Natural Selection in Spatially Distributed Populations



Jonathan Alexander Chetwynd-Diggle

Mathematical Institute

University of Oxford

This dissertation is submitted for the degree of

Doctor of Philosophy

Pembroke College

Trinity, 2019

Para Priscila

Eu te amo

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original. The results of Chapter 2 are based on joint work with Alison Etheridge and resulted in the paper Chetwynd-Diggle and Etheridge (2018). The results of Chapter 3 are based on joint work with Aleksander Klimek and resulted in the paper Chetwynd-Diggle and Klimek (2019). The results of Chapter 4 are based on joint work with Bjarki Eldon and Alison Etheridge. The joint work in Chapter 3 also formed a part of Aleksander Klimek's thesis which he submitted in Trinity 2018. This thesis is submitted to the University of Oxford for the degree Doctor of Philosophy.

Jonathan Alexander Chetwynd-Diggle

Trinity, 2019

Acknowledgements

First, I would like to thank my supervisor, Alison Etheridge, from whom I have learned a great deal, from basic analysis at the very start of my mathematical career to far more complex topics. I appreciate her taking the time to decipher my written work and for being an exceptional supervisor. In particular, I would like to thank her for many insightful, constructive and highly enjoyable discussions. Her kind and supportive mentorship have been formative to both my personal and professional life.

I also thank my colleagues within the Mathematical Institute. In particular, I would like to thank those with whom I have enjoyed many arguments across a whiteboard. Though these discussions have ranged from the mundane to the substantial I have always taken much from them. A special thanks to my office mates, who have made my working environment so enjoyable.

I would like to thank my whole family for their warmth and support. My parents have always been a great source of inspiration to me, including my choice of study (even though they always said there were other options). Lastly, to my partner, words can not express how thankful I am for your daily reminder that the most important things in life are not mathematically tractable.

Abstract

We introduce the spatial Lambda-Fleming-Viot model for natural selection within a population distributed on a spatial continua. We then analyse the behaviour of this model when one considers the appearance of a rare mutation. We investigate how the mechanism for reproduction will effect the resulting limiting behaviour. In particular, we not only show convergence to classical superBrownian motion but also, under suitable assumptions, to superBrownian motion with stable branching mechanism. We then look at the effect of a randomly fluctuating environment on the spatial Lambda-Fleming-Viot model. This is motivated by the changing effects of natural selection driven by the environment such as the El Niño effect. The final part of the thesis is devoted to an investigation of the genealogies within a population evolving under sweepstakes reproduction. This is related to the stable branching mechanism discussed earlier, but with a focus on the practical implications when sampling from a population one might expect to be modeled with a stable branching mechanism.

Table of contents

List of figures **xv**

1 Introduction **1**

1.1 Early population models 1

1.1.1 Fixed population models 4

1.1.2 Lineages, coalescents and duality 7

1.1.3 Spatial structure 10

1.1.4 Some spatial models 11

1.2 Thesis Overview 13

2 SuperBrownian motion and the spatial Lambda-Fleming-Viot process **15**

2.1 Background 15

2.2 Definitions and statement of results 19

2.2.1 SuperBrownian motion 19

2.2.2 The Spatial Lambda-Fleming-Viot Process 20

2.2.3 Main results 23

2.2.4 Heuristics 27

2.2.4.1 Fixed radius 27

2.2.4.2 Variable radii 28

2.3 Martingale characterisation of the process $\{X_t^N\}_{t\geq 0}$ 31

2.4 Identifying the limit 34

2.4.1 Spatial motion 34

2.4.2 Fixed radius case 36

2.4.3 Variable radius case 38

2.5 Tightness 41
2.5.1 Verification of Condition $2'$ of Corollary 2.5.16 43
2.5.2 Verification of Condition $1'$ of Corollary 2.5.16 46
2.5.2.1 Tightness of projections: fixed radius case 48
2.5.2.2 Tightness of projections: variable radius case 51
2.6 Proofs of Lemmas 2.4.13 and 2.4.14 58
2.6.1 A coupling of ancestral lineages 58
2.6.2 Coalescence Probabilities 62
2.7 Showing convergence 65
2.7.1 Proof of Theorem 2.2.6 65
2.7.2 Proof of Theorem 2.2.7 66

3 Spatial Lambda-Fleming-Viot Model in fluctuating environment 67
3.1 Introduction 67
3.1.1 Statement of main results 70
3.2 Scaling limits of the LFV - dynamics of the rare type 76
3.2.1 Scaling 81
3.2.2 Main result of this section 82
3.2.2.1 Projected model 85
3.2.3 Convergence of the non-spatial model 86
3.2.3.1 Intensity estimate 91
3.2.3.2 Neutral model - proof of Theorem 3.2.12 94
3.2.3.3 Selective model - proof of Theorem 3.2.13 107
3.2.3.4 Model with selection in fluctuating environment: proof of Theorem 3.2.10 113
3.3 SuperBrownian motion in a random environment 114
3.3.1 Definitions 114
3.3.2 Lookdown representation for BBMRE and SBMRE 117
3.4 Scaling limits of the SLFV - dynamics of the rare type 123
3.4.1 Scaling and statement of main results 125
3.4.2 Intensity estimate 127

3.4.3 Sketch of the proof of Theorem 3.1.4 . . . 130
3.5 Supplementary results . . . 133
3.5.1 Poisson random measures . . . 133
3.5.2 Markov Mapping Theorem . . . 135
3.5.3 Kurtz-Rodrigues' Martingale Lemma . . . 137
3.5.4 Proofs of Theorem 3.3.41 and Theorem 3.3.42 . . . 138

4 Multiple merger coalescents: truncated offspring distributions and large sample sizes 145

4.1 Introduction . . . 145
4.2 Background . . . 149
4.2.1 Mutation and the site frequency spectrum . . . 149
4.3 Beta coalescents and a modified Schweinsberg model . . . 153
4.3.1 Model definition . . . 156
4.4 Results . . . 157
4.5 Large sample size . . . 162
4.5.1 Large Sample Size in a modified Wright-Fisher Model . . . 162
4.5.2 Large Sample Size in the Lambda Coalescent . . . 164
4.6 Convergence proofs . . . 166
4.6.1 Key Lemmas . . . 166
4.6.2 Asymptotic rate of the coalescent: calculating c_N . . . 168
4.6.2.1 Preliminary lemmas . . . 168
4.6.2.2 Proof of Proposition 4.6.3 . . . 173
4.6.3 Convergence to the Kingman Coalescent . . . 177
4.6.4 Convergence to the Lambda coalescent . . . 178
4.7 Large sample size calculations . . . 180
4.7.1 Large Sample Size and the Kingman coalescent . . . 180
4.7.2 Large Sample Size and the Lambda Coalescent . . . 182

Bibliography 189

List of figures

Chapter 1

Introduction

1.1 Early population models

There has been a great deal of interest in how populations evolve over long time periods. One of the first mathematically rigorous analyses of this was by Galton and Watson in 1875. Francis Galton had been investigating the extinction of English family names, as the rate of extinction of aristocratic family names caused concern among Victorians at the time. Galton therefore proposed the question of what was the probability that a given family name would become extinct, which was replied to by the Reverend Henry William Watson, and resulted in the paper Galton and Watson (1875). The model they proposed can be formulated as a discrete time $\mathbb{N}$-valued Markov process:

$$\begin{cases} X_{n+1} := \sum_{i=1}^{X_n} \xi_i^n & \text{if } X_n \geq 1, \\ X_{n+1} := 0 & \text{if } X_n = 0, \end{cases}$$

where $\{\xi_i^n : i, n \in \mathbb{N}\}$ is a family of independent identically distributed N-valued random variables with finite mean, m, and variance, γ. The law of ξ_1^1 is referred to as the offspring distribution. In this model, X_n represents the number of male members in generation n of the family in question, and is called a Galton-Watson branching process. We will say that extinction has occurred by generation n if $X_n = 0$. The work of Galton and Watson had an algebraic error. However, if this is corrected they could show, for

$\gamma > 0$,

$$\begin{cases} \lim_{n\to\infty} \mathbb{P}[X_n = 0] = 1 & \text{if } m \leq 1, \\ \lim_{n\to\infty} \mathbb{P}[X_n = 0] < 1 & \text{if } m > 1. \end{cases}$$

This means extinction of a family name is guaranteed to occur if the expected number of legitimate sons a man will have is less than or equal to one.

This model only looks at the total number of individuals present in the population. However, one can also look at a version of the model which has many different types of individuals present. When considering individuals of multiple types, we will have a 'type space' of possible types which can be present in a given population. The multi-type Galton-Watson branching process with N types is the $\mathbb{N}^N$-valued Markov process $X_n := (X_n^1, \ldots, X_n^N)$, where X_n^i is the number of individuals if type 'i' in generation n. We let

$$X_{n+1}^j := \sum_{i:X_n^i \neq 0} \sum_{k=1}^{X_n^i} \xi_{i,j}^{k,n},$$

where $\{\xi_{i,j}^{k,n} : k, n \in \mathbb{N}\ i, j \in \{1, \ldots, N\}\}$ is a collection of independent $\mathbb{N}$-valued random variables with $\{\xi_{i,j}^{k,n} : k, n \in \mathbb{N}\}$ identically distributed for each pair $\{i, j\}$. The law of $\xi_{i,j}^{1,1}$ is the distribution of the number of type j individuals produced by an individual of type i. This can be seen as a version of the previous model when there are several distinct types of individuals in the population. The original model can be thought of as the special case when the number of types, N, is one. There has been interest in the limiting behaviour of this multi-type model. For instance, an early limiting theorem for this multi-type model was proved in Kesten and Stigum (1966) and a conceptual proof was later presented in Kurtz et al. (1997). The proof shows that under very minor conditions on the distributions of $\xi_{i,j}^{k,n}$, there exists a real number, ρ, a real vector, b, and a positive random variable, W with $\mathbb{E}[W] > 0$, such that

$$\lim_{n\to\infty} \frac{X_n}{\rho^n} = Wb \text{ a.s.}$$

In the following work we will write 'type space' to refer to the space of possible types in a population. In this case the type space is finite and can be represented by $\{1, \ldots, N\}$. In later work in Chapter 2 and Chapter 3 the type space will be assumed to have only two types.

Galton and Watson were interested in the total size of a population. In the multi-type model we then track the total number of a finite number of different types in a population. We will restrict ourselves to considering a haploid population in our models rather than a diploid population. We will also not consider sexual reproduction in our models. In the following we will continue to refer to individuals. However, these will no longer be members of species but instead represent a small section of DNA within an individual, referred to in the literature as a locus. This is because a small section of DNA is highly likely to remain intact over many generations and in particular, even in sexual reproduction, will come from only one parent. Therefore, even if a population undergoes sexual reproduction, if we are interested only in a single locus, then we can use models which assume the type present at a locus is determined only by the type at a single parent's locus.

The single and multi-type Galton-Watson branching processes are some of the earliest random population models with both being considered in discrete time. The complete analysis of these processes required the later development of far more sophisticated machinery, in particular the developments in stochastic analysis made by Kolmogorov, Lévy, Itô, Feller and many others in the first half of the 20th century. We draw attention to the semi-group characterisation of Markov processes and their generators. For an E-valued Markov process, X, and a continuous function, $f : E \to \mathbb{R}$, the semi-group, $\{T_t : t \geq 0\}$, associated with X is defined by

$$T_t f(x) := \mathbb{E}_x[f(X_t)],$$

where $\mathbb{E}_x$ denotes the expectation under the condition $X_0 = x$. We then define a generator $\mathcal{L}$ which acts on a subset of functions $f \in D(\mathcal{L}) \subset C(E)$ with the relation

$$\mathcal{L}(f)(x) := \lim_{t \to 0} \frac{T_t f(x) - f(x)}{t}, \tag{1.1.1}$$

when the limit exists for every x and defines a continuous function. A key result for such analysis is the Hille-Yosida theorem proved independently by Hille (1948) and Yosida (1948). This theorem gives conditions under which a semi-group gives rise to a generator and vice versa. These tools are primarily used for continuous time models and, using work by many previous authors, in particular Ethier and Kurtz (1986), one can prove powerful convergence results for Markov processes by looking only at the generator associated with them.

The above models have no inherent limit on the population sizes and in fact can show both exponential growth in the number of individuals and extinction events. However, it is not natural to see exponential growth over extended time periods. The total population of individuals will be limited by the availability of food and space. We therefore turn our attention to models which have a fixed total population.

1.1.1 Fixed population models

For an established species population sizes will tend to remain stable over long periods of time. Therefore, when looking over an extended time period it is often useful to look at the different types of individuals present in a fixed total population. As we are interested in the types present in an assumed very large population we will consider the frequencies of different types rather than the total number of individuals of each type. We now introduce the Wright-Fisher model.

Definition 1.1.1 (Wright-Fisher model)**.** *If the space of possible types in our population is $\mathcal{T}$ then the Wright-Fisher model is a discrete time $\mathcal{T}^N$-valued process, $X_n = (X_1^n, \ldots, X_N^n)$. In this model X_n represents a list of the types of the N individuals present in the population in generation n. In each generation all individuals present will die and are replaced by individuals who select parents uniformly at random from the previous generation. They take their type to be that of their parent.*

If we were to start with N distinct types at time 0 then an individual's type at time n will tell us which individual was its ancestor at time 0. Conversely, if we know an individual's ancestor at time 0 and that ancestor's type then we will know the type of the individual. We can therefore trace the ancestors of individuals rather than their

types. It is for this reason that we are often interested in the most recent common ancestor (MRCA) of a pair of individuals. This way of thinking about the Wright-Fisher model gives rise to a dual process which can help analyse the model's behaviour. We will discuss this further in Section 1.1.2. We now present some heuristic limiting arguments following the work of Etheridge (2011).

In the model above, if we pick any two individuals at random at a given generation, $n > 0$, then the probability that they have the same parent in generation $n-1$ will be $\frac{1}{N}$. Therefore the number of generations since the MRCA will have a geometric distribution. The probability that the most recent ancestor was exactly m generations ago is therefore given by $\frac{1}{N}\Big(1-\frac{1}{N}\Big)^{m-1}$. We now give a rough calculation for the limiting behaviour of the Wright-Fisher model. The appropriately scaled limit of the Wright-Fisher model is called the Wright-Fisher diffusion. We look at a formal way of obtaining the form of this diffusion using infinitesimal generators. We recall (1.1.1) and look at a population with only two different genetic types. The above suggests that if we have N individuals in the Nth level of the scaling then by time t we would like to see $\lfloor tN \rfloor$ generations. We will consider a population with only two types: type 1 and type 0. We denote the proportion of type 1 in the population at time t in the Nth level of rescaling by $w^N(t)$. By looking at one generation we can see that

$$\mathbb{E}[f(w^N(1/N))] = \sum_{k=0}^{N} \binom{N}{k} w^N(0)^k \Big(1-w^N(0)\Big)^k f(k/N).$$

From standard properties of the binomial distribution, it follows that

$$\mathbb{E}[w^N(1/N) - w^N(0)] = 0,$$

and

$$\mathrm{var}(w^N(1/N) - w^N(0)) = \frac{1}{N} w^N(0)\Big(1-w^N(0)\Big).$$

Then with an abuse of notation and recalling (1.1.1) we see, by using the Taylor expansion, that we would expect a limit to satisfy

$$\begin{aligned}\frac{\mathbb{E}[f(w(1/N)] - w(0)]}{1/N} \sim & N\Big(\mathbb{E}[w^N(1/N) - w^N(0)]f'(w(0)) \\ & + \mathbb{E}[(w^N(1/N) - w^N(0))^2]f''(w(0))\Big) \\ \sim & w(0)(1-w(0))f''(w(0)).\end{aligned}$$

This suggests a limiting process will satisfy the SDE

$$\mathrm{d}w_t = \sqrt{w_t(1-w_t)}\mathrm{d}B_t,$$

where B_t is a Brownian motion. The solution to this SDE is the Wright-Fisher diffusion.

The Wright-Fisher model is a specific example of the more general fixed population model introduced in Cannings (1974) and Cannings (1975). In a Cannings model, within a total population of N individuals an individual 'i' will have $v_{i,N}$ offspring in the next generation. In order to maintain a fixed population size it is assumed $v_{1,N} + \cdots + v_{N,N} = N$. We also assume that for each N $(v_{1,N}, \ldots, v_{N,N})$ are exchangeable and identically distributed. The Wright-Fisher model is one of the simplest Cannings models and is often the first case analysed in the literature. The other simple model often used as a test case is the Moran model. The Moran model is often considered in continuous time but we include a definition of the discrete time Moran model as then it appears in the form of a Cannings model.

Definition 1.1.2 (Discrete time Moran model)**.** *If the space of possible types in our population is $\mathcal{T}$ then the Moran model is a discrete time $\mathcal{T}^N$-valued process, $X_n = (X_1^n, \ldots, X_N^n)$. In this model X_n represents a list of the types of the N individuals present in the population in generation n. In each generation two individuals present will be chosen uniformly at random to die. They are replaced by two individuals who select parents uniformly at random from the two individuals who died. They take their type to be that of their parent.*

A key difference between the Wright-Fisher and Moran models comes in the different scalings associated with them. In particular we can see that, in a total population size

of N, the probability that a given two individuals share the same parent in the previous generation is $1/\binom{N}{2}$. Therefore, in the Nth level of the scaling by time t we would like to see $\lfloor \binom{N}{2} t \rfloor$ generations.

As in the Wright-Fisher case we can see

$$\mathbb{E}\left[w^N\left(1 \Big/ \binom{N}{2}\right) - w^N(0) \right] = 0,$$

and

$$\begin{aligned} \operatorname{var}\left(w^N\left(1 \Big/ \binom{N}{2}\right) - w^N(0) \right) =& \frac{1}{N^2} 2 \frac{w^N(0)(N - N w^N(0))}{N-1} \\ =& \frac{2}{N(N-1)} w^N(0)(1 - w^N(0)). \end{aligned}$$

Therefore, exactly as in the Wright-Fisher case we would expect a limiting process to satisfy the SDE

$$\mathrm{d}w_t = \sqrt{w_t(1-w_t)}\mathrm{d}B_t.$$

1.1.2 Lineages, coalescents and duality

It is often convenient to consider population models through their lineages. Commonly lineages are considered in order to use a duality argument as we will in Chapter 2. A lineage of an individual is a record of the individual's ancestors. We note that under the models discussed above each individual has only one parent. Also, as generations do not overlap, at any point in time the population will contain exactly one ancestor for each individual. When the ancestors at time t in the past of two individuals are the same we say the lineages have coalesced by time t. Once two lineages have coalesced then, under these models, they will remain coalesced. In the Wright-Fisher model the probability that three or more lineages coalesce at the same time will be $\mathcal{O}(1/N^2)$. Therefore, when N is large, if we are tracing a finite number of lineages they will have all coalesced into one lineage through pairwise coalescences before we see a coalescence involving three or more lineages. If we consider tracing k lineages, in a population of size N with N generations in each unit of time as $N \to \infty$, a pair of traced lineages will coalesce

exponentially at rate $\binom{k}{2}$. This is because there are $\binom{k}{2}$ pairs of lineages that each have equal chance of coalescing and the choice of scaling will mean that a given two lineages coalesce at rate one.

We now introduce Kingman's coalescent, which was first described in Kingman (1982c). This is a Markov process taking values in partitions on $\mathbb{N}$, which we denote by $\mathcal{P}(\mathbb{N})$.

Definition 1.1.3 (Kingman's coalescent)**.** *For $\pi = \{\pi_t : t \geq 0\}$, a $\mathcal{P}(\mathbb{N})$-valued stochastic process, we denote by π^n its restrictions to $\{1, \ldots, n\}$, defined by*

$$\pi_t^n := \{p \cap \{1, \ldots, n\} \ : \ p \in \pi_t\} \setminus \{\emptyset\}.$$

Note that π^n is a $\mathcal{P}(\{1, \ldots, n\})$-valued process. Kingman's coalescent is defined to be the $\mathcal{P}(\mathbb{N})$-valued process π such that for every $n \in \mathbb{N}$, π_t^n makes transitions to $\{A_1, \ldots, A_k\} \in \mathcal{P}(\{1, \ldots, n\})$ at rate 1 whenever $\{A_1, \ldots, A_k\}$ can be formed by combining exactly two elements of π_t^n.

Our previous heuristic arguments show that, under the scalings discussed, Kingman's coalescent will be the limit of both the Wright-Fisher coalescent and the Moran coalescent. As Kingman's coalescent is a limit of both the Wright-Fisher coalescent and the Moran coalescent it is used as a model for lineages present in large populations. However, interest in models where more than three lineages may coalesce at the same time gave rise to the study of Λ-coalescents.

Definition 1.1.4 (Λ-coalescent)**.** *The Λ-coalescent is defined as in Kingman's coalescent except that its restriction to $\{1, \ldots, n\}$ makes transitions to $\{A_1, \ldots, A_k\}$ at rate $\beta_{m,j}$ if $\{A_1, \ldots, A_k\}$ can be formed by combining exactly j elements of π_t^n and there are m elements present in π_t^n. This process does not exist for arbitrary choices of the rates. It was proved in Pitman (1999) that it will exist if and only if $\beta_{m,j}$ can be expressed by*

$$\beta_{m,j} = \int_{[0,1]} u^{j-2}(1-u)^{m-j}\Lambda(\mathrm{d}u), \tag{1.1.2}$$

where Λ is a finite measure on $[0, 1]$.

We note that one could attempt to consider a more general Λ-coalescent model where $\beta_{m,j}$ does not have this form. However any such model would be expected to follow what is known as a consistency condition

$$\beta_{m,j} = \beta_{m+1,j} + \beta_{m+1,j+1},$$

which in fact enforces (1.1.2), Pitman (1999). This model can be further generalised when we allow not only multiple lineages to coalesce but also multiple coalescence events to occur simultaneously. This has led to the study of Ξ-coalescents which are also characterised through a finite measure Ξ in Schweinsberg (2000). These have also been studied as limits of Cannings models for example in Möhle and Sagitov (2001). We will follow the notation in Schweinsberg (2003).

Definition 1.1.5 (Ξ-coalescent)**.** *The Ξ-coalescent is also defined through its restrictions to* $\{1, \dots, n\}$*. It will make transitions to* $\{A_1, \dots, A_k\}$ *at rate* $\lambda_{b;k_1,\dots,k_r,s}$*, where* $b = s + \sum_i k_i$*, if* π_t^n *contains* b *elements and* $\{A_1, \dots, A_k\}$ *can be formed by, for each* i*, combining exactly* k_i *distinct elements of* π_t^n*.*

In species that produce extremely high numbers of offspring, such as Pacific oysters which were analysed in Eldon and Wakeley (2006), Kingman's coalescent is not a good approximation for the lineages. In several marine species, individuals will have extremely high numbers of offspring and large proportions of the next generation can be descended from a single mating pair. This is often referred to as sweepstakes reproduction in the literature. This means that one will occasionally see a large decrease in genetic diversity over a single generation without unusual external factors. Therefore, the Λ-coalescent, or Ξ-coalescent, is of particular biological interest in these species. In Birkner et al. (2013b) they analyse if genealogies can be described by β-coalescents, which are a special class of Λ-coalescents. We will investigate a model in Chapter 4 where both Kingman's coalescent and β-coalescents will be the limit of the model we are interested in.

We note that taking $\Lambda = \delta_0$ corresponds to Kingman's coalescent. The Λ-coalescent is dual to the Λ-Fleming-Viot process, which takes its values in probability measures on the type space. This process was first introduced in Donnelly and Kurtz (1999) then later in Bertoin and Le Gall (2003) where it is called a generalised Fleming-Viot

process. When $\Lambda(\{0\}) = 0$ there is a natural way of thinking of the Λ-Fleming-Viot process. Suppose you have a Poisson point process on $[0, \infty) \times (0, 1]$ with intensity measure $\mathrm{d}t \otimes u^{-2}\Lambda(\mathrm{d}u)$. We will denote the Λ-Fleming-Viot process at time t by w_t and will call the points of the Poisson point process reproduction events. A point in this Poisson point process then represents the time and the size of a reproduction event. At a reproduction event (t, u) we select a type according to w_{t-} and then kill a proportion u of the population and replace them with copies of the selected type. In particular if e is the type selected according to the probability distribution w_{t-} then

$$w_t = (1 - u)w_{t-} + u\delta_e.$$

The duality can be seen by tracing a finite number of lineages backwards in time. When we hit an event (t, u) we independently mark each lineage that we are tracing with probability u and then merge all the marked lineages together into a single lineage.

1.1.3 Spatial structure

These models all assume that individuals within the population will interact with each other equally. However, populations are typically spread over a large area. In general, individuals will be more closely related to those geographically nearer to them. One way of modelling this is to assume that a population is subdivided into a number of 'demes' where reproduction takes place within a deme and migration takes place between demes. A deme represents a portion of the population where all individuals present will reproduce with each other but will rarely reproduce with those outside their own deme. One example is the stepping stone model introduced by Kimura and discussed in Kimura and Weiss (1964). In some circumstances demes can be natural, but it is often more desirable to have a model for analysis in spatial continua.

Interaction between demes is modelled as migration and this can be dependent on the spatial distance between them. In order to keep the population constant in each deme, the migration must be modelled by swapping individuals between demes. Typically one will model demes in a lattice structure with migration only between adjacent lattice points. If one imagines following lineages in a population, they will

coalesce according to one of the processes described above if two or more lineages are within the same deme. Lineages will independently make jumps to adjacent demes which, if all demes have the same size, will be at a constant rate. Tracing a single lineage on the lattice, we see a simple random walk with exponentially distributed times between jumps. If we were to scale the migration rate and the distance between demes in the appropriate way, in the limit lineages will move according to Brownian motion. Then when the lineages are in same location they will coalesce according to one of the methods discussed above. The approximation of an individual's movement by Brownian motion is very natural as Brownian motion is thought of as the typical continuous space approximation for random movement, particularly as a limit of simple random walks. However, in two dimensions Brownian motion is not point recurrent and three and higher dimensions Brownian motion is non-recurrent. If lineages move in spatial continua according to Brownian motion we will not see lineages in the same location and therefore no interaction between lineages in limiting arguments in dimensions two and higher.

