A non-localisation property of a binary branching random walk

Supervisor: Cécile Mailler

Benjamin Holland Young

2024/25

Contents

1 Introduction

The main goal of this project is to explain the behaviour of a branching process called the "catalytic branching random walk". One of the aims is to explain [DR12], which gives the behaviour of the asymptotics of the moments. We start by giving an introduction to renewal theory and spine techniques (from [HR14]), as well as adding extra detail to make the content more accessible to an MMath student. Another aim of the project is to apply some of the techniques of [DR12] to a more general situation. The model discussed in [DR12] works as follows: the particles move about a countable state space S according to some continuous-time Markov process with Q-matrix $\mathcal{A}$; each particle carries an exponential clock of rate 1 (when at the origin); every time it rings the particle dies and is replaced by a random number of particles according to some distribution μ. This model will be referred to as the DR model. In Section 4, we expand on this by branching everywhere (instead of just at the origin), whilst each particle still moves about according to some Q-matrix $\mathcal{A}$ and has offspring distribution μ. The branching occurs at rate $\lambda_0 > 0$ at the origin, and rate $\lambda > 0$ outside of it. We are also interested in a specific case when we almost-surely branch into two particles (binary branching) and move uniformly around $S = \mathbb{Z}^d$, for $d \in \mathbb{N}$. We refer to this model as the MS model, and it is studied in [MS24].

1.1 Background from the literature

The main result of [DR12] gives the asymptotics of the moments of the number of particles in the system at some time $t > 0$, and the number of particles at any given point $y \in S$ at some time $t > 0$.

The analysis of these moments was first studied for $\mathcal{A} = \Delta$ the discrete Laplacian, which represents uniform movement around the d-dimensional lattice $\mathbb{Z}^d$, in [ABY98] with PDEs and Tauberian theorems (see [Tru09]). Several other moment analyses have been conducted using techniques such as Bellman-Harris branching processes (see [TV03]), operator theory (see [Yar10]), and renewal theory (see [HVT10]). These approaches relied on four assumptions on the movement of the particles governed by $\mathcal{A}$, namely: irreducibility, spatial homogeneity, symmetry, finite variance of jump sizes. This approach is not taken in [DR12] and instead only the assumption of irreducibility of the motion is assumed. In addition the techniques used here rely solely on general probability theory (including spine techniques) and elementary algebra/analysis. Spine techniques were first used by Kesten and Stigum in the paper [KS66], this project uses [HR14] to give an introduction to this topic.

1.2 The binary branching process or MS model

The MS model behaves as follows: each particle (which exists in $\mathbb{Z}^d$, $d \in \mathbb{N}$) carries two clocks. One Exp(λ_x) clock, for $x \in \mathbb{Z}^d$ the position of the particle, every time it rings the particles branches. Another Exp(1) clock, each time it rings the particle move to one of neighbours, uniformly, i.e. in $\mathbb{Z}^d$ each neighbour has probability $\frac{1}{2d}$ of being moved to. Where for all $x \in \mathbb{Z}^d$ and $0 < \lambda < \lambda_0$:

$$\lambda_x = \begin{cases} \lambda_0 & x = 0 \\ \lambda & x \neq 0. \end{cases}$$

Note that, we start with one particle at the origin in $S = \mathbb{Z}^d$.

We would like to analyse the distribution of particles through time, so the object of interest for this problem is the empirical measure of particles over the space $\mathbb{Z}^d$. Let $X_i(t)$ be the position in S of the ith for $1 \leq i \leq N(t)$, where $N(t)$ is the total number of particles present at time $t > 0$. We define for $t > 0$, the measure $\pi_t = \sum_{i=1}^{N(t)} \delta_{X_i(t)}$ where the $\delta_{X_i(t)}$ are Dirac measures. For example with $S = \mathbb{Z}$, this means that $\pi_t(\{0, 1\})$ tells us how many particles (at time $t > 0$) are at either position 0 or 1.

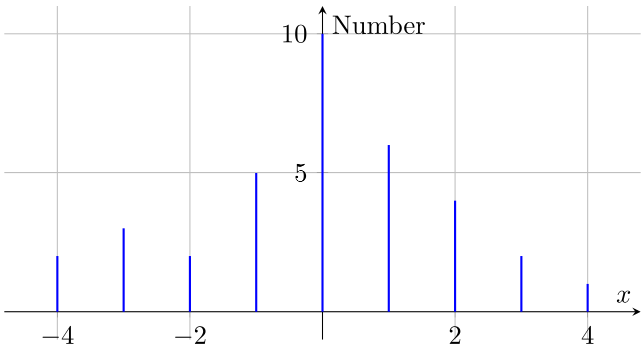


Figure 1: An example empirical measure π_t for some $t > 0$.

1.3 The localisation and non-localisation regimes

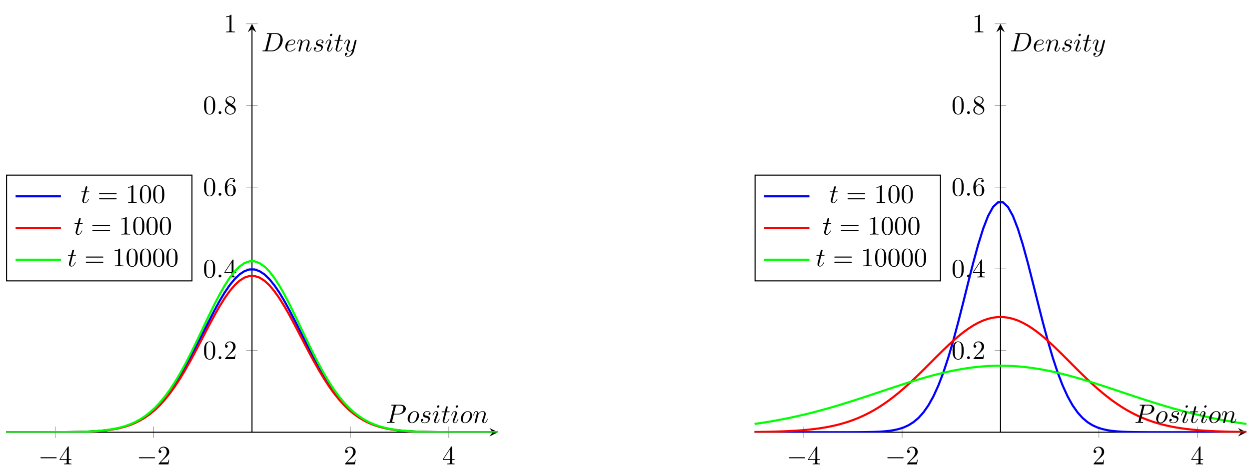
It was shown in [MS24] that the following regimes occur for the MS model (depending on the difference $\lambda_0 - \lambda$).

Definition 1.1 (Localisation)**.** *We say that* ***localisation*** *occurs when* $\frac{\pi_t}{N(t)} \xrightarrow[t\to\infty]{} \nu$ *almost-surely and where the convergence is in the weak topology*[1] *on probability measures. Note that* ν *is the fixed distribution to which we are converging.*

Definition 1.2 (Non-localisation)**.** *We say that* ***non-localisation*** *occurs when these distributions flatten/diffuse out with time (i.e. less and less concentrated at any point). This can be written as* $\pi_t(\{x\}) = o(N(t))$ *for all* $x \in \mathbb{Z}^d$ *as* $t \to \infty$.

The results in [DS10] show a similar but distinct behaviour of the moments for the DR model. One of the goals of this project is to show that we have similar results for the MS model. To this end, we would like to apply the ideas of the proofs in [DR12] to show these similar results for the MS model.

Here are two examples of these two modes of behaviour, note this is in the $d = 1$ case of our model. We know that localisation only occurs in the MS model with $d \geq 3$, therefore these are just theoretical illustrations, and the true distributions may not be Gaussian as in Figure 2.



(a) Localisation

(b) Non-Localisation

Figure 2: Comparison of Localisation and Non-Localisation. (ChatGPT assisted with making the LaTex for these originally.)

[1]For a sequence of probability measures on $\mathbb{Z}^d$, for $d \in \mathbb{N}$, we say that $\pi_t \to \pi$ in the weak topology of probability measures, if for all continuous and bounded $f : \mathbb{Z}^d \to \mathbb{R}$ we have that $\int f \mathrm{d}\pi_t \to \int f \mathrm{d}\pi$.

2 Pre-requisites

2.1 Renewal theory

In this section, a summary of useful results from renewal theory is given, since these are used in [DR12] for proving results about the asymptotics of the DR model. These results come from the standard references [Fel71] and [MO14].

2.1.1 Basic definitions

A renewal process is a process where the time between "renewal epochs" (which is another word for a recurring event, such as a component breaking down in a machine) are i.i.d. waiting times.

Definition 2.1 (Renewal process). *A* ***renewal process*** *is a stochastic process* $(S_n)_{n\geq 0}$ *such that* $S_0 = 0$ *and for all* $n \in \mathbb{N}$, $S_n = \sum_{i=1}^{n} T_i$ *where the* $(T_i)_{i\geq 1}$ *are i.i.d. non-negative random variables, whose common expectation* $\mu = \mathbb{E}[T_1]$ *is called the* ***mean recurrence time****.*

Note that S_n represents the time of the nth renewal epoch, and T_n represents the time between the $(n-1)$th and nth renewal epochs. Further, observe that $\mu \in (0, \infty]$, so we allow infinite-mean recurrence times.

We also have that, for all $t \geq 0$

$$N(t) = \sup\{n \geq 0 : S_n \leq t\} \quad \text{and} \quad M(t) = \inf\{n \geq 1 : S_n > t\},$$

where $N(t)$ is the number of renewal epochs *up to time* t, and $M(t)$ is the number of number of renewal epochs *up to time* t counting $S_0 = 0$ as the first renewal epoch. Hence we have the relation $M(t) = N(t)+1$, $\forall t \geq 0$.

Example 2.2 (Poisson process). *Suppose the waiting times* $(T_i)_{i\geq 1}$ *are i.i.d.* $\mathrm{Exp}(\lambda)$ *random variables, then* $(S_n)_{n\geq 0}$ *is a renewal process and* $(N(t))_{t\geq 0}$ *is called a Poisson process.*

Definition 2.3 (Renewal measure/function). *The* ***renewal measure*** *is defined in the following way: given an interval of time* $I \subset [0, \infty)$*, we define:*

$$U(I) = \sum_{n=0}^{\infty} \mathbb{P}(S_n \in I),$$

where $\mathbb{P}(S_n \in I)$ *represents the probability that the* n*th renewal epoch happens within time interval* I*. In particular, we define, for all* $t \in (0, \infty)$*, the* ***renewal function*** *as:*

$$U(t) = U([0, t]),$$

which is the mean number of renewal epochs up to time t *i.e.*

$$U(t) = \mathbb{E}[M(t)] = \mathbb{E}[N(t)] + 1.$$

We now give the definition of what it means to take an integral with respect to a cumulative distribution function F.

Definition 2.4 (Integrating with respect to F). *Suppose* $F : \mathbb{R} \to [0, 1]$ *is a differentiable distribution function, then for a given integrable function* $h : \mathbb{R} \to \mathbb{R}$ *and* $A \subset \mathbb{R}$ *we define:*

$$\int_A h(x)\mathrm{d}F(x) = \int_A h(x)f(x)\mathrm{d}x,$$

where $f = F'$*.*

Definition 2.5 (Renewal equation)**.** *A* ***renewal equation*** *is an integral equation of the form, for $x \in (0, \infty)$, and F differentiable:*

$$Z(x) = z(x) + \int_0^x Z(x-y)\mathrm{d}F(y) = z(x) + \int_0^x Z(x-y)F'(y)\mathrm{d}y,$$

where we are interested in the behaviour of $Z(x)$, and the integral is with respect to the distribution function in the sense of Definition 2.4. This is often rewritten with $*$ as notation for convolution as follows:

$$Z = z + F * Z.$$

Whilst the intuitive interpretation of "recurrent events" is useful, we sometimes only have expressions with the right analytical form and properties where we can apply these results, which we see in Section 3.4.

2.1.2 Renewal theorems for finite-mean recurrence time

We first give some important renewal theorems in the case of a finite-mean recurrence time.

Theorem 2.6 (Elementary Renewal Theorem)**.** *Suppose that $F(0) < 1$ and $F(\infty) = 1$ then we have:*

$$\lim_{t\to\infty} \frac{U(t)}{t} = \begin{cases} \mu^{-1} & \mu < \infty \\ 0 & \mu = \infty. \end{cases} \tag{2.1}$$

This theorem is not entirely surprising, given that Theorem 1.1 of [MO14, page 3] is a simpler theorem which says that $\frac{N(t)}{t} \to \frac{1}{\mu}$ or zero when $\mu = \infty$, which can be easily proven with the definitions from Section 2.1.1 and the *Central Limit Theorem*.

In order to introduce the next theorem, we require the following definition.

Definition 2.7 (Lattice distribution function)**.** *A distribution function F is said to be* ***lattice*** *with span $a > 0$ if the support of the distribution determined by F lies within the set $\{x : x = b + na, n \in \mathbb{Z}\}$, where a is the largest positive number (called the* ***span****) for which this condition holds, and where $b \in \mathbb{R}$ is called the shift. A distribution without this property is said to be* ***nonlattice****.*

This definition makes the distinction between distributions whose support is part of a lattice, and those that do not. For example the distributions of any any Binomial or Poisson random variables are lattice, but a distribution on any continuous random variable is not lattice.

Theorem 2.8 (Blackwell's Renewal Theorem)**.** *Suppose that the distribution F is such that $F(0) < 1$ and that $F(\infty) = \lim_{t\to\infty} F(t) = \infty$:*

$$\mu = \int_0^\infty t\mathrm{d}F(t) \in (0, \infty)$$

If F is ***nonlattice****, then for any fixed $h > 0$:*

$$\lim_{t\to\infty} [U(t) - U(t-h)] = \frac{h}{\mu}$$

If F is ***lattice*** *with span a, then for any fixed integer $\ell > 0$:*

$$\lim_{t\to\infty} [U(t) - U(t-\ell a)] = \frac{\ell a}{\mu}.$$

If $\mu = \infty$, then Theorem 2.8 still holds but we define the right-hand side of these equations as zero. We once again require a technical definition to understand the next theorem, that is the definition of a function being **directly Riemann integrable**, from [Car12].

Definition 2.9 (Directly Riemann Integrable)**.** *Given a non-negative function $z : \mathbb{R} \to [0, \infty)$ we say it is* ***directly Riemann integrable*** *if for any $h > 0$, we have:*

$$\lim_{\delta\downarrow 0} \delta \sum_{m\in\mathbb{Z}} \sup_{m\delta \le z \le (m+1)\delta} g(z) = \lim_{\delta\downarrow 0} \delta \sum_{m\in\mathbb{Z}} \inf_{m\delta \le z \le (m+1)\delta} g(z) < \infty$$

For a general function $f : \mathbb{R} \to \mathbb{R}$ it is directly Riemann integrable, if the positive and negative parts f^+ and f^- are.

This is saying that the upper and lower Riemann sums over the **unbounded domain** $\mathbb{R}$ (not on just a finite closed interval) converge to the same finite limit as the mesh size δ tends to 0. A useful sufficient condition is simply being Riemann integrable over the interval $(0, \infty)$.

Theorem 2.10 (The Key Renewal Theorem)**.** *Let $z : \mathbb{R} \to \mathbb{R}$ be a function that vanishes on $(-\infty, 0)$. Suppose F is a distribution function satisfying $F(0) < 1$ and $F(\infty) := \lim_{t\to\infty} F(t) = \infty$, and let $\mu \in (0, \infty)$.*

- *If F is **nonlattice**, then*

$$Z(t) = (U * z)(t) \xrightarrow[t\to\infty]{} \frac{1}{\mu} \int_0^\infty z(t)\, \mathrm{d}t.$$

- *If F is **lattice** with span a, then*

$$Z(t + na) \xrightarrow[n\to\infty]{} \frac{a}{\mu} \sum_{k=0}^{\infty} z(t + ka).$$

We can state similar results to these for renewal processes with infinite mean recurrent times, but the technical setup needed would be too much of a tangent for this project.

2.1.3 Examples of renewal processes

Example 2.11. *As mentioned in Example 2.2, the **Poisson process** arises from a renewal process. A Poisson process $(X_t)_{t\geq 0}$ of rate λ is a continuous-time process defined with $S_0 = 0$, $T_1, T_2, \ldots \sim \mathrm{Exp}(\lambda)$ and $S_n = T_1 + \ldots + T_n$, where, for all $t > 0$:*

$$X_0 = 0 \quad \textit{and} \quad X_t = \sup\{n \geq 0 : S_n \leq t\}.$$

In other words, the renewal process has i.i.d. interarrival times which are all exponential rate λ distributed and $X_t = N(t)$, with starting point $X_0 = 0$. We can actually see a direct verification of the Elementary Renewal Theorem (Theorem 2.6, as the renewal function has a nice form. Recall that $\mathbb{E}[T_1] = \frac{1}{\lambda} = \mu$, and by the definition of a Poisson process, for all $t > 0$, $\mathbb{E}[X_t] = \lambda t$:

$$U(t) = \mathbb{E}[N(t)] + 1 = \mathbb{E}[X_t] + 1 = \lambda t + 1.$$

Therefore, as $t \to \infty$,

$$\frac{U(t)}{t} = \frac{\lambda t + 1}{t} = \lambda + \frac{1}{t} \to \lambda = (\frac{1}{\lambda})^{-1} = \mu^{-1},$$

which confirms Theorem 2.6 in this case.

Example 2.12. *Suppose we have a sequence $\boldsymbol{X} = (X_1, X_2, \ldots)$ of $\mathrm{Bernoulli}(p)$ random variables with probability $p \in (0, 1)$. This can be seen as a sequence of success/failure trials, where a success is considered a renewal epoch in the notation of section 2.1.1. The partials sums of $\boldsymbol{X}$ give rise to a sequence $\boldsymbol{Y}$, note that $\forall n \in \mathbb{N}_0$, $Y_n \sim \mathrm{Bin}(n, p)$. Another process which derived from $\boldsymbol{X}$ is $\boldsymbol{U}$ where for any $n \in \mathbb{N}_0$ we have U_n is the number of trials needed to go from the $n-1$th success to the nth. Notice that $U_n \sim \mathrm{Geo}(p)$. The partial sums of the process $\boldsymbol{U}$ gives rise to a process $\boldsymbol{V}$, where for all $n \in \mathbb{N}_0$ we have that V_n is the number of trials required to obtain n successes and thus that $V_n \sim \mathrm{NegBin}(n, p)$.*

This a renewal process where the successes are the renewal epochs, thus $(U_n)_{n\geq 0}$ are the interarrival times, and thus $(V_n)_{n\geq 0}$ is the sequence of arrival/renewal times (which corresponds to $(S_n)_{n\geq 0}$. Note that for $n \geq 0$, $N(n)$ is number of successes up to time n, and thus is equal to Y_n.

The renewal function is straightforward to compute in this instance since $U(n) = \mathbb{E}(N(n)) = \mathbb{E}(Y_n) = np$, for any $n \in \mathbb{N}_0$.

2.2 Spine techniques

We are interested in spine techniques as they give us a way to reduce the complexity of calculations. Specifically, some expectations require dealing with an arbitrarily large number of particles (for example if we have binary branching), this can lead to highly complex and intractable calculations. However, the use of spine-techniques allows us (via a change of measure) to weight some of the particles in a certain way, making calculation easier as we only have to deal with a smaller number of particles. The main result of this chapter is the "many-to-few" lemma whose name indicates the idea of dealing with fewer particles which eases computation. This discussion starts with an easier discrete-time case for Galton-Watson trees, which follows [Shi11], as it serves as intuition for the more complex continuous-time case.

2.2.1 Galton-Watson trees

Recall that a **Galton-Watson tree** is a discrete-time stochastic process $(Z_n)_{n\geq 0}$, where Z_n is the number of individuals in generation $n \in \mathbb{N}_0$. The process evolves as follow, for $n \geq 0$:

$$Z_{n+1} = \sum_{i=1}^{Z_n} X_i^{(n)} \quad \text{and} \quad Z_0 = 1, \tag{2.2}$$

where $(X_i^{(n)})_{1\leq i\leq Z_n}$ are i.i.d. non-negative integer-valued random variables. We start with one particle at generation 0, it then has a random number of offspring according to the distribution of $X = X_1^{(0)}$, giving generation 1. Subsequently, each of it's offspring independently have a random number of offspring, each still following the distribution of X due to the i.i.d. condition.

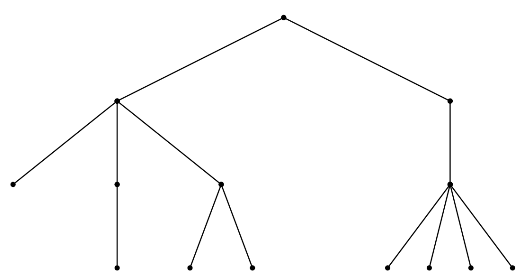


Figure 3: An example of a Galton-Watson tree (taken from [Shi11]).

Further recall that, if at any $n \in \mathbb{N}$ we have $Z_n = 0$, then the system has gone extinct and thus $Z_m = 0$ for all $m \geq n$. We now turn to the probability of this event occurring, which is called the **extinction probability**. We define it as:

$$q = \mathbb{P}\left(\exists n \in \mathbb{N} : Z_n = 0\right).$$

One important quantity in the analysis of q is the expected number of offspring $m = \mathbb{E}[X] \in (0, \infty]$. We can write the distribution of X as $(p_i)_{i\geq 0}$, where p_i is the probability of obtaining $i \in \mathbb{N}_0$ offspring. We state here a particularly insightful theorem (without proof) which gives an idea of the behaviour of q.

