\infty, (\mathcal{F}_t)_{t\geq 0}, P^\cdot)$ is the canonical model for $\mathbb{X}_t$, the branching Markov process described in definition 2.1.*

For details of how the measures P^x are formally constructed on the underlying space of trees, we refer the reader to the work of Neveu [24] and Chauvin [4, 2].

All spine approaches rely on building a measure $\tilde{P}^x$ under which the spine is a single genealogical line of descent chosen from the underlying tree. If we are given a sample tree (τ, M) for the branching process, it is easy to verify that if at each fission we make a uniform choice amongst the offspring to decide which line of descent continues the spine ξ, then if $u \in \tau$ we have

$$\mathrm{Prob}(u \in \xi) = \prod_{v<u} \frac{1}{1+A_v}. \tag{6}$$

This observation is the key to our method for extending the measures, and for this we make use of the following representation found in Lyons [22].

Theorem 4.2 *If f is a $\tilde{\mathcal{F}}_t$-measurable function then we can write:*

$$f = \sum_{u\in N_t} f_u \mathbf{1}_{(\xi_t = u)} \tag{7}$$

where f_u is $\mathcal{F}_t$-measurable.

As an example of this, in the case of the finite-typed branching diffusion of section 2.1, such a representation would be:

$$e^{\int_0^t R(\eta_s)\,\mathrm{d}s} v_\lambda(\eta_t)\, e^{\lambda\xi_t - E_\lambda t} = \sum_{u\in N_t} e^{\int_0^t R(Y_u(s))\,\mathrm{d}s} v_\lambda(Y_u(t))\, e^{\lambda X_u(t) - E_\lambda t}\, \mathbf{1}_{(\xi_t=u)}. \tag{8}$$

Definition 4.3 *Given the measure P^x on $(\tilde{\mathcal{T}}, \mathcal{F}_\infty)$ we extend it to the probability measure $\tilde{P}^x$ on $(\tilde{\mathcal{T}}, \tilde{\mathcal{F}}_\infty)$ by defining*

$$\int_{\tilde{\mathcal{T}}} f \,\mathrm{d}\tilde{P}^x := \int_{\mathcal{T}} \sum_{u\in N_t} f_u \prod_{v<u} \frac{1}{1+A_v}\,\mathrm{d}P^x, \tag{9}$$

for each $f \in m\tilde{\mathcal{F}}_t$ with representation like (7).

Previous approaches to spines, exemplified in Lyons [22], used the idea of *fibres* to get a measure analogous to our $\tilde{P}$ that could measure the spine. However, a weakness in this approach was that the corresponding measure had time-dependent and unbounded total mass and therefore was not normalized to become a probability measure. Our new idea of using the down-weighting term $1/(1+A_v)$ in the definition of $\tilde{P}$ is crucial in ensuring that we have a *probability* measure with a clear intuitive construction (look back at equation (6) and ahead to Lemma 6.6) and also where *all* subsequent measure changes in our formulation are carried out by *martingales*.

Theorem 4.4 *This measure $\tilde{P}^x$ really is an extension of P^x in that $P = \tilde{P}|_{\mathcal{F}_\infty}$.*

Proof: If $f \in m\mathcal{F}_t$ then the representation (7) is trivial and therefore by definition

$$\int_{\tilde{\mathcal{T}}} f\,\mathrm{d}\tilde{P} = \int_{\tilde{\mathcal{T}}} f \times \Big(\sum_{u\in N_t}\prod_{v<u}\frac{1}{1+A_v}\Big)\,\mathrm{d}P.$$

However, it can be shown that $\sum_{u\in N_t}\prod_{v<u}\frac{1}{1+A_v}=1$ by retracing the sum back through the lines of ancestors to the original ancestor $\emptyset$, factoring out the product terms as each generation is passed. Thus

$$\int_{\tilde{\mathcal{T}}} f\,\mathrm{d}\tilde{P}=\int_{\tilde{\mathcal{T}}} f\,\mathrm{d}P.$$

□

Definition 4.5 *The filtered probability space $(\tilde{\mathcal{T}},\tilde{\mathcal{F}}_\infty,(\tilde{\mathcal{F}}_t)_{t\geq 0},\tilde{P})$ together with $(\mathbb{X}_t,\xi_t)$ will be referred to as the* ***canonical model with spines****.*

In the single-particle model of section 2.1 we assumed the existence of a separate measure $\mathbb{P}$ and a process (ξ_t,η_t) that behaved stochastically like a 'typical' particle in the typed branching diffusion $\mathbb{X}_t$. In our formalization the *spine* is exactly the single-particle model:

Definition 4.6 *We define the measure $\mathbb{P}$ on $\big(\tilde{\mathcal{T}},\mathcal{G}_\infty\big)$ as the projection of $\tilde{P}$:*

$$\mathbb{P}|_{\mathcal{G}_t}:=\tilde{P}|_{\mathcal{G}_t}.$$

Under the measure $\mathbb{P}$ the spine process ξ_t has exactly the same law as Ξ_t.

Definition 4.7 *The filtered probability space $(\tilde{\mathcal{T}},\mathcal{G}_\infty,(\mathcal{G}_t)_{t\geq 0},\mathbb{P})$ together with the spine process ξ_t will be referred to as the* ***single-particle model****.*

5 Martingales

Starting with the single Markov process Ξ_t that lives in $(J,\mathcal{B})$ we have built $(\mathbb{X}_t,\xi_t)$, a branching Markov process with spines, in which the spine ξ_t behaves stochastically like the given Ξ_t. In this section we are going to show how *any* given martingale for the spine $\zeta(t)$ leads to a corresponding additive martingale for the whole branching model.

We have actually seen an example of this already. For the finite-type model of section 2.1 we met two martingales:

$$Z_\lambda(t):=\sum_{u\in N_t} v_\lambda\big(Y_u(t)\big)e^{\lambda X_u(t)-E_\lambda t},\tag{10}$$

$$\zeta_\lambda(t):=e^{\int_0^t R(\eta_s)\,\mathrm{d}s}v_\lambda(\eta_t)e^{\lambda\xi_t-E_\lambda t}.\tag{11}$$

Just from their very form it has always been clear that they are closely related. What we shall be demonstrating in full generality in this section is that the key to their relationship comes through the following $\tilde{\mathcal{F}}_t$-measurable martingale.

Definition 5.1 *We define an $\tilde{\mathcal{F}}_t$-measurable martingale:*

$$\tilde{\zeta}_\lambda(t):=\prod_{u<\xi_t}(1+A_v)\times v_\lambda(\eta_t)e^{\lambda\xi_t-E_\lambda t}.\tag{12}$$

The important result that we show in this article, in a more general form, is that $Z_\lambda(t)$ and $\zeta_\lambda(t)$ are just conditional expectations of this new martingale $\tilde{\zeta}_\lambda$:

- $Z_\lambda(t)=\tilde{P}\big(\,\tilde{\zeta}_\lambda(t)\,|\,\mathcal{F}_t\big)$,
- $\zeta_\lambda(t)=\tilde{P}\big(\,\tilde{\zeta}_\lambda(t)\,|\,\mathcal{G}_t\big)$.

We emphasize that this relationship has not previously been formalized, and that it is only *possible* because of our new approach to the definition of $\tilde{P}$ as a *probability* measure, and of our using filtrations to capture the different knowledge generated by the spine and the branching particles.

Furthermore, in the general form that we present below it provides a consistent methodology for using well-known martingales for a single process ξ_t to get new additive martingales for the related branching process. In Hardy and Harris [10, 11] we use these powerful ideas to give substantially easier proofs of large-deviations problems in branching diffusions than have previously been possible.