1.1.4 Some spatial models

In Chapter 2 we will focus on the analysis of the Spatial Lambda-Fleming-Viot model which will be introduced within that chapter. In order to frame this result we present some other models which have been analysed in the literature.

We first introduce the Voter model analysed in Cox et al. (2000). This is a Markov process based on the lattice, $\mathbb{Z}^d$. Each point on the lattice will hold an opinion which for simplicity is often assumed to be in $\{0, 1\}$. For instance when considering electorates this can represent a preference for the incumbent or a challenger. Each individual on the lattice has an independent exponential clock and at rate one will adopt the opinion of another individual in the lattice. The spatial structure is given through the 'voting kernel', a function $k : \mathbb{Z}^d \times \mathbb{Z}^d \to [0, \infty)$. The 'voting kernel' is chosen so that $k(x, y) = k(y, x) = k(0, y - x)$ and $\sum_{x \in \mathbb{Z}^d} k(0, x) = 1$. When the exponential clock of an individual at position x rings the individual will, with probability $k(x, y)$ adopt the opinion of the individual at position y. A key point in the analysis in Cox et al. (2000) comes through looking at the voter model in a different way. They consider the spread

of one type of voter by considering points on the lattice as spreading their 'ideology' rather than points adopting ideologies of other points. This allows them to use the rich literature of branching random walks in their analysis.

The next model we introduce has clearer analogies with population models. This is the contact process which was introduced in Harris (1974) and analysed in Van der Hofstad and Sakai (2010). The contact process is similar to the voter model. However, the two types of individuals are fundamentally different. Each point of the lattice is either infected or healthy and an infected individual at point y will infect a healthy site at point x at rate $\lambda k(y-x)$. We note that k is a kernel with the same constraints as in the voter model and $\lambda > 0$ is an infection rate. We will also have infected points recover at rate 1. We note that although one of these two models has exchangeable types and the other has distinct types they are both shown to have the same scaling limit.

The third related lattice model is a model for competition between species. The model was introduced in Neuhauser and Pacala (1999). This is a stochastic spatial version of the Lotka-Volterra model, also known as predator-prey equations

$$\begin{aligned}\frac{\mathrm{d}x}{\mathrm{d}t} =& \alpha_1 x - \beta_1 x^2 - \gamma_1 xy \\ \frac{\mathrm{d}y}{\mathrm{d}t} =& \alpha_2 y - \beta_2 y^2 - \gamma_2 xy,\end{aligned}$$

introduced by Volterra (1928) and Lotka (1932). The model introduced by Neuhauser and Pacala (1999) is based on the lattice as above. They consider an individual at each point on the lattice. At exponentially distributed times the individual dies and is replaced by a new individual who takes their type depending on the surrounding configuration of other individuals. It can be seen that the voter model is a particular class of this Lotka-Volterra type model. Work by Cox and Perkins (2005) extended the the previous results with the voter model to this more general class of models.

The final model has clear analogies with the contact process. This is a branching random walk, in which a number of individuals live on the lattice and each move according to an independent simple random walk on $\mathbb{Z}^d$. In the simplest case, at exponential times each individual present will die and will either be replaced by two individuals at the same location or will not be replaced with equal probability.

1.2 Thesis Overview

This thesis will be structured as follows. In each subsequent chapter we will first introduce a specific problem along with some additional material relevant to the chapter. We will then set out the main results of the chapter followed by the more detailed proofs in later sections. In Chapter 2 we present some joint work with Alison Etheridge which resulted in a paper, Chetwynd-Diggle and Etheridge (2018). This work looks at the limiting behaviour of a spatial population model when considering the appearance of a mutation in the population carried by a very small number of individuals. As discussed in Section 1.1.1 the dynamics of a subpopulation of individuals of a rare type in a Wright-Fisher model can be approximated by a Feller branching process. In Chapter 2 we establish an analogue of that result for a spatially distributed population whose dynamics are described by a spatial Lambda-Fleming-Viot process (SLFV), which we will introduce in Section 2.2.2. The subpopulation of rare individuals is then approximated by a superBrownian motion, which we define in Section 2.2.1. This result mirrors Cox et al. (2000), where it is shown that when suitably rescaled, sparse voter models converge to superBrownian motion. The main novelty of this Chapter is that by choosing the dynamics of the SLFV appropriately we can recover superBrownian motion with stable branching in an analogous way. This is a spatial analogue of (a special case of) results of Bertoin and Le Gall (2006), who show that the generalised Fleming-Viot process, when suitably rescaled, converges to a continuous state branching process with stable branching mechanism.

In Chapter 3, we present joint work with Aleksander Klimek with resulted in a paper, Chetwynd-Diggle and Klimek (2019). In this work we looked at an extension of the work in Chapter 2 where we consider the same model, but, with the introduction of a rapidly changing environment. Although at first glance a very simple extension to the work in Chapter 2 using results from Biswas et al. (2018), the work required the use of much more complicated machinery in order to complete the proof. In Section 3.1 we will introduce the Lambda-Fleming-Viot and spatial Lambda-Fleming-Viot with fluctuating selection. This environment will determine which of the two types of individuals present in the population has a selective advantage. We will again investigate the behaviour

of an establishing mutation under this Lambda-Fleming-Viot model and show that under a suitable scaling it converges to the Feller diffusion in a random environment. In Section 3.2 we will prove this, non-spatial, result. The proof is broken down into several parts. We will have to reprove some of the results in Chapter 2, then extend those techniques to accommodate the effect of the environment. We then extend these results to a population that is distributed across a spatial continuum in Section 3.4. In this setting the scaling limit is the SuperBrownian motion in a random environment, which we introduce in Section 3.3. The scaling results for the behaviour of the rare allele in this Chapter are achieved via particle representations, which belong to the family of 'lookdown constructions'. Lookdown constructions were introduced in Donnelly and Kurtz (1996) and Donnelly and Kurtz (1999) but we will be using techniques and a construction more similar in flavor to those presented in Kurtz and Rodrigues (2011).

In Chapter 4, we present joint work with Alison Etheridge and Bjarki Eldon where we looked at a modification of the population model analysed in Schweinsberg (2003). The Chapter was motivated by recent surveys of Icelandic cod. We will consider the impact of the large sample sizes involved in these surveys by considering a model for gene genealogies in a highly fecund haploid population with sweepstakes reproduction. The model will be a modification of Schweinsberg's model of genetic reproduction from Schweinsberg (2003), which incorporates a cut-off to the random number of juveniles contributed by a given individual. We introduce the model in Section 4.3. Depending on how the cut-off behaves relative to the population size, we obtain the Kingman-coalescent, a truncated Beta-coalescent, or the original Beta-coalescent as in Schweinsberg (2003). The truncated Beta-coalescent and Beta-coalescent are both examples of Λ-coalescents, which often appear as limits of population models. We use the cutoff to estimate the error of the coalescent approximation. The error estimates reveal that convergence can be very slow, and that a very small sample size can be sufficient to invalidate convergence, especially if the cut-off is of a specific form. In Section 4.5 we will look at a common statistical tool and show that the impact on inference will only be noticed at larger sample size than that at which convergence breaks down.

Chapter 2

SuperBrownian motion and the spatial Lambda-Fleming-Viot process

2.1 Background

Our aim in this chapter is to establish a relationship between two, at first sight, very different classes of measure-valued processes. The first, the spatial Lambda-Fleming-Viot processes, is a collection of models for the evolution of frequencies of different genetic types in a population that is dispersed across a spatial continuum. The second is the (finite and infinite variance) superBrownian motions. Our motivation is two-fold. On the one hand, we add to the panoply of processes that converge to superBrownian motion; on the other, we address a question of some interest in population genetics: how does the frequency of a rare neutral mutation evolve in a spatially distributed population?

SuperBrownian motion, or the Dawson-Watanabe superprocess, was introduced independently by Watanabe (1968) and Dawson (1975) as a continuous time and space approximation to systems of branching Brownian motions. In this way it can be thought of as a spatial analogue of the Feller diffusion approximation to critical (or near-critical) Galton-Watson branching processes. We shall recall its definition in Section 2.2.1 below.

In addition to the huge literature exploring the rich mathematical structure of superBrownian motion, over the last two decades an increasing body of evidence has emerged that it is a universal scaling limit of critical interacting particle systems above a critical dimension. It has been obtained as a limit of lattice trees (above 8 dimensions, e.g. Holmes (2008)), oriented percolation (above 4 dimensions, Van der Hofstad and Slade (2003)), the contact process (above 4 dimensions, e.g. Van der Hofstad and Sakai (2010)), the voter model (in two or more dimensions, e.g. Cox et al. (2000)) and the Lotka-Volterra model (Cox and Perkins (2005)). By changing the range of the interaction with the scaling, one can also obtain it from the contact process in lower dimensions (Cox et al. (1999)). These analyses prove convergence of finite-dimensional distributions; Van der Hofstad et al. (2017) provide a tightness criterion that allows the extension to convergence on path space and apply it to the example of sufficiently spread out lattice trees above 8 dimensions. We also refer to that paper for a more complete list of references.

For populations that are not spatially distributed, one classically models frequencies of different genetic types (usually refered to as alleles) through a Wright-Fisher or a Cannings model. Suppose that we are interested in the proportion of individuals of a particular type, that we shall call type 1. Under the Wright-Fisher model, when suitably scaled, this proportion converges to the Wright-Fisher diffusion. If type 1 is rare, the absolute number of type 1 individuals evolves approximately according to a branching process which, under the same scaling, converges to a Feller diffusion. This branching process approximation for the rare type has been used extensively in the population genetics literature and so it is natural to try to establish analogous results for spatially distributed populations.

In one spatial dimension, the Wright-Fisher diffusion has a stochastic pde counterpart:

$$\mathrm{d}w_t(x) = \frac{1}{2}\Delta w_t(x)\mathrm{d}t + \sqrt{\frac{1}{K}w_t(x)(1-w_t(x))}W(\mathrm{d}t,\mathrm{d}x), \tag{2.1.1}$$

where $w_t(x)$ denotes the proportion of the population at spatial position x at time t that is of type 1, K is the local population density, and $W(\mathrm{d}t,\mathrm{d}x)$ is a space time white

noise. Formally at least, if type 1 is rare, this reduces to

$$\mathrm{d}w_t(x) = \frac{1}{2}\Delta w_t(x)\mathrm{d}t + \sqrt{\frac{1}{K}w_t(x)}W(\mathrm{d}t, \mathrm{d}x),$$

and if we set $X_t = Kw_t$ to recover absolute numbers rather than proportions, the type 1 population is modelled by

$$\mathrm{d}X_t(x) = \frac{1}{2}\Delta X_t(x)\mathrm{d}t + \sqrt{X_t(x)}W(\mathrm{d}t, \mathrm{d}x),$$

which is the stochastic pde governing the density with respect to Lebesgue measure of the (finite variance) superBrownian motion, and so it is certainly reasonable to hope to describe establishment of rare alleles in one dimensional populations using superBrownian motion.

In dimensions two and higher, equation (2.1.1) has no solution and so we need an alternative approach to modelling allele frequencies in higher dimensional spatial continua. The obstructions to finding such an approach, often refered to as 'the pain in the torus', are well documented. We refer to Barton et al. (2013) for a survey. The spatial Lambda-Fleming-Viot process (SLFV) introduced in Etheridge (2008), overcomes the pain in the torus to provide a class of models for allele frequencies in populations distributed across spatial continua of any dimension. The first rigorous construction is in Barton et al. (2010). The SLFV can be thought of as the spatial counterpart of the 'generalised Fleming-Viot process' (that we shall refer to as the Lambda-Fleming-Viot process in what follows) of Bertoin and Le Gall (2003) and, just as for their model, comes with a consistent 'backwards in time' model (a spatial analogue of the Lambda-coalescent) for the genealogies describing relatedness between genes in individuals sampled from the population. We recall the definition of the process in Section 2.2.2.

As a special case of the results in Etheridge et al. (2014), in one spatial dimension one can recover (2.1.1) as a scaling limit of a particular SLFV. The corresponding scaling in higher dimensions leads to the (deterministic) heat equation. This can perhaps best be understood as a 'law of large numbers' effect. In particular, the initial conditions

taken in that paper don't correspond to 'rare' alleles. As we shall see, we can recover superBrownian motion from the SLFV in arbitrary spatial dimensions, but only if we take a sufficiently 'sparse' initial condition. This should of course be compared to the results of Cox et al. (2000), who recover superBrownian motion from sparse voter models and our analysis in the finite variance case owes a great deal to that paper. We should also mention the work of Freeman (2010), in which he introduces a very close relative of the SLFV, which he calls a 'bursting process', on $\mathbb{Z}^d$ and shows that for $d \geq 3$, started from sparse initial conditions and suitably scaled, that process too converges to a superBrownian limit.

In the discussion up to this point we have (implicitly) considered the finite variance superBrownian motion. Where our work diverges from the body of work described above is that we are also able to obtain superBrownian motions with stable branching mechanisms from particular choices of the SLFV. Such superprocesses are the spatial analogue of the continuous state branching processes sometimes known as stable branching processes. In Birkner et al. (2005a), it is shown that the special class of Lambda-Fleming-Viot processes that are dual to the so-called Beta-coalescents can be obtained as time-changed stable branching processes, revealing a deep connection between the two classes of processes. Bertoin and Le Gall (2006) show that in much the same way as the Feller diffusion describes evolution of a rare allele in a population evolving according to the Wright-Fisher diffusion, stable branching describes the evolution of a rare allele under this Lambda-Fleming-Viot process (see Lambert and Schertzer (2016) for a 'backwards in time' analogue). We provide the 'back of the envelope' calculation that explains Bertoin and Le Gall's result in Section 2.2.4. What is more surprising is that we can extract a superBrownian motion with stable branching mechanism from a sequence of rescaled SLFVs. First, the conditions on the 'Lambda'-measure under which we can construct the SLFV are more restrictive than those under which we can construct the (non-spatial) Lambda-Fleming-Viot process. Second, the spatial motion of individuals in the SLFV is intricately connected to the reproduction mechanism, yet we are trying to produce a limit in which spatial motion is continuous and reproduction is discontinuous. On the other hand, in Etheridge et al. (2014) the analogue of (2.1.1) with the Laplacian replaced by the generator of a symmetric stable process is obtained as a

scaling limit of an SLFV. In that case, in the limit the spatial motion is discontinuous and the reproduction mechanism continuous.

The rest of this chapter is laid out as follows. In Section 2.2 we remind the reader of the definitions of the superBrownian motion and the SLFV before stating our main results. We also give a heuristic explanation of our results. In Section 2.3 we provide martingale characterisations of the scaled SLFVs, from which, in Section 2.4, we formally identify the limiting objects, deferring tightness to Section 2.5, and the proof of some key estimates to Section 2.6. The proof of convergence follows in Section 2.7.

2.2 Definitions and statement of results

Before stating our results in Section 2.2.3 below, we fix notation and define our two classes of processes.

2.2.1 SuperBrownian motion

We shall characterise superBrownian motion through a martingale problem. It is convenient to use distinct formulations in the finite and infinite variance cases. For an introduction to superprocesses and, in particular, their construction as scaling limits of branching particle systems, we refer to Dawson (1993), Etheridge (2000) and Perkins (2002).

A complete filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ will be implicit throughout. We write $\mathcal{M}_F(\mathbb{R}^d)$ for the space of finite measures on $\mathbb{R}^d$, equipped with the topology of weak convergence, and $C_0^k(\mathbb{R}^d)$ for the space of k times differentiable functions $\phi : \mathbb{R}^d \to \mathbb{R}$, vanishing at infinity, and such that ϕ and its derivatives up to kth order are bounded with norm

$$\|\phi\|_{C^k} = \max_{0\leq l\leq k} \left\| \frac{\partial^l \phi}{\partial x^l} \right\|_\infty .$$

Definition 2.2.1 (Finite variance superBrownian motion)**.** *The* finite variance superBrownian motion *is the unique $\mathcal{M}_F(\mathbb{R}^d)$-valued Markov process $\{X_t\}_{t\geq 0}$ with continuous sample paths such that for each non-negative $\phi \in C_0^3(\mathbb{R}^d)$, the process*

$$M_t(\phi) := \langle X_t, \phi\rangle - \langle X_0, \phi\rangle - \int_0^t \langle X_s, \frac{m}{2}\Delta\phi\rangle \mathrm{d}s,$$

is a continuous, square integrable martingale with quadratic variation given by

$$\langle M(\phi)\rangle_t = 2\kappa \int_0^t \langle X_s, \phi^2\rangle \mathrm{d}s,$$

where $m, \kappa > 0$ are constants.

We shall not consider the most general possible superBrownian motions. Instead, we restrict ourselves to those that arise as scaling limits of branching Brownian motions with offspring distribution in the domain of attraction of a stable law. These are naturally parametrised by a parameter $\beta \in (0, 1)$ (with $\beta = 1$ corresponding to the finite variance case).

Definition 2.2.2 (SuperBrownian motion with stable branching law)**.** *The* super-Brownian motion with stable branching law of parameter $\beta \in (0, 1)$ *is the unique $\mathcal{M}_F(\mathbb{R}^d)$-valued Markov process $\{X_t\}_{t\geq 0}$ with càdlàg sample paths such that for each non-negative $\phi \in C_0^3(\mathbb{R}^d)$, the process*

$$M_t(\phi) := \exp(-\langle X_t, \phi\rangle) - \exp(-\langle X_0, \phi\rangle) - \int_0^t \langle X_s, -\frac{m}{2}\Delta\phi + \kappa\phi^{1+\beta}\rangle \exp(-\langle X_s, \phi\rangle)\mathrm{d}s, \quad (2.2.1)$$

is a martingale, where $m, \kappa > 0$ are constants.

2.2.2 The Spatial Lambda-Fleming-Viot Process

We now introduce the SLFV processes. In fact there is a much richer class of these processes than those we consider here, incorporating, for example, various forms of natural selection. For a (somewhat out of date) survey we refer to Barton et al. (2013). We restrict ourselves to a population in which there are just two genetic types which we label by $\{0, 1\}$. At each time t, the random function $\{w_t(x),\, x \in \mathbb{R}^d\}$ is defined, up to a Lebesgue null set of $\mathbb{R}^d$, by

$$w_t(x) := \text{ proportion of type 1 at spatial position } x \text{ at time } t.$$

A construction of an appropriate state space for $x \mapsto w_t(x)$ can be found in Véber and Wakolbinger (2015). Using the identification between M and w given by

$$\int_{\mathbb{R}^d \times \{0,1\}} f(x,\alpha) M(\mathrm{d}x, \mathrm{d}\alpha) = \int_{\mathbb{R}^d} \{w(x) f(x,1) + (1 - w(x)) f(x,0)\} \, \mathrm{d}x,$$

this state space is in one-to-one correspondence with the space $\mathcal{M}_\lambda$ of measures on $\mathbb{R}^d \times \{0,1\}$ with 'spatial marginal' Lebesgue measure, which we endow with the topology of vague convergence. By a slight abuse of notation, we also denote the state space of the process $(w_t)_{t \in \mathbb{R}_+}$ by $\mathcal{M}_\lambda$.

Definition 2.2.3 (SLFV)**.** *Let μ be a finite measure on $(0,\infty)$ and, for each $r \in (0,\infty)$, let ν_r be a probability measure on $(0,1]$. Further, let Π be a Poisson point process on $\mathbb{R}^d \times (0,\infty) \times (0,\infty) \times (0,1]$ with intensity measure*

$$\mathrm{d}x \otimes \mathrm{d}t \otimes \mu(\mathrm{d}r) \nu_r(\mathrm{d}\rho). \tag{2.2.2}$$

The spatial Lambda-Fleming-Viot process (SLFV) *driven by* (2.2.2)*, when it exists, is the $\mathcal{M}_\lambda$-valued process $(w_t)_{t \in \mathbb{R}_+}$ with dynamics given as follows.*

If $(x,t,r,\rho) \in \Pi$, a reproduction event occurs at time t within the closed ball $\mathcal{B}_r(x)$ of radius r, centred on x, in which case:

1. *Choose a parental location z uniformly at random within $\mathcal{B}_r(x)$, and a parental type, α, according to $w_{t-}(z)$; that is $\alpha = 1$ with probability $w_{t-}(z)$ and $\alpha = 0$ with probability $1 - w_{t-}(z)$.*
2. *For every $y \in \mathcal{B}_r(x)$, set $w_t(y) = (1-\rho) w_{t-}(y) + \rho \mathbf{1}_{\{\alpha = 1\}}$.*
3. *For $y \notin \mathcal{B}_r(x)$, $w_t(y) = w_{t-}(y)$.*

We shall refer to ρ as the impact *of the event.*

Before providing conditions under which the process exists, it is convenient to introduce the dual process of coalescing lineages that plays the rôle for the SLFV played by the Lambda-coalescents for the (non-spatial) Lambda-Fleming-Viot processes. The idea is that these lineages trace out the ancestry of a sample from the population. The dual will also play a crucial role in establishing the estimates of Section 2.6.

The dynamics of the dual are driven by the same Poisson process of events Π that drives the SLFV. This driving process is reversible and we shall abuse notation by indexing events by 'backwards time' when discussing our dual. We suppose that at time 0, 'the present', we sample k individuals from locations $x_1, \ldots, x_k$ and we write $\xi_s^1, \ldots, \xi_s^{N_s}$ for the locations of the N_s 'ancestors' that make up our dual at time s before the present.

Definition 2.2.4 (Dual to the SLFV)**.** *The coalescing dual process $(\Xi_t)_{t\geq 0}$ is the $\bigcup_{n\geq 1}(\mathbb{R}^d)^n$-valued Markov process with dynamics defined as follows. At each event $(x, t, r, \rho) \in \Pi$:*

1. *For each $\xi_{t-}^i \in \mathcal{B}_r(x)$, independently mark the corresponding ancestral lineage with probability ρ;*

2. *if at least one lineage is marked, all marked lineages disappear and are replaced by a single ancestor, whose location is drawn uniformly at random from within $\mathcal{B}_r(x)$.*

If no particles are marked, then nothing happens.

Assuming that the SLFV and its dual exist, the duality is expressed through the following proposition.

Proposition 2.2.5. *The SLFV is dual to the process $(\Xi_t)_{t\geq 0}$ in the sense that for every $k \in \mathbb{N}$ and $\psi \in C((\mathbb{R}^d)^k) \cap L^1((\mathbb{R}^d)^k)$, we have*

$$\mathbb{E}_{w_0}\Big[\int_{(\mathbb{R}^d)^k} \psi(x_1, \ldots, x_k)\Big\{\prod_{j=1}^{k} w_t(x_j)\Big\}\,\mathrm{d}x_1 \ldots \mathrm{d}x_k\Big]$$
$$= \int_{(\mathbb{R}^d)^k} \psi(x_1, \ldots, x_k)\mathbb{E}_{\{x_1,\ldots,x_k\}}\Big[\prod_{j=1}^{N_t} w_0(\xi_t^j)\Big]\,\mathrm{d}x_1 \ldots \mathrm{d}x_k, \tag{2.2.3}$$

where the subscripts on the expectations denote the initial values of the corresponding processes. In particular, $\mathbb{E}_{\{x_1,\ldots,x_k\}}$ denotes expectation under the distribution of the dual process started from $\Xi_0 = \{x_1, \ldots, x_k\}$.

In Barton et al. (2010), through a powerful result of Evans (1997), it is shown that existence of the SLFV can be deduced from existence of the dual. In that paper, by

assuming that

$$\int_{[0,1]\times(0,\infty)} \rho r^d \nu_r(\mathrm{d}\rho)\mu(\mathrm{d}r) < \infty, \tag{2.2.4}$$

one guarantees that started from any finite number of individuals, the jump rate in the dual is finite, and so Definition 2.2.4 gives rise to a well-defined process. Ancestral lineages in the dual process move around according to (dependent) compound Poisson processes which can coalesce if they are affected by the same event. Although one can write down more general conditions under which the SLFV exists, see Etheridge and Kurtz (2018), Condition 2.2.4 is trivially satisfied for the processes considered below.

2.2.3 Main results

We are going to extract superBrownian motion from the SLFV through a scaling and a passage to the limit. There will be two cases, the first leading to the finite variance superprocess and the second to a superprocess with a stable branching law.

At the Nth stage of our scaling, the local population density will be $K = K(N)$. We shall denote our scaled SLFV by w^N and the population of type 1 individuals by $X^N = Kw^N$, which is defined Lebesgue almost everywhere. We shall think of X^N as a measure-valued process and abuse notation by writing, for any Borel measurable ϕ,

$$\langle X_t^N, \phi\rangle = K\int_{\mathbb{R}^d} \phi(x) w_t^N(x)\mathrm{d}x = \int_{\mathbb{R}^d} \phi(x) X_t^N(x)\mathrm{d}x.$$

Before scaling, the SLFV will be driven by a Poisson point process Π^N with intensity $\mathrm{d}x\otimes\mathrm{d}t\otimes\mu^N(\mathrm{d}r)\nu_r(\mathrm{d}\rho)$. Each $(x,t,r,\rho)\in\Pi^N$ signals a reproduction event for the scaled process associated with the quadruple $(\frac{x}{M}, \frac{t}{N}, \frac{r}{M}, \frac{\rho}{J})$, where $M := M(N)$, $J := J(N)$, are some positive increasing functions of N. In other words, for the scaled process, time is sped up by a factor N, space is shrunk by $M(N)$, and the impact of each event is reduced by a factor $J(N)$. Moreover, the local population density is increased to $K = K(N)$ where $K(N)$ is another increasing function of N.