Theorem 2.13. *Given the moment generating function f of X, i.e. $f(s) = \sum_{n=0}^{\infty} s^i p_i$ for $s \in [0, 1]$. We have that q is the smallest non-negative solution to the equation $f(s) = s$, for $s \in [0, 1]$. In particular:*

$$q = \begin{cases} 1 & \text{if} \quad m \leq 1, \\ < 1 & \text{if} \quad m > 1. \end{cases}$$

In words, if the expected number of offspring is greater than one, then we survive with strictly positive probability, otherwise we go extinct almost-surely. We introduce the stochastic process $(W_n = \frac{Z_n}{m^n})_{n\geq 0}$ to give some more insight into q, which is well defined when $m < \infty$. It is easy to see that $(W_n)_{n\geq 0}$ is a martingale, with respect to the natural filtration of $(Z_n)_{n\geq 0}$. Since it is non-negative, we have by the *martingale convergence theorem* for non-negative martingales that:

$$W_n \to W \quad \text{as} \quad n \to \infty,$$

where the convergence is almost-surely, and W is a non-negative random variable. This random variable is useful since we know that if $W = 0$ then the system dies out, so it is useful to know when it is **not** zero. We can see quickly that by *Fatou's lemma*, $\mathbb{E}[W] \leq \liminf_{n\to\infty} \mathbb{E}[W_n] = 1$. Combining this with Theorem 2.13 gives us that $W = 0$ almost-surely when $m \leq 1$, but we would like to know what happens when $m > 1$. The connection to the extinction probability is made clear by the following result.

Lemma 2.14. *Assuming $m < \infty$, then $\mathbb{P}(W = 0)$ equals either q or 1.*

Proof. If $m \leq 1$, then the tree dies out almost-surely, thus $\mathbb{P}(W = 0) = 1$. Thus suppose that $m \in (1, \infty)$. Looking at only the first step, we get that $Z_{n+1} = \sum_{i=1}^{Z_1} Z_n^{(i)}$, where $Z_n^{(i)}$ is the number of offspring at the nth generation (starting again at each offspring in the first generation). Note that if $Z_1 = 0$, we define this sum to be 0. These are copies of Z_n and all independent of each other. Then we have that:

$$Z_{n+1} = \sum_{i=1}^{Z_1} Z_n^{(i)}, \tag{2.3}$$

$$\frac{Z_{n+1}}{m^n} = \sum_{i=1}^{Z_1} \frac{Z_n^{(i)}}{m^n}, \tag{2.4}$$

$$\frac{mZ_{n+1}}{m^{n+1}} = \sum_{i=1}^{Z_1} \frac{Z_n^{(i)}}{m^n}. \tag{2.5}$$

Thus we see that:

$$\lim_{n\to\infty} \frac{mZ_{n+1}}{m^{n+1}} = \lim_{n\to\infty} \sum_{i=1}^{Z_1} \frac{Z_n^{(i)}}{m^n}, \tag{2.6}$$

$$mW = \sum_{i=1}^{Z_1} W^{(i)}, \tag{2.7}$$

where $W^{(i)}$ for all i are independent (of each other and Z_1) copies of W. We can then see that, by independence:

$$\mathbb{P}(W = 0) = \mathbb{E}\left[\mathbb{P}(W^{(1)} = 0, \dots, W^{(i)} = 0)\right] = \mathbb{E}\left[\mathbb{P}(W = 0)^{Z_1}\right]. \tag{2.8}$$

Thus we have that $\mathbb{P}(W = 0)$ is a solution to $f(s) = s$, for $s \in [0, 1)$ (note not equal to 1 since $m > 1$), and hence we have that $q = \mathbb{P}(W = 0)$. $\square$

I conclude this section with an insightful theorem from [KS66] which gives us information as to how W behaves, whose proof was the first use of spine techniques

Theorem 2.15 (Kesten-Stigum)**.** *Given that $m \in (1, \infty)$, then:*

$$\mathbb{E}[W] = 0 \iff \mathbb{P}(W > 0 | \textit{non-extinction}) = 1 \iff \mathbb{E}[X_0 \log(X_0^+)] < \infty,$$

where $X_0^+ = \max\{X_0, 1\}$.

The proof of this theorem uses size-biased Galton-Watson trees, which are a spine technique and the topic of Section 2.2.2.

Remark (Long term behaviour of Galton-Watson trees)**.** *Recalling notation from Equation 2.2, the asymptotics of the different regimes with Galton-Watson trees are well-known. If we are in the subcritical case, i.e. $m < 1$ then we have, $\forall k \in \mathbb{N}$:*

$$\lim_{t\to\infty} \frac{\mathbb{E}[Z_t^k]}{\mathrm{e}^{k(m-1)t}} = C_1 \in (0, \infty). \tag{2.9}$$

Hence we see that:

$$\lim_{t\to\infty} \frac{\mathbb{E}[Z_t^k]}{C_1\mathrm{e}^{k(m-1)t}} = 1, \tag{2.10}$$

thus $\mathbb{E}[Z_t^k] \sim C_1\mathrm{e}^{k(m-1)t}$, hence we have exponential decay. Similarly, for the critical case $m = 1$ and $\forall k \in \mathbb{N}$ we have:

$$\lim_{t\to\infty} \frac{\mathbb{E}[Z_t^k]}{t^{k-1}} = C_2 \in (0, \infty), \tag{2.11}$$

i.e. we have that:

$$\lim_{t\to\infty} \frac{\mathbb{E}[Z_t^k]}{C_2 t^{k-1}} = 1, \tag{2.12}$$

thus $\mathbb{E}[Z_t^k] \sim C_2 t^{k-1}$, hence we have polynomial growth. Finally for the supercritical case $m > 1$ and $\forall k \in \mathbb{N}$ we have that:

$$\lim_{t\to\infty} \frac{\mathbb{E}[Z_t^k]}{\mathrm{e}^{k(m-1)t}} = C_3 \in (0, \infty). \tag{2.13}$$

This is in turn equivalent to $\mathbb{E}[Z_t^k] \sim C_3\mathrm{e}^{k(m-1)t}$, which is exponential growth in t (since $m > 1$).

2.2.2 Size-biased Galton-Watson trees

We switch perspective here, instead of building the GW-tree iteratively, we see the entire tree as a random outcome of a probability space $(\Omega, \mathcal{F}, \mathbb{P})$[2]. Define $\mathcal{U} := \{\emptyset\} \cup \bigcup_{k=1}^{\infty}(\mathbb{N}^k)$. The elements of will be finite size tuples of natural numbers, for example $(1, 2, 4)$, $(1, 3, 10, 5)$ and we denote the tuple with no elements as $\emptyset$, which is the root. An element such as $(1, 2, 4)$ will represent the 4th offspring, of the 2nd offspring, of the 1st offspring of the root offspring.

We can concatenate elements of $\mathcal{U}$ by just concatenating the sequences. For $u, v \in \mathcal{U}$, we denote their concatenation by $uv \in \mathcal{U}$. Note that, $u\emptyset = \emptyset u = u$, for all $u \in \mathcal{U}$.

Definition 2.16 (Tree)**.** *A **tree** ω is defined as a subset of $\mathcal{U}$ satisfying:*

i. $\emptyset \in \omega$,

ii. *if $uj \in \omega$ for some $j \in \mathbb{N}$ then $u \in \omega$,*

iii. *if $u \in \omega$ then $uj \in \omega$ if and only if $1 \le j \le N_u(\omega)$ for some non-negative integer $N_u(\omega)$.*

Suppose that $u \in \mathcal{U}$, then u is a vertex of the tree, and the random variable $N_u(\omega)$ is the number of children it has, for outcome $\omega \in \Omega$. The vertices are labelled as $u = i_1 \dots i_n$, for $i_1, \dots, i_n \in \mathbb{N}$, which means that: u is the i_nth child of the i_{n-1} child, ..., of the i_1th child of the root $\emptyset$. For example if $u = 132$, as in Figure 4, then it is the 2nd child, of the 3rd child, of the 1st child of $\emptyset$. In the case of Figure 4, $N_1 = 3$ as the vertex $u = 1$ has 3 children, for this specific outcome $\omega \in \Omega$.

We can thus see that each $\omega \in \Omega$ corresponds to some tree. Thus we can see Ω as the space of trees. As in [Shi11] we construct a sigma-algebra for it, since we would like to answer probabilistic questions about these trees. Recalling that $\mathcal{U} = \{\emptyset\} \cup \bigcup_{k=1}^{\infty}(\mathbb{N}^k)$. For $u \in \mathcal{U}$, we define the subset $\Omega_u \subset \Omega$ as $\Omega_u := \{\omega \in \Omega : u \in \mathcal{U}\}$. Intuitively, Ω_u is the set of trees which contain the vertex u, i.e. with the line of descent determined by the label of u.

We can then define the sigma-algebra on Ω as the sigma-algebra generated by all such subsets:

$$\mathcal{F} = \sigma(\mathcal{F}_u : u \in \mathcal{U}),$$

[2] This new system is called the Ulam-Harris labelling system.

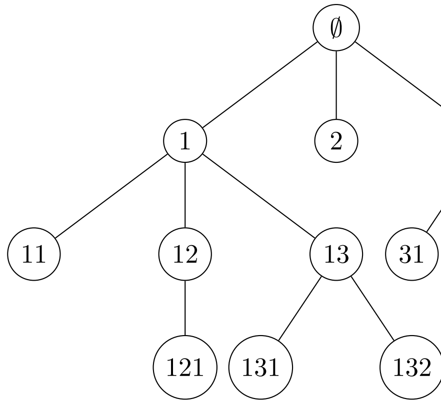


Figure 4: Example of a tree

where $\mathcal{F}_u$ represents the information about trees containing vertex u. Let us define a random variable $T : \Omega \to \Omega$, T stands for tree. The form of it looks odd, as it is essentially the identity function. However we interpret the Ω on each side differently. The domain Ω is the outcome space, as is normal for random variables, but the codomain space is the **space of trees**. Thus we have a tree-valued random variable. But since each $\omega \in \Omega$ corresponds to a tree, we can use the identity map.

We let $(p_k)_{k\geq 0}$ be a probability distribution on $\mathbb{N}_0$. As in [Shi11], by [Nev86] there exists a probability measure $\mathbb{P}$ on Ω such that T outputs trees which follow a Galton-Watson process with offspring distribution $(p_k)_{k\geq 0}$.

We now define sigma-algebras which are essentially restrictions of $\mathcal{F}$ up to some given generation $n \in \mathbb{N}$ of the trees. We define

$$\mathcal{F}_n := \sigma(\Omega_u : u \in \mathcal{U}, |u| \leq n),$$

where $|u|$ represents the length of $u \in \Omega$ i.e generation. For example $\mathcal{F}_3$ represents the information of all possible trees up to the third generation. The example tree from Figure 4 would be included, and thus we would have all the associated information about said tree in the sigma-algebra $\mathcal{F}_3$. Note that $\mathcal{F}$ is the sigma-algebra generated by all of these sigma-algebras, i.e. the smallest sigma-algebra containing them all.

We define now a random variable $Z_n : \Omega \to \mathbb{N}_0$, for $n \in \mathbb{N}$ as the number of individuals present at generation n. This is the same as the previous definition (Equation 2.2), but defined non-recursively. For the example tree in Figure 4, we have that $Z_3(\omega) = 3$, where $\omega \in \Omega$ is the tree.

We now give the key result of this section, the martingale $(W_n)_{n\geq 0}$ defined in Section 2.2.1 transforms the probability measure $\mathbb{P}$ to give a new probability measure $\hat{\mathbb{P}}$.

Proposition 2.17. *There exists such a measure $\hat{\mathbb{P}}$ such that, for any $n \in \mathbb{N}$:*

$$\hat{\mathbb{P}}_{|\mathcal{F}_n} = W_n \bullet \mathbb{P}_{|\mathcal{F}_n},$$

where we are restricting the measures to the sigma-algebra $\mathcal{F}_n$. Note that $\bullet$ means that for any $n \in \mathbb{N}$ and $A \in \mathcal{F}_n$:

$$\hat{\mathbb{P}}_{|\mathcal{F}_n}(A) = \int_A W_n \mathrm{d}\mathbb{P}_{|\mathcal{F}_n}.$$

This is often written in the form of a Radon-Nikodym derivative:

$$\frac{\mathrm{d}\hat{\mathbb{P}}_{|\mathcal{F}_n}}{\mathrm{d}\mathbb{P}_{|\mathcal{F}_n}} = W_n. \tag{2.14}$$

Combining the fact that $(W_n)_{n\geq 0}$ is a martingale, and since Proposition 2.17 gives us the necessary finite-dimensional distributions, we can safely apply *Kolmogorov's Extension Theorem* to get the existence of $\hat{\mathbb{P}}$. Note that for $n \in \mathbb{N}$ and $A \in \mathcal{F}_n$ we can write this (without the restrictions) as:

$$\hat{\mathbb{P}}(A) = \int_A W_n \mathrm{d}\mathbb{P} = \mathbb{E}[\mathbb{1}_A W_n].$$

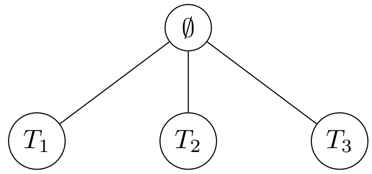


Figure 5: Example of $N = 3$, with subtrees T_1, T_2, T_3.

Observe that, for any $n \in \mathbb{N}$, we have that:

$$\hat{\mathbb{P}}(Z_n > 0) = \int_{\{\omega : Z_n(\omega) > 0\}} W_n \mathrm{d}\mathbb{P} = \int_\Omega W_n \mathbb{1}_{\{Z_n > 0\}} \mathrm{d}\mathbb{P} = \mathbb{E}[W_n \mathbb{1}_{\{Z_n > 0\}}].$$

Recall that $W_n = \frac{Z_n}{m^n}$, then $W_n = W_n \mathbb{1}_{\{Z_n > 0\}}$. Hence $\mathbb{E}[W_n \mathbb{1}_{\{Z_n > 0\}}] = \mathbb{E}[W_n]$. This equals 1 by Section 2.2.1. Thus we have that, for all $n \in \mathbb{N}$:

$$\hat{\mathbb{P}}(Z_n > 0) = 1.$$

This is a significant result, as it says that under $\hat{\mathbb{P}}$ we have survival almost surely of the Galton-Watson tree T. We therefore call the tree T under $\hat{\mathbb{P}}$ the **size-biased Galton-Watson tree**. Let $N = N_\emptyset$, the number of offspring of the root vertex $\emptyset$. We only have a non-trivial tree if $N \geq 1$, thus supposing that is the case, we consider the N subtrees (rooted at each offspring). For example we see what this looks like with $N = 3$ in Figure 5.

Proposition 2.18 (Size-biasing)**.** *Suppose $A_1, \ldots, A_k \in \mathcal{F}$, then we have:*

$$\hat{\mathbb{P}}(N = k, T_1 \in A_1, \ldots, T_k \in A_k) = \frac{kp_k}{m} \frac{1}{k} \sum_{i=1}^{k} \mathbb{P}(A_1) \cdots \mathbb{P}(A_{i-1}) \hat{\mathbb{P}}(A_i) \mathbb{P}(A_{i+1}) \cdots \mathbb{P}(A_k)$$

Proof. Given $A_1, \ldots, A_k \in \mathcal{F}$, we have that:

$$\hat{\mathbb{P}}(N = k, T_1 \in A_1, \ldots, T_k \in A_k) = \hat{\mathbb{P}}(T_1 \in A_1, \ldots, T_k \in A_k \mid N = k) \hat{\mathbb{P}}(N = k).$$

We have that $\hat{\mathbb{P}}(N = k) = \mathbb{E}(\mathbb{1}_{\{N=k\}} W_1)$, as $\mathbb{1}_{\{N=k\}}$ is $\mathcal{F}_1$-measurable, since we know how many offspring $\emptyset$ had at the first generation. Note that $W_1 = \frac{Z_1}{m}$ and $Z_1 = N$ thus we have $\mathbb{E}(\mathbb{1}_{\{N=k\}} W_1) = \frac{kp_k}{m}$. It still remains to show that:

$$\hat{\mathbb{P}}(T_1 \in A_1, \ldots, T_k \in A_k \mid N = k) = \frac{1}{k} \sum_{i=1}^{k} \mathbb{P}(A_1) \cdots \mathbb{P}(A_{i-1}) \hat{\mathbb{P}}(A_i) \mathbb{P}(A_{i+1}) \cdots \mathbb{P}(A_k).$$

Given that we know that $Z_1 = k$, we use the fact that:

$$Z_n = \sum_{i=1}^{k} Z_{n-1}^{(i)}, \quad \text{for} \quad n \geq 2$$

where $Z_{n-1}^{(i)}$ is the number of offspring in the $(n-1)$th generation of the ith child/subtree of $\emptyset$. Note that each of these offspring is a root of a new tree (the trees are independent copies of the original GW tree), thus at each offspring we "restart". Thus we have that:

$$W_n = \frac{Z_n}{m^n} = \frac{\sum_{i=1}^{k} Z_{n-1}^{(i)}}{m^n} = \frac{1}{m} \sum_{i=1}^{k} \frac{Z_{n-1}^{(i)}}{m^{n-1}} = \frac{1}{m} \sum_{i=1}^{k} W_{n-1}^{(i)}, \tag{2.15}$$

letting $W_{n-1}^{(i)} = \frac{Z_{n-1}^{(i)}}{m^{n-1}}$. We require the following definition and lemma to complete the proof.

Definition 2.19 (π-system)**.** *A **π-system** is a collection of sets $\mathcal{U}$ such that for any two $A, B \in \mathcal{U}$ we have that $A \cap B \in \mathcal{U}$.*

Lemma 2.20 (Agreement on a π-system)**.** *Suppose that $\mathcal{F} = \sigma(\mathcal{U})$ for $\mathcal{U}$ a π-system, and we have two measures μ_1 and μ_2 which agree on $\mathcal{U}$, then they agree on $\mathcal{F}$.*

Let us start by defining the two measures μ_1 and μ_2.

$$\mu_1 = \hat{\mathbb{P}}(T_1 \in A_1, \ldots, T_k \in A_k \mid N = k), \tag{2.16}$$

$$\mu_2 = \frac{1}{k}\sum_{i=1}^{k} \mathbb{P}(A_1)\cdots\mathbb{P}(A_{i-1})\hat{\mathbb{P}}(A_i)\mathbb{P}(A_{i+1})\cdots\mathbb{P}(A_k). \tag{2.17}$$

Recall that $\mathcal{F} = \sigma(\bigcup_{n\geq 0}\mathcal{F}_n)$, we claim that $\bigcup_{n\geq 0}\mathcal{F}_n$ is a π-system, and thus Lemma 2.20 applies. Suppose $A, B \in \bigcup_{n\geq 0}\mathcal{F}_n$, then w.l.o.g. we have that $A \in \mathcal{F}_i$ and $B \in \mathcal{F}_j$ where $i < j$, then since $\mathcal{F}_i \subset \mathcal{F}_j$ we have that $A, B \in \mathcal{F}_j$. Thus $A \cap B \in \mathcal{F}_j$ and thus is in $\bigcup_{n\geq 0}\mathcal{F}_n$, hence we have a π-system, and Lemma 2.20 applies.

We can weaken the condition on $A_1, \ldots, A_k$ to instead be that $A_1, \ldots, A_k \in \bigcup_{n\geq 0}\mathcal{F}_n$, and thus taking $n \in \mathbb{N}_0$ as the largest $\mathcal{F}_n$ such that $A_i \in \mathcal{F}_n$ for all $i \in \{1, \ldots, k\}$. Thus we can now work with the assumption that $A_1, \ldots, A_k \in \mathcal{F}_n$. Hence we have that

$$\hat{\mathbb{P}}(T_1 \in A_1, \ldots, T_k \in A_k \mid N = k) = \mathbb{E}\left[\mathbb{1}_{T_1\in A_1,\ldots,T_k\in A_k} W_n \mid N = k\right]. \tag{2.18}$$

Substituting in the expression from Equation 2.15 gives:

$$\hat{\mathbb{P}}(T_1 \in A_1, \ldots, T_k \in A_k \mid N = k) = \mathbb{E}\left[\mathbb{1}_{T_1\in A_1,\ldots,T_k\in A_k} W_n \mid N = k\right], \tag{2.19}$$

$$= \mathbb{E}\left[\mathbb{1}_{T_1\in A_1,\ldots,T_k\in A_k} \frac{1}{m}\sum_{i=1}^{k} W_{n-1}^{(i)}\right], \tag{2.20}$$

$$= \frac{1}{m}\sum_{i=1}^{k}\mathbb{E}\left[\mathbb{1}_{T_1\in A_1,\ldots,T_k\in A_k} W_{n-1}^{(i)}\right]. \tag{2.21}$$

As the subtrees behave independently, we can split this up into the product:

$$\frac{1}{m}\sum_{i=1}^{k}\mathbb{E}\left[\mathbb{1}_{T_1\in A_1,\ldots,T_k\in A_k} W_{n-1}^{(i)}\right] = \frac{1}{m}\sum_{i=1}^{k}\mathbb{E}\left[\mathbb{1}_{T_1\in A_1}\right]\cdot\mathbb{E}\left[\mathbb{1}_{T_2\in A_2}\right]\cdots\mathbb{E}\left[\mathbb{1}_{T_i\in A_i} W_{n-1}^{(i)}\right]\cdots\mathbb{E}\left[\mathbb{1}_{T_k\in A_k}\right], \tag{2.22}$$

$$= \frac{1}{m}\sum_{i=1}^{k}\mathbb{P}(T_1 \in A_1)\cdot\mathbb{P}(T_2 \in A_2)\cdots\mathbb{E}\left[\mathbb{1}_{T_i\in A_i} W_{n-1}^{(i)}\right]\cdots\mathbb{P}(T_k \in A_k). \tag{2.23}$$

We claim that, for all $i \in \{1, \ldots, k\}$ $\mathbb{E}\left[\mathbb{1}_{T_i\in A_i} W_{n-1}^{(i)}\right] = \frac{m}{k}\hat{\mathbb{P}}(A_i)$. Note that, recalling that $(W_{n-1}^{(j)})_{1\leq j\leq k}$ are i.i.d.

$$\hat{\mathbb{P}}(T_i \in A_i) = \mathbb{E}\left[\mathbb{1}_{T_i\in A_i} W_n\right] = \frac{1}{m}\sum_{j=1}^{k}\mathbb{E}\left[\mathbb{1}_{T_i\in A_i} W_{n-1}^{(j)}\right] = \frac{k}{m}\mathbb{E}\left[\mathbb{1}_{T_i\in A_i} W_{n-1}^{(i)}\right], \tag{2.24}$$

which justifies the claim. Substituting Equation 2.24 into Equation 2.23 gives:

$$\frac{1}{m}\sum_{i=1}^{k}\mathbb{P}(T_1 \in A_1)\cdot\mathbb{P}(T_2 \in A_2)\cdots\frac{m}{k}\hat{\mathbb{P}}(A_i)\cdots\mathbb{P}(T_k \in A_k).$$

Simplifying and noting that for $1 \leq j \leq k$, due to "restarting" at each subtree that $\mathbb{P}(T_j \in A_j) = \mathbb{P}(A_j)$, this gives the final expression:

$$\hat{\mathbb{P}}(T_1 \in A_1, \ldots, T_k \in A_k \mid N = k) = \frac{1}{k}\sum_{i=1}^{k}\mathbb{P}(A_1)\cdots\mathbb{P}(A_{i-1})\hat{\mathbb{P}}(A_i)\mathbb{P}(A_{i+1})\cdots\mathbb{P}(A_k),$$

thus the two measures μ_1, μ_2 agree on $\bigcup_{n\geq 0}\mathcal{F}_n$ and therefore the result extends to $\mathcal{F}$.