Suppose that $\zeta(t)$ is a $\big(\tilde{\mathcal{T}}, (\mathcal{G}_t)_{t\geq 0}, \tilde{P}\big)$-martingale, which is to say that it is a $\mathcal{G}_t$-measurable function that is a martingale with respect to the measure $\tilde{P}$. For example, in the case of our finite-type branching diffusion this could be the martingale $\zeta_\lambda(t)$ which is $\mathcal{G}_t$-measurable since it refers only to the spine process (ξ_t, η_t).

Definition 5.2 *We shall call $\zeta(t)$ a **single-particle martingale**, since it is $\mathcal{G}_t$-measurable and thus depends only to the spine ξ.*

Any such single-particle martingale can be used to define an additive martingale for the whole branching process via the representation (7).

Definition 5.3 *Suppose that we can represent the martingale $\zeta(t)$ as*

$$\zeta(t) = \sum_{u\in N_t} \zeta_u(t)\mathbf{1}_{(\xi_t=u)}, \tag{13}$$

for $\zeta_u(t) \in m\mathcal{F}_t$, as at (7). We can then define an $\mathcal{F}_t$-measurable process $Z(t)$ as

$$Z(t) := \sum_{u\in N_t} e^{-\int_0^t m(X_u(s))R(X_u(s))\,\mathrm{d}s}\zeta_u(t),$$

*and refer to $Z(t)$ as the **branching-particle martingale**.*

The martingale property $Z(t)$ will be established in Lemma 5.7 after first building another martingale, $\tilde{\zeta}(t)$, from the single-particle martingale $\zeta(t)$. First, for clarity, we take a moment to discuss this definition of the additive martingale and the terms like $\zeta_u(t)$.

If we return to our familiar martingales (10) and (11), it is clear that

$$\zeta_\lambda(t) = \sum_{u\in N_t} e^{\int_0^t R(Y_u(s))\,\mathrm{d}s} v_\lambda\big(Y_u(t)\big)e^{\lambda X_u(t)-E_\lambda t}\,\mathbf{1}_{(\xi_t=u)}. \tag{14}$$

The 'ζ_u' terms of (13) could be here replaced with a more descriptive notation $\zeta_\lambda[(X_u, Y_u)](t)$, where

$$\zeta_u(t) = \zeta_\lambda[(X_u, Y_u)](t) := e^{\int_0^t R(Y_u(s))\,\mathrm{d}s} v_\lambda\big(Y_u(t)\big)e^{\lambda X_u(t)-E_\lambda t},$$

can be seen to essentially be a functional of the space-type path $(X_u(t), Y_u(t))$ of particle u. In this way the original single-particle martingale ζ_λ would be understood as a functional of the space-type path (ξ_t, η_t) of the spine itself and we could write

$$\zeta_\lambda(t) = \zeta_\lambda[(\xi, \eta)](t) = \sum_{u\in N_t} \zeta_\lambda[(X_u, Y_u)](t)\,\mathbf{1}_{(\xi_t=u)}.$$

This is the idea behind the representation (13), and in those typical cases where the single-particle martingale is essentially a functional of the paths of the spine ξ_t, as is the case for our $\zeta_\lambda(t)$, we should just think of ζ_u as being that same functional but evaluated over the path $X_u(t)$

of particle u rather than the spine ξ_t. The representation (13) is used as a more general way of treating single-particle martingales that perhaps are not such a simple functional of the spine path.

Finally, from (14) it is clear that the additive martingale being defined by definition 5.3 is our familiar $Z_\lambda(t)$:

$$\begin{aligned} Z_\lambda(t) &= \sum_{u\in N_t} v_\lambda\big(Y_u(t)\big)e^{\lambda X_u(t)-E_\lambda t} \\ &= \sum_{u\in N_t} e^{-\int_0^t R(Y_u(s))\,\mathrm{d}s}\zeta_\lambda[(X_u,Y_u)](t). \end{aligned}$$

Although definition 5.3 will work in general, in the main the spine approach is interested in martingales that can act as Radon-Nikodym derivatives between probability measures, and therefore we suppose from now on that $\zeta(t)$ is *strictly positive*, and therefore that the additive martingale $Z(t)$ is strictly positive.

The work of Lyons *et al* [22, 19, 23], that of Chauvin and Rouault [5] and more recently of Kyprianou [20] suggests that when a change of measure is carried out with a branching-diffusion additive martingale like $Z(t)$ it is typical to expect three changes: the spine will gain a drift, its fission times will be increased and the distribution of its family sizes will be size-biased. In section 6.1 we shall confirm this, and now take a separate look at the martingales that could perform these changes, and which we shall combine to obtain a martingale $\tilde{\zeta}(t)$ that will ultimately be used to change the measure $\tilde{P}$.

Theorem 5.4 *The expression*

$$\prod_{v<\xi_t}\big(1+m(\xi_{S_v})\big)e^{-\int_0^t m(\xi_s)R(\xi_s)\,\mathrm{d}s}$$

is a $\tilde{P}$-martingale that will increase the rate at which fission times occur along the spine from $R(\xi_t)$ to $(1+m(\xi_t))R(\xi_t)$:

$$\frac{\mathrm{d}\mathbb{L}_t^{((1+m(\xi_t))R(\xi_t))}}{\mathrm{d}\mathbb{L}_t^{(R(\xi_t))}} = \prod_{v<\xi_t}\big(1+m(\xi_{S_v})\big)e^{-\int_0^t m(\xi_s)R(\xi_s)\,\mathrm{d}s}$$

where $\mathbb{L}^{(R(\xi_t))}$ is the law of the Poisson (Cox) process with rate $R(\xi_t)$ at time t.

Theorem 5.5 *The term $\prod_{v<\xi_t}\frac{1+A_v}{1+m(\xi_{S_v})}$ is a $\tilde{P}$-martingale that will change the measure by size-biasing the family sizes born from the spine:*

$$\textit{if } v<\xi_t,\ \textit{then} \qquad \textit{Prob}(A_v=k) = \frac{(1+k)p_k(\xi_{S_v})}{1+m(\xi_{S_v})}.$$

The product of these two martingales with the single-particle martingale $\zeta(t)$ will simultaneously perform the three changes mentioned above.

Definition 5.6 *We define a $\tilde{\mathcal{F}}_t$-measurable martingale as*

$$\begin{aligned} \tilde{\zeta}(t) &:= \prod_{v<\xi_t}(1+A_v)e^{-\int_0^t m(\xi_s)R(\xi_s)\,\mathrm{d}s}\times\zeta(t) \\ &= \prod_{v<\xi_t}\frac{1+A_v}{1+m(\xi_{S_v})}\times\prod_{v<\xi_t}\big(1+m(\xi_{S_v})\big)e^{-m\int_0^t R(\xi_s)\,\mathrm{d}s}\times\zeta(t). \end{aligned} \tag{15}$$

This martingale is the general form of $\tilde{\zeta}_\lambda(t)$ that we defined earlier for our finite-type model.

The real importance of the size-biasing and fission-time-increase operations is that they introduce the correct terms into $\tilde{\zeta}(t)$ so that the following relationships hold:

Lemma 5.7 *Both $Z(t)$ and $\zeta(t)$ are projections of $\tilde{\zeta}(t)$ onto their filtrations: for all t,*

- $Z(t) = \tilde{P}\big(\tilde{\zeta}(t)\,|\,\mathcal{F}_t\big)$,
- $\zeta(t) = \tilde{P}\big(\tilde{\zeta}(t)\,|\,\mathcal{G}_t\big)$.