We shall consider two different scenarios:

1. **Fixed radius case:** Here we shall take $\mu^N(\mathrm{d}r) = \delta_r$, independent of N, and $\nu_r(\mathrm{d}\rho) = \delta_u$ where $u\in(0,1]$ is fixed.

2. **Variable radius case:** Here we shall take

$$\mu^N(\mathrm{d}r) := r^\alpha \mathbf{1}_{\left\{J^{\frac{-1}{\gamma}}<r<1\right\}}\mathrm{d}r, \tag{2.2.5}$$

where α is a real constant and γ is a positive constant. We then take $\nu_r := \delta_{r^{-\gamma}}$.

The lower bound on r in the variable radius case ensures that after scaling the impact of each event is at most 1.

Theorem 2.2.6 (Fixed radius case)**.** *In the notation above, in the fixed radius case, suppose that X_0^N is absolutely continuous with respect to Lebesgue measure, that the support* $\mathrm{supp}(X_0^N) \subseteq D$*, where D is a compact subset of $\mathbb{R}^d$ (independent of N), and that X_0^N converges weakly to $X_0 \in \mathcal{M}_F(\mathbb{R}^d)$. Moreover suppose that*

1. *$M \to \infty$ as $N \to \infty$,*
2. *$\frac{N}{JM^2} \to C_1$ as $N \to \infty$, and*
3. *$\frac{KN}{J^2M^d} \to C_2$ as $N \to \infty$,*

for some $C_1, C_2 \in (0, \infty)$. Then if

$$\begin{cases} \frac{M}{J} \to 0 & \text{if } d = 1, \\ \frac{\log M}{J} \to 0 & \text{if } d = 2, \\ \frac{1}{J} \to 0 & \text{if } d \geq 3, \end{cases} \tag{2.2.6}$$

the sequence $\{X^N\}_{N\geq 1}$ converges weakly in $D([0,\infty), \mathcal{M}_F(\mathbb{R}^d))$, the space of cádlág functions endowed with the Skorohod topology, to finite variance superBrownian motion with initial condition X_0 and parameters

$$m = 2C_1 u r^{d+2} \int_{|x|\leq 1} x^2 \mathrm{d}x, \qquad \kappa = \frac{1}{2} C_2 u^2 |\mathcal{B}_r|^2.$$

Conditions 1-3 guarantee tightness; (2.2.6) will ensure that type 1 is sufficiently 'sparse' that, asymptotically, descendants of different type 1 'individuals' evolve independently (they don't sense that total population density is constrained) and we recover a

branching structure. Notice in particular that (2.2.6), combined with Conditions 2 and 3 implies that $K \to \infty$ as $N \to \infty$. In Section 2.2.4 we present a heuristic argument which suggests that these conditions are in some sense optimal. We note that the conditions of Theorem 2.2.6 are analogous to those of Cox et al. (2000), Theorem 1.1.

It isn't hard to convince oneself that if we fix the radius of events, then it is not possible to find a sequence of impact distributions ν^N and a scaling under which the limiting process is superBrownian motion with a stable branching mechanism of infinite variance, which is why we turn to $\mu^N(\mathrm{d}r)$. Theorem 2.2.7 provides conditions under which we do then have convergence to superBrownian motion with a stable branching mechanism. Even with our special choice of μ^N, in $d = 1$ these are considerably more technical than those in the fixed radius case. However, we shall see them emerge in a natural way from our calculations.

Theorem 2.2.7 (Variable radius case)**.** *In the notation above, suppose that X_0^N is absolutely continuous with respect to Lebesgue measure, that the support* $\mathrm{supp}(X_0^N) \subseteq D$*, where D is a compact subset of $\mathbb{R}^d$ (independent of N) and that X_0^N converges weakly to $X_0 \in \mathcal{M}_F(\mathbb{R}^d)$. We work in the variable radius case with $\mu^N(\mathrm{d}r)$ given by (2.2.5) and $\nu_r = \delta_{r^{-\gamma}}$. Fix $\beta \in (0,1)$ and take α and γ such that*

1. $0 < \gamma - d < \frac{1}{1-\beta}$ *and* $\gamma > 2$ *if* $d = 1$,
2. $\alpha + 1 = (\beta + 1)(\gamma - d)$.

Suppose that $M \to \infty$ as $N \to \infty$ and that

3. $\frac{N}{JM^2} \to C_1$,
4. $\frac{N}{J}\left(\frac{K}{JM^d}\right)^\beta \to C_2$ *and* $K \to \infty$,
5. $J^{\frac{\gamma-d}{\gamma}} M^{-\frac{2}{\beta}} \to \infty$.

In addition, we require:

$$\begin{cases} \left(\frac{1}{M^2}\right)^{(1-\beta)/\beta} J^{2(1-\beta)(\gamma-1)/\gamma} \frac{M^2}{J} \to 0 & \text{if } d = 1; \\ \frac{\log M}{J} \to 0 & \text{if } d = 2; \\ \frac{1}{J} \to 0 & \text{if } d \geq 3. \end{cases} \tag{2.2.7}$$

Then the sequence $\{X^N\}_{N\geq 1}$ *converges to superBrownian motion with stable branching law with parameter* β*, initial condition* X_0*, and*

$$m = 2C_1 \int_{|x|\leq 1} x^2 \mathrm{d}x \int_0^1 r^{\alpha+d+2-\gamma} \mathrm{d}r, \qquad \kappa = \frac{C_2}{\gamma - d} \int_0^\infty (e^{-\upsilon} + \upsilon - 1)\upsilon^{-(\beta+2)} \mathrm{d}\upsilon.$$

Once again Conditions 3–5 guarantee tightness of the sequence, whereas (2.2.7) ensures 'sparsity'. For $d = 1$ the condition in (2.2.7) is not necessary, even our proof shows that it could be improved a little, but in this form it is easy to check: since $\gamma > 2$ and $(1 - \beta)(\gamma - 1) < 1$, we only have to make sure that $J \to \infty$ as $N \to \infty$ sufficiently quickly compared to M.

Example 2.2.8. *The conditions of Theorem 2.2.7 are satisfied if*

1. $d \geq 2$ *and*
   (a) $\alpha = \beta$, $\gamma = d + 1$,
   (b) $J = M^\eta$ *with* $\eta \geq 2\gamma/\beta$, $M^{2+\eta} = N$, $K = JM^{d+2/\beta}$.
2. $d = 1$ *and*
   (a) $\beta = 3/4$, $\gamma = 3$, $\alpha = 5/2$,
   (b) $J = M^\eta$ *where* $\eta > 4$, $M^{2+\eta} = N$, $K = JM^{11/3}$.

The structure of the proofs of Theorems 2.2.6 and 2.2.7 will come as no surprise. We establish tightness of the sequences of rescaled processes and then check that all limit points satisfy an appropriate martingale problem. The first step will be to write down the martingale characterisation of the the scaled SLFV and manipulate it into a form that resembles the desired limit. Tightness for the fixed radius case is then highly reminiscent of the arguments in Cox et al. (2000). To prove tightness in the variable radius case, we modify the arguments used in constructing superBrownian motion with a stable branching mechanism as the limit of a sequence of branching Brownian motions (although the calculations here are somewhat more involved). In both cases, a key step in identifying the limit is to establish control over the probability that two individuals sampled from the same small region in the SLFV are close relatives.

This is also reminiscent of Cox et al. (2000), being based on estimates for the coalescing dual of the SLFV.

2.2.4 Heuristics

Before proceeding to the proofs, let us try to motivate the scalings in Theorem 2.2.6 and (at least some of those in) Theorem 2.2.7.

2.2.4.1 Fixed radius

First consider the fixed radius case. If superBrownian motion really is a good approximation for the type 1 population, then, in particular, we expect that the motion of a single ancestral lineage in the SLFV should converge to Brownian motion. Events that affect regions in which a given lineage lies fall according to a Poisson process with rate proportional to N, and, for each such event, the chance that the lineage is affected by it is u/J. Thus the lineage will jump at rate proportional to N/J. Each jump is mean zero, finite variance, and $\mathcal{O}(1/M)$. In order to obtain a Brownian limit, we seek a diffusive rescaling; that is $N/(JM^2)$ should converge.

Next recall that in $d \geq 2$ superBrownian motion is a two-dimensional object, whereas the support of the SLFV has the same dimension as the state space. The scaling of the impact of each event dictates that an 'atom' of mass is $\mathcal{O}(1/J)$, so in order that the number of atoms in a region of diameter M scale like KM^2 (as it would for a two-dimensional object) we take $KM^2 \sim JM^d$, which if $N/(JM^2)$ converges says that $KN/(J^2M^d)$ should converge.

The large parameter K controls the total population density, but we must still ensure that the population of rare alleles in our scaled SLFV is sufficiently 'sparse' if we are to recover superBrownian motion. In the SLFV, the density of the population is strictly regulated, creating a strong dependence between the mass born during a reproduction event and that which dies. In contrast, in superBrownian motion, once born, 'individuals' reproduce and die independently of one another. In order to ensure that the dependence inherent in the SLFV is not apparent to us when we follow just a single (rare) type, we should like to know that if we sample individuals from the same small region they are not likely to be close relatives. In this way we can guarantee

that individuals are not victims of reproduction events in which their own close family reproduces.

To check whether two individuals sampled from very close to one another are close relatives, we follow the dual process of ancestral lineages. We should like them to move apart to a distance of $\mathcal{O}(1)$ in the scaled process (without coalescing). If both lineages are in the region affected by an event, then the chance that they are both affected (and therefore coalesce) given that at least one of them jumps is of order $1/J$. On the other hand, it only takes a finite number of events in which only one of the lineages jumps before they are sufficiently far apart that they cannot be affected by the same event and so evolve independently. We then think of them as making an excursion away from one another, before they once again come close enough that they are susceptible to coalescence. The number of such excursions before we see one in which they move apart to a distance of $\mathcal{O}(1)$ (after scaling) has mean $\mathcal{O}(M)$ in $d = 1$, $\mathcal{O}(\log M)$ in $d = 2$ and $\mathcal{O}(1)$ in $d \geq 3$. Since at the end of each excursion, the chance that the lineages will coalesce rather than starting the next excursion is proportional to $1/J$, we see that our 'sparsity' conditions (2.2.6) ensure that the probability that they successfully 'escape' to a distance of $\mathcal{O}(1)$ from one another tends to one. This is the intuition underlying the calculations in Section 2.6.

2.2.4.2 Variable radii

Now we turn to the case of variable radii, from which we are trying to extract a superBrownian motion with stable branching mechanism.

In the non-spatial setting, Bertoin and Le Gall (2006) recover a stable branching process with parameter β from a Lambda-Fleming-Viot process in much the same way as we recovered the Feller branching process from the Wright-Fisher diffusion in the introduction. The Lambda-Fleming-Viot process is driven by a Poisson point process $\widetilde{\Pi}$ on $[0, \infty) \times (0, 1]$. A point $(t, \rho) \in \widetilde{\Pi}$ signals a reproduction event at time t in which a proportion ρ of the population is replaced by offspring of a randomly chosen parent. The intensity of $\widetilde{\Pi}$ is $\mathrm{d}t \otimes \Lambda(\mathrm{d}\rho)/\rho^2$, where Λ is such that the tail $\Lambda([\varepsilon, 1])$ is regularly varying with index $-(1 + \beta)$ as $\varepsilon \to 0$. A simple 'back of the envelope' calculation illustrates why Bertoin and Le Gall's result should hold. To be completely concrete,

we present it in the special case $\Lambda(\mathrm{d}\rho) = C(\beta)\rho^{-\beta}(1-\rho)^{\beta}\mathrm{d}\rho$, corresponding to the Lambda-Fleming-Viot process which is dual to a so-called Beta-coalescent and, as shown in Birkner et al. (2005a), is a timechange of the stable branching process.

Once again let K be the total population size and consider a rare allele that makes up a proportion w of the population. We are interested in the absolute number, $X = Kw$ of rare alleles. We apply the infinitesimal generator of X to a test function of the form $\exp(-\theta X)$, where $\theta \geq 0$. This yields

$$\begin{aligned}
\mathcal{L}^X(e^{-\theta X}) &= \int_0^1 \left\{ \frac{X}{K} e^{-\theta(X+K\rho(1-X/K))} + (1-\frac{X}{K}) e^{-\theta X(1-\rho)} - e^{-\theta X} \right\} \frac{\Lambda(\mathrm{d}\rho)}{\rho^2} \\
&= C(\beta) \int_0^K e^{-\theta X} \left\{ \frac{X}{K} e^{-\theta v(1-X/K)} + (1-\frac{X}{K}) e^{\theta X v/K} - 1 \right\} \frac{(1-v/K)^{\beta}}{(v/K)^{\beta+2}} \frac{1}{K} \mathrm{d}v \\
&\approx C(\beta) K^{\beta} X e^{-\theta X} \int_0^{\infty} \left\{ e^{-\theta v} + \theta v - 1 \right\} \frac{1}{v^{\beta+2}} \mathrm{d}v \\
&= C K^{\beta} X e^{-\theta X} \theta^{\beta+1},
\end{aligned}$$

which we recognise as the infinitesimal generator of the stable branching process, timechanged by a factor proportional to K^{β}. Recalling the construction of this Lambda-Fleming-Viot process from individual based models, for example as in Schweinsberg (2003), we see that the evolution of a population of size K should be compared to the Lambda-Fleming-Viot process on the timescale $1/K^{\beta}$ (just as we see a factor $1/K$ in the Wright-Fisher diffusion (2.1.1)). This precisely cancels the K^{β} we see here. This calculation confirms that the emergence of the stable branching process was dictated by the behaviour of the measure $\Lambda(\mathrm{d}\rho)$ close to $\rho = 0$.

If we are to extract infinite variance superBrownian motion from an SLFV in an analogous way, we must have random event radii and, since the spatial motion is bound up in reproduction events, if in the limit the spatial motion is to be continuous, the impact will depend on the event size in a nontrivial way. In order to make the calculations tractable, we fix the impact to be a negative power of the radius of events and we take $\mu^N(\mathrm{d}r)$ to be a truncated power law. The purpose of the truncation is two-fold: first, by bounding the radii below we force the impact of each event to lie in $(0, 1]$; second, by bounding radii above, we ensure that the jumps of lineages in the unscaled SLFV have finite moments of all orders.

As for the fixed radius case, we should like the motion of a single ancestral lineage to converge to Brownian motion. A lineage will fall in the region affected by an event of (scaled) radius r/M at rate $Nr^d\mu^N(\mathrm{d}r)$ in which case, with probability ρ/J (with ρ sampled from $\nu_r(\mathrm{d}\rho)$), it will make a mean zero, finite variance jump of size of order r/M. Substituting our chosen form of $\mu^N(\mathrm{d}r)\nu_r(\mathrm{d}\rho)$, we see that in order to obtain a Brownian limit for the motion of lineages, we should require convergence of

$$\int_0^\infty \frac{Nu(r)r^{d+2}}{JM^2}\,\mu^N(\mathrm{d}r) = \int_{J^{-1/\gamma}}^1 \frac{N}{JM^2} r^{d+2+\alpha-\gamma}\mathrm{d}r.$$

Under our conditions on α, γ, this implies that $N/(JM^2)$ should converge, just as in the fixed radius case.

We now turn to recovery of the stable branching mechanism. When an event of radius r falls, the total mass of the offspring in the rescaled population process is $K/(JM^d)$ times $v(r) = C(d)r^{d-\gamma}$, where $C(d) := \int_{\mathcal{B}_1(0)} |x|^2\mathrm{d}x$. Such events fall on a given point x, and an 'individual' at that point is selected as parent of the event, at rate $(N/K)r^\alpha\mathrm{d}r\mathrm{d}x$ (a factor of $C(d)r^d$ in the rate at which events of radius r cover the point x has cancelled with the reciprocal of the same factor in the probability that x is the point selected uniformly at random from within the ball to be the location of the parent). Observing that $\mathrm{d}v(r) = C(d)(d-\gamma)r^{d-\gamma-1}\mathrm{d}r$ (and thus substituting for $r^\alpha\mathrm{d}r$), we see that for a given 'individual' at x, events in which it produces offspring with mass proportional to v occur at rate proportional to $v^{-(\gamma-d+\alpha+1)/(\gamma-d)}\mathrm{d}v\mathrm{d}x$. We need to match this with the $v^{-\beta-2}$ of the non-spatial case close to $v = 0$, and so we take $\alpha + 1 = (\beta+1)(\gamma-d)$.

A more careful version of the argument in the previous paragraph leads to Conditions 4 and 5. In fact, Condition 4 can be understood in terms of the dimension of the limit in much the same way as the corresponding condition for the fixed radius case. In the stable case, the Hausdorff dimension of the support of the limiting superBrownian motion will be $d \wedge (2/\beta)$, so at least in high enough dimensions we expect to need $JM^d \sim KM^{2/\beta}$, which combined with convergence of $N/(JM^2)$ leads to Condition 4. Conditons (2.2.7), which follow from the same considerations as in the fixed radius case, ensure sufficient 'sparsity'.

2.3 Martingale characterisation of the process $\{X_t^N\}_{t\geq 0}$

In this section we characterise the distribution of the scaled process $\{X_t^N\}_{t\geq 0}$ as a solution to a martingale problem. It will be convenient to use different formulations of the martingale problem for our two scalings. First we need some notation. Suppose that $\phi \in C_0^3(\mathbb{R}^d)$ and that $f \in C_0^2(\mathbb{R})$. We use the notation

$$\mathcal{L}_f^N(\phi)(X_0) := \lim_{t\to 0} \frac{\mathbb{E}^{X_0}[f(\langle X_t^N, \phi\rangle)] - f(\langle X_0^N, \phi\rangle)}{t},$$

for the infinitesimal generator of the measure-valued process X_t^N applied to test functions of the form $F(X_t^N) = f(\langle X_t^N, \phi\rangle)$. Though we define $\mathcal{L}_f^N$ for $f \in C_0^2(\mathbb{R})$ we can use the compact support of X_0^t to extend this to $f(x) \equiv x$ and $f(x) \equiv x^2$. With a slight abuse of notation we use $\mathcal{L}^N(\phi)(X_0^N)$ for the corresponding quantity when $f(x) \equiv x$.

Notice that we have not assumed that ϕ has compact support. Because X_0^N has compact support, the rate at which that support is overlapped by a reproduction event is bounded, and such an event can only increase the volume of the support of X^N by an amount bounded above by the volume of the event (which in turn is uniformly bounded). Iterating this argument and comparing to a pure birth process, it is evident that the support will remain bounded (indeed compact) up to any finite time, and so the rate of events affecting X^N is bounded.

Writing down the generator is now standard as our process evolves according to a series of jumps of finite rate. Recall that at each point $(x, t, r, \rho) \in \Pi^N$, the scaled process w^N is subject to the reproduction event associated with the quadruple $(\frac{x}{M}, \frac{t}{N}, \frac{r}{M}, \frac{\rho}{J})$. At such an event, within the ball $\mathcal{B}_{r/M}(x)$ we have two possibilities:

$$\begin{cases} w_t^N(y) - w_{t-}^N(y) = -\frac{\rho}{J} w_{t-}^N(y) + \frac{\rho}{J} & \text{with probability } w_{t-}^N(z), \\ \\ w_t^N(y) - w_{t-}^N(y) = -\frac{\rho}{J} w_{t-}^N(y) & \text{with probability } 1 - w_{t-}^N(z). \end{cases}$$

We write $\mathcal{B}_r^M(x)$ for the ball $\mathcal{B}_{r/M}(x)$ and $|\mathcal{B}_r^M| = |\mathcal{B}_r|/M^d$ for its volume. At the time of the first event to affect X^N, we have $w_{t-}^N = w_0^N$ and since, moreover, we only consider

ν_r of the form $\delta_{u(r)}$, we find

$$\mathcal{L}_f^N(\phi)(X_0^N) = NM^d\Bigg[\int_{\mathbb{R}^d}\int_0^\infty\int_{\mathcal{B}_r^M(x)}\frac{1}{|\mathcal{B}_r^M|}w_0^N(z)f\Bigg(K\frac{u(r)}{J}\int_{\mathcal{B}_r^M(x)}\phi(y)\mathrm{d}y$$
$$+K\Big(1-\frac{u(r)}{J}\Big)\int_{\mathcal{B}_r^M(x)}\phi(y)w_0^N(y)\mathrm{d}y+K\int_{\mathbb{R}^d\setminus\mathcal{B}_r^M(x)}\phi(y)w_0^N(y)\mathrm{d}y\Bigg)$$
$$+(1-w_0^N(z))f\Bigg(K\Big(1-\frac{u(r)}{J}\Big)\int_{\mathcal{B}_r^M(x)}\phi(y)w_0^N(y)\mathrm{d}y+K\int_{\mathbb{R}^d\setminus\mathcal{B}_r^M(x)}\phi(y)w_0^N(y)\mathrm{d}y\Bigg)$$
$$-f\Big(K\int_{\mathbb{R}^d}\phi(y)w_0^N(y)\mathrm{d}y\Big)\mathrm{d}z\,\mu^N(\mathrm{d}r)\mathrm{d}x\Bigg].$$

Substituting $Kw_0^N = X_0^N$ this becomes

$$\mathcal{L}_f^N(\phi)(X_0^N) = NM^d\Bigg[\int_{\mathbb{R}^d}\int_0^\infty\int_{\mathcal{B}_r^M(x)}\frac{1}{|\mathcal{B}_r^M|}\frac{X_0^N(z)}{K}f\Bigg(K\frac{u(r)}{J}\int_{\mathcal{B}_r^M}\phi(y)\mathrm{d}y$$
$$+\Big(1-\frac{u(r)}{J}\Big)\int_{\mathcal{B}_r^M(x)}\phi(y)X_0^N(y)\mathrm{d}y+\int_{\mathbb{R}^d\setminus\mathcal{B}_r^M(x)}\phi(y)X_0^N(y)\mathrm{d}y\Bigg)$$
$$+(1-\frac{X_0^N(z)}{K})f\Bigg(\Big(1-\frac{u(r)}{J}\Big)\int_{\mathcal{B}_r^M(x)}\phi(y)X_0^N(y)\mathrm{d}y+\int_{\mathbb{R}^d\setminus\mathcal{B}_r^M(x)}\phi(y)X_0^N(y)\mathrm{d}y\Bigg)$$
$$-f\Big(\int_{\mathbb{R}^d}\phi(y)X_0^N(y)\mathrm{d}y\Big)\mathrm{d}z\,\mu^N(\mathrm{d}r)\mathrm{d}x\Bigg]. \quad (2.3.1)$$

Since $\{X_t^N\}_{t\geq 0}$ is a pure jump Markov process, driven by a Poisson process of jumps, it follows immediately that for f and ϕ as above,

$$f(\langle X_t^N,\phi\rangle) - f(\langle X_0^N,\phi\rangle) - \int_0^t \mathcal{L}_f(\phi)(X_s^N)\mathrm{d}s$$

defines a mean zero local martingale. In the variable radius case, we shall exploit this with $f(x) = \exp(-x)$ and non-negative ϕ. In the fixed radius case, the following lemma, which follows immediately on setting $f(x) = x$ and $f(x) = x^2$, will provide a more convenient tool.

Lemma 2.3.9. *The quantity $\langle X_t^N,\phi\rangle$ has the semimartingale decomposition:*

$$\langle X_t^N,\phi\rangle = \langle X_0^N,\phi\rangle + \int_0^t \mathcal{L}^N(\phi)(X_s^N)\mathrm{d}s + M_t^N(\phi), \quad (2.3.2)$$

where

$$\mathcal{L}^N(\phi)(X_s^N) = \int_{\mathbb{R}^d}\int_0^\infty \int_{\mathcal{B}_r^M(x)} \frac{Nu(r)M^{2d}}{J|\mathcal{B}_r|} \Big\{\Big(\int_{\mathcal{B}_r^M(x)}\phi(y)\mathrm{d}y\Big)X_s^N(z) - \int_{\mathcal{B}_r^M(x)}\phi(y)X_s^N(y)\mathrm{d}y\Big\}\mathrm{d}z\,\mu^N(\mathrm{d}r)\mathrm{d}x.$$

The local martingale $M_t^N(\phi)$ has quadratic variation process

$$\langle M^N(\phi)\rangle_t = \int_0^t\int_{\mathbb{R}^d}\int_{\mathcal{B}_r^M(x)} \frac{NM^{2d}u(r)^2}{J^2|\mathcal{B}_r|}\Big\{(1-\frac{X_s^N(z)}{K})\Big(\int_{\mathcal{B}_r^M(x)}\phi(y)X_s^N(y)\mathrm{d}y\Big)^2 + \frac{X_s^N(z)}{K}\Big(\int_{\mathcal{B}_r^M(x)}K\phi(y)\mathrm{d}y - \int_{\mathcal{B}_r^M(x)}\phi(y)X_s^N(y)\mathrm{d}y\Big)^2\Big\}\mathrm{d}z\mathrm{d}x\mathrm{d}s. \quad (2.3.3)$$

It will be convenient to extend the semimartingale decomposition to $\phi = \phi(t,x) \in C_b^{2,3}(\mathbb{R}\times\mathbb{R}^d)$. Writing

$$f(\langle X_t^N,\phi_t\rangle) - f(\langle X_0^N,\phi_0\rangle) = f(\langle X_t^N,\phi_0\rangle) - f(\langle X_0^N,\phi_0\rangle) + f'(\langle X_t^N,\phi_0\rangle)\langle X_t^N,\phi_t-\phi_0\rangle + \mathcal{O}(\langle X_t^N,\phi_t-\phi_0\rangle^2\|f''\|_\infty),$$

we deduce that this extension simply results in the additional term

$$f'(\langle X_0^N,\phi_0\rangle)\langle X_0^N,\dot{\phi}_0\rangle$$

in the generator. In particular, we arrive at the following lemma.