$\square$

This is a highly intuitive result, as the first term $\frac{kp_k}{m}$ corresponds to the size-biased probability of $N = k \in \mathbb{N}_0$ subtrees of the root node $\emptyset$. This is called size-biasing as we are making a higher number of offspring more likely, and then in order to still have a probability distribution we normalise using the first moment. Then among these subtrees, one is chosen at random with probability $\frac{1}{k}$, i.e. uniformly over the options, to be the special "size-biased" subtree. We finally sum over all the possible special "size-biased" subtrees, where the non-special subtrees are independent ordinary Galton-Watson trees. We can see this as the non-special subtrees act under $\mathbb{P}$, but the special subtree acts under $\hat{\mathbb{P}}$.

Following [Shi11], we can see this as a spine decomposition. Where we denote the spine by $(w_i)_{i\geq 0}$, where $w_0 = \emptyset$. From the subtrees rooted at all the offspring of w_0, we uniformly choose one to be the next element of the spine w_1, whereas all others are independent (non size-biased) Galton-Watson trees. We can iterate this process to get our infinite spine (recall that we do not go extinct almost-surely), as in Figure 6.

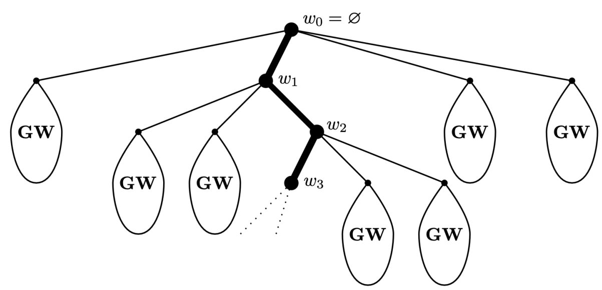


Figure 6: An example of a size-biased Galton-Watson tree (taken from [Shi11]).

2.2.3 Spines in continuous time

In this chapter, we explore the more general spine setup of [HR14], before specialising in Section 3. We work with a continuous-time branching process, where particles move on a countable state space S, and we start with one particle at $x \in S$ which moves according to a Markov process with *generator*[3] $\mathcal{M}$. When a particle is at position $y \in S$, then it branches at rate $R(y)$, according to the function $R : S \to \mathbb{R}_0^+$. If a particle is at $y \in S$ and it branches, it dies and is replaced by a random number of offspring, according to the distribution $\mu^{(y)}$ on $\mathbb{N}_0$. Note that the time this happens at is a generalisation of the exponential distribution (where the rate is variable) and thus the probability that a particle has not branched by time $t > 0$, conditional on $(X(s))_{0\leq s\leq t}$ is:

$$\mathrm{e}^{-\int_0^t R(X(s))\mathrm{d}s}, \tag{2.25}$$

where $X(s)$ gives the position in S of the particle at time $s > 0$. Equation 2.25 simplifies down to one minus the familiar exponential distribution CDF formula when $R \equiv \lambda > 0$, which gives the expected probability:

$$\mathrm{e}^{-\lambda t}.$$

We denote by $N(t)$ is the set of particles alive at time $t > 0$, when this instead means the number will be clear.

Definition 2.21 (Time of birth/death)**.** *Given a time $t > 0$ and a particle $v \in N(t)$, we denote by σ_v the **birth time** of the particle, and respectively by τ_v the **death time**. We also define the following for $t > 0$:*

$$\sigma_v(t) = \min\{t, \sigma_v\} \quad \textit{and} \quad \tau_v(t) = \min\{t, \tau_v\},$$

which become fixed once the birth/death respectively of v has occurred.

These quantities are useful as we often work with t which is between the birth and death of a particle and such quantities will then be useful for calculation.

For $s > 0$ we define the position $X_v(s)$ of a particle $v \in N(s)$ as follows:

[3]For the sake of brevity, we can understand the generator as simply a way of defining the Markov process. In this case it is the same as the Q-matrix of the process.

Definition 2.22 (Position of a particle v). *We define the position of a particle v at time s as:*

$$X_v(s) = \begin{cases} \text{position of unique ancestor of } v \text{ alive at time } s, \\ \Delta \text{ when } v \text{ has zero children and } s \geq \tau_v, \end{cases}$$

where $\Delta \notin S$ is the graveyard/dead state.

We now define two probability measures $\mathbb{P}_x^k$ and $\mathbb{Q}_x^k$ on this setup, with $k \in \mathbb{N}$ referring to the number of marks or spines the root particle carries or is an ancestor of at time 0. These marks can spread throughout system as branching and movement occurs, defining the lines of descent of the spines. The measure $\mathbb{P}_x^k$ is simply the measure $\mathbb{P}_x$ but with the extra information (i.e. a larger σ-algebra) of where the marks are. On the other hand, particles with a mark will behave **differently** under $\mathbb{Q}_x^k$. Observe that we will have exactly k distinct lines of descent, particles in these spines will have their behaviour modified.

Summary of behaviour of particles under the probability measure $\mathbb{P}_x^k$:

- At time 0, one particle is placed at $x \in S$, which carries k marks. This will be called the **root particle**.
- Particles move according to the Markov process with generator $\mathcal{M}$, independently of each other and the marks.
- Each of the k marks gives us a specific line of descent or "spine", as the marks spread throughout the system the particles alive are all descendants of the root particle, as well their other ancestors. We define ξ_t^i for $1 \leq i \leq k$ to be the position of the (unique) particle carrying the ith mark at time $t > 0$.
- A particle at position $y \in S$ carrying j marks at time $t > 0$ branches at rate $R(y)$, when it branches it dies and replaced by a random number of offspring according to the distribution $\mu^{(y)}$, independently of the rest of the system.
- Given that n particles are born at a branching event, the marks will be distributed to the offspring. Each mark moves to an offspring, each of which has uniform probability $\frac{1}{n}$ of being chosen, and independently of each other and the rest of the system.
- If a particle dies and is replaced by 0 offspring, then it moves into the graveyard state Δ and the marks remain with it.

As previously said, particles move the same under $\mathbb{P}_x^k$ as under $\mathbb{P}_x$, but as we have seen we have the extra information about the movement of the k marks. Specifically, we denote by $(\mathcal{F}_t^k)_{t\geq 0}$ the filtration containing information about the system up to time $t > 0$ (including the k marks). We have that $\mathbb{P}_x^k$ is defined on the sigma-algebra $\mathcal{F}_\infty^k = \sigma\left(\bigcup_{t\geq 0} \mathcal{F}_t^k\right)$.

Definition 2.23 (Skeleton). *We will denote by* $\text{skel}(t)$ *for $t > 0$ the* ***skeleton*** *at time t. The skeleton is the collection of particles which have carried at least one mark up to time t. In other words, all ancestors and current particles in the k lines of descent.*

Definition 2.24 (Size-biased distribution). *For $n \in \mathbb{N}_0$, and $y \in S$ we have: $m_n(y) = \sum_{a\in\mathbb{N}_0} a^n \mu^{(y)}(a)$, as the nth moment of the offspring distribution at $y \in S$. We now define, for $n \in \mathbb{N}_0$, and $a \in \mathbb{N}_0$:*

$$\mu_n^{(y)}(a) = \frac{a^n \mu^{(y)}(a)}{m_n(y)},$$

as the nth ***size-biased distribution*** *of $\mu^{(y)}$.*

Note that this is very similar to the size-biased distribution that was introduced in Section 2.2.2. The only difference here is that we have introduced n in order to use the nth moments to normalise the distribution, whereas in Section 2.2.2 we used $n = 1$.

We also define the random variable $T(i, j)$ for $1 \leq i, j \leq k$ which is the first time the marks i and j are carried by different particles, note that this is well-defined as all k marks start off on the root particle. In addition to this, we denote by $D(v)$ the number of marks carried by particle v.

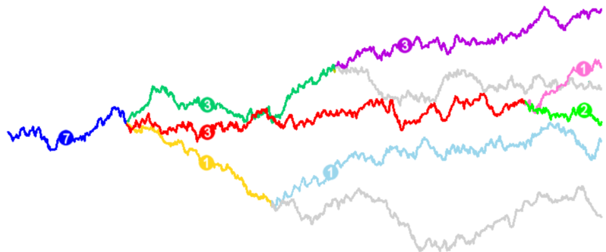


Figure 7: An example of a skeleton, each colour represents a different particle in the skeleton. Grey represents particles not in the skeleton. The number of each particle is the number of marks it carries (taken from [DR12]).

Let $\zeta(X,t)$ be a function of a process $(X_t)_{t\geq 0}$ such that if $(X_t)_{t\geq 0}$ is a Markov process with generator $\mathcal{M}$, then $\zeta(X,t)$ is a non-negative unit-mean martingale, with respect to the natural filtration of $(X_t)_{t\geq 0}$. Note that the constant martingale $\zeta \equiv 1$ will always work for this purpose.

We can now give the behaviour of particles under $\mathbb{Q}^k_x$.

- We start with a root particle at $x \in S$ carrying k marks.
- As for $\mathbb{P}^k_x$. we think of each mark as denoting a spine or line of descent, we write ψ^i_t for the particle carrying mark $1 \leq i \leq k$ at time $t > 0$, we continue to write ξ^i_t for the position of this particle.
- A particle with only mark i at time $t > 0$ moves as if under the change-of-measure $\mathbb{Q}^k_x\,|_{\sigma((\xi^i_s)_{s\leq t})} := \zeta(\xi^i,t)\mathbb{P}^k_x\,|_{\sigma((\xi^i_s)_{s\leq t})}$.
- A particle at position $y \in S$, carrying j marks at time $t > 0$ branches at a different rate $m_j(y)R(y)$, as opposed to $R(y)$. When a particle branches, it dies and is replaced by a random number of offspring according to the jth size biased distribution $\mu^{(y)}_j$, independently of the rest of the system.
- As for $\mathbb{P}^k_x$, given that n particles are born at a branching event, the marks will be distributed to the offspring in the same way.
- Particles carrying no marks, i.e. not in the skeleton, behave exactly as under $\mathbb{P}^k_x$. In other words, their movement and branching behaviours are not weighted or size-biased respectively. They branch at rate $R(y)$ at $y \in S$, and give birth to a random number of offspring according to the distribution $\mu^{(y)}$.

In summary, under $\mathbb{Q}^k_x$ the movement of particles in the skeleton is weighted according to ζ and their offspring distributions are size-biased according to the number of marks they carry. However, particles not in the skeleton move and branch as they normally would. In particular, one can define a martingale [4] $(\tilde{\xi}_t)_{t\geq 0}$ such that $\mathbb{Q}^k$ and $\mathbb{P}^k$ are related via the Radon-Nikodym derivative:

$$\frac{\mathrm{d}\mathbb{Q}^k_{|\mathcal{F}^k_t}}{\mathrm{d}\mathbb{P}^k_{|\mathcal{F}^k_t}} = \tilde{\xi}_t. \tag{2.26}$$

This setup is analogous to the setup in Section 2.2.2, in particular Equation 2.26 here is analogous to Equation 2.14. Finally, we introduce the **excess branching rate** at $y \in S$ for nth size-biased distribution:

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

[4] See [HR14] for the details of this construction.

2.2.4 The many-to-few lemma

Recall that $(\mathcal{F}_t^k)_{t\geq 0}$ is the filtration containing information about the system up to time $t > 0$, including information about the k marks. Suppose that Y is a random variable which is $\mathcal{F}_t^k$-measurable for some $t > 0$, then we can write Y in the form:

$$Y = \sum_{v_1,\ldots,v_k \in N(t)} Y(v_1,\ldots,v_k)\mathbb{1}_{\{\psi_t^1=v_1,\ldots,\psi_t^k=v_k\}}, \tag{2.27}$$

where for any $v_1,\ldots,v_k$, we have that $Y(v_1,\ldots,v_k)$ is $\mathcal{F}_t$-measurable. To explain this form in words, it is decomposed according to which particles currently carry the k marks, the indicators partitions the space of possibilities. We get this form because the k marks must be on exactly one combination of the particles present in the system at time t, i.e. $N(t)$. Thus, exactly one of the indicators will be non-zero. Note also that Y can depend on all particles not just the spines, but adding the $v_1,\ldots,v_k$ in brackets are a convenient notation for that specific random variable. This is an example of the situation where $(A_i)_{1\leq i\leq \ell}$ form a partition of the sample space Ω, and thus for a random variable X defined on the probability space:

$$X = \sum_{i=1}^{\ell} X_i \mathbb{1}_{A_i},$$

where $(X_i)_{1\leq i\leq \ell}$ are a "decomposition" of X.

We now state the many-to-few "lemma" where we can use carefully chosen choices of Y to reduce the complexity of expectations. We do this by choosing Y to have the form in the left hand side expectation of the following theorem.

Theorem 2.25 (General many-to-few lemma)**.** *For any $k \geq 1$, $x \in S$ and $\mathcal{F}_t^k$-measurable Y as setup in section 2.2.3,*

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

Example 2.26. *Suppose we are working with $R(y) = \mathbb{1}_{\{y=0\}}$ where $S = \mathbb{Z}$ and wish to simplify the calculation of:*

$$\mathbb{P}_0\left[\#\{v \in N(t) : |X_v(t)| \leq 10\}\right], \tag{2.28}$$

at some time $t > 0$. We can use $Y = \mathbb{1}_{\{|\xi_t^1|\leq 10\}}$, then we can write in the form of Equation 2.27 in the following way:

$$Y = \sum_{v\in N(t)\cup D(t)} \mathbb{1}_{\{|X_v(t)|\leq 10\}}\mathbb{1}_{\{\psi_t^1=v\}}. \tag{2.29}$$

We can then note that $Y(v) = \mathbb{1}_{\{|X_v(t)|\leq 10\}}$, thus we can now apply Theorem 2.25 with the constant functional $\zeta \equiv 1$ to obtain:

$$\mathbb{P}_0\left[\sum_{v\in N(t)} Y(v)\right] = \mathbb{Q}_0^1\left[\mathbb{1}_{\{|\xi_t^1|\leq 10\}}\prod_{v\in \mathrm{skel}(t)}\exp\left(\int_{\sigma_v(t)}^{\tau_v(t)}\alpha_{D(v)}(X_v(s))\mathrm{d}s\right)\right]. \tag{2.30}$$

Note that $\#\{v \in N(t) : |X_v(t)| \leq 10\} = \sum_{v\in N(t)} Y(v)$, thus we have that:

$$\mathbb{P}_0\left[\#\{v \in N(t) : |X_v(t)| \leq 10\}\right] = \mathbb{Q}_0^1\left[\mathbb{1}_{\{|\xi_t^1|\leq 10\}}\prod_{v\in \mathrm{skel}(t)}\exp\left(\int_{\sigma_v(t)}^{\tau_v(t)}\alpha_{D(v)}(X_v(s))\mathrm{d}s\right)\right]. \tag{2.31}$$

This further simplifies to:

$$\begin{aligned}\mathbb{P}_0\left[\#\{v \in N(t) : |X_v(t)| \leq 10\}\right] &= \mathbb{Q}_0^1\left[\mathbb{1}_{\{|\xi_t^1|\leq 10\}}\exp\left(\int_0^t \alpha_1(\xi_s)\mathrm{d}s\right)\right], &(2.32)\\ &= \mathbb{Q}_0^1\left[\mathbb{1}_{\{|\xi_t^1|\leq 10\}}\exp\left(\int_0^t \mathbb{1}_0(\xi_s)\mathrm{d}s\right)\right], &(2.33)\end{aligned}$$

as we only have one spine particle ξ. The right-hand expression is simpler to deal with.

3 The Döring-Roberts paper

3.1 Assumptions and notation

In [DR12], the branching process begins with one particle, denoted by ξ, at some starting position. The motion of this particle follows an **irreducible** continuous-time Markov process with Q-matrix $\mathcal{A}$ on a countable state space S. When a particle is at $0 \in S$, then the particle branches with rate 1. If a particle branches, then it dies and is replaced by a random number of offspring, with distribution μ. These particles then continue to move around S according to the same Markov process $\mathcal{A}$ (independently of each other) and branch independently (still only at rate 1 at the origin). We assume that μ has finite moments of all orders. This is called the **catalytic branching random walk**. This contrasts to the MS model, the key difference being the fact that branching occurs here occurs only at the origin, whereas in the MS model it occurs everywhere but at one rate at the origin, and a smaller rate outside the origin. For $y \in S$ and $t > 0$ we define $N_t(y)$ to be the number of particles at site y at time t, and write N_t for the total number of particles in the system at time t, i.e.

$$N_t = \sum_{y \in S} N_t(y).$$

Definition 3.1 (Moments). *For $x, y \in S$ and $k \in \mathbb{N}$:*

$$M^k(t, x, y) = \mathbb{E}[N_t(y)^k | \xi_0 = x] \quad \textit{and} \quad M^k(t, x) = \mathbb{E}[N_t^k | \xi_0 = x],$$

are the kth moments of the number of particles at position $y \in S$, as well as the kth moments of the total number of particles respectively.

Note that, conditioning on $\xi_0 = x$ means we place the initial particle ξ at state $x \in S$.

Definition 3.2 (Green's function). *Given the transition probability $p_t(x, y) = \mathbb{P}_x(\xi_t = y)$ for $x, y \in S$, Green's function is defined as follows:*

$$G_\infty(x, y) = \int_0^\infty p_t(x, y) \mathrm{d}t.$$

This definition can be better interpreted as the continuous analogue to the expected number of visits to y, starting from x. For $x, y \in S$ and $t > 0$, we see that:

$$\begin{aligned}
\sum_{t \in \mathbb{N}} p_t(x, y) &= \sum_{t \in \mathbb{N}} \mathbb{E}_x \left[\mathbb{1}_{\{\text{the walk is at } y \text{ at time } t\}} \right], \\
&= \mathbb{E}_x \left[\sum_{t \in \mathbb{N}} \mathbb{1}_{\{\text{the walk is at } y \text{ at time } t\}} \right], \\
&= \mathbb{E}_x[\# \text{ of visits to } y \text{ starting from } x].
\end{aligned}$$

Note that, swapping the infinite sum and expectation is allowed by *Fubini-Tonelli's theorem.*

Definition 3.3 (Time spent at site y). *This is defined as:*

$$L_t(y) = \int_0^t \mathbb{1}_{\{\xi_s = y\}} \mathrm{d}s.$$

As well as these definitions, we introduce a specialised version of the many-to-few lemma, i.e. Theorem 2.25.

Lemma 3.4 (Specialised Many-to-few). *Recalling the notation used in Section 2.2.3, and that $B_v \equiv D(v), \forall v$. Suppose that $f : \mathbb{R} \to \mathbb{R}$ is measurable. Then, for any $k \geq 1$ and $t > 0$,*

$$\mathbb{E}\left[\sum_{v_1, \ldots, v_k \in N_t} f(X_{v_1}(t)) \cdots f(X_{v_k}(t)) \right] = \mathbb{Q}^k \left[f(\chi_t^1) \cdots f(\chi_t^k) \prod_{v \in \mathrm{skel}(t)} \exp\left((m_{B_v} - 1) \int_{\sigma_v(t)}^{\tau_v(t)} \mathbb{1}_0(X_v(s)) \, \mathrm{d}s \right) \right].$$

Proof. We shall use Theorem 2.25 to deduce this version of the many-to-few lemma. We start by using the constant martingale $\zeta \equiv 1$ and setting $Y(v_1, \dots, v_k) = f(X_{v_1}(t)) \cdots f(X_{v_k}(t))$. This satisfies the measurability requirements as f is measurable. We also note that the

$$\alpha_{D(v)}(X_v(s)) = (m_{D(v)}(X_v(s)) - 1)R(X_v(s)) = (m_{B_v} - 1)\mathbb{1}_0(X_v(s)),$$

recalling that $D(v) \equiv B_v$, and that we only branch at rate 1 at the origin in the context of the DR model. $\square$

Following the convention of [DR12] I use both $\mathbb{E}[\cdot]$ and $\mathbb{P}[\cdot]$ for the expectation under $\mathbb{P}$, but only use $\mathbb{Q}[\cdot]$ for the expectation under $\mathbb{Q}$.

3.2 The main result

Recall that $m = \mathbb{E}[X]$, for $X \sim \mu$, the expected number of offspring. Further recall that an irreducible Markov process governed by $\mathcal{A}$ is **transient** if it returns to some state only finitely-many times almost surely, otherwise it is **recurrent**. The first part of the main result gives us asymptotics in the subcritical regime, i.e. when $m < 1$.

Theorem 3.5 (Subcritical Case)**.** *For all integers $k \geq 1$, $t > 0$ and $m < 1$, we have:*

$$\lim_{t\to\infty} M^k(t,x) \begin{cases} \in (0,\infty) \text{ when } \mathcal{A} \text{ is transient,} \\ = 0 \text{ when } \mathcal{A} \text{ is recurrent.} \end{cases}$$

In both cases we have:

$$\lim_{t\to\infty} M^k(t,x,y) = 0,$$

for any $x, y \in S$.

This result is quite intuitive, as particles which branch are on average dying (i.e. having zero offspring), so it implies that the total number of particles we should see the total settle to some limit (especially as particles move away from zero and stop branching and thus dying), and that the number of particles at any position goes to zero. The second part of the main result gives us asymptotics in the critical regime, $m = 1$.

Theorem 3.6 (Critical Case)**.** *For all integers $k \geq 1$, $t > 0$ and $m = 1$, we have:*

$$\lim_{t\to\infty} \frac{M^k(t,x)}{\mathbb{E}_x[L_t(0)^{k-1}]} \in (0,\infty) \quad \text{and} \quad M^1(t,x,y) = p_t(x,y).$$

These two results are interesting by themselves, but the final part of this final result shows a similar behaviour to localisation/non-localisation phase transition occurring.