Proof: We use the representation (7) of $\tilde{\zeta}(t)$:

$$\tilde{\zeta}(t) = \sum_{u\in N_t}\prod_{v<u}(1+A_v)e^{-\int_0^t m(X_u(s))R(X_u(s))\,\mathrm{d}s}\zeta_u(t)\mathbf{1}_{(\xi_t=u)}. \tag{16}$$

From this it follows that

$$\begin{aligned}\tilde{P}\big(\tilde{\zeta}(t)|\mathcal{F}_t\big) &= \sum_{u\in N_t} e^{-\int_0^t m(X_u(s))R(X_u(s))\,\mathrm{d}s}\zeta_u(t)\times\prod_{v<u}(1+A_v)\,\tilde{P}\big(\mathbf{1}_{(\xi_t=u)}|\mathcal{F}_t\big)\\ &= \sum_{u\in N_t} e^{-\int_0^t m(X_u(s))R(X_u(s))\,\mathrm{d}s}\zeta_u(t) = Z(t),\end{aligned}$$

since $\tilde{P}\big(\mathbf{1}_{(\xi_t=u)}|\mathcal{F}_t\big) = \mathbf{1}_{(u\in N_t)}\times\prod_{v<u}(1+A_v)^{-1}$.

On the other hand, the martingale terms in (15) imply

$$\tilde{P}\big(\tilde{\zeta}(t)|\mathcal{G}_t\big) = \zeta(t)\times\tilde{P}\Big(\prod_{v<\xi_t}(1+A_v)e^{-\int_0^t m(\xi_s)R(\xi_s)\,\mathrm{d}s}\big|\mathcal{G}_t\Big) = \zeta(t),$$

$\square$

6 Changing the measures

In Hardy [9] the single-particle martingale $\zeta_\lambda(t)$ defined at (5) is used to define new measures for the single-particle model, via

$$\left.\frac{\mathrm{d}\mathbb{P}_\lambda}{\mathrm{d}\mathbb{P}}\right|_{\mathcal{G}_t} = \frac{\zeta_\lambda(t)}{\zeta(0)}.$$

We have seen the close relationships between the three martingales ζ_λ, Z_λ and $\tilde{\zeta}_\lambda$:

$$Z_\lambda(t) = \tilde{P}\big(\tilde{\zeta}_\lambda(t)\,|\,\mathcal{F}_t\big), \qquad \zeta_\lambda(t) = \tilde{P}\big(\tilde{\zeta}_\lambda(t)\,|\,\mathcal{G}_t\big),$$

and in this section we show in a more general form how these close relationships mean that a new measure $\tilde{\mathbb{Q}}_\lambda$ defined in terms of $\tilde{P}$ as

$$\left.\frac{\mathrm{d}\tilde{\mathbb{Q}}_\lambda}{\mathrm{d}\tilde{P}}\right|_{\tilde{\mathcal{F}}_t} = \frac{\tilde{\zeta}_\lambda(t)}{\tilde{\zeta}_\lambda(0)},$$

will induce measure changes on the sub-filtrations $\mathcal{G}_t$ and $\mathcal{F}_t$ of $\tilde{\mathcal{F}}_t$ whose Radon-Nikodym derivatives are given by $\zeta_\lambda(t)$ and $Z_\lambda(t)$ respectively.

We recall that in our set up we have a finest filtration $(\tilde{\mathcal{F}}_t)_{t\geq 0}$ associated with the measure $\tilde{P}$, and two sub-filtrations $(\mathcal{F}_t)_{t\geq 0}$ with measure P and $(\mathcal{G}_t)_{t\geq 0}$ with measure $\mathbb{P}$. The martingale $\tilde{\zeta}$ can change the measure $\tilde{P}$:

Definition 6.1 *A measure* $\tilde{\mathbb{Q}}$ *on* $(\tilde{\mathcal{T}}, \tilde{\mathcal{F}}_\infty)$ *is defined via its Radon-Nikodym derivative with respect to* $\tilde{P}$*:*

$$\left.\frac{\mathrm{d}\tilde{\mathbb{Q}}}{\mathrm{d}\tilde{P}}\right|_{\tilde{\mathcal{F}}_t} = \frac{\tilde{\zeta}(t)}{\tilde{\zeta}(0)}.$$

Precisely, this notation means that for each event $F \in \tilde{\mathcal{F}}_t$ *we define* $\tilde{\mathbb{Q}}(F) := \tilde{P}\big(\mathbf{1}_F Z(t)\big)$.

As we did for the measures P and $\mathbb{P}$, we can restrict $\tilde{\mathbb{Q}}$ to the sub-filtrations:

Definition 6.2 *We define the measure* $\mathbb{Q}$ *on* $(\tilde{\mathcal{T}}, \mathcal{F}_\infty, (\mathcal{F}_t)_{t\geq 0})$ *via*

$$\mathbb{Q} := \tilde{\mathbb{Q}}|_{\mathcal{F}_\infty}.$$

Definition 6.3 *We define the measure* $\hat{\mathbb{P}}$ *on* $(\tilde{\mathcal{T}}, \mathcal{G}_\infty, (\tilde{\mathcal{G}}_t)_{t\geq 0})$ *via*

$$\hat{\mathbb{P}} := \tilde{\mathbb{Q}}|_{\mathcal{G}_\infty}.$$

A consequence of our new formulation in terms of filtrations and the equalities of Lemma 5.7 is that the changes of measure are carried out by $Z(t)$ and $\zeta(t)$ on their subfiltrations:

Theorem 6.4

$$\left.\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}P}\right|_{\mathcal{F}_t} = \frac{Z(t)}{Z(0)}, \qquad \textit{and} \qquad \left.\frac{\mathrm{d}\hat{\mathbb{P}}}{\mathrm{d}\mathbb{P}}\right|_{\mathcal{G}_t} = \frac{\zeta(t)}{\zeta(0)}.$$

Proof: These two results actually follow from a more general observation that if $\tilde{\mu}_1$ and $\tilde{\mu}_2$ are two measures defined on a measure space $(\Omega, \tilde{\mathcal{S}})$ with Radon-Nikodym derivative

$$\frac{\mathrm{d}\tilde{\mu}_2}{\mathrm{d}\tilde{\mu}_1} = f, \tag{17}$$

and if $\mathcal{S}$ is a sub-σ-algebra of $\tilde{\mathcal{S}}$, then the two measures $\mu_1 := \tilde{\mu}_1|_{\mathcal{S}}$ and $\mu_2 := \tilde{\mu}_2|_{\mathcal{S}}$ on $(\Omega, \mathcal{S})$ are related by the conditional expectation operation:

$$\frac{\mathrm{d}\mu_2}{\mathrm{d}\mu_1} = \tilde{\mu}_1\big(f|\mathcal{S}\big).$$

The proof of this is that if $g \in m\mathcal{S}$ and $S \in \mathcal{S}$ then

$$\begin{aligned}
\int_S g\,\mathrm{d}\mu_2 &= \int_S g\,\mathrm{d}\tilde{\mu}_2 && \text{since } g \text{ is also in } m\tilde{\mathcal{S}}, \text{ and } S \in \mathcal{S} \text{ too,}\\
&= \int_S g\,f\,\mathrm{d}\tilde{\mu}_1 && \text{by (17),}\\
&= \int_S \tilde{\mu}_1\big(gf|\mathcal{S}\big)\,\mathrm{d}\tilde{\mu}_1 && \text{by definition of the conditional expectation,}\\
&= \int_S g\,\tilde{\mu}_1\big(f|\mathcal{S}\big)\,\mathrm{d}\tilde{\mu}_1 && \text{since } g \text{ is } \mathcal{S}\text{-measurable,}\\
&= \int_S g\,\tilde{\mu}_1\big(f|\mathcal{S}\big)\,\mathrm{d}\mu_1 && \text{since everything is in } m\mathcal{S}.
\end{aligned}$$