Lemma 2.3.10. *The quantity $\langle X_t^N,\phi_t\rangle$ has the semimartingale decomposition:*

$$\langle X_t^N,\phi_t\rangle = \langle X_0^N,\phi_0\rangle + \int_0^t \mathcal{L}^N(\phi_s)(X_s^N)\mathrm{d}s + M_t^N(\phi_\cdot), \quad (2.3.4)$$

where

$$\mathcal{L}^N(\phi_s)(X_s^N) = \int_{\mathbb{R}^d}\dot{\phi}_s(y)X_s^N(y)\mathrm{d}y + \int_{\mathbb{R}^d}\int_0^\infty\int_{\mathcal{B}_r^M(x)} \frac{Nu(r)M^{2d}}{J|\mathcal{B}_r|} \Big\{\Big(\int_{\mathcal{B}_r^M(x)}\phi_s(y)dy\Big)X_s^N(z) - \int_{\mathcal{B}_r^M(x)}\phi_s(y)X_s^N(y)\mathrm{d}y\Big\}\mathrm{d}z\,\mu^N(\mathrm{d}r)\mathrm{d}x.$$

2.4 Identifying the limit

We now turn to identifying the possible limit points of the sequence of processes $\{X^N\}_{N\geq 1}$ under our chosen scalings, deferring the proof of tightness of the sequence to Section 2.5.

2.4.1 Spatial motion

Although our scaling and limits are quite different, our calculations are reminiscent of those of Berestycki et al. (2013). Our aim is to find an approximate form of the martingale problem that is close to that for superBrownian motion.

For simplicity we take ϕ to be constant in time. An interchange of integrals followed by an interchange of the roles of y and z in our notation yields

$$\begin{aligned}\int_{\mathcal{B}_r^M(x)}\int_{\mathcal{B}_r^M(x)}&\phi(y)(X_s^N(z)-X_s^N(y))\mathrm{d}y\mathrm{d}z\\ &=\int_{\mathcal{B}_r^M(x)}\int_{\mathcal{B}_r^M(x)}\phi(z)X_s^N(y)\mathrm{d}y\mathrm{d}z-\int_{\mathcal{B}_r^M(x)}\int_{\mathcal{B}_r^M(x)}\phi(y)X_s^N(y)\mathrm{d}y\mathrm{d}z\\ &=\int_{\mathcal{B}_r^M(x)}\int_{\mathcal{B}_r^M(x)}(\phi(z)-\phi(x)+\phi(x)-\phi(y))X_s^N(y)\mathrm{d}y\mathrm{d}z.\end{aligned}$$

We use the Taylor expansion

$$\phi(y)=\phi(x)+\nabla\phi(x)(y-x)+\frac{1}{2}\sum_{|\alpha|=2}\partial_\alpha\phi(x)(y-x)^\alpha+\sum_{|\alpha|=3}R_\alpha(y)((y-x)^\alpha),$$

where $\|R_\alpha\|_\infty\leq C\|\phi\|_{C^3}$, with C independent of ϕ, to obtain

$$\begin{aligned}&\int_{\mathcal{B}_r^M(x)}\int_{\mathcal{B}_r^M(x)}(\phi(z)-\phi(x))X_s^N(y)\mathrm{d}z\mathrm{d}y+\int_{\mathcal{B}_r^M(x)}\int_{\mathcal{B}_r^M(x)}(\phi(x)-\phi(y))X_s^N(y)\mathrm{d}z\mathrm{d}y\\ =&\int_{\mathcal{B}_r^M(x)}\int_{\mathcal{B}_r^M(x)}\Big(\nabla\phi(x)(z-x)+\frac{1}{2}\sum_{|\alpha|=2}\partial_\alpha\phi(x)(z-x)^\alpha\\ &\qquad+\sum_{|\alpha|=3}R_\alpha(z)(z-x)^\alpha\Big)X_s^N(y)\mathrm{d}z\mathrm{d}y+\frac{|\mathcal{B}_r|}{M^d}\int_{\mathcal{B}_r^M(x)}(\phi(x)-\phi(y))X_s^N(y)\mathrm{d}y\\ =&\Big\{\frac{C(d)r^{d+2}}{M^{d+2}}\frac{\Delta\phi(x)}{2}+\mathcal{O}\Big(\frac{\|\phi\|_{C^3}r^{d+3}}{M^{d+3}}\Big)\Big\}\int_{\mathcal{B}_r^M(x)}X_s^N(y)\mathrm{d}y\\ &\qquad+\frac{|\mathcal{B}_r|}{M^d}\int_{\mathcal{B}_r^M(x)}(\phi(x)-\phi(y))X_s^N(y)\mathrm{d}y,\end{aligned}$$

where $C(d) := \int_{\mathcal{B}_1(0)} |x|^2 \mathrm{d}x$. Moreover,

$$\begin{aligned}\int_{\mathbb{R}^d}\int_{\mathcal{B}_r^M(x)} (\phi(x)-\phi(y))X_s^N(y)\mathrm{d}y\mathrm{d}x &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \mathbf{1}_{\{|x-y|\le \frac{r}{M}\}}(\phi(x)-\phi(y))X_s^N(y)\mathrm{d}y\mathrm{d}x \\ &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \mathbf{1}_{\{|x-y|\le \frac{r}{M}\}}(\phi(x)-\phi(y))X_s^N(y)\mathrm{d}x\mathrm{d}y \\ &= \int_{\mathbb{R}^d}\int_{\mathcal{B}_r^M(y)} (\phi(x)-\phi(y))X_s^N(y)\mathrm{d}x\mathrm{d}y. \quad (2.4.1)\end{aligned}$$

Again using Taylor's Theorem (and interchanging the role of x and y) this is

$$\begin{aligned}&\int_{\mathbb{R}^d}\int_{\mathcal{B}_r^M(x)} (\phi(y)-\phi(x))X_s^N(x)\mathrm{d}y\mathrm{d}x \\ &= \int_{\mathbb{R}^d}\int_{\mathcal{B}_r^M(x)} \Big(\nabla\phi(x)(y-x) + \frac{1}{2}\sum_{|\alpha|=2}\partial_\alpha\phi(x)(z-x)^\alpha \\ &\qquad\qquad + \sum_{|\alpha|=3} R_\alpha(y)(y-x)^\alpha\Big)X_s^N(x)\mathrm{d}y\mathrm{d}x \\ &= \int_{\mathbb{R}^d}\Big\{\frac{C(d)r^{d+2}}{M^{d+2}}\frac{\Delta\phi(x)}{2} + \mathcal{O}\Big(\frac{\|\phi\|_{C^3}r^{d+3}}{M^{d+3}}\Big)\Big\}X_s^N(x)\mathrm{d}x.\end{aligned}$$

Combining the above with our expression for $\mathcal{L}^N(\phi)$ from Lemma 2.3.9, for the fixed radius case we obtain

$$\begin{aligned}\int_0^t \mathcal{L}^N(\phi)(X_s^N)\mathrm{d}s =& \frac{NM^{2d}u}{J|\mathcal{B}_r|}\int_0^t\int_{\mathbb{R}^d}\Big\{\frac{C(d)r^{d+2}}{M^{d+2}}\frac{\Delta\phi(x)}{2} + \mathcal{O}\Big(\frac{\|\phi\|_{C^3}r^{d+3}}{M^{d+3}}\Big)\Big\} \\ &\times\Big(\int_{\mathcal{B}_r^M(x)} X_s^N(y)\mathrm{d}y + \frac{|\mathcal{B}_r|}{M^d}X_s^N(x)\Big)\mathrm{d}x\mathrm{d}s \\ =& \int_0^t\int_{\mathbb{R}^d}\Big\{\frac{C(d)r^{d+2}Nu}{JM^2}\frac{\Delta\phi(x)}{2} + \mathcal{O}\Big(\frac{Nu\|\phi\|_{C^3}r^{d+3}}{JM^3}\Big)\Big\} \\ &\times\Big(\frac{M^d}{|\mathcal{B}_r|}\int_{\mathcal{B}_r^M(x)} X_s^N(y)\mathrm{d}y + X_s^N(x)\Big)\mathrm{d}x\mathrm{d}s.\end{aligned}$$

We note that, using the same manipulation as in (2.4.1),

$$\int_{\mathbb{R}^d}\int_{\mathcal{B}_r^M(x)} \Delta\phi(x)X_s^N(y)\mathrm{d}y\mathrm{d}x = \int_{\mathbb{R}^d}\int_{\mathcal{B}_r^M(x)} \Delta\phi(y)X_s^N(x)\mathrm{d}y\mathrm{d}x,$$

and so since $\phi \in C_b^3$, a Taylor expansion yields,

$$\int_0^t \mathcal{L}^N(\phi)(X_s^N)\mathrm{d}s = \int_0^t\int_{\mathbb{R}^d}\Big\{\frac{C(d)r^{d+2}Nu}{JM^2}\Delta\phi(x) + \mathcal{O}\Big(\frac{Nu\|\phi\|_{C^3}r^{d+3}}{JM^3}\Big)\Big\}X_s^N(x)\mathrm{d}x\mathrm{d}s$$

$$= \int_0^t \langle X_s^N, \frac{C(d) r^{d+2} N u}{JM^2} \Delta\phi \rangle + \mathcal{O}\Big(\frac{Nu\|\phi\|_{C^3} r^{d+3}}{JM^3}\Big) \langle X_s^N, \mathbf{1} \rangle \mathrm{d}s.$$

For the variable radius case, we integrate this expression against $\nu_r(\mathrm{d}u)\mu^N(\mathrm{d}r)$.

Definition 2.4.11. *We denote by $\mathcal{A}^N$ the operator*

$$\mathcal{A}^N(\phi) := \frac{m(N)}{2} \Delta\phi,$$

where in the fixed radius case

$$m(N) := \frac{2C(d) N u r^{d+2}}{JM^2},$$

and in the variable radius case

$$m(N) := \frac{2C(d) N}{JM^2} \int_0^\infty u(r) r^{d+2} \mu^N(\mathrm{d}r),$$

with $C(d) := \int_{\mathcal{B}_1(0)} x^2 \mathrm{d}x$ and $u(r) = r^{-\gamma}$.

Evidently it is straightforward to extend the calculation above to suitable time dependent $\phi_\cdot$. We record the result as a lemma.

Lemma 2.4.12. *For $\phi_s(x) : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R} \in C_0^{2,3}$,*

$$\int_0^t \mathcal{L}^N(\phi_s)(X_s^N) \mathrm{d}s = \int_0^t \langle X_s^N, \dot{\phi}_s \rangle + \langle X_s^N, \mathcal{A}^N(\phi_s) \rangle \mathrm{d}s + \zeta_t^N(\phi),$$

where $|\zeta_t^N(\phi)| \leq \mathcal{O}\Big(\frac{N \sup_{0 \leq s \leq t} \|\phi_s\|_{C^3}}{JM^3} \int_0^\infty u(r) r^{d+3} \mu^N(\mathrm{d}r)\Big) \int_0^t \langle X_s^N, \mathbf{1} \rangle \mathrm{d}s$.

2.4.2 Fixed radius case

We now turn to identification of the limit as $N \to \infty$ of the quadratic variation $\langle M^N(\phi) \rangle_t$ of (2.3.3) in the fixed radius case. Recall that

$$\langle M^N(\phi) \rangle_t = \int_0^t \int_{\mathbb{R}^d} \int_{\mathcal{B}_r^M(x)} \frac{N M^{2d} u^2}{J^2 |\mathcal{B}_r(x)|} \Big\{ (1 - \frac{X_s^N(z)}{K}) \Big(\int_{\mathcal{B}_r^M(x)} \phi(y) X_s^N(y) \mathrm{d}y \Big)^2 + \frac{X_s^N(z)}{K} \Big(K \int_{\mathcal{B}_r^M(x)} \phi(y) \mathrm{d}y - \int_{\mathcal{B}_r^M(x)} \phi(y) X_s^N(y) \mathrm{d}y \Big)^2 \Big\} \mathrm{d}z \mathrm{d}x \mathrm{d}s.$$

Expanding the brackets we see that

$$\langle M^N(\phi)\rangle_t = \int_0^t \int_{\mathbb{R}^d} \int_{\mathcal{B}_r^M(x)} \frac{KNM^{2d}u^2}{J^2|\mathcal{B}_r|} \Big[X_s^N(z) \Big(\int_{\mathcal{B}_r^M(x)} \phi(y)\mathrm{d}y \Big)^2 - 2\frac{X_s^N(z)}{K} \Big(\Big(\int_{\mathcal{B}_r^M(x)} \phi(y)\mathrm{d}y \Big) \Big(\int_{\mathcal{B}_r^M(x)} \phi(y) X_s^N(y)\mathrm{d}y \Big) \Big) + \frac{1}{K} \Big(\int_{\mathcal{B}_r^M(x)} \phi(y) X_s^N(y)\mathrm{d}y \Big)^2 \Big] \mathrm{d}z\mathrm{d}x\mathrm{d}s,$$

which can be rearranged to yield

$$\langle M^N(\phi)\rangle_t = \frac{KN|\mathcal{B}_r|^2u^2}{J^2M^d} \int_0^t \int_{\mathbb{R}^d} \frac{1}{|\mathcal{B}_r^M|} \int_{\mathcal{B}_r^M(x)} \Big[X_s^N(z) \Big(\frac{1}{|\mathcal{B}_r^M|} \int_{\mathcal{B}_r^M(x)} \phi(y)\mathrm{d}y \Big)^2 - 2\frac{X_s^N(z)}{K} \Big(\Big(\frac{1}{|\mathcal{B}_r^M|} \int_{\mathcal{B}_r^M(x)} \phi(y)\mathrm{d}y \Big) \Big(\frac{1}{|\mathcal{B}_r^M|} \int_{\mathcal{B}_r^M(x)} \phi(y) X_s^N(y)\mathrm{d}y \Big) \Big) + \frac{1}{K} \Big(\frac{1}{|\mathcal{B}_r^M|} \int_{\mathcal{B}_r^M(x)} \phi(y) X_s^N(y)\mathrm{d}y \Big)^2 \Big] \mathrm{d}z\mathrm{d}x\mathrm{d}s. \quad (2.4.2)$$

Mimicking the manipulation that gave us (2.4.1), and using the regularity of ϕ, this can be written

$$\langle M^N(\phi)\rangle_t = \frac{KN|\mathcal{B}_r|^2u^2}{J^2M^d} \int_0^t \int_{\mathbb{R}^d} \Big[X_s^N(x)\phi^2(x) - \frac{1}{K} \Big(\frac{1}{|\mathcal{B}_r^M|} \int_{\mathcal{B}_r^M(x)} \phi(y) X_s^N(y)\mathrm{d}y \Big)^2 + \mathcal{O}\Big(\frac{\|\phi\|_{C^1}^2}{M} \Big) X_s^N(x) \Big] \mathrm{d}x\mathrm{d}s, \quad (2.4.3)$$

where we used that if $z \in \mathcal{B}_r^M(x)$ and $y \in \mathcal{B}_r^M(x)$, then $|z-y| < 2/M$ and $X/K \leq 1$ to estimate the error in replacing the second term in (2.4.2) by twice the third. The necessity of Conditions 1–3 of Theorem 2.2.6 is already evident from (2.4.3). This result will be sufficient for the proof of tightness, but more work will be needed to check that the second term on the right tends to zero as $N \to \infty$, and thus identify the limiting quadratic variation as that corresponding to superBrownian motion. The proof of the following lemma, which we defer to Section 2.6, rests on the duality of Proposition 2.2.5.

Lemma 2.4.13. *Under the conditions of Theorem 2.2.6, for any $\phi \in C_0^3(\mathbb{R}^d)$,*

$$\mathbb{E}\Big[\int_0^t \int_{\mathbb{R}^d} \frac{1}{K} \Big(\frac{1}{|\mathcal{B}_r^M|} \int_{\mathcal{B}_r^M(x)} \phi(y) X_s^N(y)\mathrm{d}y \Big)^2 \mathrm{d}x\mathrm{d}s \Big] \to 0$$

as $N \to \infty$.

If we can prove that our sequence $\{X^N\}_{N\geq 1}$ of processes is tight and that all limit points are martingales, then granted Lemma 2.4.13 (and some uniform integrability), Lemma 2.4.12 and (2.4.3) allows us to identify the limit points as solutions to the martingale problem for finite variance superBrownian motion with m and κ as in the statement of Theorem 2.2.6.

2.4.3 Variable radius case

We now turn to the variable radius case. Since, if we are to have convergence to the superBrownian motion with stable branching, the quadratic variation of the previous subsection must be unbounded as $N \to \infty$, we instead turn our attention to $\mathcal{L}_f^N(\phi)$ with $f(x) = e^{-x}$. We suppose that ϕ is non-negative and, for simplicity, independent of time. Substituting in (2.3.1), we obtain

$$\begin{aligned}\mathcal{L}_{\exp(-\cdot)}(\phi)(X_0^N) = NM^d \exp(-\langle X_0^N, \phi_0\rangle) \int_{\mathbb{R}^d}\int_0^\infty \frac{1}{|\mathcal{B}_r^M|}\int_{\mathcal{B}_r^M(x)}\Bigg[\\ \frac{X_0^N(z)}{K}\exp\Bigg(-K\frac{u(r)}{J}\int_{\mathcal{B}_r^M(x)}\phi(y)\mathrm{d}y + \frac{u(r)}{J}\int_{\mathcal{B}_r^M(x)}\phi(y)X_0^N(y)\mathrm{d}y\Bigg) \\ + (1-\frac{X_0^N(z)}{K})\exp\left(\frac{u(r)}{J}\int_{\mathcal{B}_r^M(x)}\phi(y)X_0^N(y)\mathrm{d}y\right) - 1\Bigg]\mathrm{d}z\mu^N(\mathrm{d}r)\mathrm{d}x.\end{aligned}$$

Taylor expansion of the exponential function yields

$$\begin{aligned}&\frac{X_0^N(z)}{K}\exp\Bigg(-K\frac{u(r)}{J}\int_{\mathcal{B}_r^M(x)}\phi(y)\mathrm{d}y + \frac{u(r)}{J}\int_{\mathcal{B}_r^M(x)}\phi(y)X_0^N(y)\mathrm{d}y\Bigg) \\ &\qquad + (1-\frac{X_0^N(z)}{K})\exp\left(\frac{u(r)}{J}\int_{\mathcal{B}_r^M(x)}\phi(y)X_0^N(y)\mathrm{d}y\right) - 1 \\ &= \frac{X_0^N(z)}{K}\exp\Big(-K\frac{u(r)}{J}\int_{\mathcal{B}_r^M(x)}\phi(y)\mathrm{d}y\Big)\Bigg[1 + \frac{u(r)}{J}\int_{\mathcal{B}_r^M(x)}\phi(y)X_0^N(y)\mathrm{d}y \\ &\qquad + \mathcal{O}\Big(\frac{u(r)}{J}\int_{\mathcal{B}_r^M(x)}\phi(y)X_0^N(y)\mathrm{d}y\Big)^2\Bigg] \\ &+ (1-\frac{X_0^N(z)}{K})\Bigg[1 + \frac{u(r)}{J}\int_{\mathcal{B}_r^M(x)}\phi(y)X_0^N(y)\mathrm{d}y + \mathcal{O}\Big(\frac{u(r)}{J}\int_{\mathcal{B}_r^M(x)}\phi(y)X_0^N(y)\mathrm{d}y\Big)^2\Bigg] - 1 \\ &= \frac{X_0^N(z)}{K}\exp\Big(-K\frac{u(r)}{J}\int_{\mathcal{B}_r^M(x)}\phi(y)\mathrm{d}y\Big) - \frac{X_0^N(z)}{K} + \frac{u(r)}{J}\int_{\mathcal{B}_r^M(x)}\phi(y)X_0^N(y)\mathrm{d}y\end{aligned}$$

$$+\frac{X_0^N(z)}{K}\frac{u(r)}{J}\int_{\mathcal{B}_r^M(x)}\phi(y)X_0^N(y)\mathrm{d}y\;\mathcal{O}\Big(\Big(\frac{Ku(r)}{J}\int_{\mathcal{B}_r^M(x)}\phi(y)\mathrm{d}y)\Big)^2\Big)$$
$$+\mathcal{O}\Big(\frac{u(r)}{J}\int_{\mathcal{B}_r^M(x)}\phi(y)X_0^N(y)\mathrm{d}y\Big)^2$$

from which, noting that, under the assumptions of Theorem 2.2.7, $K/(JM^d)\to 0$,

$$\mathcal{L}_{\exp(-\cdot)}(\phi)(X_0^N)=NM^d\exp(-\langle X_0^N,\phi_0\rangle)\Bigg[\int_{\mathbb{R}^d}\int_0^\infty\frac{1}{|\mathcal{B}_r^M|}\int_{\mathcal{B}_r^M(x)}\Bigg\{$$
$$\frac{X_0^N(z)}{K}\exp\Bigg(-K\frac{u(r)}{J}\int_{\mathcal{B}_r^M(x)}\phi(y)\mathrm{d}y\Bigg)+\frac{u(r)}{J}\int_{\mathcal{B}_r^M(x)}\phi(y)X_0^N(y)\mathrm{d}y-\frac{X_0^N(z)}{K}\Bigg\}\mathrm{d}z\mu^N(\mathrm{r})\mathrm{d}x$$
$$+\int_{\mathbb{R}^d}\int_0^\infty\mathcal{O}\Bigg(\frac{u(r)^2}{J^2}\Bigg(\int_{\mathcal{B}_r^M(x)}\phi(y)X_0^N(y)\mathrm{d}y\Bigg)^2\Bigg)\mu^N(\mathrm{d}r)\mathrm{d}x\Bigg].$$

The analogue of Lemma 2.4.13 that we need in this context is

Lemma 2.4.14. *Under the assumptions of Theorem 2.2.7, for $\phi\in C_0^3(\mathbb{R}^d)$,*

$$\mathbb{E}\Bigg[\frac{N}{J^2M^d}\int_0^t\int_{\mathbb{R}^d}\int_0^\infty u(r)^2r^{2d}\Bigg(\frac{1}{|\mathcal{B}_r^M|}\int_{\mathcal{B}_r^M(x)}\phi(y)X_s^N(y)\mathrm{d}y\Bigg)^2\mu^N(\mathrm{d}r)\mathrm{d}x\mathrm{d}s\Bigg]\to 0$$

as $N\to\infty$.

Once again the proof, which relies on duality, is deferred to Section 2.6.

Rearranging the expression for $\mathcal{L}_{\exp(-\cdot)}(\phi)(X_0^N)$ and reversing the order of integration we find

$$\mathcal{L}_{\exp(-\cdot)}(\phi)(X_0^N)=\exp\Big(-\langle X_0^N,\phi\rangle\Big)\Bigg[\langle X_0^N,A^N\phi+B^N\phi\rangle$$
$$+\int_{\mathbb{R}^d}\int_0^\infty\mathcal{O}\Bigg(\frac{Nu(r)^2r^{2d}}{J^2M^d}\Bigg)\Bigg(\frac{1}{|\mathcal{B}_r^M|}\int_{\mathcal{B}_r^M(x)}\phi(y)X_0^N(y)\mathrm{d}y\Bigg)^2\mu^N(\mathrm{d}r)\mathrm{d}x\Bigg],$$

where

$$A^N\phi(x):=\frac{NM^d}{K}\int_0^\infty\frac{1}{|\mathcal{B}_r^M|}\int_{\mathcal{B}_r^M(x)}\exp\Bigg(-K\frac{u(r)}{J}\int_{\mathcal{B}_r^M(x)}\phi(y)\mathrm{d}y\Bigg)$$
$$-\exp\Bigg(-|\mathcal{B}_r^M|\frac{Ku(r)}{J}\phi(x)\Bigg)\mathrm{d}z\mu^N(\mathrm{d}r),$$
$$B^N\phi(x):=\frac{NM^d}{K}\int_0^\infty\Bigg[\exp\Bigg(-|\mathcal{B}_r^M|\frac{Ku(r)}{J}\phi(x)\Bigg)+|\mathcal{B}_r^M|\frac{Ku(r)}{J}\phi(x)-1\Bigg]\mu^N(\mathrm{d}r).$$

Consider $A^N\phi(x)$. Again by Taylor expansion, we have

$$\begin{aligned}A^N\phi(x) =&\frac{NM^d}{J}\left[\int_0^\infty \frac{1}{|\mathcal{B}_r^M|}\int_{\mathcal{B}_r^M(x)}\int_{\mathcal{B}_r^M(z)} -u(r)(\phi(y)-\phi(x))\mathrm{d}y\mathrm{d}z\mu^N(\mathrm{d}r)\right]\\ &+\int_0^\infty \frac{1}{|\mathcal{B}_r^M|}\int_{\mathcal{B}_r^M(x)} \mathcal{O}\left(\frac{KM^dN}{J^2}\right)u(r)^2\left(\int_{\mathcal{B}_r^M(z)}\phi(y)-\phi(x)\mathrm{d}y\right)^2\mathrm{d}y\mathrm{d}z\mu^N(\mathrm{d}r)\\ =&-\frac{NM^d}{J}\left[\int_0^\infty \frac{1}{|\mathcal{B}_r^M|}\int_{\mathcal{B}_r^M(x)}\int_{\mathcal{B}_r^M(z)} u(r)(\phi(y)-\phi(x))\mathrm{d}y\mathrm{d}z\mu^N(\mathrm{d}r)\right] \qquad (2.4.4)\\ &+\int_0^\infty \mathcal{O}\left(\frac{KN\|\phi\|_{C^1}^2}{J^2M^{d+2}}\right)u(r)^2r^{2d+2}\,\mu^N(\mathrm{d}r),\end{aligned}$$

where the last line follows by observing that $\phi(y)-\phi(x)=\mathcal{O}(r\|\phi\|_{C^1}/M)$ for $y\in\mathcal{B}_r^M(x)$.