Theorem 3.7 (Supercritical Case)**.** *For all integers $k \geq 1$, $t > 0$ and $m > 1$, there exists a* ***critical*** *constant $\beta \geq 1$ such that:*

$$\beta = \frac{1}{G_\infty(0,0)} + 1.$$

a) For $m < \beta$

$$\lim_{t\to\infty} M^1(t,x) \in (0,\infty) \quad \text{and} \quad \lim_{t\to\infty} M^1(t,x,y) = 0,$$

and we can bound $M^k(t,x)$ in the following way, i.e. $\exists c, C > 0$ such that:

$$c\mathbb{E}_x[L_t(0)^{k-1}] \leq M^k(t,x) \leq Ct^{k-1}$$

b) For $m = \beta$

$$\lim_{t\to\infty} M^k(t,x) = \infty$$

and we have:

$$\lim_{t\to\infty} M^k(t,x,y) = \infty \quad \text{when} \quad \int_0^\infty rp_r(0,0)\mathrm{d}r = \infty,$$

$$\lim_{t\to\infty} \frac{M^k(t,x,y)}{t^{k-1}} \in (0,\infty) \quad when \quad \int_0^\infty r p_r(0,0)\mathrm{d}r < \infty,$$

and either way $M^k(t,x,y)$ grows sub-exponentially.
c) For $m > \beta$

$$\lim_{t\to\infty} \mathrm{e}^{-kr(m)t} M^k(t,x,y) \in (0,\infty) \quad and \quad \lim_{t\to\infty} \mathrm{e}^{-kr(m)t} M^k(t,x) \in (0,\infty)$$

*where $r(m)$ equals the **unique** solution λ to*

$$\int_0^\infty \mathrm{e}^{-\lambda t} p_t(0,0)\mathrm{d}t = \frac{1}{m-1}.$$

This $r(m)$ from part (c) of Theorem 3.7 has certain properties, that we summarise in Corollary 3.8.

Corollary 3.8. *The function $m \mapsto r(m)$ is a convex function, with the following properties:*

$$r(1) = 0, \quad r(m) \le m-1, \quad \lim_{m\to\infty} \frac{r(m)}{m-1} = 1$$

Proof. Let us define a function $f : [0,\infty) \to [0,\infty)$ such that:

$$f(\lambda) = \int_0^\infty \mathrm{e}^{-\lambda t} p_t(0,0)\mathrm{d}t. \tag{3.1}$$

Suppose that $m > \beta$, for β defined in Theorem 3.7. We can thus see that $\int_0^\infty p_t(0,0) > \frac{1}{m-1}$. We can see that f is continuous as a Laplace transform of $\mathrm{e}^{-\lambda t} p_t(0,0)$. Notice also (by dominated convergence) that $f(\lambda_n) \to 0$, as $\lambda_n \to \infty$. Thus we can see that the function $m \mapsto r(m)$ is well defined as $r(m) = f^{-1}(\frac{1}{m-1})$. We can thus see that:

$$\frac{\mathrm{d}r}{\mathrm{d}\lambda} = \frac{1}{f'(f^{-1}(\frac{1}{m-1}))} \cdot \log(m-1), \tag{3.2}$$

and by differentiating under the integral sign we see that $f'(\lambda) \le 0$ for all $\lambda \in [0,\infty)$. Therefore we see that r is convex. Notice also that f is decreasing in λ. We see that as $m \to 1$ we have that $f^{-1}(\frac{1}{m-1}) \to 0$, thus by continuity we have that $r(1) = 0$. We have that:

$$f(m-1) = \int_0^\infty \mathrm{e}^{-(m-1)t} p_t(0,0)\mathrm{d}t, \tag{3.3}$$

$$\le \int_0^\infty \mathrm{e}^{-(m-1)t}\mathrm{d}t = \frac{1}{m-1}. \tag{3.4}$$

Thus we have that $f(m-1) \le f(r(m))$, and since f is decreasing we must therefore have that $r(m) \le m-1$. Finally note that r is increasing in m, since if one takes m larger, then one must take λ larger in order to get $f(\lambda) = \frac{1}{m-1}$, since f is decreasing. This combined with the fact that $r(m) \le m-1$ gives that $\lim_{m\to\infty} \frac{r(m)}{m-1} = 1$. □

We see a phase transition at $m = \beta$ where in case (a) of Theorem 3.7 we see a non-localisation or diffusion type of behaviour of the moments, but in case (c) we see a similar behaviour to localisation. This is an indication that perhaps something similar can be achieved in the everywhere-branching case, which is dealt with in Section 4.

3.3 Feynman-Kac representation of moments

Lemma 3.9 (Feynman-Kac representation)**.** *In the context of Section 3.2, for $t > 0$ and $x, y \in S$, we have the following:*

$$M^1(t,x) = \mathbb{E}_x\left[\mathrm{e}^{(m-1)\int_0^t \mathbb{1}_0(\xi_s)\mathrm{d}s}\right] \tag{3.5}$$

$$M^1(t,x,y) = \mathbb{E}_x[\mathrm{e}^{(m-1)\int_0^t \mathbb{1}_0(\xi_s)\mathrm{d}s}\mathbb{1}_y(\xi_t)] \tag{3.6}$$

where ξ_t represents a single particle moving with Q-matrix $\mathcal{A}$.

Proof. Suppose $t > 0$ and $x, y \in S$, we use Theorem 3.4 with $k = 1$ as we have one spine, $f \equiv 1$ (for the $M^1(t, x, y)$ we use $f(z) = \mathbb{1}_y(z)$, $z \in S$). This simplifies the skeleton, since we only have one line of descent. We refer to the only particle (at a time) which is in this spine as ξ. By Theorem 3.4, for starting position $x \in S$:

$$\mathbb{P}_x \left[\sum_{v_1 \in N(t)} 1 \right] = M^1(t, x) = \mathbb{Q}^1_x \left[\mathrm{e}^{(m-1)(\int_0^t \mathbb{1}_0(\xi_s))\mathrm{d}s)} \right]. \tag{3.7}$$

As we have one spine, the expectations under $\mathbb{Q}$ and under $\mathbb{P}$ are the same. This is because the motion of particles is the same under $\mathbb{Q}$ as under $\mathbb{P}$, since in the DR model we use the constant functional ζ and thus do not bias the movement of particles (only the branching is size-biased). Thus Equation 3.7 further simplifies to:

$$M^1(t, x) = \mathbb{P}_x \left[\mathrm{e}^{(m-1)(\int_0^t \mathbb{1}_0(\xi_s))\mathrm{d}s)} \right].$$

By the same steps but with $f(z) = \mathbb{1}_y(z)$, $z \in S$, we also get:

$$M^1(t, x, y) = \mathbb{E}_x \left[\mathrm{e}^{(m-1)(\int_0^t \mathbb{1}_0(\xi_s))\mathrm{d}s)} \mathbb{1}_y(\xi_t) \right].$$

□

Lemma 3.10. *We also have the relations:*

$$M^1(t, x) = 1 + (m-1)p_t(x, 0) * M^1(t, 0)$$
$$M^1(t, x, y) = p_t(x, y) + (m-1)p_t(x, 0) * M^1(t, 0, y)$$

where $*$ represents convolution.

We give the proof for $M^1(t, x, y)$, but the steps are exactly the same for $M^1(t, x)$ without the additional indicator function.

Proof. Let $t > 0$ and $x, y \in S$, recall from Lemma 3.9 that we have:

$$M^1(t, x, y) = \mathbb{E}_x \left[\mathrm{e}^{(m-1)(\int_0^t \mathbb{1}_0(\xi_s))\mathrm{d}s)} \mathbb{1}_y(\xi_t) \right].$$

Expanding the exponential gives:

$$\begin{aligned} M^1(t, x, y) &= \mathbb{E}_x \left[\sum_{n=0}^{\infty} \frac{(m-1)^n}{n!} \left(\int_0^t \mathbb{1}_0(\xi_s)\mathrm{d}s) \right)^n \mathbb{1}_y(\xi_t) \right], \\ &= \mathbb{E}_x \left[\mathbb{1}_y(\xi_t) + \sum_{n=1}^{\infty} \frac{(m-1)^n}{n!} \left(\int_0^t \mathbb{1}_0(\xi_s)\mathrm{d}s \right)^n \mathbb{1}_y(\xi_t) \right], \\ &= p_t(x, y) + \mathbb{E}_x \left[\sum_{n=1}^{\infty} \frac{(m-1)^n}{n!} \left(\int_0^t \mathbb{1}_0(\xi_s)\mathrm{d}s \right)^n \mathbb{1}_y(\xi_t) \right]. \end{aligned}$$

We note that since the inside function is symmetric in all the arguments we have the following, for $j \in \mathbb{N}$:

$$\begin{aligned} \left(\int_0^t \mathbb{1}_0(\xi_s)\mathrm{d}s \right)^n &= \int_0^t \cdots \int_0^t \mathbb{1}_0(\xi_{r_1}) \cdots \mathbb{1}_0(\xi_{r_n})\mathrm{d}r_n \ldots \mathrm{d}r_1, \\ &= n! \cdot \int_0^t \int_{r_1}^t \cdots \int_{r_{n-1}}^t \mathbb{1}_0(\xi_{r_1}) \cdots \mathbb{1}_0(\xi_{r_n})\mathrm{d}r_n \ldots \mathrm{d}r_2\mathrm{d}r_1. \end{aligned}$$

Substituting this into the previous expression gives:

$$p_t(x, y) + (m-1)\mathbb{E}_x \Big[\sum_{n=1}^{\infty} (m-1)^{n-1} \int_0^t \int_{r_1}^t \cdots \int_{r_{n-1}}^t \mathbb{1}_0(\xi_{r_1}) \cdots \mathbb{1}_0(\xi_{r_n}) \mathbb{1}_y(\xi_t)\mathrm{d}r_n \ldots \mathrm{d}r_2\mathrm{d}r_1 \Big].$$

We can exchange the sums and the expectation to give:

$$p_t(x,y) + (m-1)\left(\int_0^t \sum_{n=1}^{\infty}(m-1)^{n-1}\int_{r_1}^t \cdots \int_{r_{n-1}}^t \mathbb{P}_x(\xi_{r_1}=0,\ldots,\xi_{r_n}=0,\xi_t=y)\mathrm{d}r_n\ldots\mathrm{d}r_2\mathrm{d}r_1\right).$$

Using the Markov property gives:

$$p_t(x,y) + (m-1)\bigg(\int_0^t p_{r_1}(x,0)\sum_{n=1}^{\infty}(m-1)^{n-1}\int_{r_1}^t \cdots \int_{r_{n-1}}^t \mathbb{P}_0(\xi_{r_2-r_1}=0,\ldots,\xi_{r_n-r_1}=0,\xi_{t-r_1}=y)\mathrm{d}r_n\ldots\mathrm{d}r_2\mathrm{d}r_1\bigg).$$

Notice that $r_j \le r_i \le t$ is equivalent to $r_j - r_1 \le r_i - r_1 \le t - r_1$, so by a change of variables we get that:

$$\int_{r_1}^t \cdots \int_{r_{n-1}}^t \mathbb{P}_0(\xi_{r_2-r_1}=0,\ldots,\xi_{r_n-r_1}=0,\xi_{t-r_1}=y)\mathrm{d}r_n\ldots\mathrm{d}r_2 \tag{3.8}$$

$$= \int_0^{t-r_1} \cdots \int_{r_{n-1}}^{t-r_1} \mathbb{P}_0(\xi_{r_2}=0,\ldots,\xi_{r_n}=0,\xi_{t-r_1}=y)\mathrm{d}r_n\ldots\mathrm{d}r_2 \tag{3.9}$$

Let us define the function C in terms of $t>0$ and $y \in S$ such that:

$$C(t,y) = \sum_{n=1}^{\infty}(m-1)^{n-1}\int_0^t \cdots \int_{r_{n-1}}^t \mathbb{P}_0(\xi_{r_2}=0,\ldots,\xi_{r_n}=0,\xi_t=y)\mathrm{d}r_n\ldots\mathrm{d}r_2, \tag{3.10}$$

$$= \sum_{n=1}^{\infty}\frac{(m-1)^{n-1}}{(n-1)!}\int_0^t \cdots \int_0^t \mathbb{P}_0(\xi_{r_2}=0,\ldots,\xi_{r_n}=0,\xi_t=y)\mathrm{d}r_n\ldots\mathrm{d}r_2, \tag{3.11}$$

$$= \sum_{n=1}^{\infty}\frac{(m-1)^{n-1}}{(n-1)!}\int_0^t \cdots \int_0^t \mathbb{E}_0\left[\mathbb{1}_{\{\xi_{r_2}=0\}}\ldots,\mathbb{1}_{\{\xi_{r_n}=0\}}\mathbb{1}_{\{\xi_t=y\}}\right]\mathrm{d}r_n\ldots\mathrm{d}r_2, \tag{3.12}$$

$$= \mathbb{E}_0\left[\sum_{n=0}^{\infty}\frac{(m-1)^n}{(n)!}\mathbb{1}_{\{\xi_t=y\}}\left(\int_0^t \mathbb{1}_{\{\xi_r=0\}}\mathrm{d}r\right)^n\right]. \tag{3.13}$$

Then our final expression for $M^1(t,x,y)$ equals:

$$p_t(x,y) + (m-1)\int_0^t p_{r_1}(x,0)C(t-r_1,y)\mathrm{d}r_1.$$

By working the logic backwards, we can see that $C(t,y) = M^1(t,0,y)$ thus:

$$p_t(x,y) + (m-1)\int_0^t p_{r_1}(x,0)M^1(t-r_1,0,y)\mathrm{d}r_1.$$

which proves the lemma. □

There is a natural reason why relations like this occur, that is, as solutions to discrete-space heat equations with one-point potential. To break this down, "discrete-space" because particles exist only at positions in countable S, "heat equations" as we are analysing the diffusion of particles throughout the system, and "one-point potential" refers to the single source of particles being $0 \in S$ (as we can only branch there). The equation in question is of the form:

$$\begin{cases} \frac{\partial}{\partial t}u(t,x) = \mathcal{A}u(t,x) + (m-1)\mathbb{1}_0(x)u(t,x), \\ u(0,x) = \mathbb{1}_y(x), \end{cases}$$

where u is a vector giving the number of particles at each position at time $t>0$ and position $x \in S$. The $\mathcal{A}u(t,x)$ term describes the change in position of the existing particles due to the Markov process $\mathcal{A}$. The $(m-1)\mathbb{1}_0(x)u(t,x)$ term describes the contribution of the branching (note only at the origin and on average $m-1$ new particles), and proportional to the number of particles present there, i.e. more particles means more likely to branch and give birth to more offspring. As to why we should have solutions of the form of Equation 3.5, see Appendix A.

3.4 Proof for M^1

3.4.1 Subcritical case

Proof of Theorem 3.5 when $k = 1$. Notice that, the statement "$\int_0^\infty \mathbb{1}_0(\xi_r)\mathrm{d}r = \infty$ almost-surely", does not depend on the starting value $x = \xi_0$, hence the arguments are independent of the startion position. We begin with the case that $\mathcal{A}$ is transient, this means that:

$$\int_0^\infty \mathbb{1}_0(\xi_r)\mathrm{d}r < \infty.$$

From Theorem 3.9 we have that, for $x \in S$ and $t > 0$:

$$M^1(t, x) = \mathbb{E}_x[\mathrm{e}^{(m-1)\int_0^t \mathbb{1}_0(\xi_r)\mathrm{d}r}].$$

Rewriting this as an integral gives:

$$M^1(t, x) = \int_S \mathrm{e}^{(m-1)\int_0^t \mathbb{1}_0(\xi_r)\mathrm{d}r}\mathrm{d}\mathbb{P},$$

where $\mathbb{P}$ is the probability measure. In order to study the asymptotics of this quantity, we can use the *dominated convergence theorem (DOM)* . To make this clearer we define f such that:

$$f_t(\xi) = \mathrm{e}^{(m-1)\int_0^t \mathbb{1}_0(\xi_r)\mathrm{d}r} \leq \mathrm{e}^{(m-1)\int_0^\infty \mathbb{1}_0(\xi_r)\mathrm{d}r}.$$

Thus we can write M^1 as:

$$M^1(t, x) = \int_S f_t(\xi)\mathrm{d}\mathbb{P}.$$

We note that $\lim_{t\to\infty} f_t(\xi) \in (0, \infty)$ by assumption of transience, and that $f_t(\xi) \leq 1$, as $m < 1$ (subcritical case). Thus by *DOM* we have

$$\lim_{t\to\infty} M^1(t, x) = \int_S \lim_{t\to\infty} f_t(\xi)\mathrm{d}\mathbb{P} \in (0, \infty)$$

and so (i) is proved when $\mathcal{A}$ is transient. When $\mathcal{A}$ is recurrent we have this limit in the integral is 0 since the integral in the exponent tends to infinity. As for $M^1(t, x, y)$, when $\mathcal{A}$ is transient $\mathbb{1}_y(\xi_t) \to 0$ as $t \to \infty$ almost surely, thus we can apply dominated convergence to:

$$f_t(\xi) = \mathrm{e}^{(m-1)\int_0^t \mathbb{1}_0(\xi_r)\mathrm{d}r}\mathbb{1}_y(\xi_t)$$

which now tends to 0 as $t \to \infty$, and similarly when $\mathcal{A}$ is recurrent then we can use the bound $M^1(t, x, y) \leq M^1(t, x)$, and note that $M^1(t, x) \to 0$, giving $M^1(t, x, y) \to 0$.

Thus Theorem 3.5 follows in the $k = 1$ case. □

3.4.2 Critical case

Proof of Theorem 3.6 when $k = 1$. Let $t > 0$ and $x, y \in S$, then this is trivial in the $k = 1$ case, since $M^1(t, x) = 1$ and by Lemma 3.9 we have that:

$$M^1(t, x, y) = \mathbb{E}_x[\mathrm{e}^{(m-1)\int_0^t \mathbb{1}_0(\xi_s)\mathrm{d}s}\mathbb{1}_y(\xi_t)] = \mathbb{E}_x[\mathbb{1}_y(\xi_t)] = p_t(x, y).$$

Thus Theorem 3.6 follows, in the $k = 1$ case. □

3.4.3 Supercritical case - analytical arguments

Proof of Theorem 3.7 case (*a*) *when* $k = 1$. We start by recalling the Hölder and reverse Hölder inequalities. The Hölder inequality states that for random variables X, Y and $p, q > 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$ we have that:

$$\mathbb{E}(|XY|) \leq \|X\|_p \cdot \|Y\|_p, \tag{3.14}$$

where $\|Z\|_r = (\mathbb{E}(|Z|^r))^{\frac{1}{r}}$ is the $\mathcal{L}^r$-norm of Z. Similarly, the reverse Hölder inequality is:

$$\|XY\|_1 \geq \|X\|_{\frac{1}{p}} \cdot \|Y\|_{\frac{1}{1-p}}. \tag{3.15}$$

Let $p > 1$. Applying Equation 3.14 to $X = \mathrm{e}^{(m-1)\int_0^t \mathbb{1}_0(\xi_s)\mathrm{d}s}$ and $Y = \mathbb{1}_y(\xi_t)$ gives the upper bound:

$$M^1(t,x,y) \leq \mathbb{E}_x[\mathrm{e}^{p(m-1)\int_0^t \mathbb{1}_0(\xi_s)\mathrm{d}s}]^{\frac{1}{p}} p_t(x,y)^{\frac{p-1}{p}}. \tag{3.16}$$

Then applying Equation 3.15 gives the lower bound:

$$M^1(t,x,y) \geq \mathbb{E}_x[\mathrm{e}^{\frac{-1}{p-1}(m-1)\int_0^t \mathbb{1}_0(\xi_s)\mathrm{d}s}]^{-(p-1)} p_t(x,y)^p. \tag{3.17}$$

Recall further we define the constant $\beta = \frac{1}{G_\infty(0,0)} + 1 \geq 1$. When $\mathcal{A}$ is recurrent, we have $G_\infty(0,0) = \infty$, and thus $\beta = 1$. Hence for case (a) where $m < \beta = 1$, we simply apply Theorem 3.5. Thus we assume $\mathcal{A}$ is transient, i.e. $\int_0^\infty \mathbb{1}_0(\xi_r)\mathrm{d}r < \infty$ with positive probability.

Since the exponent in Equation 3.17 is negative, and since $\int_0^\infty \mathbb{1}_0(\xi_r)\mathrm{d}r < \infty$ with positive probability, then the expectation converges to a finite constant (as we either have $\int_0^\infty \mathbb{1}_0(\xi_r)\mathrm{d}r = \infty$ with some probability which gives expectation of e^0, or $\int_0^\infty \mathbb{1}_0(\xi_r)\mathrm{d}r < \infty$ which gives a small but positive expectation). Thus there exists $C > 0$ such that $Cp_t(x,y)^p \leq M^1(t,x,y)$. In order to show the existence of an upper bound of $M^1(t,x,y)$ of the form $C'p_t(x,y)^{\frac{p-1}{p}}$ for some $C' > 0$, we refer to [DS10].

Since $m < \beta$, we have that $m-1 < \beta$, and thus there exists $p > 1$ such that $p(m-1) < \beta$, and with this choice of p we can apply Theorem 1 part (3) from [DS10], and this gives the required result for $M^1(t,x,y)$. We obtain the result for $M^1(t,x)$ directly from Theorem 1 of [DS10]. □

3.4.4 Supercritical case - renewal theory arguments

The proofs for parts (b) and (c) rely on renewal theorems. Note that, by Lemma 3.9, $M^1(t,x,y)$ is only on both sides of the relation for $x = 0$. We use this to formulate a renewal equation. We start with the case for $x = 0$, and then later generalise to when $x \neq 0$.