Applying this general result (17) using the relationships between the general martingales given in Lemma 5.7 concludes the proof. □

6.1 Understanding the measure $\tilde{\mathbb{Q}}$

As the name suggests, we should be able to think of the spine as the backbone of the branching process. This is made precise by the following decomposition originally given by Chauvin and Rouault [5], and which now has become a standard part of most spine-based studies. We state this in a form taken from Kyprianou's spine-proofs for BBM [20]:

Theorem 6.5 *The measure $\tilde{P}$ on $\tilde{\mathcal{F}}_t$ can be decomposed as:*

$$\mathrm{d}\tilde{P}(\tau, M, \xi) = \mathrm{d}\mathbb{P}(\xi)\mathrm{d}\mathbb{L}^{(R(\xi_t))}(n_t) \prod_{v<\xi_t} \frac{1}{1+A_v} \prod_{v<\xi_t} p_{A_v}(\xi_{S_v}) \prod_{j=1}^{A_v} \mathrm{d}P\big((\tau, M)_j^v\big), \tag{18}$$

where $\mathbb{L}^{(R(\xi_t))}$ is the law of the Poisson (Cox) process with rate $R(\xi_t)$ at time t, and we remember that n_t counts the number of fission times on the spine before time t.

We can offer a clear intuitive picture of this decomposition, which we summarize in the following lemma.

Lemma 6.6 *The decomposition of measure $\tilde{P}$ at (18) enables the following construction:*

1. *the spine's motion is determined by the single-particle measure $\mathbb{P}$;*
2. *the spine undergoes fission at time t at rate $R(\xi_t)$;*
3. *at the fission time of node v on the spine, the single spine particle is replaced by $1 + A_v$ children, with A_v being chosen independently and distributed according to the location-dependent random variable $A(\xi_{S_v})$ with probabilities $(p_k(\xi_{S_v}) : k = 0, 1, \ldots)$;*
4. *the spine is chosen uniformly from the $1 + A_v$ children at the fission point v;*
5. *each of the remaining A_v children gives rise to the independent subtrees $(\tau, M)_j^v$, for $1 \leq j \leq A_v$, which are not part of the spine and which are each determined by an independent copy of the original measure P shifted to their point and time of creation.*

This decomposition of $\tilde{P}_t$ given at (18) will allow us to interpret the measure $\tilde{\mathbb{Q}}$ if we appropriately factor the components of the change-of-measure martingale $\tilde{\zeta}(t)$ across this representation. On $\tilde{\mathcal{F}}_t$,

$$\begin{aligned}
\mathrm{d}\tilde{\mathbb{Q}} &= \tilde{\zeta}(t)\,\mathrm{d}\tilde{P} \\
&= \zeta(t) \times e^{-\int_0^t R(\xi_s)\mathrm{d}s}\big(1 + m(\xi_{S_v})\big)^{n_t} \times \prod_{v<\xi_t} \frac{1 + A_v}{1 + m(\xi_{S_v})} \times \mathrm{d}\tilde{P} \\
&= \mathrm{d}\hat{\mathbb{P}}(\xi)\,\mathrm{d}\mathbb{L}^{((1+m(\xi_t))R(\xi_t))}(n) \prod_{v<\xi_t} \frac{1}{1+A_v} \prod_{v<\xi_t} \frac{1+A_v}{1+m(\xi_{S_v})} p_{A_v} \prod_{j=1}^{A_v} \mathrm{d}P\big((\tau, M)_j^v\big).
\end{aligned} \tag{19}$$

Just as we did for $\tilde{P}$, we can offer a clear interpretation of this decomposition:

Lemma 6.7 *Under the measure $\tilde{\mathbb{Q}}$,*

1. *the spine process ξ_t moves as if under the* changed *measure $\hat{\mathbb{P}}$;*
2. *the fission times along the spine occur at an* accelerated *rate $(1 + m(\xi_t))R(\xi_t)$;*
3. *at the fission time of node v on the spine, the single spine particle is replaced by $1 + A_v$ children, with A_v being chosen independently according to the* tilted *offspring distribution $((1 + k)p_k(\xi_{S_v})/(1 + m(\xi_{S_v})) : k = 0, 1, \ldots)$;*

4. *the spine is chosen uniformly from the* $1 + A_v$ *particles at the fission point* v*;*
5. *each of the remaining* A_v *children gives rise to the independent subtrees* $(\tau, M)^v_j$*, for* $1 \leq j \leq A_v$*, which are not part of the spine and evolve as independent processes determined by the measure* P *shifted to their point and time of creation.*

Such an interpretation of the measure $\tilde{\mathbb{Q}}$ was first given by Chauvin and Rouault [5] in the context of BBM, allowing them to come to the important conclusion that under the new measure $\mathbb{Q}$ the branching diffusion remains largely unaffected, except that the brownian particles of a single (random) line of descent in the family tree are given a changed motion, with an accelerated birth rate – they did not have random family sizes, so the size-biasing aspect was not seen. In the context of spines, size-biasing was first introduced in the Lyons *et al* papers [22, 19, 23].

7 The spine decomposition

One of the most important results introduced in Lyons [22] was the so-called spine decomposition, which in the case of the additive martingale

$$Z_\lambda(t) = \sum_{u \in N_t} v_\lambda(Y_u(t)) e^{\lambda X_u(t) - E_\lambda t},$$

from the finite-type branching diffusion would be:

$$\tilde{\mathbb{Q}}_\lambda\big(Z_\lambda(t) | \tilde{\mathcal{G}}_\infty\big) = \sum_{u < N_t} v_\lambda(\eta_{S_u}) e^{\lambda \xi_{S_u} - E_\lambda S_u} + v_\lambda(\eta_t) e^{\lambda \xi_t - E_\lambda t}. \tag{20}$$

To prove this we start by decomposing the martingale as

$$Z_\lambda(t) = \sum_{u \in N_t, u \notin \xi} v_\lambda(Y_u(t)) e^{\lambda X_u(t) - E_\lambda t} + v_\lambda(\eta_t) e^{\lambda \xi_t - E_\lambda t},$$

which is clearly true since one of the particles $u \in N_t$ must be the in the line of descent that makes up the spine ξ. Recalling that the σ-algebra $\tilde{\mathcal{G}}_\infty$ contains all information about the line of nodes that makes up the spine, all about the spine diffusion (ξ_t, η_t) for all times t, and also contains all information regarding the fission times on the spine, it is useful to partition the particles $v \in \{u \in N_t, u \notin \xi\}$ into the distinct subtrees $(\tau, M)^u$ that were born at the fission times S_u from the particles that made up the spine before time t, or in other words those nodes in the $\{u < \xi_t\}$ of ancestors of the current spine node ξ_t. Thus:

$$Z_\lambda(t) = \sum_{u < \xi_t} e^{\lambda \xi_{S_u} - E_\lambda S_u} \Bigg\{ \sum_{v \in N_t, v \in (\tau, M)^u} v_\lambda(Y_v(t)) e^{\lambda (X_u(t) - \xi_{S_u}) - E_\lambda (t - S_u)} \Bigg\} + v_\lambda(\eta_t) e^{\lambda \xi_t - E_\lambda t}.$$

If we now take the $\tilde{\mathbb{Q}}_\lambda$-conditional expectation of this, we find

$$\tilde{\mathbb{Q}}_\lambda\big(Z_\lambda(t) | \tilde{\mathcal{G}}_\infty\big) = v_\lambda(\eta_t) e^{\lambda \xi_t - E_\lambda t} + \\ \sum_{u < \xi_t} e^{\lambda \xi_{S_u} - E_\lambda S_u} \tilde{\mathbb{Q}}_\lambda \Big(\sum_{v \in N_t, v \in (\tau, M)^u} v_\lambda(Y_v(t)) e^{\lambda (X_u(t) - \xi_{S_u}) - E_\lambda (t - S_u)} \Big| \tilde{\mathcal{G}}_\infty \Big).$$