We note that the conditions $\gamma-d<\frac{1}{1-\beta}$ and $\alpha+1=(\beta+1)(\gamma-d)$ will imply $\int_0^\infty u(r)^2r^{2d+2}\mu^N(\mathrm{d}r)<\infty$. The calculations of Section 2.4.1 then allow us to write

$$\begin{aligned}A^N\phi(x)=\int_0^\infty\Bigg[&-\frac{C(d)N}{JM^2}u(r)r^{d+2}\Delta\phi+\mathcal{O}\left(\frac{N\|\phi\|_{C^3}}{JM^3}\right)u(r)r^{d+3}\\ &+\mathcal{O}\left(\frac{KN\|\phi\|_{C^3}^2}{J^2M^{d+2}}\right)u(r)^2r^{2d+2}\Bigg]\mu^N(\mathrm{d}r),\end{aligned}$$

where $C(d):=\int_{B_1(0)}x^2\mathrm{d}x$.

We now specialise to $u(r)=r^{-\gamma}$ and $\mu^N(\mathrm{d}r):=r^\alpha\mathbf{1}_{\left\{J^{\frac{-1}{\gamma}}<r<1\right\}}\mathrm{d}r$ as specified in Theorem 2.2.7. However, it should be clear that other choices would result in nontrivial scaling limits.

Substituting into our expression for $B(x)$ and making the change of variable $\upsilon=|\mathcal{B}_r^M|Ku(r)/J$

$$B^N\phi(x):=\frac{NM^d}{K(\gamma-d)}\left(\frac{K|\mathcal{B}_1|}{JM^d}\right)^{\frac{\alpha+1}{\gamma-d}}\int_{\frac{K|\mathcal{B}_1|}{JM^d}}^{J^{\frac{\gamma-d}{\gamma}}\frac{K|\mathcal{B}_1|}{JM^d}} g(\upsilon\phi(x))\upsilon^{-\left(\frac{\alpha+1}{\gamma-d}+1\right)}\mathrm{d}\upsilon, \qquad (2.4.5)$$

where $g(\upsilon)=\exp(-\upsilon)+\upsilon-1$.

Now

$$\int_0^\infty g(\upsilon\phi(x))\upsilon^{-\beta-2}\mathrm{d}\upsilon=\kappa_0\phi(x)^{\beta+1},$$

where κ_0 is a constant, and so to recover the superBrownian motion with stable branching law with index β in this limit, we choose $\alpha+1=(\beta+1)(\gamma-d)$ (to get the right exponent)

and $J^{\frac{\gamma-d}{\gamma}} K/(JM^d) \to \infty$, to ensure that the upper limit of integration tends to infinity. The lower limit of integration $K/(JM^d) \to 0$ was imposed already in order for $A^N\phi(x)$ to converge to a nontrivial limit. To ensure that the term premultiplying the integral in (2.4.5) is positive and finite, we take $\gamma - d > 0$ and assume that $\frac{N}{J}\Big(\frac{K}{JM^d}\Big)^\beta \to C_2$.

Granted Lemma 2.4.14, we have now shown that under the conditions of Theorem 2.2.7, for any $\phi \in C_0^3(\mathbb{R}^d)$, as $N \to \infty$,

$$\mathcal{L}_{\exp(-\cdot)}(\phi)(X_t^N) \to \langle X_t^N, -\frac{m}{2}\Delta\phi + \kappa\phi^{\beta+1}\rangle \exp\Big(-\langle X_t^N, \phi\rangle\Big),$$

where $m, \kappa \in (0, \infty)$ are as in the statement of the Theorem.

2.5 Tightness

In this section we turn to the proof of tightness of the sequence $\{X^N\}_{N\geq 1}$ in the space of càdlàg $\mathcal{M}_F(\mathbb{R}^d)$-valued processes. In the fixed radius case, we shall also check that all limit points are actually continuous processes.

To prove tightness, we appeal to a specialised version of Jakubowski's general criterion that we have taken from Cox et al. (2000). For a Borel set A, let $X_t^N(A) := \langle X_t^N, \mathbf{1}_A\rangle$.

Proposition 2.5.15 (Cox et al. (2000), Proposition 3.1)**.** *Let $\Phi \subset C_b(\mathbb{R}^d)$ be a separating class which is closed under addition. A sequence of càdlàg $\mathcal{M}_F(\mathbb{R}^d)$-valued processes $\{X^N\}_{N\geq 1}$ is tight if and only if the following conditions hold:*

1. *For each T, $\varepsilon > 0$ there is a compact set $K_{T,\varepsilon} \subset \mathbb{R}^d$ such that*

$$\sup_N \mathbb{P}\left[\sup_{t\leq T} X_t^N\Big(K_{T,\varepsilon}^c\Big) > \varepsilon\right] < \varepsilon. \tag{2.5.1}$$

2. *For each $T > 0$, $\lim_{H\to\infty} \sup_N \mathbb{P}[\sup_{t\leq T}\langle X_t^N, \mathbf{1}\rangle \geq H] = 0$.*

3. *For each $\phi \in \Phi$, $\{\langle X_\cdot^N, \phi\rangle\}_{N\geq 1}$ is tight.*

In fact it is convenient to slightly modify the statements of Conditions 1 and 2.

Corollary 2.5.16. *The conclusion of Proposition 2.5.15 remains valid if Condition 1 and 2 are replaced by:*

$1'$. *For each T, $\varepsilon > 0$ there is a compact set $K_{T,\varepsilon} \subset \mathbb{R}^d$ such that*

$$\limsup_{N\to\infty} \mathbb{P}\left[\sup_{t\le T} X_t^N\left(K_{T,\varepsilon}^c\right) > \varepsilon\right] < \varepsilon, \tag{2.5.2}$$

$2'$ *For each $T > 0$,*

$$\lim_{H\to\infty} \limsup_{N\to\infty} \mathbb{P}[\sup_{t\le T}\langle X_t^N, \mathbf{1}\rangle \ge H] = 0. \tag{2.5.3}$$

Proof We mimic the proof of Ethier and Kurtz (1986), Chapter 3, Corollary 7.4.

Suppose that (2.5.2) is satisfied, then given $\varepsilon > 0$, there exists a compact set $K_{T,\varepsilon}^0$ and an integer N_0 such that for all $N > N_0$

$$\mathbb{P}\left[\sup_{t\le T} X_t^N\left((K_{T,\varepsilon}^0)^c\right) > \varepsilon\right] < \varepsilon.$$

For each $N \le N_0$ (c.f. the argument at the beginning of Section 2.3 that X_t^N has compact support for all $t \le T$), there is a set $K_{T,\varepsilon}^N$ such that

$$\mathbb{P}\left[\sup_{t\le T} X_t^N\left((K_{T,\varepsilon}^N)^c\right) > \varepsilon\right] < \varepsilon.$$

Set $K_{T,\varepsilon} = \bigcup_{N=0}^{N_0} K_{T,\varepsilon}^N$ and Condition 1 of Proposition 2.5.15 is satisfied.

The proof that $2'$ implies 2 is similar. □

We borrow a convenient formulation of the additional criterion for the limit points to be continuous from the same paper.

Corollary 2.5.17 (Cox et al. (2000), Corollary 3.2)**.** *If a sequence of measure valued processes satisfy the conditions of Proposition 2.5.15 with $\Phi = C_0^\infty(\mathbb{R}^d)$ and for each $\phi \in \Phi$, every limit point of $\{\langle X_\cdot^N, \phi\rangle\}_{N\ge1}$ is supported on the space of continuous functions, then $\{X^N\}_{N\ge1}$ is tight and all limit points are continuous.*

These two results reduce much of the work in proving tightness to an examination of the one-dimensional projections $\{\langle X_\cdot^N, \phi\rangle\}_{N\ge1}$. For this we shall make use of the calculations of Section 2.4 which showed, in particular, that $\mathcal{L}^N(\phi)(X_t^N)$ is close to $\langle X_t^N, \mathcal{A}^N\phi\rangle$ where $\mathcal{A}^N$, defined in Definition 2.4.11, is $m(N)\Delta/2$. This will allow us to

exploit the following elementary properties of the heat equation which, for convenience, we record as a lemma.

Lemma 2.5.18. *Suppose that $\psi : \mathbb{R}^d \to \mathbb{R} \in C_0^3$ and that v is a classical solution to*

$$\frac{\partial v}{\partial t} = \frac{m}{2}\Delta v, \qquad v(0,x) = \psi(x), \tag{2.5.4}$$

where the diffusion coefficient is a constant $m \in (0,\infty)$.

1. *If ψ is uniformly Lipschitz continuous with Lipschitz-constant M, then for all t, $v(t,\cdot)$ is Lipschitz continuous with Lipschitz constant M.*
2. *For any $k \in \mathbb{N}$, for each t, the C^k norm of $v(t,x)$ (as a function of x) is bounded above by that of ψ.*
3. *If v solves the equation with initial data ψ then Δv solves the same equation with initial data $\Delta\psi$.*

Notation 2.5.19. *We denote by $P_t^{(m)}$ the heat semi-group with diffusion coefficient m, so that the solution to (2.5.4) can be written*

$$v(t,x) = P_t^{(m)}(\psi)(x) = \mathbb{E}_x[\psi(B_{mt})],$$

where B_t is a standard Brownian Motion.

2.5.1 Verification of Condition $2'$ of Corollary 2.5.16

Lemma 2.5.20. *Under the conditions of either Theorem 2.2.6 or Theorem 2.2.7, for each $T > 0$,*

1. $$\mathbb{E}[\langle X_t^N, \mathbf{1}\rangle] \le \langle X_0^N, \mathbf{1}\rangle \exp\Big(\mathcal{O}(\frac{N}{JM^3})t\Big), \qquad \forall t \le T; \tag{2.5.5}$$

2. $$\lim_{H\to\infty} \sup_N \mathbb{P}[\sup_{t\le T}\langle X_t^N, \mathbf{1}\rangle \ge H] = 0.$$

Proof

1. We set $\chi^N(t) = \mathbb{E}[\langle X_t^N, \mathbf{1}\rangle]$.

Let $\varepsilon > 0$ and let $\{h^R\}_{R\geq 1}$ be a sequence of smooth, compactly supported functions on $\mathbb{R}^d$ such that

$$h^R|_{B_R(0)} = 1, \quad h^R|_{B_{2R}(0)^c} = 0, \quad \Delta h^R \leq \epsilon, \quad |h^R| \leq 1, \quad h^R \leq h^{R+1}.$$

We can further arrange that $\|h^R\|_{C^3} \leq D$ for all R.

In the notation of Lemma 2.4.12,

$$\langle X_t^N, h^R\rangle = \langle X_0^N, h^R\rangle + \int_0^t \langle X_s^N, \mathcal{A}^N(h^R)\rangle \mathrm{d}s + \zeta_t^N(h^R) + M_t^N(h^R).$$

Taking expectations,

$$\begin{aligned}\mathbb{E}[\langle X_t^N, h^R\rangle] &= \langle X_0^N, h^R\rangle + \mathbb{E}\Big[\int_0^t \langle X_s^N, \mathcal{A}^N(h^R)\rangle \mathrm{d}s\Big] + \mathbb{E}[\zeta_t^N(h^R)] \\ &\leq \chi^N(0) + C\|\Delta h^R\|_\infty \int_0^t \chi^N(s)\mathrm{d}s + \mathcal{O}\Big(\frac{N\|h^R\|_{C^3}}{JM^3}\Big)\int_0^t \chi^N(s)\mathrm{d}s.\end{aligned}$$

We have used the fact that, under our assumptions, $m(N)$ converges to bound the constant in front of the Laplacian, and the bound on the error $\mathbb{E}[\zeta_t^N(h^R)]$ from Lemma 2.4.12. Letting $R \uparrow \infty$ and applying the Monotone Convergence Theorem,

$$\chi^N(t) \leq \chi^N(0) + \Big(C\epsilon + \mathcal{O}\big(\frac{N\|h^R\|_{C^3}}{JM^3}\big)\Big)\int_0^t \chi^N(s)\mathrm{d}s,$$

and applying Gronwall's inequality and using that ε was arbitrary,

$$\chi^N(t) \leq \chi^N(0)\exp\Big(\mathcal{O}\big(\frac{N}{JM^3}\big)t\Big),$$

as required.

2. Define a sequence of stopping times by

$$\tau^N := \inf\{t \geq 0 : \langle X_t^N, \mathbf{1}\rangle \geq H\},$$

and set $\widehat{X}_t^N := X^N_{t\wedge\tau^N}$. Then

$$\langle \widehat{X}_t^N, \phi\rangle = \langle \widehat{X}_0^N, \phi\rangle + \int_0^{t\wedge\tau^N} \langle \widehat{X}_s^N, \mathcal{A}^N(\phi)\rangle \mathrm{d}s + \zeta^N_{t\wedge\tau^N}(\phi) + \widehat{M}_t^N(\phi),$$

where $\widehat{M}_t^N(\phi) = M^N_{t\wedge\tau^N}$. Write $\widehat{\chi}^N(t) = \mathbb{E}[\langle \widehat{X}_t^N, \mathbf{1}\rangle]$, and proceed as above, but with the stopped martingale problem, to obtain

$$\begin{aligned}
\mathbb{E}[\langle \widehat{X}_t^N, h^R\rangle] =& \langle \widehat{X}_0^N, h^R\rangle + \mathbb{E}\left[\int_0^{t\wedge\tau^N} \langle \widehat{X}_s^N, \mathcal{A}(h^R)\rangle \mathrm{d}s\right] + \mathbb{E}[\zeta^N_{t\wedge\tau^N}(h^R)] + \mathbb{E}[\widehat{M}_t^N(h^R)] \\
\leq& \widehat{\chi}^N(0) + C\epsilon\Big[\int_0^{t\wedge\tau^N} \langle \widehat{X}_s^N, \mathbf{1}\rangle \mathrm{d}s\Big] + \mathcal{O}\Big(\frac{NT\|h^R\|_{C^3}}{JM^3}\Big)\mathbb{E}[\sup_{s\leq t}\langle \widehat{X}_s^N, \mathbf{1}\rangle] \\
\leq& \chi^N(0) + C\epsilon\int_0^t \chi^N(s)\mathrm{d}s + \mathcal{O}\left(\frac{HTN\|h^R\|_{C^3}}{JM^3}\right).
\end{aligned}$$

The Monotone Convergence Theorem gives $\mathbb{E}[\langle \widehat{X}_t^N, h^R\rangle] \to \mathbb{E}[\langle \widehat{X}_t^N, \mathbf{1}\rangle]$ and so $\mathbb{E}[\langle \widehat{X}_t^N, \mathbf{1}\rangle] \leq \chi^N(0) + \mathcal{O}\Big(\frac{HTND}{JM^3}\Big)$. then Gronwall's inequality implies

$$\mathbb{E}[\langle \widehat{X}_t^N, \mathbf{1}\rangle] \leq \left(\mathbb{E}[\langle \widehat{X}_0^N, \mathbf{1}\rangle] + \mathcal{O}\Big(\frac{HTND}{JM^3}\Big)\right)\exp(Ct).$$

Markov's inequality now gives

$$\mathbb{P}[\langle \widehat{X}_t^N, \mathbf{1}\rangle \geq H] \leq \frac{\mathbb{E}[\langle \widehat{X}_t^N, \mathbf{1}\rangle]}{H} \leq \left(\frac{\mathbb{E}[\langle X_0^N, \mathbf{1}\rangle]}{H} + \mathcal{O}\left(\frac{NT\|h^R\|_{C^3}}{JM^3}\right)\right)\exp(Ct). \quad (2.5.6)$$

Noting that $\mathbb{P}[\sup_{t\leq T}\langle X_t^N, \mathbf{1}\rangle \geq H] = \mathbb{P}[\langle \widehat{X}_T^N, \mathbf{1}\rangle \geq H]$, we can now conclude, since for any T, $\varepsilon > 0$ we can choose H_1, N_1 such that for $H \geq H_1$, $N \geq N_1$ and $t \leq T$ the right hand side of (2.5.6) is less than ε. □

Remark 2.5.21. *An elegant proof of the first part of Lemma 2.5.20 takes a sequence $\psi_n \uparrow \mathbf{1}$ in the duality relation (2.2.3) with $k = 1$. Symmetry of the jump distribution of ancestral lineages, Fubini's Theorem, and monotone convergence will yield the result. One can also show, under our particular choice of initial conditions and event structure, that $\langle X_t^N, \mathbf{1}\rangle$ is a martingale. However, one needs to take care to show that the support of X_t^N remains in a compact set.*

2.5.2 Verification of Condition $1'$ of Corollary 2.5.16

Lemma 2.5.22. *Under the conditions of either Theorem 2.2.6 or Theorem 2.2.7, for each T, $\varepsilon > 0$ there is a compact set $K_{T,\varepsilon} \subset \mathbb{R}^d$ such that*

$$\limsup_{N\to\infty} \mathbb{P}\left[\sup_{t\le T} X_t^N\left(K_{T,\varepsilon}^c\right) > \varepsilon\right] < \varepsilon \tag{2.5.7}$$

Proof

Let $h \in C^\infty(\mathbb{R}^d)$ satisfy

$$h|_{\mathcal{B}_1(0)} = 0, \quad \text{and} \quad h|_{\mathcal{B}_2(0)^c} = 1.$$

Define $h_n(x) := h(x/n)$. It will evidently suffice to show that for sufficiently large n,

$$\limsup_{N\to\infty} \mathbb{P}\left[\sup_{t\le T}\langle X_t^N, h_n\rangle > \varepsilon\right] < \varepsilon. \tag{2.5.8}$$

We note that the C^3 norms of Δh_n are uniformly bounded in n, so that by Lemma 2.5.18 the C^3 norms of $\Delta P_t^{(m)}(h_n)$ are uniformly bounded in n and t.

With $m(N)$ as in Definition 2.4.11 we set $\phi_s := \Delta P_{t-s}^{(m(N))}(h_n)$, so that

$$\dot{\phi}_s(x) + \mathcal{A}^N(\phi_s)(x) = 0.$$

By Lemma 2.4.12 and Lemma 2.5.18

$$\left|\int_0^t \mathcal{L}^N(P_{t-s}^{(m(N))}(h_n))(X_s^N)\mathrm{d}s\right| = \mathcal{O}\left(\frac{N\|h\|_{C^3}}{JM^3}\right)\int_0^t \langle X_s^N, \mathbf{1}\rangle \mathrm{d}s, \tag{2.5.9}$$

and this bound is uniform in n. In particular, using (2.5.5),

$$\begin{aligned}\mathbb{E}[\langle X_t^N, h_n\rangle] &= \mathbb{E}[\langle X_0^N, P_t^{(m(N))}h_n\rangle] + \mathbb{E}\left[\mathcal{O}\left(\frac{N\|h\|_{C^3}}{JM^3}\right)\int_0^t \langle X_s^N, \mathbf{1}\rangle \mathrm{d}s\right] \\ &\le \mathbb{E}[\langle X_0^N, P_t^{(m(N))}h_n\rangle] + \mathcal{O}\left(\frac{N}{JM^3}\right) T\exp\left(\mathcal{O}(\frac{N\|h\|_{C^3}}{JM^3})T\right)\langle X_0^N, \mathbf{1}\rangle.\end{aligned} \tag{2.5.10}$$

We recall that

$$\langle X_t^N, h_n\rangle = \langle X_0^N, h_n\rangle + \int_0^t \langle X_s^N, \mathcal{A}^N(h_n)\rangle \mathrm{d}s + \zeta_t^N(h_n) + M_t^N(h_n). \tag{2.5.11}$$

We shall consider each term on the right hand side separately. The first term will be zero for sufficiently large n, since $\mathrm{supp}(X_0^N) \subseteq D$ for all N, where $D \subseteq \mathbb{R}^d$ is compact. The bound (2.5.10) will help us to control the integral term. First note that, since $m(N)$ converges,

$$\begin{aligned}\mathbb{E}\Big[\sup_{t\leq T}\Big|\int_0^t \langle X_s^N, \mathcal{A}^N(h_n)\rangle \mathrm{d}s\Big|\Big] \leq& \mathbb{E}\Big[\int_0^T \langle X_s^N, |\mathcal{A}^N(h_n)|\rangle \mathrm{d}s\Big] \\ \leq& C\mathbb{E}\Big[\int_0^T \langle X_s^N, \mathbf{1}_{\mathcal{B}_{n-1}(0)^c}\rangle \mathrm{d}s\Big],\end{aligned}$$

where C is independent of n and N. Now set $m := \lfloor \frac{n-1}{2} \rfloor$.

$$\begin{aligned}&\mathbb{E}\Big[\int_0^T \langle X_s^N, \mathbf{1}_{\mathcal{B}_{n-1}(0)^c}\rangle \mathrm{d}s\Big] \leq \mathbb{E}\Big[\int_0^T \langle X_s^N, h_m\rangle \mathrm{d}s\Big] \leq \int_0^T \mathbb{E}[\langle X_s^N, h_m\rangle] \mathrm{d}s \\ &\leq \int_0^T \langle X_0^N, P_t^{(m(N))}(h_m)\rangle \mathrm{d}s + \mathcal{O}\Big(\frac{N\|h_m\|_{C^3}}{JM^3}\Big) T^2 \exp\Big(\mathcal{O}(\frac{N}{JM^3})T\Big)\langle X_0^N, \mathbf{1}\rangle,\end{aligned} \tag{2.5.12}$$

where we used (2.5.10) to obtain (2.5.12). Under the assumptions of Theorem 2.2.6 or Theorem 2.2.7, $N/(JM^2) \to C_1$ and $M \to \infty$ as $N \to \infty$ and so we can take N sufficiently large that the error term is less than $\varepsilon^2/32$.

Denoting standard Brownian motion by $\{B_t\}_{t\geq 0}$, consider now

$$\begin{aligned}\int_0^T \langle X_0^N, P_s^{(m(N))}(h_m)\rangle \mathrm{d}s =& \int_0^T \langle X_0^N, \mathbb{E}^x[h_m(B_{m(N)s})]\rangle \mathrm{d}s \\ \leq& \int_0^T \langle X_0^N, \mathbb{E}^x[\mathbf{1}_{\mathcal{B}_m(0)^c}(B_{m(N)s})]\rangle \mathrm{d}s \\ \leq& T\Big[X_0^N(\mathcal{B}_{m/2}(0)^c) + X_0^N(\mathcal{B}_{m/2}(0)) \sup_{s\leq T} \mathbb{P}^0[|B_{m(N)s}| \geq m/2]\Big] \\ \leq& T\Big[X_0^N(\mathcal{B}_{m/2}(0)^c) + X_0^N(\mathbb{R}^d) \sup_{s\leq T} \mathbb{P}^0[|B_{m(N)s}| \geq m/2]\Big],\end{aligned} \tag{2.5.13}$$

which (by our assumptions that $\mathrm{supp}(X_0^N) \subseteq D$ for all N and that $m(N)$ converges) tends to 0 as $m \to \infty$ uniformly in N, so in particular is $< \varepsilon^2/32$ for sufficiently large m.

Combining the estimates above, we have shown that given $\varepsilon > 0$, there exist N_0, n_0 such that for $N \geq N_0$ and $n \geq n_0$,

$$\mathbb{E}\left[\sup_{t\leq T}\left|\int_0^t \langle X_s^N, \mathcal{A}^N(h_n)\rangle \mathrm{d}s\right|\right] < \varepsilon^2/16.$$

From Lemma 2.4.12,

$$\sup_{t\leq T}|\zeta_t^N(h_n)| \leq \mathcal{O}\Big(\frac{N\|h_n\|_{C^3}}{JM^3}\Big)\int_0^T \langle X_s^N, \mathbf{1}\rangle \mathrm{d}s$$

and we can estimate the right hand side through (2.5.5) from Lemma 2.5.20 and see that for sufficiently large N it is bounded above by $\varepsilon^2/16$.

In order to control the martingale term we shall control $\mathbb{E}[|M_T^N(h_n)|]$ and then apply Doob's inequality.

For sufficiently large N, the argument that gave us (2.5.13) allows us to bound (2.5.10) with $t = T$ by $\varepsilon^2/8$. Rearranging (2.5.11) and using the triangle inequality, given $\varepsilon > 0$, there exist N_0, n_0, such that uniformly in $n \geq n_0$, for $N \geq N_0$, $\mathbb{E}[|M_T^N(h_n)|] \leq \varepsilon^2/4$. Therefore by Doob's maximal inequality, for any such n and N, we have that

$$\mathbb{P}[\sup_{t\leq T}|M_t^N(h_n)| > \frac{\varepsilon}{2}] \leq \frac{2\mathbb{E}[|M_T^N(h_n)|]}{\varepsilon} < \frac{\varepsilon}{2}.$$

Combining the estimates above with another application of Markov's inequality to the sum of the remaining terms on the right hand side of (2.5.11) we have verified (2.5.8) and the proof is complete. □

2.5.2.1 Tightness of projections: fixed radius case

We shall verify Condition 3 of Proposition 2.5.15 separately for our two cases. In the fixed radius case it is relatively straightforward as the quadratic variation of the martingale part of our semimartingale decomposition (2.5.11) will remain bounded as $N \to \infty$. In this section we shall deal with that case and moreover check the conditions

of Corollary 2.5.17 so that we can deduce that all the limit points of $\{X^N\}_{N\geq 1}$ are in fact continuous processes. We shall exploit some standard results that we record here for convenience.