Proof of Theorem 3.7 case (c) when $k = 1$. We start with part (c), i.e. $m > \beta$, as it is simpler. The following arguments work whether $\mathcal{A}$ is recurrent or transient. Let us define the function $f : [0,\infty) \to \mathbb{R}$ by:

$$f(\lambda) = \int_0^\infty \mathrm{e}^{-\lambda t} p_t(0,0)\mathrm{d}t, \quad \lambda \in [0,\infty).$$

Recalling that $m > \beta$, this is equivalent to $m > \frac{1}{G_\infty(0,0)} + 1$, which is in turn equivalent to $\int_0^\infty p_t(0,0)\mathrm{d}t > \frac{1}{m-1}$. We now show that there exists a unique positive root λ of the equation:

$$f(\lambda) = \frac{1}{m-1},$$

which we shall refer to as $r(m)$. Which defines a function r whose input m gives the necessary root λ. We will show there exists a unique positive root by showing that $\lim_{\lambda\to\infty} f(\lambda) = 0$ and that f is strictly decreasing in λ. Observe that, $f(0) = \int_0^\infty \mathrm{e}^{-0\cdot t} p_t(0,0)\mathrm{d}t = \int_0^\infty p_t(0,0)\mathrm{d}t > \frac{1}{m-1}$. Subsequently, assuming enough regularity to differentiate under the integral:

$$f'(\lambda) = \frac{d}{d\lambda} f(\lambda) = \frac{d}{d\lambda}\int_0^\infty \mathrm{e}^{-\lambda t} p_t(0,0)\mathrm{d}t = \int_0^\infty \frac{\partial}{\partial\lambda}\mathrm{e}^{-\lambda t} p_t(0,0)\mathrm{d}t = -\int_0^\infty t\mathrm{e}^{-\lambda t} p_t(0,0)\mathrm{d}t < 0.$$

Thus f is decreasing in λ, and now all that is left to be shown is that $\lim_{\lambda\to\infty} f(\lambda) = 0$. For this we shall use the *dominated convergence theorem (DOM)*. Note that, in order for *DOM* we verify that we have a function g:

$$f(\lambda) = \int_0^\infty \mathrm{e}^{-\lambda t} p_t(0,0)\mathrm{d}t \leq \int_0^\infty p_t(0,0)\mathrm{d}t = g(\lambda),$$

since $\mathrm{e}^{-\lambda t} \leq 1$ for $t > 0$ and $\lambda \geq 0$. In this case g is constant. Thus we can apply *DOM* to get:

$$\lim_{\lambda\to\infty} f(\lambda) = \lim_{\lambda\to\infty} \int_0^\infty \mathrm{e}^{-\lambda t} p_t(0,0)\mathrm{d}t = \int_0^\infty \lim_{\lambda\to\infty} \mathrm{e}^{-\lambda t} p_t(0,0)\mathrm{d}t = 0,$$

because the $\mathrm{e}^{-\lambda t}$ dominates as λ increases and the integral decreases. Thus we have our unique positive root of $f(\lambda) = \frac{1}{m-1}$ which we denote $r(m)$. This result tells us that the measure so defined:

$$U(\mathrm{d}t) = (m-1)\mathrm{e}^{-r(m)t} p_t(0,0)\mathrm{d}t \iff \frac{\mathrm{d}U(t)}{\mathrm{d}t} = (m-1)\mathrm{e}^{-r(m)t} p_t(0,0) \tag{3.18}$$

in fact defines a change of measure in the Radon-Nikodym sense, with density $(m-1)\mathrm{e}^{-r(m)t}p_t(0,0)$. We can also thus see that $\mathrm{e}^{-r(m)t}p_t(0,y)$ is directly Riemann integrable, as in Definition 2.9. This is because as $|\mathrm{e}^{-r(m)t}p_t(0,y)| \leq \mathrm{e}^{-r(m)t}$, which is integrable. Recalling the variation of constants formula for $M^1(t,x,y)$ with $x = 0$ from Lemma 3.9:

$$M^1(t,0,y) = p_t(0,y) + (m-1)p_t(0,0) * M^1(t,0,y). \tag{3.19}$$

Multiplying both sides of Equation 3.19 by $\mathrm{e}^{-r(m)t}$ gives:

$$\mathrm{e}^{-r(m)t} M^1(t,0,y) = \mathrm{e}^{-r(m)t} p_t(0,y) + (m-1)\mathrm{e}^{-r(m)t} p_t(0,0) * M^1(t,0,y). \tag{3.20}$$

Let us define the two following functions f and g on $(0,\infty)$:

$$f(t) = \mathrm{e}^{-r(m)t} M^1(t,0,y), \tag{3.21}$$

$$g(t) = \mathrm{e}^{-r(m)t} p_t(0,y). \tag{3.22}$$

Substituting this into Equation 3.20 gives:

$$f(t) = g(t) + (f * U)(t), \tag{3.23}$$

which is a complete[5] renewal equation. We can now apply Theorem 2.10, which gives:

$$\lim_{t\to\infty} f(t) = \mu^{-1} \int_0^\infty g(s)\mathrm{d}s.$$

We recall that μ is the mean recurrence time, and claim (without proof) that $\mu = \int_0^\infty U((s,\infty))\mathrm{d}s$. This is analogous to the result for X a non-negative real-valued random variable:

$$\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X \geq s)\mathrm{d}s.$$

This gives the final result:

$$\lim_{t\to\infty} f(t) = \frac{\int_0^\infty g(s)\mathrm{d}s}{\int_0^\infty U((s,\infty))\mathrm{d}s}.$$

The claim for $M^1(t,0,y)$ thus follows. The claim for $M^1(t,0)$ follows analogously using the corresponding relation in Theorem 3.10 for $M^1(t,x)$. □

Proof of Theorem 3.7 case (b) when $k = 1$. We now address part part *b*). In this case we have that $m = \beta$, thus $m - 1 = \frac{1}{G_\infty(0,0)}$, hence:

$$\int_0^\infty p_r(0,0)\mathrm{d}r = \frac{1}{m-1} \iff \int_0^\infty (m-1)p_r(0,0)\mathrm{d}r = 1. \tag{3.24}$$

The expression $(m-1)p_r(0,0)\mathrm{d}r$ is the same as $U(\mathrm{d}r)$ from Equation 3.18, but with $r(m) = 0$. In order to apply the previous results we need U to have finite mean μ (in order for Theorem 2.10 to apply). Thus we need:

$$\mu = \int_0^\infty tU(\mathrm{d}t) = \int_0^\infty t(m-1)p_t(0,0)\mathrm{d}t < \infty.$$

[5] In this case the renewal equation is called *complete* because of the non-trivial $G(t)$ term.

Hence in this case we assume that:

$$\int_0^\infty t p_t(0,0)\mathrm{d}t < \infty, \tag{3.25}$$

in order to apply Theorem 2.10, we also need the function $t \mapsto p_t(0,y)$ to be directly Riemann integrable, but this follows from the fact that $p_t(0,0) \geq p_{t-1}(0,y)p_1(y,0)$ and thus:

$$\int_1^\infty p_{t-1}(0,y)p_1(y,0)\mathrm{d}t \leq \int_1^\infty p_t(0,0)\mathrm{d}t < \infty,$$

and since we have finite mean μ (by Equation 3.25) and since $0 \leq p_t(0,y) \leq 1 \ \forall t > 0, y \in S$ this gives:

$$\int_0^\infty p_t(0,y)\mathrm{d}t < \infty.$$

Hence in the case of a finite mean for U we can apply the previous results. In order to see this, let us recall the definitions of f and g in this case:

$$f(t) = \mathrm{e}^{-r(m)t} M^1(t,0,y) = M^1(t,0,y), \tag{3.26}$$
$$g(t) = \mathrm{e}^{-r(m)t} p_t(0,y) = p_t(0,y). \tag{3.27}$$

The final result gives us:

$$\lim_{t\to\infty} f(t) = \frac{\int_0^\infty g(s)\mathrm{d}s}{\int_0^\infty U((s,\infty))\mathrm{d}s} \iff \lim_{t\to\infty} M^1(t,0,y) = \frac{\int_0^\infty p_t(0,y)\mathrm{d}s}{\mu} = \frac{1}{\mu(m-1)} \in (0,\infty),$$

as required. If instead, U does not have a finite mean, i.e. Equation 3.25 does not hold. Then we need to use a renewal theorem which works with infinite-mean variables. Recall the variation of constants relation for $M^1(t,0,y)$:

$$M^1(t,0,y) = p_t(0,y) + (m-1)p_t(0,0) * M^1(t,0,y). \tag{3.28}$$

Then we can iterate this once to get:

$$\begin{aligned} M^1(t,0,y) &= p_t(0,y) + (m-1)p_t(0,0) * \left(p_t(0,y) + (m-1)p_t(0,0) * M^1(t,0,y)\right), \qquad (3.29)\\ &= p_t(0,y) * p_t(0,y)^{*0} + p_t(0,0) * (m-1)p_t(0,y)^{*1} + (m-1)^2 p_t(0,0) * p_t(0,y)^{*2} * M^1(t,0,y), \qquad (3.30)\end{aligned}$$

where we use $*n$ to denote the nth convolution, and $*0$ to denote no convolution and therefore equal to 1. Continuing the iteration in Equation 3.29 gives the expansion:

$$M^1(t,0,y) = p_t(0,y) * \sum_{n\geq 0} (m-1)^n p_t(0,0)^{*n}. \tag{3.31}$$

This is a justified expansion since, for all $n \in \mathbb{N}_0$:

$$(m-1)^n p_t(0,0)^{*n} \leq \left((m-1)\int_0^t p_r(0,0)\mathrm{d}r\right)^n.$$

Which holds by *Young's Inequality* for convolutions:

$$p_t(0,0) * p_t(0,0) = \int_0^t p_r(0,0)p_{t-r}(0,0)\mathrm{d}r \leq \left(\int_0^t p_r(0,0)\mathrm{d}r\right)^2,$$

since $p_t \in [0,1]$, and we then apply induction to get $p_t(0,0)^{*n} \leq \left(\int_0^t p_r(0,0)\mathrm{d}r\right)^n$. Which gives that:

$$M^1(t,0,y) = p_t(0,y) * \sum_{n\geq 0}(m-1)^n p_t(0,0)^{*n} \leq p_t(0,y) * \sum_{n\geq 0}\left((m-1)\int_0^t p_r(0,0)\mathrm{d}r\right)^n = \sum_{n\geq 0}(\alpha)^n,$$

where $\alpha = (m-1)\int_0^t p_r(0,0)\mathrm{d}r < (m-1)\int_0^\infty p_r(0,0)\mathrm{d}r = 1$, as $m = \beta$ and $t \mapsto p_t$ is strictly decreasing (but non-zero for all $t > 0$). The upper-bound sum converges as a geometric sum with $\alpha < 1$. Now we define the function $F : \mathbb{R}^+ \to \mathbb{R}$ defined by:

$$F(r) = \sum_{n\geq 0}(m-1)^n p_r(0,0)^{*n},$$

for $r > 0$. Hence we can rewrite Equation 3.31 as:

$$M^1(t,0,y) = p_t(0,y) * F(t) = \int_0^t p_{t-r}(0,y)F(r)\mathrm{d}r. \tag{3.32}$$

Using Lemma 1 of [Eri73] with distribution function $U(\mathrm{d}t) = (m-1)p_t(0,0)\mathrm{d}t$ we get that:

$$F(t) \approx \frac{t}{(m-1)\int_0^t \int_s^\infty p_r(0,0)\mathrm{d}r\mathrm{d}s}. \tag{3.33}$$

To be precise, this means that $F(t) = \Theta\left(\frac{t}{(m-1)\int_0^t \int_s^\infty p_r(0,0)\mathrm{d}r\mathrm{d}s}\right)$. Now observe that as $t \to \infty$, as $(m-1)\int_s^\infty p_r(0,0)\mathrm{d}r \to 0$ as $s \to \infty$ because $(m-1)p_r(0,0)$ is a probability density in r by Equation 3.24. To show that $M^1(t,0,y) \to \infty$ we will find upper and lower bounds which both go to infinity. We start with the simple upper bound:

$$M^1(t,0,y) = \int_0^t p_{t-r}(0,y)F(r)\mathrm{d}r \leq \int_0^t F(r)\mathrm{d}r, \tag{3.34}$$

where $\int_0^t F(r)\mathrm{d}r$ clearly goes to ∞ as $t \to \infty$. For the lower bound we use the irreducibility and continuity of $t \mapsto p_t$, i.e. it is always non-zero and is continuous in t. Thus there exists $0 < t_0 < t_1$ and $\epsilon > 0$ such that for $t_0 \leq r \leq t_1$ we have $p_r(0,y) > \epsilon$. Then we note that:

$$M^1(t,0,y) = \int_0^t p_{t-r}(0,y)F(r)\mathrm{d}r \geq \int_{t-t_0}^{t-t_1} p_{t-r}(0,y)F(r)\mathrm{d}r, \tag{3.35}$$

and notice that for $t - t_0 \leq r \leq t - t_1 \iff t_0 \leq t - r \leq t_1$, and thus in the interval $[t-t_0, t-t_1]$ we have $p_{t-r}(0,y) > \epsilon$ and thus we have:

$$M^1(t,0,y) \geq \epsilon \int_{t-t_0}^{t-t_1} F(r)\mathrm{d}r, \tag{3.36}$$

where the lower bound tends to infinity as $t \to \infty$. We can thus conclude that $M^1(t,0,y) \to \infty$ as $t \to \infty$, in the infinite mean case. There now only remains for dealing with cases (b) and (c) when $x \neq 0$. These results come from a combination of Lemma 3.10 and very similar upper/lower bounds to above. We start with case (b) and some $x \neq 0$, we start in the infinite renewal mean case. We know that $\lim_{t\to\infty} M^1(t,0,y) = \infty$ and aim to show that $\lim_{t\to\infty} M^1(t,x,y) = \infty$. Recall from Lemma 3.10 that:

$$M^1(t,x,y) = p_t(x,y) + (m-1)p_t(x,0) * M^1(t,0,y). \tag{3.37}$$

Thus one can easily see that since $M^1(t,0,y) \to \infty$ as $t \to \infty$, then we have that $M^1(t,x,y) \to \infty$. As for the result $\lim_{t\to\infty} M^1(t,x) = \infty$ this in the case when we have infinite renewal measure as $M^1(t,x,y) \leq M^1(t,x), \forall y \in S$, giving the result. In the finite renewal measure mean case we have polynomial growth of $M^1(t,x,y)$ for all $y \in S$, thus we once again get $\lim_{t\to\infty} M^1(t,x) = \infty$ in that case. □

It remains to deal with the results for $M^1(t,x)$ which follow from the following bounds and the asymptotics of $M^1(t,0)$ since there exists such an $\epsilon > 0$ and $0 < t_0 < t_1$:

$$1 + \epsilon \int_{t-t_1}^{t-t_0} M^1(r,0)\mathrm{d}r \leq M^1(t,x) \leq 1 + \int_0^t M^1(r,0)\mathrm{d}r, \tag{3.38}$$

which is derived essentially the same as in Equations 3.35 and 3.36 but with the expression for $M^1(t,x)$. We then combine this with case 2 of Theorem 1 from [DS10] to give the results for $M^1(t,x)$ in case (c).

3.5 Moment simplification lemma

This lemma will be key in generalising the proof we give in Section 3.4 to the general kth moments case, as it relates the kth moments of N_t and $N_t(y)$ to the $1, \ldots, k-1$th moments, giving us a recursive relation.

Lemma 3.11 (Moment simplification). *We have that for $k \geq 2$:*

$$M^k(t,x) = M^1(t,x) + M^1(t,x,0) * g_k\left(M^1(t,0), \ldots, M^{k-1}(t,0)\right),$$
$$M^1(t,x,y) = M^1(t,x,y) + M^1(t,x,0) * g_k\left(M^1(t,0,y), \ldots, M^{k-1}(t,0,y)\right),$$

where the function g_k is defined as follows [6]

$$g_k(M^1, \ldots, M^{k-1}) = \sum_{j=2}^{k} \mathbb{E}\left[\binom{X}{j}\right] \sum_{\substack{i_1,\ldots,i_j>0 \\ i_1+\ldots+i_j=k}} \frac{k!}{i_1!\ldots i_j!} M^{i_1}\ldots M^{i_j}.$$

Proof. We recall that we start with a particle ξ at $x \in S$ carrying all $k \in \mathbb{N}$ marks, which is behaving according to the $\mathbb{Q}_x^k$ measure and branches at rate $m_k = \mathbb{E}[X^k]$ when at the origin. We first ask at what rate the marks first split over two particles, as we could have the situation where we split into j particles but all the marks stay on one particle. We drop the x from $\mathbb{Q}_x^k$ notation for simplicity, thus we shall only use $\mathbb{P}$ and $\mathbb{Q}^k$. We denote $p \in [0,1]$ the probability of this happening:

$$\begin{aligned} p &= \sum_{j=1}^{\infty} \mathbb{Q}^k(\xi \text{ branches into } j \text{ particles})\mathbb{Q}^k(\text{the k spines all stay at } \xi \text{ — } \xi \text{ branches into } j \text{ particles}), \\ &= \sum_{j=1}^{\infty} j^k p_j m_k^{-1} \cdot (\frac{1}{j})^{k-1} = \sum_{j=1}^{\infty} j p_j m_k^{-1}. \end{aligned}$$

Thus $m_k * p = \sum_{j=1}^{\infty} jp_j = m$, so the rate at which the k marks distribute over 2 or more particles is $m_k - m$. We denote the time this happens T.

We then consider $i_1, \ldots, i_j > 0$ such that $i_1 + \cdots + i_j = k$ and define the event $A_k(j; i_1, \cdots, i_j)$ that at time T: i_1 particles follow the first particle, i_2 the second particle, … and i_j particles follow the jth particle. Suppose the initial particle ξ splits into $a \in \mathbb{N}$ particles, under $\mathbb{Q}_x^k$ this happens with probability $a^k p_a m_k^{-1}$. Then, conditioning on splitting into a particles, the probability of $A_k(j; i_1, \ldots, i_j)$ under $\mathbb{Q}_x^k$ is:

$$\frac{1}{a^{i_1}} \cdot \frac{1}{a^{i_2}} \cdots \frac{1}{a^{i_j}} \cdot \binom{a}{j} \cdot \frac{k!}{i_1! \cdots i_j!} = \frac{1}{a^k} \cdot \binom{a}{j} \cdot \frac{k!}{i_1! \cdots i_j!}.$$

The first term comes from the probability of i_1 marks choosing the first particle, then i_2 marks choosing the second particle, etc until the i_j marks choose the jth particle. This happens independently and uniformly with $\frac{1}{a}$ since each particle chooses uniformly from the a available particles. The second term is the number of ways of choosing the designated j particles to spread the k marks over. Then the final term is the number of ways of arranging the k marks into the groups of size $i_1, \ldots, i_j$, by the multinomial formula. Then, by the law of total probability, we have that the probability of $A_k(j; i_1, \ldots, i_j)$ under $\mathbb{Q}^k$ is:

$$\sum_{a=0}^{\infty} a^k p_a m_k^{-1} \cdot \frac{1}{a^k} \cdot \binom{a}{j} \cdot \frac{k!}{i_1! \cdots i_j!} = \frac{1}{m_k} \sum_{a=0}^{\infty} \binom{a}{j} p_a \frac{k!}{i_1! \cdots i_j!} = \frac{1}{m_k} \mathbb{E}\left[\binom{X}{j}\right] \frac{k!}{i_1! \cdots i_j!}.$$

Taking $j \geq 2$, if we condition on having a separation event (i.e. the marks do not all stick to one particle), then $A_k(j; i_1, \ldots, i_j)$ has probability:

[6] These g_k appear in Faà di Bruno's differentiation formula.

$$\begin{aligned}
\mathbb{Q}^k\left(A_k(j;i_1,\dots,i_j)\mid \text{separation event}\right) &= \frac{\mathbb{Q}^k\left(A_k(j;i_1,\dots,i_j),\text{separation event}\right)}{\mathbb{Q}^k\left(\text{separation event}\right)}\\
&= \frac{\mathbb{Q}^k\left(A_k(j;i_1,\dots,i_j)\right)}{\frac{m_k-m}{m_k}}\\
&= \mathbb{Q}^k\left(A_k(j;i_1,\dots,i_j)\right)\frac{m_k}{m_k-m}\\
&= \frac{1}{m_k}\mathbb{E}\left[\binom{X}{j}\right]\frac{k!}{i_1!\cdots i_j!}\left(\frac{m_k}{m_k-m}\right).
\end{aligned}$$

Recall that T is the separation time, when the k marks first split off into 2 or more particles. Let χ_t for $t\in[0,T)$ be the position of the single particle carrying all k marks, note that several branching events may have occurred, but all the k marks stayed on the same particle. We further define $\mathcal{F}_t$ to be the filtration containing all the information about χ up to time t, including the information about the spines until to time t. Recall from Section 2.2.3 that skel(t) refers to the tree generated by all the particles having carried at least one spine mark by time t, and further that skel$(s;t)$ for $0<s<t$ is the same tree of particles which between time s and t which have carried at least one mark.