We know from the decomposition (19) that the under the measure $\tilde{\mathbb{Q}}_\lambda$ the subtrees coming off the spine evolve as if under the measure P, and therefore

$$\tilde{\mathbb{Q}}_\lambda \Big(\sum_{v \in N_t, v \in (\tau, M)^u} v_\lambda(Y_v(t)) e^{\lambda (X_u(t) - \xi_{S_u}) - E_\lambda (t - S_u)} \Big| \tilde{\mathcal{G}}_\infty \Big) \\ = \tilde{P} \Big(\sum_{v \in N_t, v \in (\tau, M)^u} v_\lambda(Y_v(t)) e^{\lambda (X_u(t) - \xi_{S_u}) - E_\lambda (t - S_u)} \Big| \tilde{\mathcal{G}}_\infty \Big) = v_\lambda(\eta_{S_u}),$$

since the additive expression being evaluated on the subtree is just a shifted form of the martingale Z_λ itself.

This concludes the proof of (20), but before we go move on to give a similar proof for the general case, for easier reference through the cumbersome-looking general proof it is worth recalling that

$$\zeta_\lambda(t) = e^{\int_0^t R(\eta_s)\,\mathrm{d}s} v_\lambda(\eta_t) e^{\lambda \xi_t - E_\lambda t},$$

and therefore noting that (20) can alternatively be written as

$$\tilde{\mathbb{Q}}_\lambda\big(Z_\lambda(t)|\tilde{\mathcal{G}}_\infty\big) = \sum_{u<N_t} e^{-\int_0^{S_u} R(\eta_s)\,\mathrm{d}s} \zeta_\lambda(S_u) + e^{-\int_0^t R(\eta_s)\,\mathrm{d}s} \zeta_\lambda(t).$$

Also, in the general model we are supposing that each particle u in the spine will give birth to a total of A_u subtrees that go off from the spine – the one remaining other offspring is used to continue the line of descent that makes up the spine. This explains the appearance of A_u in the decomposition.

Theorem 7.1 (Spine decomposition) *We have the following* ***spine decomposition*** *for the additive branching-particle martingale:*

$$\tilde{\mathbb{Q}}^x\big(Z(t)|\tilde{\mathcal{G}}_\infty\big) = \sum_{u<\xi_t} A_u\, e^{-\int_0^{S_u} m(\xi_s)R(\xi_s)\,\mathrm{d}s} \zeta(S_u) + e^{-\int_0^t m(\xi_s)R(\xi_s)\,\mathrm{d}s} \zeta(t).$$

Proof: In each sample tree one and only one of the particles alive at time t is the spine and therefore:

$$\begin{aligned} Z(t) &= \sum_{u\in N_t} e^{-\int_0^t m(X_u(s))R(X_u(s))\,\mathrm{d}s} \zeta_u(t), \\ &= e^{-\int_0^t m(\xi_s)R(\xi_s)\,\mathrm{d}s} \zeta(t) + \sum_{u\in N_t, u\neq \xi_t} e^{-\int_0^t m(X_u(s))R(X_u(s))\,\mathrm{d}s} \zeta_u(t). \end{aligned}$$

The other individuals $\{u \in N_t, u \neq \xi_t\}$ can be partitioned into subtrees created from fissions along the spine. That is, each node u in the spine ξ_t (so $u < \xi_t$) has given birth at time S_u to one offspring node uj (for some $1 \le j \le 1 + A_u$) that was chosen to continue the spine whilst the other A_u individuals go off to make the subtrees $(\tau, M)^u_j$. Therefore,

$$Z(t) = e^{-\int_0^t m(\xi_s)R(\xi_s)\,\mathrm{d}s} \zeta(t) + \sum_{u<\xi_t} e^{-\int_0^{S_u} m(\xi_s)R(\xi_s)\,\mathrm{d}s} \sum_{\substack{j=1,\dots,1+A_u \\ uj\notin\xi}} Z_{uj}(S_u;t), \tag{21}$$

where for $t \ge S_u$,

$$Z_{uj}(S_u;t) := \sum_{v\in N_t, v\in(\tau,M)^u_j} e^{-\int_{S_u}^t m(X_v(s))R(X_v(s))\,\mathrm{d}s} \zeta_v(t),$$

is, conditional on $\tilde{\mathcal{G}}_\infty$, a $\tilde{P}$-martingale on the subtree $(\tau, M)^u_j$, and therefore

$$\tilde{P}\big(Z_{uj}(S_u;t)|\tilde{\mathcal{G}}_\infty\big) = \zeta(S_u).$$

Thus taking $\tilde{\mathbb{Q}}$-conditional expectations of (21) gives

$$\begin{aligned} \tilde{\mathbb{Q}}^x\big(Z(t)|\tilde{\mathcal{G}}_\infty\big) &= e^{-\int_0^t m(\xi_s)R(\xi_s)\,\mathrm{d}s} \zeta(t) + \sum_{u<\xi_t} e^{-\int_0^{S_u} m(\xi_s)R(\xi_s)\,\mathrm{d}s} \tilde{P}\Big(\sum_{\substack{j=1,\dots,1+A_u \\ uj\notin\xi}} Z_{uj}(S_u;t)\big|\,\tilde{\mathcal{G}}_\infty\Big), \\ &= e^{-\int_0^t m(\xi_s)R(\xi_s)\,\mathrm{d}s} \zeta(t) + \sum_{u<\xi_t} e^{-\int_0^{S_u} m(\xi_s)R(\xi_s)\,\mathrm{d}s} A_u\, \zeta(S_u), \end{aligned}$$

which completes the proof. □

This representation was first used in the Lyons *et al* [22, 19, 23] papers and has become the standard way to investigate the behaviour of Z under the measure $\tilde{\mathbb{Q}}$. We observe that the two measures $\tilde{P}$ and $\tilde{\mathbb{Q}}$ for the general model are equal when conditioned on $\tilde{\mathcal{G}}_\infty$ since this factors out their differences in the spine diffusion ξ_t, the family sizes born from the spine and the fission times on the spine. Therefore it follows that same argument as used above applies for $\tilde{P}$ to give:

Corollary 7.2

$$\tilde{P}\big(Z(t)|\tilde{\mathcal{G}}_\infty\big) = \sum_{u<\xi_t} A_u\, e^{-\int_0^{S_u} m(\xi_s)R(\xi_s)\,\mathrm{d}s}\zeta(S_u) + e^{-\int_0^t m(\xi_s)R(\xi_s)\,\mathrm{d}s}\zeta(t).$$

As a matter of fact, this above representation for $\tilde{P}\big(Z(t)|\tilde{\mathcal{G}}_\infty\big)$ has not received any attention in the literature, and in the context of our particular finite-type branching diffusion we can use it to explain a subtle difference between the spine and branching pictures. We note that the following remarks are not needed elsewhere in this article and are just to be seen as a side observation.

We proved at Lemma 5.7 that the martingales Z_λ and ζ_λ are both projections of the martingale $\tilde{\zeta}_\lambda$:

$$Z_\lambda(t) = \tilde{P}\big(\tilde{\zeta}_\lambda(t)\,|\,\mathcal{F}_t\big), \quad \text{and}$$
$$\zeta_\lambda(t) = \tilde{P}\big(\tilde{\zeta}_\lambda(t)\,|\,\mathcal{G}_t\big).$$

Given that $Z(t) \in m\mathcal{F}_t \subset m\tilde{\mathcal{F}}_t$, it is natural to wonder if we could also have the equality $\zeta_\lambda(t) = \tilde{P}\big(Z_\lambda(t)\,|\,\mathcal{G}_t\big)$?