Proposition 2.5.23 (Jacod and Shiryaev (2003), Chapter VI, Part of Proposition 3.26)**.** *A sequence $\{Y^N\}_{N\geq 1}$ of càdlàg $\mathbb{R}^d$-valued processes is tight and all limit points of the sequence of laws of Y^N are laws of continuous processes if and only if the following two conditions are satisfied:*

1. *$\forall T \in \mathbb{N}$, $\varepsilon > 0$, $\exists N_0 \in \mathbb{N}$, $\Gamma \in \mathbb{R}$ such that*

$$N \geq N_0 \Rightarrow \mathbb{P}\Big[\sup_{t\leq T} |Y_t^N| > \Gamma\Big] \leq \varepsilon.$$

2. *$\forall T \in \mathbb{N}$, $\varepsilon > 0$, $\eta > 0$, $\exists N_0 \in \mathbb{N}$, $\delta > 0$ such that*

$$N \geq N_0 \Rightarrow \mathbb{P}\Big[\sup_{0\leq t\leq T-\delta} \Big(\sup_{s_1,s_2\in[t,t+\delta]} |Y_{s_1}^N - Y_{s_2}^N|\Big) > \eta\Big] \leq \varepsilon.$$

Proposition 2.5.24 (Jacod and Shiryaev (2003), Chapter VI, Part of Proposition 3.26)**.** *A sequence $\{Y^N\}_{N\geq 1}$ is tight and all limit points of the sequence of laws of Y^N are laws of continuous processes if and only if $\{Y^N\}_{N\geq 1}$ is tight and for all $T > 0$, $\varepsilon > 0$,*

$$\lim_{N\to\infty} \mathbb{P}[\sup_{t\leq T} |Y_t^N - Y_{t-}^N| > \varepsilon] = 0.$$

Theorem 2.5.25 (Jacod and Shiryaev (2003), Chapter VI, Theorem 4.13)**.** *If for each N, M^N is a locally square integrable local martingale, then sufficient conditions for the sequence $\{M^N\}_{N\geq 1}$ to be tight are:*

1. *The sequence $\{M_0^N\}_{N\geq 1}$ is tight.*

2. *The sequence $\{\langle M^N\rangle\}_{N\geq 1}$ is tight with all limit points being continuous.*

Lemma 2.5.26. *Under the assumptions of Theorem 2.2.6, for each $\phi \in C_0^3(\mathbb{R}^d)$, $\{\langle X_\cdot^N, \phi\rangle\}_{N\geq 1}$ is tight and all the limit points are continuous.*

Proof Consider $Y_t^N := \langle X_t^N, \phi\rangle$ with $\phi \in C_0^3(\mathbb{R}^d)$. Since $|Y_t^N| \leq \|\phi\|_\infty \langle X_t^N, \mathbf{1}\rangle$, the first condition of Proposition 2.5.23 follows immediately from our verification of Condition $2'$

of Corollary 2.5.16. To check the second condition of Proposition 2.5.23, we once again use the semimartingale decomposition

$$\langle X_t^N, \phi\rangle = \langle X_0^N, \phi\rangle + \int_0^t \langle X_s^N, \mathcal{A}^N(\phi)\rangle \mathrm{d}s + \zeta_t^N(\phi) + M_t^N(\phi).$$

Evidently

$$\left|\int_0^{s_1} \langle X_s^N, \mathcal{A}^N(\phi)\rangle \mathrm{d}s - \int_0^{s_2} \langle X_s^N, \mathcal{A}(\phi)\rangle \mathrm{d}s\right| \leq C(\phi)|s_1 - s_2| \sup_{t\leq T}\langle X_t^N, \mathbf{1}\rangle$$

and combining with

$$|\zeta_t^N(\phi)| \leq \mathcal{O}\left(\frac{Nur^{d+3}\|\phi\|_{C^3}}{JM^3}\right) t \sup_{s\leq t}\langle X_s^N, \mathbf{1}\rangle$$

and (2.5.6) we see that, under the conditions of Theorem 2.2.6, both these terms satisfy Condition 2 of Proposition 2.5.23 and our problem is reduced to showing tightness with continuous limits of the martingale part. Theorem 2.5.25 tells us that it suffices to consider the quadratic variation.

Recall from (2.4.3) that

$$\begin{aligned}\langle M^N(\phi)\rangle_t = \int_0^t \int_{\mathbb{R}^d} \frac{KN|\mathcal{B}_r|u^2}{J^2M^d}\Big[\Big(\phi(x)^2 X_s^N(x) - \frac{1}{K}\Big(\frac{M^d}{|\mathcal{B}_r|}\int_{\mathcal{B}_r^M(x)} \phi(y) X_s^N(y)\mathrm{d}y\Big)^2\Big) \\ + \mathcal{O}(\frac{\|\phi\|_{C^1}^2}{M}) X_s^N(x)\Big]\mathrm{d}x\mathrm{d}s \\ \leq \int_0^t \int_{\mathbb{R}^d} \frac{KN|\mathcal{B}_r|u^2}{J^2M^d}\Big[\langle X_s^N, \phi^2\rangle + \mathcal{O}(\frac{\|\phi\|_{C^1}^2}{M})\langle X_s^N, 1\rangle\Big]\mathrm{d}s. \quad (2.5.14)\end{aligned}$$

Using Proposition 2.5.23 and our bounds on $\langle X_s^N, \mathbf{1}\rangle$ from Lemma 2.5.20, the tightness of $\{\langle M^N(\phi)\rangle\}_{N\geq 1}$ is immediate.

We now conclude through an application of Proposition 2.5.24. The probability that the support of X^N, which is compact, is simultaneously overlapped by two events of the SLFV is zero, and ϕ is bounded, so

$$\sup_{t\leq T}|Y_t^N - Y_{t-}^N| \leq \frac{C(d)K}{JM^d}, \quad \mathbb{P}-a.s..$$

□

Proposition 2.5.27. *Under the conditions of Theorem 2.2.6, $\{X^N\}_{N\geq 1}$ is tight and all limit points are continuous.*

Proof This is now immediate from Corollary 2.5.17. □

2.5.2.2 Tightness of projections: variable radius case

Our proof in the fixed radius case breaks down in the variable radius setting since the quadratic variation of the martingale part of $\langle X^N_\cdot, \phi\rangle$ will grow without bound as $N \to \infty$. Instead we exploit the fact that the $(1+\theta)$th moments of $\sup_{0\leq t\leq T}\langle X^N_t, \phi\rangle$ will remain bounded for any $\theta < \beta$. We follow Dawson (1993) who used essentially the same argument to show that branching Brownian motion with stable branching law, suitably rescaled, converges to superBrownian motion with stable branching, although there are some extra layers of estimation in our setting.

We exploit the following elementary lemma, which we learned from a preliminary version of Dawson (1993), but we record here since it does not appear in the published version.

Lemma 2.5.28. *There exist $c_1, c_2, c_3 \in (0,\infty)$ such that*

1. $1 - x + \frac{x^2}{2} - \exp(-x) \geq 0$, *for* $x \geq 0$.
2. $\frac{1}{x}\Big(1 - x + \frac{x^2}{2} - \exp(-x)\Big) \geq c_1 > 0$ *for* $x \geq 2$.
3. $0 \leq \exp(-x) + x - 1 \leq c_2 x^{1+\beta}$, *for* $x \geq 0$ *where* $0 < \beta \leq 1$.
4. *For a non-negative random variable Y and $\theta \neq -1$,*

$$\mathbb{E}[Y^{1+\theta}] \leq 2 + (1+\theta)\int_1^\infty y^\theta \mathbb{P}[Y \geq y]\mathrm{d}y.$$

5. *For a non-negative random variable Y and any $y \geq 1$,*

$$\mathbb{P}[Y \geq y] \leq c_3 y \int_0^{\frac{2}{y}} \mathbb{E}[\exp(-\lambda Y) - 1 + \lambda Y]\mathrm{d}\lambda.$$

Proof(sketch) For 4, note that

$$x^{1+\theta} = 1 + (1+\theta)\int_1^x y^\theta \mathrm{d}y,$$

and so

$$\begin{aligned}\mathbb{E}[Y^{1+\theta}] &\leq 1 + \int_1^\infty \Big(1 + (1+\theta)\int_1^x y^\theta \mathrm{d}y\Big)\mathrm{d}\mathbb{P}[Y \leq x] \\ &\leq 2 + \int_1^\infty (1+\theta)\int_1^x y^\theta \mathrm{d}y \mathrm{d}\mathbb{P}[Y \leq x] \\ &= 2 + (1+\theta)\int_1^\infty y^\theta \mathbb{P}[Y \geq y]\mathrm{d}y,\end{aligned}$$

where the last line follows from reversing the order of integration.

For 5, observe that

$$\begin{aligned}y\int_0^{\frac{2}{y}} \mathbb{E}[\exp(-\lambda Y) - 1 + \lambda Y]\mathrm{d}\lambda &= \int_0^\infty y\int_0^{\frac{2}{y}} \Big(\exp(-\lambda x) - 1 + \lambda x\Big)\mathrm{d}\lambda \mathrm{d}\mathbb{P}[Y \leq x] \\ &= \int_0^\infty \frac{y}{x}\left(1 - \frac{2x}{y} + \frac{1}{2}\Big(\frac{2x}{y}\Big)^2 - \exp\Big(-\frac{2x}{y}\Big)\right)\mathrm{d}\mathbb{P}[Y \leq x] \\ &\geq \int_y^\infty \frac{y}{x}\left(1 - \frac{2x}{y} + \frac{1}{2}\Big(\frac{2x}{y}\Big)^2 - \exp\Big(-\frac{2x}{y}\Big)\right)\mathrm{d}\mathbb{P}[Y \leq x] \\ &\geq 2c_1\mathbb{P}[Y \geq y],\end{aligned}$$

where we have used 1 and 2. □

In the classical setting of Dawson, at this stage one exploits the fact that the approximating processes are branching processes, in exact duality with the solution to a deterministic evolution equation. This is no longer true in our setting, so instead our first task is to write down an *approximate* evolution equation, whose solution we denote by v_ϕ^N, with the property that

$$\mathbb{E}[\exp(-\langle X_t^N, \phi\rangle)] = \exp(-\langle X_0^N, v_\phi^N(t,\cdot)\rangle) + \mathcal{O}(\varepsilon^N)||\phi||_\infty^2\langle X_0^N, \mathbf{1}\rangle,$$

where $\varepsilon^N \to 0$ as $N \to \infty$.

From our calculations in Section 2.4.3 (in particular equations (2.4.4) and (2.4.5)), and assuming the result of Lemma 2.4.14, we can write

$$\mathcal{L}_{\exp(-\cdot)}(\phi)(X_0^N) = \exp\Big(-\langle X_0^N, \phi\rangle\Big)\left[\langle X_0^N, -L^N\phi + B^N\phi\rangle + \mathcal{O}(\widetilde{\varepsilon}^N)||\phi||_\infty^2\langle X_0^N, \mathbf{1}\rangle\right] \tag{2.5.15}$$

with $\widetilde{\varepsilon}^N \to 0$, where

$$L^N\phi(x) = \frac{NM^d}{J}\left[\int_0^\infty \frac{1}{|\mathcal{B}_r^M|}\int_{\mathcal{B}_r^M(x)}\int_{\mathcal{B}_r^M(z)} u(r)(\phi(y)-\phi(x))\mathrm{d}y\mathrm{d}z\mu^N(\mathrm{d}r)\right]$$

and

$$B^N\phi(x) := \frac{NM^d}{K(\gamma-d)}\Big(\frac{K|\mathcal{B}_1|}{JM^d}\Big)^{\frac{\alpha+1}{\gamma-d}}\int_{\frac{K|\mathcal{B}_1|}{JM^d}}^{J^{\frac{\gamma-d}{\gamma}}\frac{K|\mathcal{B}_1|}{JM^d}} g(\upsilon\phi(x))\upsilon^{-\left(\frac{\alpha+1}{\gamma-d}+1\right)}\mathrm{d}\upsilon,$$

Now let us define $v_\phi^N(t,x)$ by

$$\begin{cases} \frac{\partial v_\phi^N(t,x)}{\partial t} & = L^N v_\phi^N(t,\cdot)(x) - B^N v_\phi^N(t,\cdot)(x), \\ v_\phi^N(0,x) & = \phi(x). \end{cases} \tag{2.5.16}$$

This is the evolution equation corresponding to a superprocess in which the spatial motion is the compound Poisson process with generator L^N and the branching mechanism is determined by B^N. The existence of such a process (which then, by duality, guarantees the uniqueness of the solution of the evolution equation) follows from, for example, El Karoui and Roelly (1991).

Let us write $\{S_t^N\}_{t\geq 0}$ for the semigroup generated by L^N. Then the solution to (2.5.16) can be written

$$v_\phi^N(t,x) := S_t^N(\phi)(x) - \int_0^t S_{t-s}^N\Big(B^N v_\phi^N(s,\cdot)\Big)(x)\mathrm{d}s. \tag{2.5.17}$$

An immediate consequence of (2.5.15) is that we have an approximate duality between X^N and v_ϕ^N since

$$\Big|\mathcal{L}_{\exp(-\cdot)}(v_\phi^N(t-s,\cdot)(X_s^N)\Big| \leq \mathcal{O}(\varepsilon^N)||v_\phi^N(t-s,\cdot)||_\infty^2\langle X_s^N, \mathbf{1}\rangle\exp\Big(-\langle X_s^N, v_\phi^N(t-s,\cdot)\rangle\Big)$$

$$\leq \mathcal{O}(\varepsilon^N)||v_\phi^N(t-s,\cdot)||_\infty^2\langle X_s^N,\mathbf{1}\rangle,$$

and so integrating over $[0,t]$,

$$\begin{aligned}\mathbb{E}[\exp\Big(-\langle X_t^N,\phi\rangle\Big)] &= \exp\Big(-\langle X_0^N, v_\phi^N(t,\cdot)\rangle\Big) + \mathcal{O}(\varepsilon^N)\int_0^t \|v_\phi^N(t-s,\cdot)\|_\infty^2 \mathrm{d}s\ \mathbb{E}[\langle X_0^N,\mathbf{1}\rangle] \\ &= \exp\Big(-\langle X_0^N, v_\phi^N(t,\cdot)\rangle\Big) + \mathcal{O}(\varepsilon^N)\|\phi\|_\infty^2\langle X_0^N,\mathbf{1}\rangle, \qquad (2.5.18)\end{aligned}$$

where we used Lemma 2.5.20 to replace $\mathbb{E}[\int_0^t\langle X_s^N,\mathbf{1}\rangle \mathrm{d}s]$ by $\langle X_0^N,\mathbf{1}\rangle$ at the expense of replacing $\widetilde{\varepsilon}^N$ by ε^N, but still with $\varepsilon^N\to 0$ as $N\to\infty$, and we used (2.5.17) to see that $\|v_\phi^N(t-s)\|_\infty \leq \|\phi\|_\infty$. In particular, applying this with $\lambda\phi$ in place of ϕ and differentiating with respect to λ at $\lambda = 0$ gives

$$\mathbb{E}[\langle X_t^N,\phi\rangle] = \langle X_0^N, S_t^N(\phi)\rangle.$$

The key step in proving tightness is the following:

Lemma 2.5.29. *For $\{X^N\}_{N\geq 1}$ satisfying the conditions of Theorem 2.2.7 and $\phi\in C_0^\infty$ we will have that, for any fixed $0<\theta<\beta$,*

$$\mathbb{E}[\sup_{s\leq T}\langle X_s^N,\phi\rangle^{1+\theta}] \leq H,$$

where H is independent of N.

Proof In the proof of this lemma, C is a constant which is independent of N, but may change from line to line. First observe that, using Lemma 2.5.28 parts 4 and 5,

$$\begin{aligned}\mathbb{E}[\langle X_t^N,\phi\rangle^{1+\theta}] &\leq 2 + C\int_1^\infty y^{1+\theta}\int_0^{\frac{2}{y}} \mathbb{E}[\exp(-\langle X_t^N,\lambda\phi\rangle) - 1 + \langle X_t^N,\lambda\phi\rangle]\mathrm{d}\lambda\mathrm{d}y \\ &= 2 + C\int_1^\infty y^{1+\theta}\int_0^{\frac{2}{y}} \Big[\exp(-\langle X_0^N, v_{\lambda\phi}^N(t,\cdot)\rangle) - 1 + \langle X_0^N, S_t^N(\lambda\phi)\rangle \\ &\qquad + \mathcal{O}(\varepsilon^N)\|\lambda\phi\|_\infty^2 t\langle X_0^N,\mathbf{1}\rangle\Big]\mathrm{d}\lambda\mathrm{d}y, \qquad (2.5.19)\end{aligned}$$

where we have used (2.5.18) and (2.5.5) and the constant C is determined by those in Lemma 2.5.28. Using Part 3 of Lemma 2.5.28,

$$\exp(-\langle X_0^N, v_{\lambda\phi}^N(t,\cdot)\rangle) - 1 + \langle X_0^N, S_t^N(\lambda\phi)\rangle$$
$$\leq c_2\langle X_0^N, \lambda S_t^N(\phi)\rangle^{1+\beta} + \exp(-\langle X_0^N, v_{\lambda\phi}^N(t,\cdot)\rangle) - \exp(-\langle X_0^N, S_t^N(\lambda\phi)\rangle.$$

The first term is at most $c_2\|\lambda\phi\|_\infty^{1+\beta}\langle X_0^N, \mathbf{1}\rangle^{1+\beta}$. To control the remaining terms, note that

$$\exp(-\langle X_0^N, v_{\lambda\phi}^N(t,\cdot)\rangle) - \exp(-\langle X_0^N, S_t^N(\lambda\phi)(\cdot)\rangle \leq \langle X_0^N, |v_{\lambda\phi}^N(t,\cdot) - S_t^N(\lambda\phi)(\cdot)|\rangle$$
$$= \langle X_0^N, \Big| \int_0^t S_{t-s}^N\Big(B^N v_{\lambda\phi}^N(s,\cdot)\Big)\mathrm{d}s\Big|\rangle,$$

(where we have used (2.5.17)). Now for non-negative ϕ, under the assumptions of Theorem 2.2.7, using (2.4.5),

$$B^N\phi(x) \leq \frac{NM^d}{K(\gamma-d)}\Big(\frac{K\phi(x)}{JM^d}\Big)^{\frac{\alpha+1}{\gamma-d}} \int_0^\infty (\exp(-v)+v-1)v^{-\frac{\alpha+1}{\gamma-d}-1}\mathrm{d}v$$
$$\leq C\phi(x)^{1+\beta},$$

so that using (2.5.17) again,

$$\langle X_0^N, \Big| \int_0^t S_{t-s}^N\Big(B^N v_{\lambda\phi}^N(s,\cdot)\Big)\mathrm{d}s\Big|\rangle \leq C\langle X_0^N, \int_0^t S_{t-s}^N\Big(v_{\lambda\phi}^N(s,\cdot)^{1+\beta}\Big)\mathrm{d}s\rangle$$
$$\leq C\langle X_0^N, \int_0^t S_{t-s}^N\Big(S_s^N(\lambda\phi)^{1+\beta}\Big)\mathrm{d}s\rangle$$
$$\leq C\|\phi\|_\infty^{1+\beta}\lambda^{1+\beta}t\langle X_0^N, \mathbf{1}\rangle.$$

Substituting in (2.5.19),

$$\mathbb{E}[\langle X_t^N, \phi\rangle^{1+\theta}] \leq 2 + C\int_1^\infty y^{1+\theta}\int_0^{\frac{2}{y}} \Big[||\phi||_\infty^{1+\beta}\lambda^{1+\beta}\Big(\langle X_0^N, \mathbf{1}\rangle^{1+\beta} + t\langle X_0^N, \mathbf{1}\rangle\Big)$$
$$+ \mathcal{O}(\varepsilon^N)\|\lambda\phi\|_\infty^2 t\langle X_0^N, \mathbf{1}\rangle\Big]\mathrm{d}\lambda\mathrm{d}y,$$

and since $\theta < \beta$ the right hand side is finite.

An application of the Monotone Convergence Theorem yields

$$\mathbb{E}[\langle X_t^N, \mathbf{1}\rangle^{1+\beta}] \leq H', \tag{2.5.20}$$

with $H' < \infty$ independent of $t \leq T$.

We have now established a bound of the form

$$\mathbb{E}[\langle X_t^N, \phi\rangle^{1+\theta}] \leq 2 + C\int_1^\infty y^{1+\theta}\int_0^{\frac{2}{y}} \lambda^{1+\beta} + \lambda^2 \mathrm{d}\lambda\mathrm{d}y\Big(\langle X_0^N, \mathbf{1}\rangle^{1+\beta} + \langle X_0^N, \mathbf{1}\rangle\Big) \leq H''.$$

The proof now follows the pattern established in Lemmas 2.5.22 and 2.5.26. Taking the terms in the semimartingale decomposition of $\langle X_t^N, \phi\rangle$ one by one, first observe that by Jensen's inequality,

$$\mathbb{E}\left[\left(\frac{1}{t}\int_0^t \langle X_s^N, |\mathcal{A}^N(\phi)|\rangle \mathrm{d}s\right)^{1+\theta}\right] \leq \mathbb{E}\left[\frac{1}{t}\int_0^t \langle X_s^N, |\mathcal{A}^N\phi|\rangle^{1+\theta}\mathrm{d}s\right],$$

which we can bound using (2.5.20) and the fact that, from Lemma 2.5.18, $|A^N\phi| \leq m(N)\|\phi\|_{C^3}$. Rearranging the semimartingale decomposition to give an expression for $M_t^N(\phi)$ and combining with Minkowski's inequality we can bound $\mathbb{E}[|M_T^N(\phi)|^{1+\theta}]$, and then by Doob's martingale inequality we see that

$$\begin{aligned}&\mathbb{E}[\sup_{s\leq T}\langle X_s^N, \phi\rangle^{1+\theta}]\\ &\qquad \leq C\left(\mathbb{E}[\langle X_0^N, \phi\rangle^{1+\theta}] + \mathbb{E}\left[\left(\int_0^T |\mathcal{L}^N(\phi_s)(X_s^N)|\mathrm{d}s\right)^{1+\theta}\right] + \mathbb{E}[|M_T^N(\phi)^{1+\theta}|]\right),\end{aligned}$$

and, since we have established uniform bounds on each of the terms on the right hand side of this expression, this completes the proof. □

To conclude the proof of tightness of $\{X^N\}_{N\geq 1}$ we shall use a criterion due to Aldous.

Theorem 2.5.30 (Aldous (1978))**.** *A sequence of càdlàg real-valued processes $\{Y^N\}_{N\geq 1}$ is tight if and only if the following two conditions hold:*

1. *for each fixed t, $\{Y_t^N\}_{N\geq 1}$ is tight in $\mathbb{R}$;*
2. *for any $\varepsilon > 0$, given a sequence of stopping times τ_N bounded above by T and a sequence of real numbers $\delta_N \to 0$ as $N \to \infty$,*

$$\lim_{N\to\infty} \mathbb{P}[|Y^N_{\tau_N+\delta_N} - Y^N_{\tau_N}| > \varepsilon] = 0.$$

The first condition is trivially satisfied, and so it remains to prove the following lemma.