We apply the specialised version of the many-to-few lemma (Theorem 3.4) with $f=1$, and the fact that before T all spines lie on the same particle. From this we obtain:

$$\begin{aligned}
\mathbb{P}\left[N_t^k\right] &= \mathbb{Q}^k\left[\prod_{v\in\text{skel}(t)} \mathrm{e}^{(m_{B_v}-1)\int_{\sigma_v(t)}^{\tau_v(t)}\mathbb{1}_0(X_v(s))\mathrm{d}s}\right]\\
&= \mathbb{Q}^k\left[\prod_{v\in\text{skel}(t)} \mathrm{e}^{(m_{B_v}-1)\int_{\sigma_v(t)}^{\tau_v(t)}\mathbb{1}_0(X_v(s))\mathrm{d}s}\mathbb{1}_{\{T\le t\}}\right]+\mathbb{Q}^k\left[\prod_{v\in\text{skel}(t)} \mathrm{e}^{(m_{B_v}-1)\int_{\sigma_v(t)}^{\tau_v(t)}\mathbb{1}_0(X_v(s))\mathrm{d}s}\mathbb{1}_{\{T> t\}}\right].
\end{aligned}$$

In the case where $T>t$, this means that the k marks haven't separated by time t yet, and hence the skeleton is just one particle with k marks. Recall that this has position χ_t. Hence,

$$\mathbb{Q}^k\left[\prod_{v\in\text{skel}(t)} \mathrm{e}^{(m_{B_v}-1)\int_{\sigma_v(t)}^{\tau_v(t)}\mathbb{1}_0(X_v(s))\mathrm{d}s}\mathbb{1}_{\{T> t\}}\right]=\mathbb{Q}^k\left[\mathrm{e}^{(m_k-1)\int_0^t\mathbb{1}_0(\chi_s)\mathrm{d}s}\mathbb{1}_{\{T>t\}}\right]. \tag{3.39}$$

In the case where $T\le t$, we can break down the first summand in the following way, using the *tower rule*,

$$\mathbb{Q}^k\left[\prod_{v\in\text{skel}(t)} \mathrm{e}^{(m_{B_v}-1)\int_{\sigma_v(t)}^{\tau_v(t)}\mathbb{1}_0(X_v(s))\mathrm{d}s}\mathbb{1}_{\{T\le t\}}\right]=\mathbb{Q}^k\left[\mathbb{Q}^k\left[\prod_{v\in\text{skel}(t)} \mathrm{e}^{(m_{B_v}-1)\int_{\sigma_v(t)}^{\tau_v(t)}\mathbb{1}_0(X_v(s))\mathrm{d}s}\mathbb{1}_{\{T\le t\}}\mid\mathcal{F}_T\right]\right].$$

We split skel(t) into two components, the part before time T, skel(T) and the part after time T, but before t which is: skel$(T;t)$. We do this since skel(T) has just one particle (as no separation has occurred yet). Thus:

$$\begin{aligned}
\prod_{v\in\text{skel}(t)} \mathrm{e}^{(m_{B_v}-1)\int_{\sigma_v(t)}^{\tau_v(t)}\mathbb{1}_0(X_v(s))\mathrm{d}s} &= \prod_{v\in\text{skel}(T)} \mathrm{e}^{(m_{B_v}-1)\int_{\sigma_v(t)}^{\tau_v(t)}\mathbb{1}_0(X_v(s))\mathrm{d}s}\times\prod_{v\in\text{skel}(T;t)} \mathrm{e}^{(m_{B_v}-1)\int_{\sigma_v(t)}^{\tau_v(t)}\mathbb{1}_0(X_v(s))\mathrm{d}s},\\
&= \mathrm{e}^{(m_k-1)\int_0^T\mathbb{1}_0(\chi_s)\mathrm{d}s}\times\prod_{v\in\text{skel}(T;t)} \mathrm{e}^{(m_{B_v}-1)\int_{\sigma_v(t)}^{\tau_v(t)}\mathbb{1}_0(X_v(s))\mathrm{d}s}.
\end{aligned}$$

Noting that both the first multiplicand and $\mathbb{1}_{\{T\le t\}}$ are $\mathcal{F}_T$-measurable. Thus we apply *TOWIK* to get:

$$\mathbb{Q}^k\left[\prod_{v\in\text{skel}(t)} \mathrm{e}^{(m_{B_v}-1)\int_{\sigma_v(t)}^{\tau_v(t)}\mathbb{1}_0(X_v(s))\mathrm{d}s}\mathbb{1}_{\{T\le t\}}\right]=\mathbb{Q}^k\left[\mathbb{Q}^k\left[\prod_{v\in\text{skel}(t)} \mathrm{e}^{(m_{B_v}-1)\int_{\sigma_v(t)}^{\tau_v(t)}\mathbb{1}_0(X_v(s))\mathrm{d}s}\mathbb{1}_{\{T\le t\}}\mid\mathcal{F}_T\right]\right], \tag{3.40}$$

$$=\mathbb{Q}^k\left[\mathrm{e}^{(m_k-1)\int_0^T\mathbb{1}_0(\chi_s)\mathrm{d}s}\mathbb{1}_{\{T\le t\}}\mathbb{Q}^k\left[\prod_{v\in\text{skel}(T;t)} \mathrm{e}^{(m_{B_v}-1)\int_{\sigma_v(t)}^{\tau_v(t)}\mathbb{1}_0(X_v(s))\mathrm{d}s}\mid\mathcal{F}_T\right]\right]. \tag{3.41}$$

In total, combining 3.39 and 3.41 gives:

$$\mathbb{P}\left[N_t^k\right] = \mathbb{Q}^k\left[\mathrm{e}^{(m_k-1)\int_0^T \mathbb{1}_0(\chi_s)\mathrm{d}s}\mathbb{1}_{\{T\leq t\}}\mathbb{Q}^k\left[\prod_{v\in\mathrm{skel}(T;t)}\mathrm{e}^{(m_{B_v}-1)\int_{\sigma_v(t)}^{\tau_v(t)}\mathbb{1}_0(X_v(s))\mathrm{d}s}\mid\mathcal{F}_T\right]\right] + \mathbb{Q}^k\left[\mathrm{e}^{(m_k-1)\int_0^t \mathbb{1}_0(\chi_s)\mathrm{d}s}\mathbb{1}_{\{T>t\}}\right].$$

Note that T is not exactly exponential, as the rate is variable (it depends on the behaviour of χ). I do not go into the background here, but one can at least see intuitively why the following is true:

$$\mathbb{Q}^k\left(T>t\mid(\chi_s)_{0\leq s\leq t}\right) = \mathrm{e}^{-(m_k-m)\int_0^t \mathbb{1}_0(\chi_s)\mathrm{d}s},$$

by analogy to the exponential CDF. Notice that, by the *tower rule* and then *TOWIK*:

$$\begin{aligned}\mathbb{Q}^k\left[\mathrm{e}^{(m_k-1)\int_0^t \mathbb{1}_0(\chi_s)\mathrm{d}s}\mathbb{1}_{\{T>t\}}\right] &= \mathbb{Q}^k\left[\mathbb{Q}^k\left[\mathrm{e}^{(m_k-1)\int_0^t \mathbb{1}_0(\chi_s)\mathrm{d}s}\mathbb{1}_{\{T>t\}}\mid(\chi_s)_{0\leq s\leq t}\right]\right],\\ &= \mathbb{Q}^k\left[\mathrm{e}^{(m_k-1)\int_0^t \mathbb{1}_0(\chi_s)\mathrm{d}s}\mathbb{Q}^k\left[\mathbb{1}_{\{T>t\}}\mid(\chi_s)_{0\leq s\leq t}\right]\right],\\ &= \mathbb{Q}^k\left[\mathrm{e}^{(m_k-1)\int_0^t \mathbb{1}_0(\chi_s)\mathrm{d}s}\mathrm{e}^{-(m_k-m)\int_0^t \mathbb{1}_0(\chi_s)\mathrm{d}s}\right],\\ &= \mathbb{Q}^k\left[\mathrm{e}^{(m-1)\int_0^t \mathbb{1}_0(\chi_s)\mathrm{d}s}\right].\end{aligned}$$

This is a very useful, form as by Theorem 3.4 we get that:

$$\mathbb{Q}^k\left[\mathrm{e}^{(m-1)\int_0^t \mathbb{1}_0(\chi_s)\mathrm{d}s}\right] = \mathbb{P}_x[N_t].$$

Now we deal with the first summand of $\mathbb{P}[N_t^k]$. Integrating out by T, gives the following expression:

$$\begin{aligned}&\mathbb{Q}^k\left[\mathrm{e}^{(m_k-1)\int_0^T \mathbb{1}_0(\chi_s)\mathrm{d}s}\mathbb{1}_{\{T\leq t\}}\mathbb{Q}^k\left[\prod_{v\in\mathrm{skel}(T;t)}\mathrm{e}^{(m_{B_v}-1)\int_{\sigma_v(t)}^{\tau_v(t)}\mathbb{1}_0(X_v(s))\mathrm{d}s}\mid\mathcal{F}_T\right]\right],\\ &= \int_0^t \mathbb{Q}^k\left[\mathrm{e}^{(m_k-1)\int_0^u \mathbb{1}_0(\chi_s)\mathrm{d}s}\mathbb{1}_{\{u\leq t\}}\mathbb{Q}^k\left[\prod_{v\in\mathrm{skel}(u;t)}\mathrm{e}^{(m_{B_v}-1)\int_{\sigma_v(t)}^{\tau_v(t)}\mathbb{1}_0(X_v(s))\mathrm{d}s}\mid\mathcal{F}_u, T=u\right]\right] f(u\mid(\chi_s)_{0\leq s\leq u})\mathrm{d}u,\end{aligned}$$

where f is the density function of T, which is a function of $(\chi_s)_{0\leq s\leq t}$. For $u>0$, this is given by:

$$f(u\mid(\chi_s)_{0\leq s\leq u}) = \mathbb{1}_0(\chi_u)(m_k-m)\mathrm{e}^{-(m_k-m)\int_0^u \mathbb{1}_0(\chi_s)\mathrm{d}s}.$$

Taking f into the expectation by linearity, and since $\mathbb{1}_{\{u\leq t\}}=1$ gives:

$$\begin{aligned}\int_0^t \mathbb{Q}^k\Big[&\mathrm{e}^{(m_k-1)\int_0^u \mathbb{1}_0(\chi_s)\mathrm{d}s}\mathbb{1}_0(\chi_u)(m_k-m)\mathrm{e}^{-(m_k-m)\int_0^u \mathbb{1}_0(\chi_s)\mathrm{d}s}\\ &\times\mathbb{Q}^k\left[\prod_{v\in\mathrm{skel}(u;t)}\mathrm{e}^{(m_{B_v}-1)\int_{\sigma_v(t)}^{\tau_v(t)}\mathbb{1}_0(X_v(s))\mathrm{d}s}\mid\mathcal{F}_u, T=u\right]\Bigg]\mathrm{d}u.\end{aligned}$$

Simplifying this gives:

$$\int_0^t \mathbb{Q}^k\left[\mathrm{e}^{(m-1)\int_0^u \mathbb{1}_0(\chi_s)\mathrm{d}s}\mathbb{1}_0(\chi_u)(m_k-m)\mathbb{Q}^k\left[\prod_{v\in\mathrm{skel}(u;t)}\mathrm{e}^{(m_{B_v}-1)\int_{\sigma_v(t)}^{\tau_v(t)}\mathbb{1}_0(X_v(s))\mathrm{d}s}\mid\mathcal{F}_u, T=u\right]\right]\mathrm{d}u.$$

We split the sample space (use law of total probability) into how many particles the initial particle with k marks splits into, and what the distribution of marks between them is, i.e. events of the form $A_k(j;i_1,\ldots,i_j)$

for $j \geq 2$, for $i_1 + \cdots + i_j = k$ happen. This gives us a final expression for $\mathbb{P}[N_t^k]$ as:

$$\begin{aligned}
\mathbb{P}[N_t^k] = &\int_0^t \mathbb{Q}^k \left[\mathrm{e}^{(m-1)\int_0^u \mathbb{1}_0(\chi_s)\mathrm{d}s} \mathbb{1}_0(\chi_u)(m_k - m)\mathbb{Q}^k \left[\prod_{v \in \mathrm{skel}(u;t)} \mathrm{e}^{(m_{B_v}-1)\int_{\sigma_v(t)}^{\tau_v(t)} \mathbb{1}_0(X_v(s))\mathrm{d}s} \mid \mathcal{F}_u, T = u \right]\right] \mathrm{d}u \\
&+ \mathbb{Q}^k \left[\mathrm{e}^{(m-1)\int_0^t \mathbb{1}_0(\chi_s)\mathrm{d}s} \right], \\
&= \int_0^t \mathbb{Q}^k \left[\mathrm{e}^{(m-1)\int_0^u \mathbb{1}_0(\chi_s)\mathrm{d}s} \mathbb{1}_0(\chi_u)(m_k - m)\mathbb{Q}^k \left[\prod_{v \in \mathrm{skel}(u;t)} \mathrm{e}^{(m_{B_v}-1)\int_{\sigma_v(t)}^{\tau_v(t)} \mathbb{1}_0(X_v(s))\mathrm{d}s} \mid \mathcal{F}_u, T = u \right]\right] \mathrm{d}u \\
&+ \mathbb{P}_x[N_t], \\
&= \int_0^t \sum_{j=2}^k \sum_{\substack{i_1,\ldots,i_j > 0 \\ i_1+\ldots+i_j = k}} \mathbb{Q}^k(A_k(j; i_1, \ldots, i_j) \mid \text{separation event}) \mathbb{Q}^k \left[\mathrm{e}^{(m-1)\int_0^u \mathbb{1}_0(\chi_s)\mathrm{d}s} \mathbb{1}_0(\chi_u)(m_k - m) \right. \\
&\left. \mathbb{Q}^k \left[\prod_{v \in \mathrm{skel}(u;t)} \mathrm{e}^{(m_{B_v}-1)\int_{\sigma_v(t)}^{\tau_v(t)} \mathbb{1}_0(X_v(s))\mathrm{d}s} \mid \mathcal{F}_u, T = u, A_k(j; i_1, \ldots, i_j) \right]\right] \mathrm{d}u + \mathbb{P}_x[N_t].
\end{aligned}$$

We know from previously that:

$$\mathbb{Q}^k(A_k(j; i_1, \ldots, i_j) \mid \text{separation event}) = \frac{1}{m_k} \mathbb{E}\left[\binom{X}{j} \right] \frac{k!}{i_1! \cdots i_j!} \left(\frac{m_k}{m_k - m} \right).$$

Let us concentrate for now on the inside expectation:

$$\mathbb{Q}^k \left[\prod_{v \in \mathrm{skel}(u;t)} \mathrm{e}^{(m_{B_v}-1)\int_{\sigma_v(t)}^{\tau_v(t)} \mathbb{1}_0(X_v(s))\mathrm{d}s} \mid \mathcal{F}_u, T = u, A_k(j; i_1, \ldots, i_j) \right].$$

The separation time is u, at which point the k marks split off into j particles with $i_1, \ldots, i_j$ assigned to each particle, respectively. Hence we can partition up $\mathrm{skel}(u;t)$ into j groups of descendants of each of the j particles and their paths up to time t. As we have the separation time, and the distribution of marks these particles act independently and thus we split up the sum into a product of j terms. We can 'restart' the system using the Markov property and thus instead consider $\mathrm{skel}(t-u)$ for each particle, as the particles all act independently and thus time restarts at u. In mathematics, we can write this as:

$$\begin{aligned}
&\mathbb{Q}^k \left[\prod_{v \in \mathrm{skel}(u;t)} \mathrm{e}^{(m_{B_v}-1)\int_{\sigma_v(t)}^{\tau_v(t)} \mathbb{1}_0(X_v(s))\mathrm{d}s} \mid \mathcal{F}_u, T = u, A_k(j; i_1, \ldots, i_j) \right] \\
&= \prod_{\ell=1}^j \mathbb{Q}^{i_\ell} \left[\prod_{v \in \mathrm{skel}(t-u)} \mathrm{e}^{(m_{B_v}-1)\int_{\sigma_v(t)}^{\tau_v(t)} \mathbb{1}_0(X_v(s))\mathrm{d}s} \right]
\end{aligned}$$

Substituting this into our expression for $\mathbb{P}[N_t^k]$ gives:

$$\mathbb{P}[N_t^k] = \int_0^t \sum_{j=2}^k \sum_{\substack{i_1,\ldots,i_j>0 \\ i_1+\ldots+i_j=k}} \mathbb{Q}^k(A_k(j;i_1,\ldots,i_j) \mid \text{separation event})\mathbb{Q}^k\Bigg[\mathrm{e}^{(m-1)\int_0^u \mathbb{1}_0(\chi_s)\mathrm{d}s}\mathbb{1}_0(\chi_u)(m_k - m)$$

$$\mathbb{Q}^k\left[\prod_{v\in\mathrm{skel}(u;t)} \mathrm{e}^{(m_{B_v}-1)\int_{\sigma_v(t)}^{\tau_v(t)} \mathbb{1}_0(X_v(s))\mathrm{d}s} \mid \mathcal{F}_u, T=u, A_k(j;i_1,\ldots,i_j)\right]\Bigg]\mathrm{d}u + \mathbb{P}_x[N_t],$$

$$= \int_0^t \sum_{j=2}^k \sum_{\substack{i_1,\ldots,i_j>0 \\ i_1+\ldots+i_j=k}} \mathbb{E}\left[\binom{X}{j}\right] \frac{k!}{i_1!\cdots i_j!}\mathbb{Q}^k\left[\mathrm{e}^{(m-1)\int_0^u \mathbb{1}_0(\chi_s)\mathrm{d}s}\mathbb{1}_0(\chi_u)\right.$$

$$\left.\times \prod_{\ell=1}^j \mathbb{Q}^{i_\ell}\left[\prod_{v\in\mathrm{skel}(t-u)} \mathrm{e}^{(m_{B_v}-1)\int_{\sigma_v(t)}^{\tau_v(t)} \mathbb{1}_0(X_v(s))\mathrm{d}s}\right]\mathrm{d}u + \mathbb{P}_x[N_t].$$

Finally, we use Theorem 3.4 backwards with $f = \mathbb{1}_0$ to give:

$$\mathbb{Q}^k\left[\mathrm{e}^{(m-1)\int_0^u \mathbb{1}_0(\chi_s)\mathrm{d}s}\mathbb{1}_0(\chi_u)\right] = \mathbb{P}_x[N_u(0)].$$

We further use $f = 1$ to give, for $\ell \in \{1,\ldots,j\}$:

$$\mathbb{Q}^{i_\ell}\left[\prod_{v\in\mathrm{skel}(t-u)} \mathrm{e}^{(m_{B_v}-1)\int_{\sigma_v(t)}^{\tau_v(t)} \mathbb{1}_0(X_v(s))\mathrm{d}s}\right] = \mathbb{P}_0[N_{t-u}^{i_\ell}].$$

We finally conclude that:

$$\mathbb{P}_x[N_t^k] = \int_0^t \sum_{j=2}^k \sum_{\substack{i_1,\ldots,i_j>0 \\ i_1+\ldots+i_j=k}} \mathbb{E}\left[\binom{X}{j}\right] \frac{k!}{i_1!\cdots i_j!}\mathbb{P}_x[N_u(0)] \cdot \prod_{\ell=1}^j \mathbb{P}_0[N_{t-u}^{i_\ell}]\mathrm{d}u + \mathbb{P}_x[N_t],$$

which proves the first part of the lemma. Note that I have added back in the starting point $x \in S$. The case for $M^1(t,x,y)$ follows exactly the same lines of reasoning but uses $f = \mathbb{1}_y$ in the original application of Theorem 3.4. $\square$

3.6 Final proof - for M^k, $k \geq 2$

The following result is used in this final proof and is given without proof, for the proof see [DR12, page 11].

Lemma 3.12. *For any non-negative integer-valued random variable Y, and any integers $a \geq b \geq 1$ we have:*

$$\mathbb{E}[Y^a]\mathbb{E}[Y] \geq \mathbb{E}[Y^b]\mathbb{E}[Y^{a-b+1}].$$

Proof of Theorem 3.5. For the general subcritical case, for the kth moment, for $k \geq 1$. We use Theorem 3.4 to get the following expression for $M^k(t,x)$:

$$M^k(t,x) = \mathbb{Q}_z^k\left[\prod_{v\in\mathrm{skel(t)}} \exp\left((m_{B_v}-1)\int_{\sigma_v(t)}^{\tau_v(t)} \mathbb{1}_0(X_v(s))\mathrm{d}s\right)\right]. \tag{3.42}$$

We apply the *dominated convergence theorem*. Thus we need to find the limit of

$$\prod_{v\in\mathrm{skel(t)}} \exp\left((m_{B_v}-1)\int_{\sigma_v(t)}^{\tau_v(t)} \mathbb{1}_0(X_v(s))\mathrm{d}s\right),$$

and also find a dominating integrable random variable of it. Firstly, notice that if we take t large enough, all of the k marks will have distributed to individual particles so there will be k particles each carrying 1 mark. We will call this set of particles C.

$$M^k(t,x) = \mathbb{Q}_z^k\left[\prod_{v\in\mathrm{skel(t)}} \exp\left((m_{B_v}-1)\int_{\sigma_v(t)}^{\tau_v(t)} \mathbb{1}_0(X_v(s))\mathrm{d}s\right)\right], \tag{3.43}$$

$$= \mathbb{Q}_z^k\left[\exp\left(\sum_{v\in\mathrm{skel}(t)} (m_{B_v}-1)\int_{\sigma_v(t)}^{\tau_v(t)} \mathbb{1}_0(X_v(s))\mathrm{d}s\right)\right], \tag{3.44}$$

$$= \mathbb{Q}_z^k\left[\exp\left(\sum_{v\in C}(m-1)\int_{\sigma_v(t)}^{t} \mathbb{1}_0(X_v(s))\mathrm{d}s + Z\right)\right], \tag{3.45}$$

where Z is a finite sum of other terms consisting of all other particles in the skeleton until all the k marks split up. I replace the $\tau_v(t)$ term with t, as particles with marks always have a positive number of offspring under $\mathbb{Q}_x^k$, thus the line of descent never ends. As $t\to\infty$, the term Z stays the same but the other either goes to $-\infty$ if $\mathcal{A}$ is recurrent or to some constant if $\mathcal{A}$ is transient. Thus if we can find a dominated random variable, this gives us the result of Theorem 3.5 for the kth moment. Let $\beta_v = \tau_v - \sigma_v$, where $\beta_v \sim \mathrm{Exp}(m_{B_v} - m_1)$, as $m_{B_v} - m_1$ is the rate at which the B_v marks split during a branching event (as in proof of Lemma 3.11). Notice that, for any $v \in \mathrm{skel}(t)$,

$$\exp\left((m_{B_v}-1)\int_{\sigma_v(t)}^{\tau_v(t)} \mathbb{1}_0(X_v(s))\mathrm{d}s\right) \le \exp\left((m_{B_v}-1)(\tau_v-\sigma_v)\right) = \exp\left((m_{B_v}-1)\beta_v\right). \tag{3.46}$$

Thus we have that:

$$\prod_{v\in\mathrm{skel(t)}} \exp\left((m_{B_v}-1)\int_{\sigma_v(t)}^{\tau_v(t)} \mathbb{1}_0(X_v(s))\mathrm{d}s\right) \le \prod_{v\in\mathrm{skel(t)}} \exp\left((m_{B_v}-1)\beta_v\right). \tag{3.47}$$

By independence we get, which gives us that the upper bound is integrable:

$$\mathbb{Q}_x^k\left[\prod_{v\in\mathrm{skel(t)}} \exp\left((m_{B_v}-1)\beta_v\right)\right] = \prod_{v\in\mathrm{skel(t)}} \mathbb{Q}_x^k\left[\exp\left((m_{B_v}-1)\beta_v\right)\right]. \tag{3.48}$$