Since the algebra $\mathcal{G}_t$ is not a sub-algebra of $\mathcal{F}_t$, this is not immediately a trivial question, and in fact the result *does not hold* – basically because the branching process conditioned on the spine contains more information than the spine itself. To see this we first note that $\mathcal{G}_t$ is a sub-algebra of $\tilde{\mathcal{G}}_\infty$, whence by the tower property

$$\tilde{P}\big(Z_\lambda(t)\,|\,\mathcal{G}_t\big) = \tilde{P}\big(\tilde{P}\big(Z_\lambda(t)|\tilde{\mathcal{G}}_\infty\big)\,|\,\mathcal{G}_t\big).$$

At the same time, since $\zeta_\lambda(t) = \tilde{P}\big(\tilde{\zeta}_\lambda|\mathcal{G}_t\big)$, it follows that the equality $\zeta_\lambda(t) = \tilde{P}\big(Z_\lambda(t)\,|\,\mathcal{G}_t\big)$ is equivalent to showing

$$\tilde{P}\big(\tilde{\zeta}_\lambda|\mathcal{G}_t\big) = \tilde{P}\big(\tilde{P}\big(Z_\lambda(t)|\tilde{\mathcal{G}}_\infty\big)\,|\,\mathcal{G}_t\big),$$

which by definition means showing that for all $G \in \mathcal{G}_t$ we have the equality

$$\int_G \tilde{\zeta}_\lambda(t)\,\mathrm{d}\tilde{P} = \int_G \tilde{P}\big(Z_\lambda(t)|\tilde{\mathcal{G}}_\infty\big)\,\mathrm{d}\tilde{P}. \tag{22}$$

However, just by comparing

$$\tilde{\zeta}_\lambda(t) = \prod_{u<\xi_t}(1+A_v)\times v_\lambda(\eta_t)e^{\lambda\xi_t - E_\lambda t},$$

with

$$\tilde{P}\big(Z_\lambda(t)|\tilde{\mathcal{G}}_\infty\big) = \sum_{u<\xi_t} v_\lambda(\eta_{S_u})e^{\lambda\xi_{S_u} - E_\lambda S_u} + v_\lambda(\eta_t)e^{\lambda\xi_t - E_\lambda t}$$

it is clear that $\tilde{P}\big(Z_\lambda(t)|\tilde{\mathcal{G}}_\infty\big)$ makes references the spine's position at the earlier times S_u, whilst $\tilde{\zeta}_\lambda(t)$ does not. Thus if the event G depends on the history of the spine up to time t then likely as not $\tilde{\zeta}_\lambda(t)$ and $\tilde{P}\big(Z_\lambda(t)|\tilde{\mathcal{G}}_\infty\big)$ will be different, and equality will not hold.

What is happening here is that the births of subtrees on the spine act as a memory of where the spine has been in the past, so that conditional on knowing which particle is currently the spine, the algebra $\mathcal{F}_t$ contains *more information* about the spine than $\mathcal{G}_t$, and therefore

$$\zeta_\lambda(t) \neq \tilde{P}\big(Z_\lambda(t)\,|\,\mathcal{G}_t\big).$$

8 New spine results

Having covered the formal basis for our spine approach, we now present some new results that follow from our spine formulation: the Gibbs-Boltzmann weights, conditional expectations, and a simpler proof of the improved Many-to-One theorem.

8.1 The *Gibbs-Boltzmann* weights of $\tilde{\mathbb{Q}}$

The Gibbs-Boltzmann weightings in branching processes are well-known, for example see Chauvin and Rouault [3] where they consider random measures on the boundary of the tree, and Harris [14] which gives convergence results for Gibbs-Boltzmann random measures. They have previously been considered via the individual terms of the additive martingale Z, but the following theorem gives a new interpretation of these weightings in terms of the spine. We recall that

$$Z(t) = \sum_{u \in N_t} e^{-\int_0^t m(X_u(s))R(X_u(s))\,\mathrm{d}s} \zeta_u(t).$$

Theorem 8.1 *Let $u \in \Omega$ be a given and fixed label. Then*

$$\tilde{\mathbb{Q}}\big(\xi_t = u | \mathcal{F}_t\big) = \mathbf{1}_{(u \in N_t)} \frac{e^{-\int_0^t m(X_u(s))R(X_u(s))\,\mathrm{d}s} \zeta_u(t)}{Z(t)}.$$

Proof: Suppose $u \in \Omega$, and $F \in \mathcal{F}_t$. We aim to show:

$$\int_F \mathbf{1}_{(\xi_t = u)} \,\mathrm{d}\tilde{\mathbb{Q}}(\tau, M, \xi) = \int_F \mathbf{1}_{(u \in N_t)} \frac{e^{-\int_0^t m(X_u(s))R(X_u(s))\,\mathrm{d}s} \zeta_u(t)}{Z(t)} \,\mathrm{d}\tilde{\mathbb{Q}}(\tau, M, \xi).$$

First of all we know that $\mathrm{d}\tilde{\mathbb{Q}}/\mathrm{d}\tilde{P} = \tilde{\zeta}(t)$ on $\mathcal{F}_t$ and therefore,

$$\mathrm{LHS} = \int_F \mathbf{1}_{(\xi_t = u)} \prod_{v < \xi_t} (1 + A_v) e^{-\int_0^t m(\xi_s)R(\xi_s)\,\mathrm{d}s} \zeta(t) \,\mathrm{d}\tilde{P}(\tau, M, \xi),$$

by definition of $\tilde{\zeta}(t)$ at (15). The definition 4.3 of the measure $\tilde{P}$ requires us to express the integrand with a representation like (7):

$$\begin{aligned}
&\mathbf{1}_{(\xi_t = u)} \prod_{v < \xi_t} (1 + A_v) e^{-\int_0^t m(\xi_s)R(\xi_s)\,\mathrm{d}s} \zeta(t) \\
&= \mathbf{1}_{(\xi_t = u)} \sum_{w \in N_t} \prod_{v < w} (1 + A_v) e^{-\int_0^t m(X_w(s))R(X_w(s))\,\mathrm{d}s} \zeta_w(t) \mathbf{1}_{(\xi_t = w)}, \\
&= \mathbf{1}_{(\xi_t = u)} \mathbf{1}_{(u \in N_t)} \prod_{v < u} (1 + A_v) e^{-\int_0^t m(X_u(s))R(X_u(s))\,\mathrm{d}s} \zeta_u(t),
\end{aligned}$$

and therefore

$$\begin{aligned}
\mathrm{LHS} &= \int_F \mathbf{1}_{(u \in N_t)} \prod_{v < u} (1 + A_v) e^{-\int_0^t m(X_u(s))R(X_u(s))\,\mathrm{d}s} \zeta_u(t) \mathbf{1}_{(\xi_t = u)}, \ \mathrm{d}\tilde{P}(\tau, M, \xi), \\
&= \int_F \mathbf{1}_{(u \in N_t)} e^{-\int_0^t m(X_u(s))R(X_u(s))\,\mathrm{d}s} \zeta_u(t) \,\mathrm{d}P(\tau, M, \xi),
\end{aligned}$$

by definition 4.3. We emphasize that now this is an integral taken with respect to the measure P over the σ-algebra $\mathcal{F}_t$, and here we know that $\mathrm{d}P/\mathrm{d}\mathbb{Q} = 1/Z(t)$, so:

$$\mathrm{LHS} = \int_F \mathbf{1}_{(u \in N_t)} e^{-\int_0^t m(X_u(s))R(X_u(s))\,\mathrm{d}s} \zeta_u(t) \ \frac{1}{Z(t)} \mathrm{d}\mathbb{Q}(\tau, M, \xi),$$

and the proof is concluded. $\square$

8.1.1 Comment

At first it may seem that result 8.1 could give rise to a paradox: suppose that two particles are at the same spatial position: $X_{u_1}(t) = X_{u_2}(t)$ but that u_1 has had more ancestors. According to the above there is an equal conditional probability that either is the spine, but we know that under $\tilde{\mathbb{Q}}$ the spine has an *increased* birth rate, and might therefore expect u_1 to have a higher probability of being the spine.