Lemma 2.5.31. *Under the conditions of Theorem 2.2.7, for non-negative $\phi \in C_0^3(\mathbb{R}^d)$, given $\varepsilon > 0$, a sequence of stopping times τ_N bounded above by T, and a sequence of real numbers $\delta_N \to 0$ as $N \to \infty$,*

$$\lim_{N\to\infty} \mathbb{P}[|\langle X^N_{\tau_N+\delta_N}, \phi\rangle - \langle X^N_{\tau_N}, \phi\rangle| > \varepsilon] = 0.$$

Proof Again following the proof of Theorem 4.6.2 in Dawson (1993), we study the joint distribution of $\langle X^N_{\tau^N+\delta^N}, \phi\rangle$ and $\langle X^N_{\tau^N}, \phi\rangle$ through the Laplace transform

$$L^N(\delta_N; \lambda_1, \lambda_2) := \mathbb{E}\Big[\exp\Big(-\lambda_1\langle X^N_{\tau_N+\delta_N}, \phi\rangle - \lambda_2\langle X^N_{\tau_N}, \phi\rangle\Big)\Big].$$

The approximate duality (2.5.18), combined with the strong Markov property, gives

$$\begin{aligned}
&L^N(\delta_N; \lambda_1,\lambda_2)\\
&\quad:=\mathbb{E}\Big[\mathbb{E}\Big[\exp\Big(-\lambda_1\langle X^N_{\tau_N+\delta_N}, \phi\rangle - \lambda_2\langle X^N_{\tau_N}, \phi\rangle\Big)|X^N_{\tau_N}\Big]\Big]\\
&\quad=\mathbb{E}\Big[\exp\Big(-\langle X^N_{\tau_N}, v^N_{\lambda_1\phi}(\delta_N, \cdot)\rangle - \langle X^N_{\tau_N}, v^N_{\lambda_2\phi}(0, \cdot)\rangle\Big)\Big] + \mathcal{O}(\varepsilon^N)\|\lambda_1\phi\|^2\langle X^N_0, \mathbf{1}\rangle\\
&\quad=\mathbb{E}\Big[\exp\Big(-\langle X^N_{\tau_N}, v^N_{\lambda_1\phi}(\delta_N, \cdot) + v^N_{\lambda_2\phi}(0, \cdot)\rangle\Big)\Big] + \mathcal{O}(\varepsilon^N)\|\lambda_1\phi\|^2\langle X^N_0, \mathbf{1}\rangle,
\end{aligned}$$

from which

$$\begin{aligned}
&|L^N(\delta_N; \lambda_1, \lambda_2) - L^N(0; \lambda_1, \lambda_2)|\\
&\quad\le \|v^N_{\lambda_1\phi}(\delta_N, \cdot) - v^N_{\lambda_1\phi}(0, \cdot)\|_\infty\mathbb{E}[\sup_{s\le T}\langle X^N_s, \mathbf{1}\rangle] + \mathcal{O}(\varepsilon^N)\|\lambda_1\phi\|^2\langle X^N_0, \mathbf{1}\rangle\\
&\quad\le \|v^N_{\lambda_1\phi}(\delta_N, \cdot) - v^N_{\lambda_1\phi}(0, \cdot)\|_\infty\mathbb{E}[1 + \sup_{s\le T}\langle X^N_s, \mathbf{1}\rangle^{1+\beta}] + \mathcal{O}(\varepsilon^N)\|\lambda_1\phi\|^2\langle X^N_0, \mathbf{1}\rangle.
\end{aligned}$$

Now from (2.5.17) it follows that $\|v^N_{\lambda_1\phi}(\delta_N, \cdot) - v^N_{\lambda_1\phi}(0, \cdot)\|_\infty \to 0$ as $N \to \infty$ and so combining with the above,

$$|L^N(\delta_N; \lambda_1, \lambda_2) - L^N(0; \lambda_1, \lambda_2)| \to 0.$$

Tightness of $\{\langle X^N_{\tau_N+\delta_N}, \phi\rangle, \langle X^N_{\tau_N}, \phi\rangle\}$ follows from Lemma 2.5.29. Taking a convergent subsequence, since $|L^N(\delta_N; \lambda_1, \lambda_2) - L^N(0; \lambda_1, \lambda_2)| \to 0$, the limit of $\{\langle X^N_{\tau_N+\delta_N}, \phi\rangle, \langle X^N_{\tau_N}, \phi\rangle\}$ must be of the form $\{Z, Z\}$ and so we see that $\langle X^N_{\tau_N+\delta_N}, \phi\rangle - \langle X^N_{\tau_N}, \phi\rangle \to 0$ in probability, which completes the proof. □

Tightness is now proved.

2.6 Proofs of Lemmas 2.4.13 and 2.4.14

In this section we exploit the duality of the SLFV with a process of coalescing lineages to prove the two key estimates provided by Lemmas 2.4.13 and 2.4.14. In the fixed radius case, it is straightforward to make the heuristic argument of Section 2.2.4 rigorous and analogous arguments can be found in, for example, Etheridge et al. (2017). However, in the variable radius case, we must work a little harder to recover the result claimed here. We present an approach that works in either setting, but to avoid repetition restrict ourselves to the variable radius case.

2.6.1 A coupling of ancestral lineages

The key to our proof is the duality of the SLFV from Proposition 2.2.5. We first consider

$$\mathbb{E}\left[\int_{\mathbb{R}^d}\left(\frac{1}{|\mathcal{B}^M_r|}\int_{\mathcal{B}^M_r(x)}\phi(y)X^N_t(y)\mathrm{d}y\right)^2\mathrm{d}x\right] \tag{2.6.1}$$

for fixed r and t.

Using Proposition 2.2.5, for integrable functions ψ, for the scaled process,

$$\mathbb{E}\left[\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\psi(y,z)w^N_t(y)w^N_t(z)\mathrm{d}y\mathrm{d}z\right] = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\psi(y,z)\mathbb{E}_{(y,z)}\left[\prod_{i=1}^{N_t}w^N_0(\xi^{N,i}_t)\right]\mathrm{d}y\mathrm{d}z,$$

where $\{\xi^{N,1}_t, \xi^{N,2}_t\}$ are the positions of two ancestral lineages started from y and z respectively at time 0. In the scaled ancestral process, events of radius r/M covering the point y fall at rate $Nr^d\mu^N(\mathrm{d}r)$. During such an event, each lineage in the region covered by the event, independently, is affected with probability $u(r)/J$. Those lineages that are affected all jump to the same new position which is chosen uniformly at random

from the affected region. To estimate (2.6.1) we take

$$\psi(y,z) = \frac{1}{|\mathcal{B}_r^M|^2}\int_{\mathbb{R}^d} \phi(y)\phi(z)\mathbf{1}_{|y-x|<r/M}\mathbf{1}_{|z-x|<r/M}\mathrm{d}x$$

and recall that $X^N = Kw^N$. The key tool is a coupling.

Proposition 2.6.32. *Let $\{\xi_t^{N,1}, \xi_t^{N,2}\}_{t\geq 0}$ be the scaled ancestral lineages above. Then*

$$(\xi_t^{N,1}, \xi_t^{N,2}) \stackrel{d}{=} \big(Y_t^{N,1} + \varepsilon_t^{N,1}, (Y_t^{N,2} + \varepsilon_t^{N,2})\mathbf{1}_{t<\tau} + (Y_t^{N,1} + \varepsilon_t^{N,1})\mathbf{1}_{t\geq\tau}\big),$$

where $Y^{N,1}$ and $Y^{N,2}$ are independent random walks and the random coalescence time τ has hazard rate bounded above when $|\xi_t^{N,1} - \xi_t^{N,2}| = 2r_0/M$ by

$$h(r_0) = C\frac{M^2}{J}\frac{1}{(r_0 \vee J^{-1/\gamma})^{\gamma-\beta(\gamma-d)}}, \tag{2.6.2}$$

with C a constant independent of N. Moreover, for $d \geq 2$

$$\mathbb{E}[\sup_{0\leq t\leq T} M^2 J^{2/\gamma}|\varepsilon_t^{N,1} - \varepsilon_t^{N,2}|^2] \to 0 \quad \text{as } N \to \infty. \tag{2.6.3}$$

whereas if $d = 1$,

$$\mathbb{E}[\sup_{0\leq t\leq T} J|\varepsilon_t^{N,1} - \varepsilon_t^{N,2}|^2] = \mathcal{O}(1). \tag{2.6.4}$$

Before proving this result, let us see why it helps. The error terms $\varepsilon_t^{N,i}$ are very far from being independent, but they are very small. Notice in particular that if $\xi^{N,1}$ and $\xi^{N,2}$ have coalesced by time t, then since $Kw_0^N = X_0^N$ converges weakly to X_0,

$$\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \psi(y,z)\mathbb{E}_{(y,z)}[w_0^N(\xi_t^{N,1})]\mathrm{d}y\mathrm{d}z$$

is of order $1/K$, which when multiplied by K^2 to give us the contribution to (2.6.1) is $\mathcal{O}(K)$. If, on the other hand, the two lineages have not coalesced, then they are close to independent random walks and

$$\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \psi(y,z)\mathbb{E}_{(y,z)}[w_0^N(\xi_t^{N,1})w_0^N(\xi_t^{N,2})]\mathrm{d}y\mathrm{d}z = \mathcal{O}\Big(\frac{1}{K^2}\Big).$$

Under the assumptions of Theorem 2.2.6, $(KN)/(J^2M^d) \to C_2$ as $N \to \infty$, and so the proof of Lemma 2.4.13 is reduced to checking that the coalescence probability tends to zero as $N \to \infty$. Under the conditions of Theorem 2.2.7,

$$\frac{N}{J^2M^d}\int_0^\infty u(r)^2 r^{2d}\mu^N(\mathrm{d}r) = \mathcal{O}\Big(\frac{N}{J^2M^d}J^{(1-\beta)(\gamma-d)/\gamma}\Big) = \mathcal{O}\Big(\frac{M^{2-d}}{J}J^{(1-\beta)(\gamma-d)/\gamma}\Big). \tag{2.6.5}$$

In $d \geq 2$, since $(1-\beta)(\gamma-d)/\gamma < 1$, we easily see that this tends to zero as $N \to \infty$. The sparsity condition (2.2.7) guarantees that the same is true in $d = 1$ and so the contribution to the expression in Lemma 2.4.14 from lineages that have not coalesced is negligible. On the other hand, multiplying the expression in (2.6.5) by K we obtain an expression of order

$$\mathcal{O}\Big(\frac{KN}{J^2M^d}J^{(1-\beta)(\gamma-d)/\gamma}\Big) = \mathcal{O}\Big(\Big(\frac{J}{N}\Big)^{(1-\beta)/\beta}J^{(1-\beta)(\gamma-d)/\gamma}\Big),$$

which grows without bound as $N \to \infty$ (because of Condition 5 of Theorem 2.2.7). To prove Lemma 2.4.14 we must show that this quantity multiplied by the coalescence probability tends to zero. In $d \geq 2$, under the conditions of Theorem 2.2.7, the coalescence probabilities will be of the same order as in the fixed radius case, and so the Lemma will follow easily (again using $(1-\beta)(\gamma-d)/\gamma < 1$). In one dimension, things are more delicate, and we shall see the need for our more stringent sparsity condition.

Proof of Proposition 2.6.32.

We rewrite the Poisson process of events Π^N that drives the dynamics of $\xi^{N,1}$ and $\xi^{N,2}$ as the sum of four components through a thinning. Each event $(x,t,r,\rho) \in \Pi$ is augmented by two independent Bernoulli random variables, η_1, η_2, each with success probability $\rho = u(r) = r^{-\gamma}$. The random variable η_i determines whether or not the ith ancestral lineage is affected by the event. We now let $\Pi^{N,1}$ be the events in Π for which $\eta_1 = 1, \eta_2 = 0$; $\Pi^{N,2}$ is the subset of events with $\eta_1 = 0, \eta_2 = 1$ and $\Pi^{N,1,2}$ is the events with $\eta_1 = \eta_2 = 1$. The remaining events won't affect the motion of either lineage.

The lineage $\xi_t^{N,i}$ is driven by $\Pi^{N,i} \bigcup \Pi^{N,1,2}$ (recalling that after coalescence there is a single lineage, $\xi^{N,1}$). Coalescence results from both lineages being in the region

covered by an event from $\Pi^{N,1,2}$. If $|\xi_t^{N,1} - \xi_t^{N,2}| = 2r_0/M$, then the rate at which they are both in the region affected by an event from $\Pi^{N,1,2}$ is bounded above by

$$C'\frac{N}{J^2}\int_{r_0}^1 r^d u(r)^2 \mu^N(\mathrm{d}r) = C'\frac{N}{J^2}\int_{r_0\vee J^{-1/\gamma}}^1 r^{d-2\gamma+\alpha}\mathrm{d}r \leq C\frac{M^2}{J}\frac{1}{(r_0\vee J^{-1/\gamma})^{\gamma-\beta(\gamma-d)}},$$

where we have approximated the volume of the centres x such that $\mathcal{B}_r(x)$ contains two points at separation $2r_0$ by $C(d)r^d$ for $r > r_0$ and we have used that $\alpha+1 = (\beta+1)(\gamma-d)$ and $N/(JM^2) \to C_1$ as $N \to \infty$.

Let $Y^{N,i}$ be the random walk driven by the events of $\Pi^{N,i}$ and $\varepsilon_t^{N,i} = \xi_t^{N,i} - Y_t^{N,i}$. Evidently, since they are driven by independent Poisson processes of events, the walks $Y^{N,1}$ and $Y^{N,2}$ are independent. It remains to bound $|\varepsilon^{N,1} - \varepsilon^{N,2}|$. First observe that we can couple our lineages $\xi_t^{N,i}$ with a system $\widetilde{\xi}_t^{N,i}$ of lineages that do not coalesce, by at each event of $\Pi^{N,1,2}$ choosing *two* parental positions independently and uniformly at random from the affected region and, if both lineages are in the affected region, the lineage $\widetilde{\xi}_t^{N,i}$ jumps to the ith parental location. Until the coalescence time τ, the processes $\widetilde{\xi}_t^{N,i}$ and $\xi_t^{N,i}$ coincide. We define $\widetilde{\varepsilon}_t^{N,i}$ in the obvious way. Evidently, it will suffice to control the supremum of $|\widetilde{\varepsilon}^{N,1} - \widetilde{\varepsilon}^{N,2}|$ over the time interval $[0,T]$.

First observe that

$$M_{\widetilde{\varepsilon}}(t) := \widetilde{\varepsilon}_t^{N,1} - \widetilde{\varepsilon}_t^{N,2}$$

is a mean zero martingale. We should like to estimate its variance at time t. It is driven entirely by events from $\Pi^{N,1,2}$ and by exploiting translation invariance of the system, we find that

$$\begin{aligned}
\mathbb{E}[\langle M_{\widetilde{\varepsilon}}\rangle_t] &\leq C\int_0^t \frac{N}{J^2}\int_{1/J^{1/\gamma}}^1 (\frac{r}{M})^2 r^{-2\gamma+d+\alpha}\mathrm{d}r\mathrm{d}s \\
&= \begin{cases} \mathcal{O}(\frac{1}{J}) & \text{if } \beta(\gamma-d)-\gamma+2>0; \\ \mathcal{O}(\frac{\log J}{J}) & \text{if } \beta(\gamma-d)-\gamma+2=0; \\ \mathcal{O}(\frac{1}{J^{(2+\beta(\gamma-d))/\gamma}}) & \text{if } \beta(\gamma-d)-\gamma+2<0. \end{cases}
\end{aligned}$$

Since $\beta(\gamma-d)-\gamma+2 = 2-d-(1-\beta)(\gamma-d)$ and under our assumptions $(1-\beta)(\gamma-d) < 1$, if $d = 1$ the first line holds, whereas if $d \geq 2$ the third line holds. In particular, for $d \geq 2$, using Condition 5 of Theorem 2.2.7, the quantity on the right is always $o(1/(M^2 J^{2/\gamma}))$ from which (2.6.3) easily follows via Doob's maximal inequality. An application of Doob's inequality also gives the result in $d = 1$. □

2.6.2 Coalescence Probabilities

It remains to estimate our coalescence probabilities.

Proposition 2.6.33. *Let the scaled ancestral lineages $\xi^{N,1}$, $\xi^{N,2}$ be as in the previous subsection. We start them from two points sampled independently at random from $\mathcal{B}_r^M(0)$. The probability that they have coalesced by time T, which we shall denote by $p_N(T)$ satisfies*

$$p_N(T) \leq \begin{cases} CJ^{(1-\beta)(\gamma-1)/\gamma} \frac{M^2}{J} & \text{in } d = 1; \\ C\frac{\log M}{J} & \text{in } d = 2; \\ C\frac{1}{J} & \text{in } d \geq 3. \end{cases} \tag{2.6.6}$$

In the light of the discussion immediately after the statement of Proposition 2.6.32, this result will complete the proof of our key lemma in the variable radius case.

Proof of Proposition 2.6.33.

Our aim is to show that

$$\mathbb{E}\Big[\int_0^T h(M|\xi_s^{N,1} - \xi_s^{N,2}|)\mathrm{d}s\Big] \tag{2.6.7}$$

is bounded by the quantities on the right hand side of (2.6.6). where $h(r)$ is given by (2.6.2). To do this, we shall use the coupling of Proposition 2.6.32 to approximate $|\xi^{N,1}-\xi^{N,2}|$ by $|Y^{N,1}-Y^{N,2}|$ and then the classical Fourier transform approach of Darling and Kac (1957) to estimate the corresponding expected integral of the hazard rate. Since we are only interested in the separation of our lineages we denote $Y^N := |Y^{N,1} - Y^{N,2}|$ and $\varepsilon^N := |\varepsilon^{N,1} - \varepsilon^{N,2}|$.

We first observe that

$$\mathbb{E}\Big[\int_0^T h\big(M(Y_s^N+\varepsilon_s^N)\big)\mathrm{d}s\Big] = \int_0^T \mathbb{E}[h(MY_s^N)+M\varepsilon_s^N h'(MY_s^N)+(M\varepsilon_s^N)^2 h''(MY_s^N+\theta(s)M\varepsilon_s^N)]\mathrm{d}s, \quad (2.6.8)$$

for some $\theta(s)\in(0,1)$ and that

$$h'(r)=Ch(r)\frac{1}{r}\mathbf{1}_{[J^{-1/\gamma},1]}(r)\le J^{1/\gamma}h(r); \quad h''(r)=C'h(r)\frac{1}{r^2}\mathbf{1}_{[J^{-1/\gamma},1]}(r)\le J^{2/\gamma}h(r). \quad (2.6.9)$$

We now begin with $d\ge 2$. Using (2.6.3), we see that

$$\mathbb{E}\Big[\int_0^T h(M(Y_s^N+\varepsilon_s^N)\mathrm{d}s\Big]\le C\int_0^T \mathbb{E}[h(MY_s^N)]\mathrm{d}s.$$

The random walk Y^N jumps at exponentially distributed times with mean $\mathcal{O}(1/M^2)$, and so will make $\mathcal{O}(M^2)$ such jumps by time T. It is evidently enough to show that

$$\int_0^{T_N}\mathbb{E}[h(MY_s^N)]\mathrm{d}s$$

is bounded by the quantities on the right of (2.6.6) where T_N is the random time at which the walk Y^N has made a geometric number of jumps with mean M^2. Let us write $p(x,y)$ for the probability density function of the unscaled displacement MY^N at a single jump and $p^{\star k}(x,y)$ for its k-fold convolution. Further write $\mathcal{T}_M$ for the time of the G^Mth jump of Y^1, where G^M is an independent geometrically distributed random variable with mean M^2. Then

$$\mathbb{E}\Big[\int_0^{T_N} h(MY_s^N)\mathrm{d}s\Big]=\frac{1}{M^2}\mathbb{E}\Big[\int_0^{\mathcal{T}_M} h(Y_s^1)\mathrm{d}s\Big] \le \frac{1}{M^2}\sum_{k=0}^{\infty}\int_{\mathbb{R}^d} z_M^k p^{\circ k}(0,y)h(|y|)\mathrm{d}y := Q_M, \quad (2.6.10)$$

where $z_M=1-1/M^2$ and on the right hand side we have bounded the expectation above by setting $Y_0=0$. Define $I_M(x):=\sum_{k=0}^{\infty} z_M^k p^{\star k}(0,x)h(|x|)$. Writing $\phi(t)$ for the characteristic function of the displacement of Y in a single jump and taking Fourier

transforms,

$$\widehat{I}_M(t) = \Big(\frac{1}{1 - z_M\phi} \circ \widehat{h}\Big)(t).$$

Since Y is rotationally symmetric with strictly bounded jumps, its characteristic function is real-valued and $\sim 1 - C|t|^2$ close to the origin and

$$\frac{1}{1 - z_M\phi(t)} \leq \frac{C'}{\frac{1}{M^2} + C|t|^2}.$$

In $d \geq 3$ we may let $M \to \infty$ in this expression (corresponding to an infinite time horizon) and using that the inverse Fourier transform of $|t|^{-2}$ is $|x|^{2-d}$ and, recalling the definition of h from (2.6.2), we can approximate Q_M by

$$\frac{1}{M^2}\int_{\mathbb{R}^d} I_M(y)\mathrm{d}y \leq \frac{C}{M^2}\frac{M^2}{J}\int_0^1 r^{2-d}r^{d-1}\frac{1}{r^{\gamma-\beta(\gamma-d)}}\mathrm{d}r = \frac{C}{J}\int_0^1 r^{1-\gamma+\beta(\gamma-d)}\mathrm{d}r \propto \frac{1}{J},$$

where we obtained an upper bound by replacing the lower limit of integration by zero.

In $d = 2$, the inverse Fourier transform of $1/(\epsilon^2 + |t|^2)$ is $K_0(\epsilon|x|)$, where K_0 is the modified Bessel function of the second kind of order zero (Eq. 17, p.365, Gel'fand and Shilov (1964)) . Recalling that $K_0(z) \sim \log(1/z)$ as $z \to 0$, and substituting in (2.6.10) yields $Q_M \leq C(\log M)/J$.

We now consider $d = 1$. Looking again at Q_M, we can no longer replace the lower limit in the integral by zero as $r^{-2\gamma+\alpha+1}$ is not integrable at the origin. Instead observe that (by a change of variables) the inverse Fourier transform of $(1/M^2 + |t|^2)^{-1}$ at x is M times that of $1/(1 + |t|^2)$ evaluated at x/M, which, using Eq. 40, p.363 of Gel'fand and Shilov (1964) and writing $K_{1/2}$ for the modified Bessel function of the second kind of order 1/2,

$$\sqrt{\frac{|x|}{M}}K_{1/2}\Big(\frac{|x|}{M}\Big) \sim 1 \quad \text{as } M \to \infty.$$

Substituting,

$$\begin{aligned} Q_M \leq \frac{CM^2}{J}\int_0^1 \frac{1}{M}\int_{r_0 \vee 1/J^{1/\gamma}}^1 r^{-2\gamma+\alpha+1}\mathrm{d}r\mathrm{d}r_0 &= C\frac{M}{J}\frac{1}{J^{1/\gamma}}\Big(\frac{1}{J^{1/\gamma}}\Big)^{-2\gamma+\alpha+2} + C'\frac{M}{J} \\ &= C''\frac{M}{J}J^{(1-\beta)(\gamma-1)/\gamma}. \end{aligned}$$

In other words the main contribution comes from separations at most $1/J^{1/\gamma}$. This bounds the contribution from $h(MY_s^N)$ in (2.6.8). From (2.6.4), $M|\widetilde{\varepsilon}^N|$ is at most $\mathcal{O}(M/\sqrt{J})$ and so if $M/\sqrt{J} \leq \mathcal{O}(J^{-1/\gamma})$, by using (2.6.9) as in two dimensions, we obtain a bound which is $1/M$ times the expression on the right hand side of (2.6.6) and we are done. If not then the order of the error is $MJ^{1/\gamma}/\sqrt{J}$ times the quantity above and, since by assumption $\gamma > 2$, this completes the proof. □

2.7 Showing convergence

Armed with tightness, it remains to identify the limit points of the sequences $\{X^N\}_{N\geq 1}$.

2.7.1 Proof of Theorem 2.2.6

Recall that for each N, using Lemma 2.4.12,

$$M^N(\phi)_t = \langle X_t^N, \phi\rangle - \langle X_0^N, \phi\rangle - \int_0^t \langle X_s^N, \mathcal{A}^N(\phi)\rangle \mathrm{d}s - \zeta_t^N(\phi)$$

is a martingale.

From the convergence of $m(N)$,

$$\lim_{N\to\infty} \mathbb{E}\left[\sup_{t\leq T}\left|\int_0^t \langle X_s^N, \mathcal{A}^N(\phi)\rangle \mathrm{d}s + \zeta_t^N(\phi) - \int_0^t \langle X_s, \frac{m}{2}\Delta\phi\rangle \mathrm{d}s\right|\right] = 0,$$

where m is as in the statement of Theorem 2.2.6. Moreover, using (2.4.3) and Lemma 2.4.13

$$\langle M^N(\phi)\rangle_t \to 2\kappa \int_0^t \langle X_s^N, \phi\rangle \mathrm{d}s$$

as $N \to \infty$.

Suppose that $\{X^{N_k}\}_{k\geq 1}$ is a convergent subsequence. To prove that the limit is the finite variance superBrownian motion, we must check that the martingale property is preserved under passage to the limit and that for each $t \leq T$, $M_t^{N_k}(\phi)^2 + \langle M^{N_k}(\phi)\rangle_t$ is uniformly integrable (see e.g. Lemma II.4.5 in Perkins (2002)). To check the first condition, suppose that $0 \leq t_1 < ... < t_k \leq s < t$, $h_1, ..., h_k \in C_b$ and $\phi \in C_0^3(\mathbb{R}^d)$. An application of the Dominated Convergence Theorem combined with taking a subsequence

yields

$$\mathbb{E}\left[\left(\langle X_t,\phi\rangle - \langle X_s,\phi\rangle - \int_s^t \langle X_u, A\Delta\phi\rangle \mathrm{d}u\right)\prod_{j=1}^k h_j(\langle X_{t_j},\phi\rangle)\right] = 0.$$

Thus, see e.g. Theorem 4.8.10 of Ethier and Kurtz (1986),

$$M_t(\phi) := \langle X_t,\phi\rangle - \langle X_0,\phi\rangle - \int_0^t \langle X_s, \frac{m}{2}\Delta\phi\rangle \mathrm{d}s,$$

is a martingale.

The uniform integrability of $\langle M^N(\phi)\rangle_t$ is immediate from our previous calculations. That of $M_t^N(\phi)^2$ follows on setting $\theta = 1$ in the arguments (based on Jensen and Minkowski's inequalities) that gave us Lemma 2.5.29.

Since the solution to the martingale problem corresponding to the finite vairance superBrownian motion is unique, the proof of Theorem 2.2.6 is complete.

2.7.2 Proof of Theorem 2.2.7

Since the solution to the martingale problem for the superBrownian motion with stable branching law is unique, it is enough to check that all limit points satisfy the martingale problem (2.2.1). Once again we must check that the martingale property is conserved under passage to the limit, for which it suffices to show that for any $k \geq 0$, $0 \leq t_1 < ... < t_k \leq s < t$, $h_1, ..., h_k \in C_b$ and $\phi \in C_0^3(\mathbb{R}^d)$ with $\phi \geq 0$,

$$\lim_{N\to\infty} \mathbb{E}\Bigg[\Bigg(\exp(-\langle X_t^N,\phi\rangle) - \exp(-\langle X_s^N,\phi\rangle) \\ - \int_s^t \left\langle X_u^N, -\frac{m}{2}\Delta\phi + \kappa\phi^{1+\beta}\right\rangle \exp(-\langle X_u^N,\phi\rangle)\mathrm{d}u\Bigg)\prod_{j=1}^k h_j(\langle X_{t_j},\phi\rangle)\Bigg] = 0,$$

with m and κ as in the statement of Theorem 2.2.7. Once again, since the h_j are bounded, this is an easy consequence of the Dominated Convergence Theorem, our results of Section 2.4.3 and Lemma 2.4.14.