Each of these is finite as for $X\sim\mathrm{Exp}(\lambda)$ we have that $\mathbb{E}[\mathrm{e}^{\alpha X}]<\infty$ if $\alpha<\lambda$. For each $v\in\mathrm{skel}(t)$ we have that $m_k - 1 < m_k - m_1$, since $m_1 = m < 1$. This gives the dominating random variable, and thus the result follows by the *dominated convergence theorem.*

□

Proof of Theorem 3.6. Let us recall that in the critical case, by Lemma 3.9 we have that $M^1(t,x)\equiv 1$ and that $M^1(t,x,y) = p_t(x,y)$ for all $t>0, x\in S$. Now by Lemma 3.11 we have that:

$$M^k(t,x) = M^1(t,x) + M^1(t,x,0) * g_k\left(M^1(t,0),\dots,M^{k-1}(t,0)\right), \tag{3.49}$$

where g_k was defined in Lemma 3.11. Taking only the $j=2$ and $i_1 = 1, i_2 = k-1$ term (resulting in a lower bound) gives the lower bound:

$$M^k(t,x) \ge M^1(t,x) + M^1(t,x,0) * \left(\mathbb{E}\left[\binom{X}{2}\right]\frac{k!}{1!(k-1)!}M^1(t,0)M^{k-1}(t,0)\right), \tag{3.50}$$

$$\ge 1 + p_t(x,y) * \left(\mathbb{E}\left[\binom{X}{2}\right] k\cdot 1\cdot M^{k-1}(t,0)\right), \tag{3.51}$$

$$\ge 1 + C_1 p_t(x,y) * M^{k-1}(t,0), \tag{3.52}$$

by letting $C_1 := k\mathbb{E}\left[\binom{X}{2}\right]$. This gives a final lower bound of:

$$M^k(t,x) \geq C_1 p_t(x,y) * M^{k-1}(t,0). \tag{3.53}$$

Now we find an upper bound of $M^k(t,x)$, as in Equation 3.49 we have:

$$\begin{aligned}
M^k(t,x) &= 1 + p_t(x,0) * g_k\left(M^1(t,0),\ldots,M^{k-1}(t,0)\right), && (3.54)\\
&= 1 + p_t(x,0) * \sum_{j=2}^{k} \mathbb{E}\left[\binom{X}{j}\right] \sum_{\substack{i_1,\ldots,i_j>0\\ i_1+\ldots+i_j=k}} \frac{k!}{i_1!\ldots i_j!} M^{i_1}(t,0)\cdots M^{i_j}(t,0), && (3.55)\\
&= 1 + p_t(x,0) * \sum_{j=2}^{k} \mathbb{E}\left[\binom{X}{j}\right] \sum_{\substack{i_1,\ldots,i_j>0\\ i_1+\ldots+i_j=k}} \frac{k!}{i_1!\ldots i_j!} \mathbb{E}_0\left[N_t^{i_1}\right]\cdots\mathbb{E}_0\left[N_t^{i_j}\right]. && (3.56)
\end{aligned}$$

We then apply Lemma 3.12 inductively on:

$$\mathbb{E}_0\left[N_t^{i_1}\right]\cdots\mathbb{E}_0\left[N_t^{i_j}\right], \tag{3.57}$$

in the following way for $i_n \leq i_m$:

$$\mathbb{E}_0\left[N_t^{i_n}\right]\mathbb{E}_0\left[N_t^{i_m}\right] \leq \mathbb{E}_0\left[N_t^{1}\right]\mathbb{E}_0\left[N_t^{i_m}\right] = \mathbb{E}_0\left[N_t^{i_n+i_m-1}\right], \tag{3.58}$$

with $a = i_n + i_m - 1, b = i_m$. If one apply this iteratively for all $j \in \{2,\ldots,k\}$ terms then we get a final exponent which is less than or equal to $k-1$. Thus returning to Equation 3.56 we get:

$$\begin{aligned}
M^k(t,x) &= 1 + p_t(x,0) * \sum_{j=2}^{k} \mathbb{E}\left[\binom{X}{j}\right] \sum_{\substack{i_1,\ldots,i_j>0\\ i_1+\ldots+i_j=k}} \frac{k!}{i_1!\ldots i_j!} \mathbb{E}_0\left[N_t^{i_1}\right]\cdots\mathbb{E}_0\left[N_t^{i_j}\right], && (3.59)\\
&\leq 1 + p_t(x,0) * \sum_{j=2}^{k} \mathbb{E}\left[\binom{X}{j}\right] \sum_{\substack{i_1,\ldots,i_j>0\\ i_1+\ldots+i_j=k}} \frac{k!}{i_1!\ldots i_j!} \mathbb{E}_0\left[N_t^{k-1}\right], && (3.60)\\
&\leq 1 + C_2 p_t(x,0) * M^{k-1}(t,0), && (3.61)
\end{aligned}$$

letting $C_2 := \sum_{j=2}^{k} \mathbb{E}\left[\binom{X}{j}\right] \sum_{\substack{i_1,\ldots,i_j>0\\ i_1+\ldots+i_j=k}} \frac{k!}{i_1!\ldots i_j!}$.

We now inductively use the lower bound in Equation 3.53 as follow (for $k \geq 2$):

$$\begin{aligned}
M^k(t,x) &\geq C_1 \int_0^t p_{s_1}(x,0) M^{k-1}(t-s_1,0)\mathrm{d}s_1, && (3.62)\\
&\geq C_1 \int_0^t p_{s_1}(x,0) C_1 \int_0^{t-s_1} p_{s_2}(0,0) M^{k-2}(t-s_1-s_2,0)\mathrm{d}s_2\mathrm{d}s_1, && (3.63)\\
&\geq C_1^{k-1} \int_0^t \int_0^{t-s_1} \cdots \int_0^{t-s_1-s_2-\ldots-s_{k-2}} p_{s_1}(x,0)p_{s_2}(0,0)\cdots p_{s_{k-1}}(0,0) && (3.64)\\
&\times M^1(t-s_1-s_2-\ldots-s_{k-2},0)\mathrm{d}s_{k-1}\cdots\mathrm{d}s_2\mathrm{d}s_1. && (3.65)
\end{aligned}$$

However as we have $m=1$, we have $M^1(t-s_1-s_2-\cdots-s_{k-2},0) \equiv 1$, thus we now focus on the expression:

$$\begin{aligned}
&\int_0^t \int_0^{t-s_1} \cdots \int_0^{t-s_1-s_2-\ldots-s_{k-2}} p_{s_1}(x,0)p_{s_2}(0,0)\cdots p_{s_{k-1}}(0,0)\mathrm{d}s_{k-1}\cdots\mathrm{d}s_2\mathrm{d}s_1, && (3.66)\\
&= \int_0^t p_{s_1}(x,0) \int_0^{t-s_1} p_{s_2}(0,0)\cdots \int_0^{t-s_1-s_2-\ldots-s_{k-2}} \cdots p_{s_{k-1}}(0,0)\mathrm{d}s_{k-1}\cdots\mathrm{d}s_2\mathrm{d}s_1. && (3.67)
\end{aligned}$$

We perform the change of variables $s_1' = s_1$, $s_2' = s_2 + s_1'$, and thus for $2 \le i \le k-1$ we have $s_i' = s_i + s_{i-1}'$. After replacing the symbols in the second expression, this gives:

$$\int_0^t p_{s_1'}(x,0) \int_{s_1'}^t p_{s_2'-s_1'}(0,0) \cdots \int_{s_{k-2}'}^t \cdots p_{s_{k-1}'-s_{k-2}'}(0,0) \mathrm{d}s_{k-1}' \cdots \mathrm{d}s_2' \mathrm{d}s_1', \tag{3.68}$$

$$= \int_0^t \mathbb{P}_x(X_{s_1} = 0) \int_{s_1}^t \mathbb{P}_0(X_{s_2-s_1} = 0) \cdots \int_{s_{k-2}}^t \cdots \mathbb{P}_0(X_{s_{k-1}-s_{k-2}} = 0) \mathrm{d}s_{k-1} \cdots \mathrm{d}s_2 \mathrm{d}s_1. \tag{3.69}$$

Then from the Markov property and definition of conditional probability one can see that:

$$\mathbb{P}_x\left(X_{s_1} = 0, X_{s_2} = 0, \ldots, X_{s_{k-1}} = 0\right) = \mathbb{P}_x(X_{s_1} = 0)\mathbb{P}_0(X_{s_2-s_1} = 0) \cdot \cdots \cdot \mathbb{P}_0(X_{s_{k-1}-s_{k-2}} = 0). \tag{3.70}$$

Thus we have that Equation 3.69 simplifies to:

$$\int_0^t \int_{s_1}^t \cdots \int_{s_{k-2}}^t \mathbb{P}_x\left(X_{s_1} = 0, X_{s_2} = 0, \ldots, X_{s_{k-1}} = 0\right) \mathrm{d}s_{k-1} \cdots \mathrm{d}s_2 \mathrm{d}s_1. \tag{3.71}$$

We can see that is the same as integrating over $\frac{1}{(k-1)!}$ of a simplex giving, and then simplifying:

$$\int_0^t \int_{s_1}^t \cdots \int_{s_{k-2}}^t \mathbb{P}_x\left(X_{s_1} = 0, X_{s_2} = 0, \ldots, X_{s_{k-1}} = 0\right) \mathrm{d}s_{k-1} \cdots \mathrm{d}s_2 \mathrm{d}s_1, \tag{3.72}$$

$$= \frac{1}{(k-1)!} \int_0^t \int_0^t \cdots \int_0^t \mathbb{P}_x\left(X_{s_1} = 0, X_{s_2} = 0, \ldots, X_{s_{k-1}} = 0\right) \mathrm{d}s_{k-1} \cdots \mathrm{d}s_2 \mathrm{d}s_1, \tag{3.73}$$

$$= \frac{1}{(k-1)!} \mathbb{E}_x\left[\left(\int_0^t \mathbb{1}_{\{X_s=0\}} \mathrm{d}s\right)^{k-1}\right], \tag{3.74}$$

$$= \frac{1}{(k-1)!} \mathbb{E}_x\left[L_t(0)^{k-1}\right]. \tag{3.75}$$

We can use the same expression with the upper bound, except we have the extra $+1$ term. If $\mathcal{A}$ is recurrent then we see that $M^k(t,x) \to \infty$ as $t \to \infty$ as we now have a lower bound which grows with t. Similarly if $\mathcal{A}$ is transient, (and doing similar algebra to above) we see that $M^k(t,x)$ converges to a constant in $(0,\infty)$. In both situations, by dividing by $\mathbb{E}_x\left[L_t(0)^{k-1}\right]$ on both sides of the bounds we get that $\lim_{t\to\infty} \frac{M^k(t,x)}{\mathbb{E}_x[L_t(0)^{k-1}]} \in (0,\infty)$. Thus giving Theorem 3.6. □

All that remains to prove is Theorem 3.7 for $k \ge 2$. The lower bound $c\mathbb{E}_x\left[L_t(0)^{k-1}\right]$ in part (a) follows similar algebra to the case for $m = 1$, the upper bound comes from applications of Lemma 3.11 and Lemma 3.12. Cases (b) and (c) follow also by induction on the $k = 1$ case with Lemma 3.11.

3.7 Moving catalyst

Instead of having a fixed branching source at $0 \in S$, it can instead move according to a continuous-time random walk on S, which we denote by $(Y_t)_{t\ge0}$, where we assume $Y_0 = 0$. To apply Theorem 3.5, Theorem 3.6 and Theorem 3.7 we change the movement process of the particles to be the difference walk between the particles and the branching source. Given a particle v, it's position is given by the process $(X_v(s))_{\sigma_v \le s \le \tau_v}$. Therefore, instead of using it's actual position, we use it's relative position to the branching source. This translates to using the difference walk $(X_v(s) - Y_s)_{\sigma_v \le s \le \tau_v}$ as the movement process of v in the Theorems above. This makes sense "as being at 0" at some time $s > 0$ in the difference walk is equivalent to $X_v(s) = Y_s$, i.e. the particle and the branching source are at the same position. Let $x, y \in S$, and $t > 0$, the meanings of $M^1(t,x,y)$ and $M^1(t,x)$ now change slightly but still give useful information.

Observe that $M^1(t,x) = \mathbb{E}_x[N_t]$ refers to the expected total number of particles in the system, when the initial particle is placed at $x - Y_0 = 0$, therefore it's meaning is unchanged from previously. However $M^1(t,x,y)$ refers to the expected number of particles at position $y - Y_t$, thus this is a slightly different quantity (as Y_t) is random. Therefore Theorem 3.5 still applies as the first two statement concern $M^1(t,x)$ and since

$M^1(t,x,y)$ goes to zero for all y, we can thus deduce that the original quantity, the "expected number of particles at position y" also goes to zero. Theorem 3.6 holds similarly. Case (b) are all statements about $M^1(t,x)$ and $M^1(t,x,y)$ for all y, and hence hold similarly. Cases (a) and (c) are particularly interesting as they exhibit non-localisation and localisation behaviour, respectively. In the case of (a) we see $M^1(t,x)$ settling to a limit (with some nice bounds) and $M^1(t,x,y)$ going to zero, for all y. For (c) we see behaviour consistent with localisation for $M^1(t,x)$. In particular we see that, for all $x, y \in S$:

$$M^k(t,x) \sim C_1 \mathrm{e}^{kr(m)t} \quad \text{and} \quad M^k(t,x,y) \sim C_2 \mathrm{e}^{kr(m)t}.$$

This behaviour is particular interesting for $M^k(t,x,y)$ as this the expected number of particles at position $y - Y_t$, so as time goes on the expected number of particles around the particle approaches this nice exponential form.

4 Adapting Döring-Roberts to an everywhere-branching process

Here we consider the everywhere-branching process, with general offspring distribution μ. This is a generalisation of the MS model, where $\mathbb{P}_\mu(X=2)=1$, i.e. binary branching. In this section some of the techniques from [DR12] are adapted to this new process and we use Theorem 2.25 from [HR14] to prove some elementary results about it. We assume for this section that $t>0$ and $x,y \in S = \mathbb{Z}^d$ for $d \in \mathbb{N}$.

4.1 Feynman-Kac representation of moments

Here we prove an analogous result to Lemma 3.9. The everywhere-branching process works as follows: it branches at some rate $\lambda_0 > 0$ at the origin, and then at another rate $\lambda > 0$ everywhere else in S. Each particle moves independently according to some irreducible continuous-time Markov process $\mathcal{A}$, on a countable state space S. The proof follows that of Lemma 3.9 closely but uses the more general many-to-few lemma Theorem 2.25.

Theorem 4.1. *The first moments can be expressed as:*

$$M^1(t,x) = \mathbb{E}_x\left(\mathrm{e}^{(m-1)(\lambda t + (\lambda_0 - \lambda)\int_0^t \mathbb{1}_0(\xi_s)\mathrm{d}s)}\right),$$
$$M^1(t,x,y) = \mathbb{E}_x\left(\mathrm{e}^{(m-1)(\lambda t + (\lambda_0 - \lambda)\int_0^t \mathbb{1}_0(\xi_s)\mathrm{d}s)} \mathbb{1}_y(\xi_t)\right).$$

for $t > 0$ and $x, y \in S$.

Proof. We use Theorem 2.25 with the following:

- $k = 1$, one spine considered;
- $\zeta \equiv 1$, constant martingale;
- $Y \equiv 1$;
- $\mathbb{E}[X \mid X \sim \mu] = m$ the expected number of offspring;
- $Y(z) = \mathbb{1}_y(z)$ when we need to calculate $M^1(t,x)$;
- $Y(z) = \mathbb{1}_y(z)$ when we need to calculate $M^1(t,x,y)$, for $z \in S$.

We recall that $D(v) \equiv 1$, for all $v \in \mathrm{skel}(t)$, since the particles in the skeleton only carry one mark. Thus for $s > 0$ we have that $\alpha_{D(v)}(X_v(s)) = (m-1)R(X_v(s))$, and $R(X_v(s)) = \lambda_0 \mathbb{1}_0(\xi_s) + \lambda \mathbb{1}_{\neq 0}(\xi_s)$. Recall that:

$$M^1(t,x) = \mathbb{P}_x\left[\sum_{v_1 \in N(t)} 1\right] \tag{4.1}$$

We refer to the only particle who follows the single line of descent in the spine as ξ. By Theorem 2.25, for starting position x:

$$\mathbb{P}_x\left[\sum_{v_1\in N(t)} 1\right] = \mathbb{Q}_x^1\left[\mathrm{e}^{(m-1)(\int_0^t \lambda_0 \mathbb{1}_0(\xi_s)+\lambda \mathbb{1}_{\neq 0}(\xi_s)\mathrm{d}s)}\right]. \tag{4.2}$$

As in the corresponding proof on page 6 of [DR12], as $k=1$, the expectation under $\mathbb{Q}$ and under $\mathbb{P}$ are the same, and thus this further simplifies to

$$\mathbb{Q}_x^1\left[\mathrm{e}^{(m-1)(\int_0^t \lambda_0 \mathbb{1}_0(\xi_s)+\lambda \mathbb{1}_{\neq 0}(\xi_s)\mathrm{d}s)}\right] = \mathbb{E}_x\left[\mathrm{e}^{(m-1)(\int_0^t \lambda_0 \mathbb{1}_0(\xi_s)+\lambda \mathbb{1}_{\neq 0}(\xi_s)\mathrm{d}s)}\right]. \tag{4.3}$$

By the same steps as in original proof we also get:

$$M^1(t,x,y) = \mathbb{E}_x\left[\mathrm{e}^{(m-1)(\int_0^t \lambda_0 \mathbb{1}_0(\xi_s)+\lambda \mathbb{1}_{\neq 0}(\xi_s)\mathrm{d}s)}\mathbb{1}_y(\xi_t)\right].$$

Now notice that:

$$\int_0^t \lambda_0 \mathbb{1}_0(\xi_s)+\lambda \mathbb{1}_{\neq 0}(\xi_s)\mathrm{d}s = \int_0^t \lambda + \mathbb{1}_0(\xi_s)(\lambda_0-\lambda)\mathrm{d}s = \lambda t + (\lambda_0-\lambda)\int_0^t \mathbb{1}_0(\xi_s)\mathrm{d}s.$$

Bringing it all together, we have our two results:

$$M^1(t,x) = \mathbb{E}_x\left[\mathrm{e}^{(m-1)(\lambda t+(\lambda_0-\lambda)\int_0^t \mathbb{1}_0(\xi_s)\mathrm{d}s)}\right],$$

$$M^1(t,x,y) = \mathbb{E}_x\left[\mathrm{e}^{(m-1)(\lambda t+(\lambda_0-\lambda)\int_0^t \mathbb{1}_0(\xi_s)\mathrm{d}s)}\mathbb{1}_y(\xi_t)\right].$$

□

4.2 Subcritical case

Theorem 4.2 (Subcritical Case). *Suppose $m<1$ then, and $t>0$, we have:*

$$\lim_{t\to\infty} M^1(t,x)\begin{cases} \in (0,\infty) & \text{when } \mathcal{A} \text{ is transient,}\\ =0 & \text{when } \mathcal{A} \text{ is recurrent.}\end{cases}$$

In both cases we have:

$$\lim_{t\to\infty} M^1(t,x,y) = 0,$$

for any $x,y\in S$.

Proof. Suppose first, that we are in the case where $\mathcal{A}$ **is transient**, i.e. $\int_0^\infty \mathbb{1}_0(\xi_s)\mathrm{d}s<\infty$. We now rewrite $M^1(t,x)$ as

$$M^1(t,x) = \int_S \mathrm{e}^{(m-1)(\lambda t+(\lambda_0-\lambda)\int_0^t \mathbb{1}_0(\xi_s)\mathrm{d}s)}\mathrm{d}\mathbb{P}$$

and since $m<1$, we have that f_t is bounded above by 1, and thus for ξ a particle, the *dominated convergence theorem* applies to $f_t(\xi) = \mathrm{e}^{(m-1)(\lambda t+(\lambda_0-\lambda)\int_0^t \mathbb{1}_0(\xi_s)\mathrm{d}s)}$. Hence we get:

$$\lim_{t\to\infty} M^1(t,x) = \int_S \lim_{t\to\infty} f_t(\xi)\mathrm{d}\mathbb{P} = 0$$

this equals 0 as the $\mathrm{e}^{\lambda t}$ portion of f_t dominates and we get the exponential of an increasingly negative sequence which goes to zero. Similarly, when $\mathcal{A}$ is **recurrent**, we get that $\int_0^\infty \mathbb{1}_0(\xi_s)\mathrm{d}s$ diverges and thus we obtain zero in the limit as the sequence in the exponent also grows increasingly negative towards $-\infty$. □

This result should make intuitive sense, as branching (on average) kills the particles (since $m<1$), so having branching occur *everywhere* will kill off everything eventually.

4.3 Critical case

Theorem 4.3 (Critical Case). *For all $t > 0$, and $x, y \in S$ we have:*

$$\lim_{t\to\infty} M^1(t,x) \in (0,\infty) \quad \textit{and} \quad M^1(t,x,y) = p_t(x,y).$$

Proof. In the **critical** case we have $m = 1$, so the formula for M^1 from Section 4.2 remains

$$M^1(t,x) = \mathbb{E}_x\left[\mathrm{e}^{(m-1)(\lambda t+(\lambda_0-\lambda)\int_0^t \mathbb{1}_0(\xi_s)\mathrm{d}s)}\right].$$

As in [DR12], substituting in $m = 1$ gives:

$$M^1(t,x) = \mathbb{E}_x\left[\mathrm{e}^{(1-1)(\lambda t+(\lambda_0-\lambda)\int_0^t \mathbb{1}_0(\xi_s)\mathrm{d}s)}\right] = \mathbb{E}_x\left[\mathrm{e}^0\right] = 1.$$

We also get that

$$M^1(t,x,y) = \mathbb{E}_x\left[1 \cdot \mathbb{1}_y(\xi_t)\right] = \mathbb{P}_x(\xi_t = y) = p_t(x,y),$$

as in the proof of Theorem 3.6 when $k = 1$, in Section 3.4.2 $\square$

5 Simulations

To illustrate some of the processes discussed in this project some simulations were performed in $S = \mathbb{Z}^3$, with uniform movement around the 3-dimensional lattice. The simulations include both branching at just the origin, and outside of the origin. Branching at the origin happens at rate λ_0, and branching outside of the origin at rate λ.