The way to resolve this is to realise that the spine is always chosen uniformly across the same generation and therefore

$$\tilde{\mathbb{Q}}\big(\xi_t = u_1 | N_t\big) = \mathbf{1}_{(u_1 \in N_t)} \prod_{v < u_1} (1 + A_v)^{-1},$$

which in fact gives u_1 a *smaller* probability of being the spine. This neutralizes the accelerated birth rate of the spine, and the two effects balance each other exactly.

8.2 Conditional expectations and the Kesten-Stigum Theorem

The above result combines with the representation (7) to show how we take conditional expectations under the measure $\tilde{\mathbb{Q}}$.

Theorem 8.2 *If $f(t) \in m\tilde{\mathcal{F}}_t$, and $f = \sum_{u \in N_t} f_u(t) \mathbf{1}_{(\xi_t = u)}$, with $f_u(t) \in m\mathcal{F}_t$ then*

$$\tilde{\mathbb{Q}}\big(f(t) | \mathcal{F}_t\big) = \sum_{u \in N_t} f_u(t) \frac{e^{-\int_0^t m(X_u(s)) R(X_u(s))\, \mathrm{d}s} \zeta_u(t)}{Z(t)}. \tag{23}$$

Proof: It is clear that

$$\tilde{\mathbb{Q}}\big(f(t) | \mathcal{F}_t\big) = \sum_{u \in N_t} f_u(t) \tilde{\mathbb{Q}}\big(\xi_t = u | \mathcal{F}_t\big),$$

and the result follows from Theorem 8.1. □

A simple corollary of this is exceptionally useful, and goes an awful long way to obtaining the Kesten-Stigum result in very general models (see [17, 16, 18] for the classical proofs of the Kesten-Stigum results),

Corollary 8.3 *If $g(\cdot, \cdot)$ is a Borel function on $\mathbb{R} \times I$ then*

$$\sum_{u \in N_t} g(X_u(t))\, e^{-\int_0^t m(X_u(s)) R(X_u(s))\, \mathrm{d}s} \zeta_u(t) = \tilde{\mathbb{Q}}\big(g(\xi_t) | \mathcal{F}_t\big) \times Z(t).$$

Proof: It is easy to show that $g(\xi_t) = \sum_{u \in N_t} g(X_u(t)) \mathbf{1}(\xi_t = u)$, and now the result follows from the above corollary. □

In the case of BBM this particular corollary would state that for any Lebesgue-measurable function f we have

$$\sum_{u \in N_t} f(X_u(t)) e^{\lambda X_u(t) - (\lambda^2 + r)t} = \tilde{\mathbb{Q}}_\lambda\big(f(\xi_t) | \mathcal{F}_t\big),$$

The Kesten-Stigum theorem gives conditions under which an operation like this converge as $t \to \infty$, and in general it is found that when it exists the limit is a multiple of the martingale limit $Z_\lambda(\infty)$ – see Lyons *et al* [19] for a proof of this based on other spine techniques. Our improved spine formulation therefore gives a previously unknown but simple meaning to this operation, in terms of a conditional expectation. Furthermore, this new interpretation might also give good grounds for alternative spine proofs of the Kesten-Stigum theorem, via a study of when the conditional expectation $\tilde{\mathbb{Q}}\big(g(\xi_t)|\mathcal{F}_t\big)$ converges.

Remark: The convergence of conditional expectations is not necessarily easy to obtain and has not yet been considered to any great extent in the literature. The *Gibbs Conditioning Principle* (see Dembo and Zeitouni [7]) is one useful example of a result for conditional expectations, and it may be possible to adapt it to this context but this is left to further research.

8.3 The *Full* Many-to-One Theorem

Much of the work of Harris and Williams [13] and Harris and Git [15] made important use of a *Many-to-One* result, which had been proved using resolvents (see Champneys *et al* [1] for a similar proof). In the context of the finite-type branching diffusion of section 2.1, the Many-to-One theorem would be stated as follows:

Theorem 8.4 *For any measurable function $f : J \to \mathbb{R}$ we have*

$$P^{x,y}\Big(\sum_{u\in N_t} f(X_u(t), Y_u(t))\Big) = \mathbb{P}^{x,y}\Big(e^{\int_0^t R(\eta_s)\,\mathrm{d}s} f(\xi_t, \eta_t)\Big).$$

Intuitively it is clear that the up-weighting term $e^{\int_0^t R(\eta_s)\,\mathrm{d}s}$ incorporates the notion of the population growing at an exponential rate, whilst the idea of $f(\xi_t, \eta_t)$ being the 'typical' behaviour of $f(X_u(t), Y_u(t))$ is also reasonable.

The main problem with the proofs given by Harris and Williams is that it applies only to functions of the above form that therefore depend only on *the time-t location* of the spine – it does not cover functions that depend on the entire *path history* of the spine up to time t.

With the spine approach we have the benefit of being able to give a much less complicated proof of the stronger version that covers the most general path-dependent functions.

Theorem 8.5 (Many-to-One) *If $g(t) \in m\mathcal{G}_t$ has the representation*

$$g(t) = \sum_{u\in N_t} g_u(t)\mathbf{1}_{(\xi_t=u)},$$

where $g_u(t) \in m\mathcal{F}_t$, then

$$P\Big(\sum_{u\in N_t} g_u(t)\Big) = \mathbb{P}\Big(e^{\int_0^t m(\xi_s)R(\xi_s)\mathrm{d}s} g(t)\Big).$$

Proof: Let $f(t)$ be any given $m\mathcal{G}_t$-measurable function. Since $\mathcal{G}_t \subset \tilde{\mathcal{F}}_t$ it therefore follows that $f(t)$ is also $\tilde{\mathcal{F}}_t$-measurable and we can use the tower property together with Theorem 8.2 to obtain

$$\begin{aligned}\tilde{\mathbb{Q}}\big(f(t)\big) = \tilde{\mathbb{Q}}\Big(\tilde{\mathbb{Q}}\big(f(t)|\mathcal{F}_t\big)\Big) &= \mathbb{Q}\Big(\tilde{\mathbb{Q}}\big(f(t)|\mathcal{F}_t\big)\Big)\\ &= \mathbb{Q}\Big(\frac{1}{Z(t)}\sum_{u\in N_t} f_u(t)e^{-\int_0^t m(X_u(s))R(X_u(s))\,\mathrm{d}s}\zeta_u(t)\Big).\end{aligned}$$

We emphasize that this is a $\mathbb{Q}$ expectation of a $\mathcal{F}_t$-measurable expression. Theorem 6.4 states that

$$\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}P}\bigg|_{\mathcal{F}_t} = \frac{Z(t)}{Z(0)},$$

and therefore we conclude from the above that

$$\tilde{\mathbb{Q}}\big(f(t)\big) = Z(0)^{-1}P\Big(\sum_{u\in N_t} f_u(t)e^{-\int_0^t m(X_u(s))R(X_u(s))\,\mathrm{d}s}\zeta_u(t)\Big).$$