5.1 The simulation algorithm

Before displaying the results of the simulation, we introduce the algorithm that was used. The process is discretised, meaning that the dynamics are slightly different to the theoretical models. This algorithm allows us to change the parameter d of $S = \mathbb{Z}^d$; as well as the movement rate (but always with uniform movement); the branching rates λ and λ_0 and the offspring distribution μ.

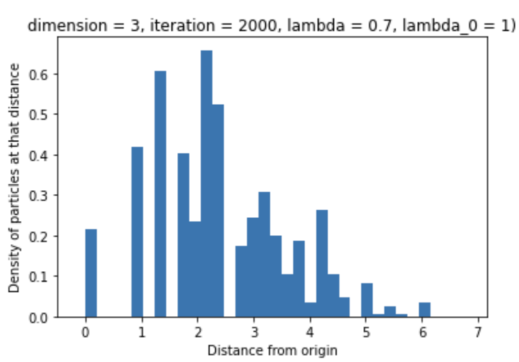
Note that "tot" stores the total number of particles, and "checkpointed" is a vector that stores the proportion of particles at the origin through time. The initial version of the LaTex code written here was written by ChatGPT (as I was unfamiliar with how to write algorithms) but it has been subsequently checked for accuracy.

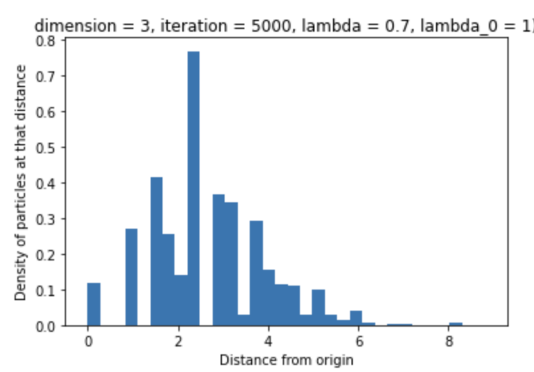
Algorithm 1 Simulate Branching Random Walk with General Offspring Distribution

```
Initialise nums ← {0 ↦ 1}                                ▷ nums[position] is the number of particles at position
Set tot ← 1, checkpointed ← [1]
Define branch_rate(x) ← branch_rate_0 if x = 0 else branch_rate_other
Let total_rate ← (move_rate + branch_rate(0)) · 1          ▷ Total rate of something happening
Store timeline ← [nums]                                  ▷ Stores some values of nums for plotting
for i = 1 to iterations do
    Filter non-zero entries in nums to get sample_from
    if sample_from is empty then
        break
    end if
    Compute sampling distribution sample_dist over positions
    Sample x from sample_from using sample_dist
    Compute prob_branch ← branch_rate(x)/(move_rate + branch_rate(x))
    Choose choice ∈ {branch, move} with probabilities [prob_branch, 1 − prob_branch]
    if choice = branch then                               ▷ Branching occurs
        Sample num_offspring from offspring distribution
        Update: nums[x] ← nums[x] − 1 + num_offspring
        Update tot ← tot − 1 + num_offspring
        Update total_rate accordingly
    else                                                  ▷ Movement occurs
        Sample a dimension and a +ve or -ve direction
        Let x′ ← x + step_vector
        Update: nums[x] ← nums[x] − 1, nums[x′] ← nums[x′] + 1
        Update total_rate accordingly
    end if
    Append nums[0]/tot to checkpointed                    ▷ Storing a version of nums to plot
    if i ∈ histogram_values then
        Append current nums to timeline
    end if
end for
return nums, checkpointed, timeline
```

5.2 Simulating the binary branching process

We start by simulating the binary branching process (the MS model) in two cases. From [MS24] we see that we can expect non-localisation/diffusion in 3 dimensions when $\lambda_0 - \lambda < 0.66$. Thus for these plots I have chosen to use $\lambda_0 = 1, \lambda = 0.7$. I have plotted a histogram of the distances of particles from the origin, as well as the proportion of particles at the origin over time, for 10000 iterations.





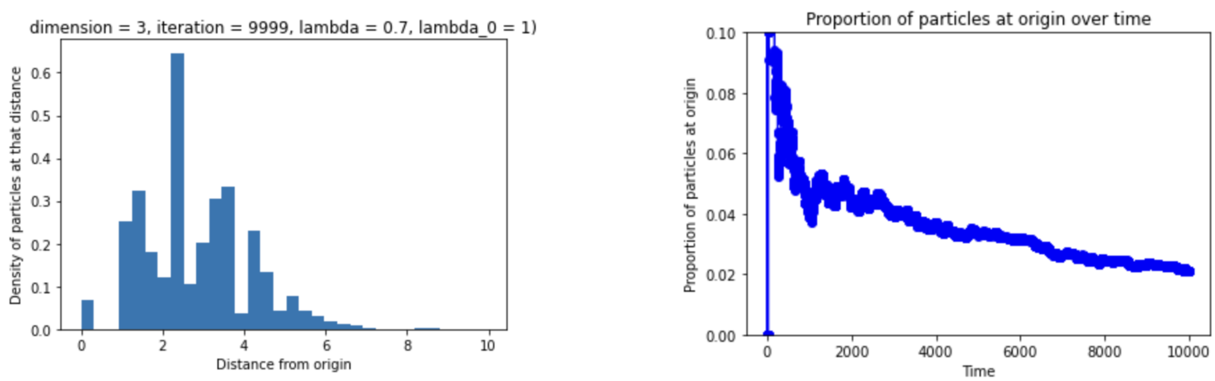


Figure 8: Distance histograms and proportion plot in non-localisation regime.

We can see the distribution slowly descending/diffusing, and the proportion of particles at the origin steadily tending to zero, this is consistent with diffusion/non-localisation behaviour

For the next case, again from [MS24] we see that we localisation behaviour in 3 dimensions when $\lambda_0 - \lambda > 0.66$. Therefore for these plots I have chosen to use $\lambda_0 = 1, \lambda = 0.2$. I have plotted a histogram of the distances of particles from the origin, as well as the proportion of particles at the origin over time, for 10000 iterations.

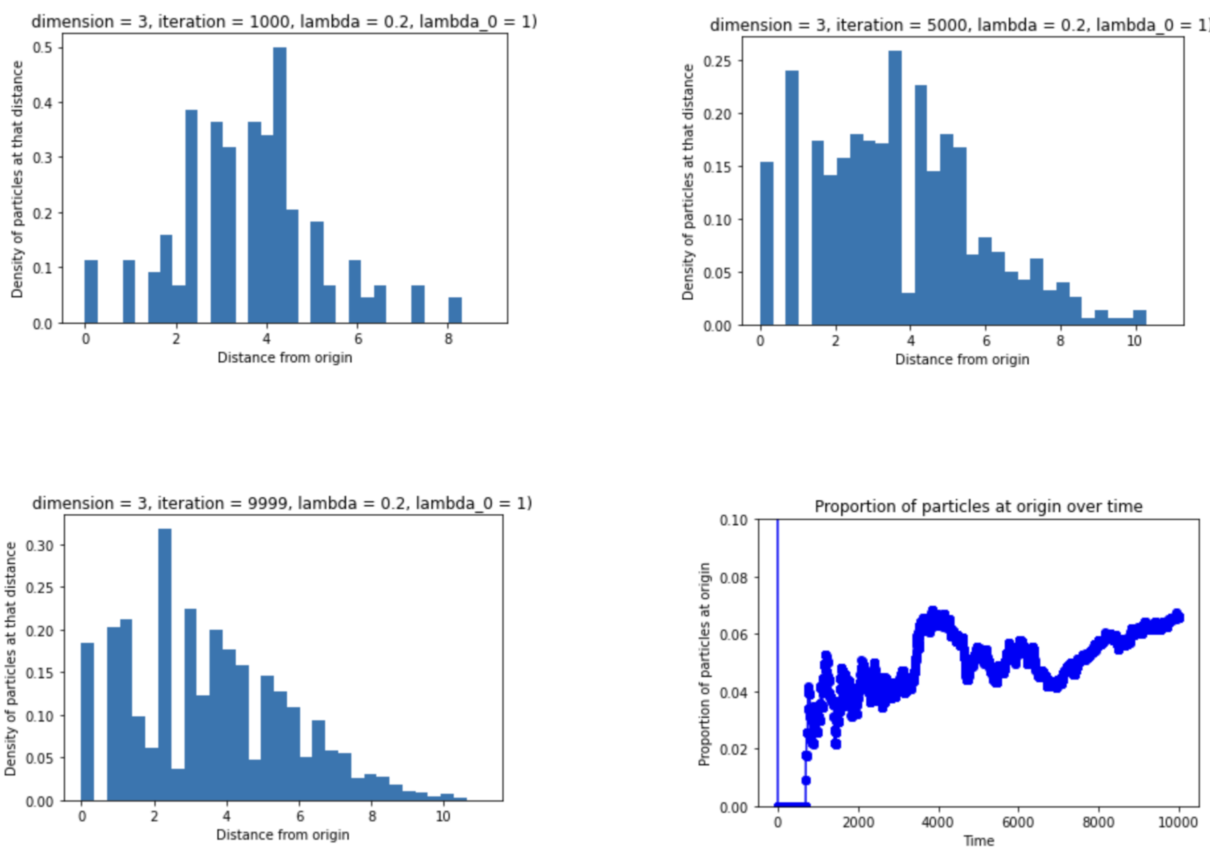


Figure 9: Distance histograms and proportion plot in localisation regime.

From these plots we can see from iterations 5000 and 9999 the distributions somewhat settling and the proportion of particles perhaps converging to a non-zero value, which is consistent with localisation behaviour. Simulations were also conducted in the subcritical and critical cases, but the particles simply moved away from the origin and thus the histograms and proportion plots were uninteresting. These plots can be recreated with the code provided in Appendix B.

6 Conclusion

In this project, we started by giving an introduction to renewal theory and spine techniques (Section 2), this was in order to explain the results and proofs of [DR12] (Section 3). We subsequently followed the ideas from [DR12] to prove some results about the asymptotics of the moments for the everywhere-branching catalytic random walk, i.e. when $\lambda > 0$. More precisely, in Section 4 we gave similar Feynman-Kac representations (Lemma 4.1), and proved the same result for M^1 in the subcritical case (Theorem 4.2) and in the critical case (Theorem 4.3). The supercritical case was attempted, but required an everywhere-branching version of Lemma 3.10. The result required a recursion which was made quite complex due to a binomial expansion, but with further work it might be possible. Finally, we conducted some simulations of the MS model which showed indications of the localisation and non-localisation modes of behaviour occurring.

A Feynman-Kac Formula

The background needed for this formula is too off-topic but I think it is useful to at least give an indication of where the Feynman-Kac representations get their functional form. To this end, I quote the following theorem from [Fle].

Theorem A.1 (Feynman-Kac). *Let $f : \mathbb{R}^d \to \mathbb{R}$, $k : \mathbb{R}^d \to \mathbb{R}^+$, and $g : [0,T] \times \mathbb{R}^d \to \mathbb{R}$ be continuous functions. Let $v : [0,T] \times \mathbb{R}^d \to \mathbb{R}$ be a function which is $C^{1,2}$ on $[0,T] \times \mathbb{R}^d$, such that*

$$-\frac{\partial v}{\partial t} + kv = \frac{1}{2}\Delta v + g; \quad \text{on } [0,T] \times \mathbb{R}^d,$$

$$v(T,x) = f(x); \quad x \in \mathbb{R}^d,$$

where Δ is the Laplace operator. Furthermore, suppose that

$$\max_{0 \le t \le T} |v(t,x)| + \max_{0 \le t \le T} |g(t,x)| \le K\mathrm{e}^{a|x|^2},$$

for all $x \in \mathbb{R}^d$ and some $K > 0$ and $0 < a < \frac{1}{2dT}$. Then we can write the solution v as

$$v(t,x) = \mathbb{E}_x \left[f(W_{T-t})\mathrm{e}^{-\int_0^{T-t} kW_s \mathrm{d}s} + \int_0^{T-t} g(t+\theta, W_\theta)\mathrm{e}^{-\int_0^\theta kW_s \mathrm{d}s}\mathrm{d}\theta \right].$$

This is a theorem which gives the representation of a solution, should it exist, in particular it does not show existence, just what the functional form must be. The reason this theorem is interesting is that the solutions to the **deterministic** PDE in the statement are written as expected values of **stochastic** expressions. In particular, the $(W_s)_{s\ge 0}$ represents a Brownian motion which is where the expression in the brackets gets its randomness from. Given that we can express the state of the DR branching process as a solution to a PDE, this is evidently useful. Through this type of theorem is how the expressions our Feynman-Kac representations of moments come about.

B Simulation Code

Below is all the code used for simulation and creation of plots. ChatGPT assisted in writing some parts of this code, but it has been checked for accuracy.

```
def run_simulation_changed(d,iterations=1000,branch_rate_0=1,
branch_rate_other=0.01,move_rate=1.0,
offspring_distribution={2:1.0}, histogram_values={0}):
```

```
nums = defaultdict(int)
# This will store the number of particles at any given position
origin = (0,)*d #tuple corresponding to d-dimensional origin
nums[origin] = 1 #initial particle at origin
tot = 1 #keeps track of total number of particles
checkpointed = [tot]
#stores checkpoint info,
#in this case keeps track of proportion of number of particles at origin
branch_rate = lambda x: branch_rate_0 if np.all(x == origin) else branch_rate_other
#function to give branching rate

#offspring distribution keys and probs (i.e the probs must sum to 1)
keys = list(offspring_distribution.keys())
probabilities = list(offspring_distribution.values())


timeline = [] #keeps track of nums over time
total_rate = (move_rate + branch_rate(origin)) * 1
#initial rate of 'something' happening due to particle starting at origin
for i in range(iterations):
    filtered_nums = {key: value for key, value in nums.items() if value != 0}
    #potential locations where 'something' could happen
    sample_from = np.array(list(filtered_nums.keys()))
    if len(sample_from) == 0:
        break
    #prob distribution for something happening to each 'x'
    sample_dist = np.array(
    [(move_rate + branch_rate(pos)) *
    nums[tuple(pos)] for pos in sample_from]) / total_rate

    #randomly choose point according to the above prob distribution
    x_index = np.random.choice(len(sample_from), size=1, p=sample_dist)[0]

    #getting the actual coordinate
    x = sample_from[x_index]

    #setting probabilities of branching vs moving (only two 'something's that can happen)
    prob_branch = branch_rate(x) / (move_rate+branch_rate(x))
    prob_move = 1 - prob_branch

    #making that choice with the above prob distribution
    choice = np.random.choice(['branch','move'], size=1, p=[prob_branch, prob_move])[0]

    if choice == 'branch':
        num_offspring = np.random.choice(keys, size=1, p=probabilities)[0]
        #number of offspring to replace the particle
        x = tuple(x) #converting x into a tuple
        nums[x] -= 1 # particle dies
        nums[x] += num_offspring
        # particle replaced by random number of offspring
        #tot += num_offspring - 1
        # total number of particles now num_offspring - 1 (as we gained num and lost one)
        total_rate -= (move_rate+branch_rate(x))
```

```
            # removed one x term from total_rate
            total_rate += (move_rate+branch_rate(x)) * num_offspring
            #update to total_rate, i.e. new summands
            tot -= 1
            tot += num_offspring
        else:

            dim_choice = np.random.choice(np.arange(d), size=1, p=[1.0/d]*d)[0]
            #of dimensions we can choose from is d, uniform prob over that
            up_down_choice = np.random.choice([1,-1], size=1, p=[0.5, 0.5])[0]
            #we can go in two directions for any given dimension

            new = np.zeros(d)
            new[dim_choice] = up_down_choice
            #gives a vector of the form (0,0,..., +- 1, ..., 0, 0)


            x = tuple(x)
            #needed to convert into right form for the code to work
            nums[x] -= 1
            # particle moves from x
            #particle moves to x+new
            nums[tuple(x+new)] += 1
            total_rate -= (move_rate+branch_rate(x))
            #removes x summand
            total_rate += (move_rate+branch_rate(x+new))
            #adds new summand (only one as no new particles just a movement)
        if tot > 0:
            checkpointed.append(float(nums[origin])/float(tot))
            #total proportion of particles at origin , at iteration i
        else:
            checkpointed.append(0.0)
        if i in histogram_values:
            timeline.append(nums.copy())

    return nums, np.array(checkpointed), timeline
    #the final config as a dictionary, and
    #the total proportion of particles


def plot_d_dim(timeline,histogram_values,branch_rate_other,branch_rate_0,d):

    for i, val in enumerate(timeline):
        # hist = defaultdict(int)
        # for key in val.keys():
        #     dist = np.linalg.norm(np.array(key))
        #     hist[dist] += val[key]
        # labels = list(hist.keys())
        # values = list(hist.values())

        # Plot histogram (bar chart)
        #plt.bar(labels, values)
        #i.e we add a distance _ times since val[key] is the
        #number of particles at position key
```

```
distances = [np.linalg.norm(np.array(key)) for key in val.keys() for _ in range(val[key])]
        plt.hist(distances, bins=30, density=True)

        # Add labels and title
        plt.xlabel('Distance from origin')
        plt.ylabel('Density of particles at that distance')
        plt.xlim(-0.5,max(distances)+1)
        plt.title(f'dimension = {d},
        iteration = {histogram_values[i]},
        lambda = {branch_rate_other}, lambda_0 = {branch_rate_0})')

        # Display plot
        plt.show()

def plot_proportion_0(checkpointed,big_axes=False):

    # Extract x and y values
    x_values, y_values = np.arange(len(checkpointed)), checkpointed

    # Create the plot
    plt.plot(x_values, y_values, marker='o', linestyle='-', color='b')

    # Labeling the axes
    plt.xlabel('Time')
    plt.ylabel('Proportion of particles at origin')
    if big_axes:
        plt.ylim(0,1)
    else:
        plt.ylim(0,0.1)
    plt.title('Proportion of particles at origin over time')
    plt.show()

#lambda = 0.7, lambda_0=1 has lambda_0-lambda = 30% < 66%, so we should
#see non-localisation/diffusion
d_2=3
branch_rate_0_2=1
branch_rate_other_2=0.7
move_rate_2=1.0
offspring_distribution_2={2: 1.0}
iterations_2 = 10000
histogram_values_2 = [2000,5000,10000-1]

_, checkpointed_2, timeline_2 = run_simulation_changed(d_2,iterations=iterations_2,
                                        branch_rate_0=branch_rate_0_2,
                                        branch_rate_other=branch_rate_other_2,
                                        move_rate=move_rate_2,
                                        offspring_distribution=offspring_distribution_2,
                                        histogram_values = histogram_values_2)

plot_d_dim(timeline_2,histogram_values_2,branch_rate_other_2,branch_rate_0_2,d_2)
plot_proportion_0(checkpointed_2)

#lambda = 0.2, lambda_0=1 has lambda_0-lambda = 80% > 66%,
#so we should see localisation
```

```
d_1=3
branch_rate_0_1=1
branch_rate_other_1=0.2
move_rate_1=1.0
offspring_distribution_1={2: 1.0}
iterations_1 = 10000
histogram_values_1 = [1000,5000,10000-1]

_, checkpointed_1, timeline_1 = run_simulation_changed(d_1,iterations=iterations_1,
                                branch_rate_0=branch_rate_0_1,
                                branch_rate_other=branch_rate_other_1,
                                move_rate=move_rate_1,
                                offspring_distribution=offspring_distribution_1,
                                histogram_values = histogram_values_1)

plot_d_dim(timeline_1,histogram_values_1,branch_rate_other_1,branch_rate_0_1,d_1)
plot_proportion_0(checkpointed_1)
```

References

[ABY98] Sergio Albeverio, Leonid Bogachev, and Elena Yarovaya. Asymptotics of branching symmetric random walk on the lattice with a single source. *Comptes Rendus de l'Académie des Sciences - Series I - Mathematics*, 326(9):975–980, 1998. In French: C. R. Acad. Sci. Paris, Série I.

[Car12] Francesco Caravenna. A note on directly riemann integrable functions, 2012.

[DR12] Leif Doering and Matthew I. Roberts. Catalytic branching processes via spine techniques and renewal theory. (arXiv:1106.5428), March 2012. arXiv:1106.5428 [math].

[DS10] L. Döring and M. Savov. An application of renewal theorems to exponential moments of local times. *Electronic Communications in Probability*, 15:263–269, 2010.

[Eri73] B. Erickson. The strong law of large numbers when the mean is undefined. *Transactions of the American Mathematical Society*, 54:371–381, 1973.

[Fel71] W. Feller. *An Introduction to Probability Theory and Its Applications. Vol. II.* John Wiley & Sons, Inc., New York, London, Sydney, 1971.

[Fle] Forrest G. Flesher. Final paper. `https://scholar.harvard.edu/files/forrestgflesher/files/final_paper_final.pdf`. Accessed: 2025-01-04.

[HR14] Simon C. Harris and Matthew I. Roberts. The many-to-few lemma and multiple spines. (arXiv:1106.4761), September 2014. arXiv:1106.4761 [math].

[HVT10] Y. Hu, V. A. Vatutin, and V. A. Topchii. Branching random walk in $\mathbb{Z}^4$ with branching at the origin only. arXiv preprint arXiv:1006.4769v1, 2010. Accessed: 2025-04-02.

[KS66] Harry Kesten and Bernt P. Stigum. A limit theorem for multidimensional galton–watson processes. *Annals of Mathematical Statistics*, 37:1211–1223, 1966.

[MO14] Kosto V. Mitov and Edward Omey. *Renewal Processes.* Springer, 2014.

[MS24] Cécile Mailler and Bruno Schapira. A localisation phase transition for the catalytic branching random walk. (arXiv:2412.11842), December 2024. arXiv:2412.11842 [math].

[Nev86] J. Neveu. Arbres et processus de galton-watson. *Annales de l'Institut Henri Poincaré Probabilités et Statistiques*, 22(2):199–207, 1986.

[Shi11] Zhan Shi. Random walks and trees. *ESAIM: Proceedings*, 31:1–39, 3 2011.

[Tru09] Timothy Trudgian. An introduction to tauberian theorems, 2009. Accessed: 2025-04-02.

[TV03] V. A. Topchii and V. A. Vatutin. Individuals at the origin in the critical catalytic branching random walk. *Discrete Mathematics and Theoretical Computer Science*, 6:325–332, 2003.

[Yar10] E. B. Yarovaya. The monotonicity of the probability of return into the source in models of branching random walks. *Moscow University Mathematics Bulletin*, 65(2):78–80, 2010.