On the other hand, since $f(t)$ is $\mathcal{G}_t$-measurable, Theorem 6.4 states that

$$\left.\frac{\mathrm{d}\hat{\mathbb{P}}}{\mathrm{d}\mathbb{P}}\right|_{\mathcal{G}_t} = \frac{\zeta(t)}{\zeta(0)},$$

and from the definitions 4.6 and 6.3 this means

$$\left.\frac{\mathrm{d}\tilde{\mathbb{Q}}}{\mathrm{d}\tilde{P}}\right|_{\mathcal{G}_t} = \frac{\zeta(t)}{\zeta(0)},$$

and therefore

$$\tilde{\mathbb{Q}}\big(f(t)\big) = \zeta(0)^{-1}\tilde{P}\big(f(t) \times \zeta(t)\big).$$

Trivially we always have $Z(0) = \zeta(0)$ and thus we have shown in the first instance that for any $f(t) \in m\mathcal{G}_t$,

$$P\Big(\sum_{u \in N_t} f_u(t) e^{-\int_0^t m(X_u(s))R(X_u(s))\,\mathrm{d}s}\zeta_u(t)\Big) = \tilde{P}\big(f(t) \times \zeta(t)\big). \tag{24}$$

Given $g(t) \in m\mathcal{G}_t$, we can define

$$f(t) := e^{\int_0^t m(\xi_s)R(\xi_s)\,\mathrm{d}s} g(t) \times \zeta(t)^{-1},$$

which is clearly $\mathcal{G}_t$-measurable and satisfies $f(t) = \sum_{u \in N_t} f_u(t)\zeta_u(t)$ with

$$f_u(t) = g_u(t) e^{\int_0^t m(X_u(s))R(X_u(s))\,\mathrm{d}s}\zeta_u(t)^{-1} \in m\mathcal{F}_t.$$

When we now use this $f(t)$ in the above (24) we arrive at the result to be proved. $\square$

In the case in which $g = g(\xi_t)$ for some Borel-measurable function $g(\cdot)$, the trivial representation

$$g(\xi_t) = \sum_{u \in N_t} g\big(X_u(t)\big)$$

leads immediately to the weaker version of the Many-to-One result that was originally proven by Harris and Williams using resolvents and the Feynman-Kac formula, expressed in terms of our more general branching Markov process $\mathbb{X}_t$:

Corollary 8.6 *If $g(\cdot) : J \to \mathbb{R}$ is $\mathcal{B}$-measurable then*

$$P\Big(\sum_{u \in N_t} g(X_u(t))\Big) = \mathbb{P}\Big(e^{\int_0^t R(\xi_s)\mathrm{d}s} g(\xi_t)\Big).$$

References

[1] A. Champneys, S. Harris, J. Toland, J. Warren, and D. Williams, *Algebra, analysis and probability for a coupled system of reaction-diffusion equations*, Philosophical Transactions of the Royal Society of London **350** (1995), 69–112.

[2] B. Chauvin, *Arbres et processus de Bellman-Harris*, Ann. Inst. H. Poincaré Probab. Statist. **22** (1986), no. 2, 209–232. MR 88a:60151

[3] B. Chauvin and A. Rouault, *Boltzmann-Gibbs weights in the branching random walk*, Classical and modern branching processes (Minneapolis, MN, 1994), IMA Vol. Math. Appl., vol. 84, Springer, New York, 1997, pp. 41–50. MR 99a:60090

[4] Brigitte Chauvin, *Product martingales and stopping lines for branching Brownian motion*, Ann. Probab. **19** (1991), no. 3, 1195–1205. MR 93b:60188

[5] Brigitte Chauvin and Alain Rouault, *KPP equation and supercritical branching Brownian motion in the subcritical speed area. Application to spatial trees*, Probab. Theory Related Fields **80** (1988), no. 2, 299–314. MR 90b:60113

[6] Brigitte Chauvin, Alain Rouault, and Anton Wakolbinger, *Growing conditioned trees*, Stochastic Process. Appl. **39** (1991), no. 1, 117–130. MR 93d:60138

[7] Amir Dembo and Ofer Zeitouni, *Large deviations techniques and applications*, Springer, 1998.

[8] Y. Git, J.W.Harris, and S.C.Harris, *Exponential growth rates & large deviations for a typed branching diffusion*, (2004), no. 0406, Mathematics Preprint, University of Bath.

[9] Robert Hardy, *Branching diffusions*, Ph.D. thesis, University of Bath Department of Mathematical Sciences, 2004.

[10] Robert Hardy and Simon C. Harris, *A spine proof of a large-deviations principle for branching Brownian motion*, (2004), no. 0407, Mathematics Preprint, University of Bath.

[11] ———, *A spine proof of a lower-bound for a typed branching diffusion*, (2004), no. 0408, Mathematics Preprint, University of Bath.

[12] ———, *Spine proofs for $\mathcal{L}^p$-convergence of branching-diffusion martingales*, (2004), no. 0405, Mathematics Preprint, University of Bath.

[13] S. C. Harris and D. Williams, *Large deviations and martingales for a typed branching diffusion. I*, Astérisque (1996), no. 236, 133–154, Hommage à P. A. Meyer et J. Neveu. MR 98c:60110

[14] Simon C. Harris, *Convergence of a 'Gibbs-Boltzmann' random measure for a typed branching diffusion*, Séminaire de Probabilités XXXIV, Lecture Notes in mathematics, vol. 1729, Springer, 2000, pp. 133–154.

[15] Simon C. Harris and Yoav Git, *Large deviations and martingales for a typed branching diffusion: II*, University of Bath Preprint, 2001.

[16] H. Kesten and B.P. Stigum, *Additional limit theorems for indecomposable multidimensional Galton-Watson processes*, Ann. Math. Stat. **37** (1966), 1463–1481.

[17] ———, *A limit theorem for multidimensional Galton-Watson processes*, Ann. Math. Stat. **37** (1966), 1211–1223.

[18] ———, *Limit theorems for decomposable multi-dimensional Galton-Watson processes*, J. Math. Anal. Applic. **17** (1967), 309–338.

[19] Thomas Kurtz, Russell Lyons, Robin Pemantle, and Yuval Peres, *A conceptual proof of the Kesten-Stigum theorem for multi-type branching processes*, Classical and modern branching processes (Minneapolis, MN, 1994), IMA Vol. Math. Appl., vol. 84, Springer, New York, 1997, pp. 181–185. MR 98j:60122

[20] Andreas Kyprianou, *Travelling wave solutions to the K-P-P equation: alternatives to Simon Harris's probabilistic analysis*, **40** (2004), no. 1, 53–72.

[21] Quansheng Liu and Alain Rouault, *On two measures defined on the boundary of a branching tree*, Classical and modern branching processes (Minneapolis, MN, 1994), IMA Vol. Math. Appl., vol. 84, Springer, New York, 1997, pp. 187–201. MR 98j:60123

[22] Russell Lyons, *A simple path to Biggins' martingale convergence for branching random walk*, Classical and modern branching processes (Minneapolis, MN, 1994), IMA Vol. Math. Appl., vol. 84, Springer, New York, 1997, pp. 217–221. MR 1 601 749

[23] Russell Lyons, Robin Pemantle, and Yuval Peres, *Conceptual proofs of $L \log L$ criteria for mean behavior of branching processes*, Ann. Probab. **23** (1995), no. 3, 1125–1138. MR 96m:60194

[24] J. Neveu, *Arbres et processus de Galton-Watson*, Ann. Inst. H. Poincaré Probab. Statist. **22** (1986), no. 2, 199–207. MR 88a:60150

[25] Peter Olofsson, *The $x \log x$ condition for general branching processes*, J. Appl. Probab. **35** (1998), no. 3, 537–544. MR 99k:60214