Chapter 3

Spatial Lambda-Fleming-Viot Model in fluctuating environment

3.1 Introduction

In this chapter we expand on the work in Chapter 2 and address a question of some biological interest: how the frequency of a rare mutation evolves in a spatially distributed population if the direction of selection on that mutation fluctuates in time? This type of question is particularly relevant in the context of the 'Court Jester hypothesis' discussed in Barnosky (2001) and Benton (2009). The 'Court Jester hypothesis' states that long term improvements in fitness may not occur, since populations must constantly evolve to keep pace with changes in their environment.

The simplest mathematical framework in which this question can be addressed is given by the observation of a single genetic locus. As in Chapter 2, we will consider a population with two genetic types. The 'common type', κ_c, is often referred to as a wild type in biological literature and the 'rare' type, κ_r. We again assume that the rare type forms only a small fraction of the total population.

We recall a simple model without selection, for example, a Moran model or a Wright-Fisher model. In the absence of spatial structure, the absolute number of rare individuals will evolve approximately according to a branching process, which, under appropriate scaling, as discussed in Section 1.1, converges to the Feller diffusion. We may ask whether a similar phenomenon occurs in the presence of selection, in particular

we are interested in when the direction of selection fluctuates rapidly in time. One expects that in the latter case the evolution approximately follows a branching process in a random environment, and, under suitable scaling, converges to the Feller diffusion in a random environment.

We also seek to establish an analogous result for a spatially distributed population. As discussed before, there are serious difficulties when trying to construct models which incorporate genetic drift in higher dimensional spatial continua, so we turn again to the spatial Lambda-Flemming-Viot model introduced in Etheridge (2008) and Barton et al. (2010).

Therefore, the spatial Lambda-Fleming-Viot model provides a reasonable framework to study the behaviour of the establishment of a mutation. The work in Chapter 2 for the model without selective advantage or disadvantage for the rare mutation shows convergence. The limiting object was superBrownian motion, a measure-valued process introduced independently by Watanabe (1968) and Dawson (1975), which is the spatial counterpart of the Feller diffusion.

Recently Biswas et al. (2018) studied a diffusion approximation of the spatial Lambda-Fleming-Viot with selection in a fluctuating environment. In contrast to their work we are interested in the behaviour of an establishing mutation within a large population rather than two established populations of comparable size. In an analogy to the non-spatial case, we show that the limiting process is the superBrownian motion in a random environment, introduced and studied in Mytnik (1996). The work of Nakashima (2015) shows that superBrownian motion in a random environment is the scaling limit of a model of branching random walks on a lattice in random environment, introduced by Birkner et al. (2005b). It is our conjecture that superBrownian motion in a random environment is a universal scaling limit for the critical interacting particle systems in random environments.

The proof of the scaling result in Chapter 2 is based on a duality method. However, as discussed in detail in Section 5 of Biswas et al. (2018), a useful dual process seems not to be available in this setting. The techniques of Biswas et al. (2018) are also not available to us as we are considering a rare mutation. Therefore, rather than adapt their techniques, we use a different approach based on a particle representation which

belongs to the family of lookdown constructions. We build on the work of Kurtz and Rodrigues (2011) and provide a new lookdown construction of superBrownian motion in a random environment. We then use ideas from Etheridge and Kurtz (2018) and a slightly modified version of their construction of the spatial Lambda-Fleming-Viot lookdown which incorporates a random environment.

For other techniques discussed above many difficulties arise with the introduction of spatial continua. However, the vast majority of our work in the non-spatial case transfers to the spatial case without difficulty. We give an explanation for this unusual phenomenon in Remark 3.1.5. For this reason, the majority of this chapter is devoted to the rigorous derivation of the non-spatial result. Our proof technique in this chapter is based on four main ingredients: lookdown representation, an averaging trick due to Kurtz (1973), a perturbation result due to Kurtz (1992) (which we recall as Theorem 3.2.16) and the Markov Mapping Theorem of Kurtz (1998) (which we recall in Section 3.5.2). Previous results using the lookdown approach have shown either the existence of processes under weak conditions, or convergence of processes which had previously been shown to converge through other means. This is, to our knowledge, the first proof of convergence using the lookdown approach in which other techniques were not available.

Lookdown constructions were introduced in Donnelly and Kurtz (1996) and Donnelly and Kurtz (1999). This approach has proved to be particularly fruitful in applications to population models. In this setting, each individual in the population is assigned a 'level' taking values in either the possitive integers, as in the original paper of Donnelly and Kurtz (1996), or in $[0, \infty)$, as introduced in Kurtz (2000). Levels typically carry information about genealogical relations between individuals. The name 'lookdown' is used as individuals usually determine their parents by 'looking down' at the sub-population with levels lower than their own.

From a practical perspective, one of the most useful properties of lookdown constructions is that when passing from individual based models to their high density limits, or continuous approximation, the genealogies are preserved. The importance of this can be seen in the examples of systems of individual based models approximated by the same diffusion processes with very different genealogies obtained in Taylor (2009). For further examples in the context of Lambda-Fleming-Viot models we refer the reader to Miller

(2015). For an excellent explanation of the general principle of lookdown constructions we refer to Kurtz and Rodrigues (2011), particularly their death process example in Section 2.1.

In the context of the Spatial Lambda-Fleming-Viot we would like to point to two different approaches. The first one is a the construction developed in Etheridge and Kurtz (2018), Section 4.1.3. This construction forms the basis for our construction of the SLFV with selection in a fluctuating environment. A special case of this construction can be found in the Appendix of Chetwynd-Diggle and Klimek (2019). The second construction, which was the first lookdown construction for the SLFV, was presented in Véber and Wakolbinger (2015), and was developed using a different approach in Etheridge and Kurtz (2018), Section 4.1.1. The later construction is much closer in the flavour to the original ideas of Donnelly and Kurtz (1996).

The rest of this chapter is laid out as follows. We start Section 3.1.1 by carefully defining the spatial Lambda-Fleming-Viot process with fluctuating selection and Super-Brownian motion in a random environment. We finish Section 3.1.1 with the statements of the main results and a brief discussion of the key difference to the results in Chapter 2. In Section 3.2 we discuss the Lambda-Fleming-Viot model, studying its scaling limits and discussing the lookdown representation. In particular, Section 3.2.3 contains the bulk of our proof. In Section 3.3, we discuss the lookdown construction for a version of Branching Brownian motion in a random environment and the lookdown representation of the superBrownian motion in a random environment. In Section 3.4, we discuss how to extend the result of Section 3.2 to the spatial setup. Section 3.5.1 contains some information of Poisson random measures, which are used extensively throughout this chapter. Section 3.5.3 briefly recalls a Lemma A.13 from Kurtz and Rodrigues (2011) which ensures our projected process is a solution to the correct martingale problem. Section 3.5.2 discusses the Markov Mapping Theorem. Section 3.5.4 contains some of the proofs of Theorems of Section 3.3.

3.1.1 Statement of main results

We suppose that the population, which is distributed across $\mathbb{R}^d$, is subdivided into two genetic types. We will denote the space of types by $\mathcal{K} = \{\kappa_c, \kappa_r\}$. Formally, the state of

the population at time t is described by a measure $M_t \in \mathcal{M}$, where $\mathcal{M}$ is the space of measures whose first marginal is Lebesgue measure on $\mathbb{R}^d \times \mathcal{K}$. We note $\mathcal{M}$ is compact when equipped with the topology of weak convergence.

At any fixed time there is a density $w(t,\cdot) : \mathbb{R}^d \to [0,1]$ such that we can use the identification

$$\int_{\mathbb{R}^d \times \{0,1\}} f(x,\kappa) M(\mathrm{d}x, \mathrm{d}\kappa) = \int_{\mathbb{R}^d} \{w(x) f(x,1) + (1 - w(x)) f(x,0)\} \, \mathrm{d}x.$$

We interpret $w(t,x)$ as the proportion of population of type κ_r at location x at time t. It is defined only up to a Lebesgue null set. For what follows, it is convenient to fix w as a representative of M_0 through the relation described above and update it according to a procedure described below. We consider two types of events - neutral and selective events. Selective events are influenced by the state of the environment.

We begin with a description of the environment, which is used for all models in this section. Our environment is modelled through a simple random field.

Definition 3.1.1. *Let Π^{env} be a Poisson process with intensity E, dictating the times of the changes in the environment. Let $\{\xi^{(m)}(\cdot)\}_{m\geq 0}$ be a family of identically distributed random fields on $\mathbb{R}^d$ such that for each*

$$\begin{aligned} \mathbb{P}\Big[\xi^{(m)}(x) = -1\Big] =& \frac{1}{2} = \mathbb{P}\Big[\xi^{(m)}(x) = +1\Big], \\ q(x,y) :=& \mathbb{E}\Big[\xi^{(m)}(x)\xi^{(m)}(y)\Big]. \end{aligned}$$

where $q(x,y)$ is the covariance function, which belongs to $C_0\Big(\mathbb{R}^d \times \mathbb{R}^d\Big)$ (continuous functions vanishing at infinity). Set $\tau_0 = 0$ and write $\{\tau_m\}_{m\geq 1}$ for the points in Π^{env} and define

$$\zeta(t,\cdot) := \sum_{m=0}^{\infty} \xi^{(m)}(\cdot) \mathbf{1}_{[\tau_m, \tau_{m+1})}(t).$$

For the construction of the random field in Definition 3.1.1 we refer to Ma (2009), especially Example 1. We observe that the generator of the process describing the evolution of the environment, A^{env}, is given by

$$A^{env} f(\zeta) = \mathbb{E}_\pi[f(\zeta)] - f(\zeta), \tag{3.1.1}$$

with domain given by continuous bounded functions, where π is the stationary distribution of the random field ζ.

The following version of the Spatial Lambda-Fleming-Viot process with fluctuating selection (SLFVFS) is a slight modification of the process discussed in Biswas et al. (2018).

Definition 3.1.2 (Spatial Lambda-Fleming-Viot process with fluctuating selection)**.** *Let μ be a σ-finite measure on $(0,\infty)$ and for each $r \in (0,\infty)$, let ν_r be a probability measure on $[0,1]$, such that the mapping $r \to \nu_r$ is measurable and*

$$\int_{(0,\infty)} r^d \int_{[0,1]} u\, \nu_r(\mathrm{d}u)\mu(\mathrm{d}r) < \infty.$$

Further, fix $\mathbf{s} \in [0,1]$ and let Π^{neu}, Π^{sel}, be independent Poisson point processes on $\mathbb{R}_+ \times \mathbb{R}^d \times (0,\infty) \times [0,1]$ with intensity measures $(1-\mathbf{s})\mathrm{d}t \otimes \mathrm{d}x \otimes \mu(\mathrm{d}r)\nu_r(\mathrm{d}u)$ and $\mathbf{s}\mathrm{d}t \otimes \mathrm{d}x \otimes \mu(\mathrm{d}r)\nu_r(\mathrm{d}u)$ respectively. Let Π^{env} be a Poisson process of Definition 3.1.1, evolving independently of Π^{neu}, Π^{sel}. Let $\sigma(\kappa,\xi) : (\mathcal{K} \times \{-1,1\}) \to \mathbb{R}^+$ be a function which satisfies the symmetry condition

$$\mathbb{E}_\pi\left[\frac{\sigma(\kappa_r,\zeta(0))}{\sigma(\kappa_c,\zeta(0))} - 1\right] = 0. \tag{3.1.2}$$

We note that the name symmetry condition could be misleading. It is not a symmetry between the rare and common type, but, rather a symmetry between the effect of the environment on the rare type within the background population of common type.

The spatial Lambda-Fleming-Viot process with fluctuating selection (SLFVFS) *with driving noises Π^{neu}, Π^{fsel}, Π^{env}, is the $\mathcal{M}_\lambda$-valued process M_t with dynamics described as follows. Let $w(t_-,\cdot)$ be a representative of the density of M_{t-} immediately before an event (t,x,r,u) from Π^{neu} or Π^{fsel}. Then the measure M_t immediately after the event has density $w(t,\cdot)$ determined by:*

1. *If $(t,x,r,u) \in \Pi^{neu}$, a neutral event occurs at time t within the closed ball $B(x,r)$. Then*
   (a) *Choose a parental location l according to the uniform distribution on $B(x,r)$.*

(b) Choose the parental type $\kappa \in \{\kappa_r, \kappa_c\}$ *according to distribution*

$$\mathbb{P}[\kappa = \kappa_r] = w(t_-, l), \quad \mathbb{P}[\kappa = \kappa_c] = 1 - w(t_-, l).$$

(c) A proportion u *of the population within* $B(x, r)$ *dies and is replaced by offspring with type* κ*. Therefore, for each point* $y \in B(x, r)$*,*

$$w(t, y) = w(t_-, y)(1 - u) + u\mathbf{1}_{\{\kappa=\kappa_r\}}.$$

2. If $(t, x, r, u) \in \Pi^{fsel}$*, a selective event occurs at time* t *within the closed ball* $B(x, r)$*. Then*

(a) Choose a parental location l *according to the uniform distribution on* $B(x, r)$*.*

(b) Choose the parental type $\kappa \in \{\kappa_r, \kappa_c\}$ *according to*

$$\mathbb{P}[\kappa = \kappa_r] = \frac{\sigma(\kappa_r, \zeta(x))w(t_-, l)}{\sigma(\kappa_r, \zeta(x))w(t_-, l) + \sigma(\kappa_c, \zeta(x))(1 - w(t_-, l))},$$
$$\mathbb{P}[\kappa = \kappa_c] = \frac{\sigma(\kappa_c, \zeta(x))(1 - w(t_-, l))}{\sigma(\kappa_r, \zeta(x))w(t_-, l) + \sigma(\kappa_c, \zeta(x))(1 - w(t_-, l))}.$$

(c) A proportion u *of the population within* $B(x, r)$ *dies and is replaced by offspring with type* κ*. Therefore, for each point* $y \in B(x, r)$*,*

$$w(t, y) = w(t_-, y)(1 - u) + u\mathbf{1}_{\{\kappa=\kappa_r\}}.$$

3. If $t \in \Pi^{env}$ *the environment evolves as in Definition 3.1.1.*

The existence of the process follows from results of Etheridge and Kurtz (2018), Section 4.1.2, Lemma 4.3. We note the symmetry condition is not required for existence of the model but is convenient for our work. This condition can be relaxed if σ depends on N, however for simplicity, we assume fixed σ and so require condition (3.1.2) to be satisfied. Furthermore, as in the fixed radius case in Chapter 2, we will fix both the impact and radius of events in our models. We assume, with a slight abuse of notation that $\mu = \delta_r$ and $\nu_r = \delta_u$. The methods used could be extended to take into account the variable radius case discussed previously.

We discuss the lookdown representation of the non-spatial version of the model in Section 3.2, while Section 3.2.2.1 explains how the lookdown relates to the underlying process. The spatial version of the lookdown representation for this process is described in Section 3.4.

We shall define the limiting process, which is a variant of superBrownian motion in a random environment with a drift, in terms of the generator. Let $\mathcal{M}_F(\mathbb{R}^d)$ we denote the space of finite measures on $\mathbb{R}^d$, again equipped with the topology of weak convergence.

Definition 3.1.3 (SuperBrownian motion in a random environment)**.** *Let $q(x,y) \in C_0\Big(\mathbb{R}^d \times \mathbb{R}^d\Big)$ be a covariance function. The superBrownian motion in a random environment with a diffusion parameter m, a growth parameter b and a quadratic variation parameter (a,c) is the unique $\mathcal{M}_f(\mathbb{R}^d)$ valued Markov process with càdlàg sample paths such that for each $f \in \overline{C}^2(\mathbb{R}), \phi \in \mathcal{D}(\Delta)$*

$$\mathcal{L}f(\langle X_t,\phi\rangle) = f'(\langle X_t,\phi\rangle)[\frac{m}{2}\langle X_t,\Delta\phi\rangle + b\langle X_t,\phi,\rangle] + \frac{1}{2}f''(\langle X_t,\phi\rangle)\Big(a\langle X_t,\phi^2\rangle + c\int_{\mathbb{R}^d\times\mathbb{R}^d} q(x,y)\phi(x)\phi(y)X_t(\mathrm{d}x)X_t(\mathrm{d}y)\Big), \quad (3.1.3)$$

is a martingale.

This model is discussed in more detail in Section 3.3. In particular, we provide a new lookdown construction of the model, which we derive from the lookdown construction of branching Brownian motion in a random environment. This follows the ideas behind the construction of superBrownian motion from branching Brownian motion.

Our results describe the scaling limit of a sequence of processes. At the Nth stage of our scaling, the local population density will be $K = K(N)$. We shall denote the representative of the density of the scaled SLFVSRE by w^N and the population of rare individuals by $X^N = Kw^N$, which is defined Lebesgue almost everywhere. We shall think of X^N as a measure-valued process and abuse notation by writing, for any Borel measurable ϕ,

$$\Big\langle X_t^N, \phi\Big\rangle = K\int_{\mathbb{R}^d}\phi(x)w_t^N(x)\mathrm{d}x = \int_{\mathbb{R}^d}\phi(x)X_t^N(x)\mathrm{d}x.$$

The scaling for the neutral part of the model is nearly the same as in Chapter 2. The modifications required by the presence of selection and fluctuations in the environment are inspired by Biswas et al. (2018). For the scaled process, time is sped up by a factor N, space is shrunk by $M(N)$, and the impact of each event is reduced by a factor $J(N)$. The rate of environmental changes is multiplied by $\widehat{S}(N)^2$ and the proportion of selective events is multiplied by $\widehat{S}(N)/S(N)$. We use $\widehat{S}(N)$ and $S(N)$ in order to consider the non-fluctuating environment case in out work. Scaling of the selective events is motivated by our inclination to model short bursts of strong selection. In the model without environmental fluctuations the rate of selective events is scaled by $1/S(N)$. The additional portion of selective events prevents the averaging out of the action of selection in the diffusive limit. The Central Limit Theorem suggests the relation between the rate of additional selective events and rate of changes of the environment. In Biswas et al. (2018), Section 3.2 and Section 4.2 they discuss this relation further. We let $V_r := |B_r(0)|$ and $C(d) := \int_{\mathbb{B}_1(0)} |x|^2 \mathrm{d}x$ in dimension d. We are now in position to state the main result of this chapter.

Theorem 3.1.4. *Suppose that X_0^N is absolutely continuous with respect to Lebesgue measure with support $\mathrm{supp}(X_0^N) \subseteq D$, where D is a compact subset of $\mathbb{R}^d$ independent of N, and X_0^N converges weakly to X_0. Moreover, suppose that, as N tends to infinity,*

$$\frac{C_d u r^{d+2} N}{JM^2} \to C_1;\ J, K, M, S, \widehat{S} \to \infty;\ \frac{K}{JM^d} \to 0;\ \frac{u^2 V_r N K}{J^2 M^d} \to a;$$
$$\frac{N^2}{K J^2 M^d} \to 0;\ \mathbb{E}_\pi\left[\left\{\frac{suNV_r}{SJ}\left(\frac{\sigma(\kappa_r, \zeta)}{\sigma(\kappa_c, \zeta)} - 1\right)\right\}^2\right] \to b^2;\ \frac{\widehat{S}}{S} \to 0.$$

If there exists an n such that, as N tends to infinity, $N(K/J)^n \to 0$ then the sequence $X_N(t)$ converges weakly to superBrownian motion in a random environment with initial condition X_0, diffusion parameter C_1, growth parameter b^2, quadratic variation parameter (a, b^2).

Remark 3.1.5. *We note that the neutral case of our model was analysed in Chapter 2. The previous scaling requires*

$$\frac{N}{JM^2} \to C_1;\ J, K, M \to \infty;\ \frac{NK}{J^2 M^d} \to C_2,$$

along with a 'sparsity' condition which we discuss below. In Chapter 2 we discuss the heuristic reason for that scaling which also gives a solid justification for our choice of scaling in Section 2.2.4.

We note that our 'sparsity' condition is the requirement $K/JM^d \to 0$*, which is stronger than the one present in Chapter 2. This is a feature of the methods we use and we do not think can be strengthened using lookdown constructions. Our sparsity condition is equivalent to* $M^2/J \to 0$ *in all dimensions in comparison to the previous conditions*

$$\frac{M}{J} \to 0 \text{ if } d = 1; \ \frac{\log M}{J} \to 0 \text{ if } d = 2; \ \frac{1}{J} \to 0 \text{ if } d \geq 3.$$

It is due to the fact that the lookdown construction 'sees' the Hausdorff dimension of the support of superBrownian motion in a way that previous work did not. Our set of test functions does not smooth out the support of the process, in sharp contrast to our previous approach. We require the intensity of levels within the ball of radius r/M *to tend to infinity. Since the limit has Hausdorff dimension two, in dimensions three and higher we will see a concentration of mass in the limit which requires an infinite intensity of levels in arbitrarily small balls. This observation shows why the proof of our results in the spatial case does not differ significantly from the proof in the non-spatial case.*

3.2 Scaling limits of the LFV - dynamics of the rare type

In this section we are interested in describing the evolution of a subpopulation with a rare mutation within a population which evolves according to the Λ-Fleming-Viot model with selection in a fluctuating environment (LFVSFE). We will again consider a type space with two types, rare and common. Which of those two types has higher fitness changes with the environment. We show that the evolution of the rare subpopulation follows a Feller diffusion in a random environment.

We provide a new description of the LFVSFE in terms of a lookdown construction. This construction is inspired by the lookdown construction of the neutral model in Section 4.1.3 of Etheridge and Kurtz (2018).

We begin with a description of the model which gives some insight into its construction before defining the model precisely in Definition 3.2.6. Each of the individuals in the population is assigned a *genetic type* κ from the set $\mathcal{K}$ and a *level* $l \in \mathbb{R}^+ \cup \{0\}$. We recall that we restrict our attention to $\mathcal{K} = \{\kappa_c, \kappa_r\}$, which we refer to as the 'common' and 'rare' type, respectively. Since in this section we consider a model without spatial structure, the state of the population can be represented as a collection of points, $\eta = \{(l, \kappa)\}$, or as a measure which can be written, with a slight abuse of notation, as

$$\eta = \sum_{(l,\kappa)\in\eta} \delta_{(l,\kappa)}, \quad \text{where } (l, \kappa) \in \Big(\mathbb{R}^+ \cup \{0\}, \mathcal{K}\Big).$$

We assume that the process with levels η is always a conditionally Poisson system with Cox measure $m_{leb} \times \Xi$, where m_{leb} is Lebesgue measure on $\mathbb{R}^+$. This is because, in our lookdown model, when our initial condition has this form then our process will have this form for all subsequent times. This is a key feature in how the lookdown construction relates to its underlying model.

For lookdown representations we consider test functions of the form

$$f(\eta) = \prod_{(l,\kappa)\in\eta} g(l, \kappa), \tag{3.2.1}$$

with the additional requirement that there exists a $\lambda_g > 0$ such that $g(l) = 1$ for all $l > \lambda_g$. This set of test functions will be used throughout this chapter for computations involving lookdown representations.

We are interested in the situation when the type which is selectively advantageous depends on the environment. We write ζ for the random process which models the state of the environment. Therefore the full state of the model at time t is given by a pair (η_t, ζ_t).

We now proceed to carefully state our non-spatial model, the scaling and the non-spatial results. The following statements are almost identical to those found in Section 3.1.1 but we include them for completeness. This further demonstrates the similarity between the spatial and non-spatial methods when using the lookdown

construction and motivates why our spatial proof mainly lies within Proposition 3.4.53 and Proposition 3.4.54.

The evolution of the population is determined by two types of reproduction events, neutral and selective. We assume that a proportion, s, of events are selective and favour one of the two types. Events are driven by independent Poisson processes Π^{neu} and Π^{sel}. For simplicity we assume that the impact of the events is fixed and equal to u.

Both neutral and selective events are composed of two elements, discrete births and thinning. The birth phase of the events differs between neutral and selective events, while the thinning phase is the same.

Whenever $t \in \Pi^{neu}$, a birth event produces offspring, with levels distributed according to an independent Poisson point process with intensity u. Let v^* be the smallest of the new levels. Let (l^*_{neu}, κ^*) denote the element of η with the smallest level greater than v^*, that is

$$l^*_{neu} = \min\{l : (l, \kappa) \in \eta, l > v^*\}.$$

The individual (l^*_{neu}, κ^*) is chosen as the parent of the event and removed from the population. All new individuals are assigned type κ^*, the type of the parent. The levels of old individuals in the population are changed. If the level of the individual was smaller than v^* it remains unaffected by the birth. If the level of the individual was larger than v^*, it is moved to $l - l^*_{neu} + v^*$.

If the event is selective, the situation is more complicated. We introduce an additional function $\sigma(\kappa, \zeta)$, which influences the likelihood of an individual of type κ being the parent, given the state of the environment, ζ. When the environment is in state ζ, the higher the value of $\sigma(\kappa, \zeta)$, the more likely an individual of type κ is to be selected as a parent during selective events in environment ζ.

Whenever $t \in \Pi^{sel}$, a birth event produces offspring, with levels distributed according to an independent Poisson process with intensity u. As before, let v^* be the smallest of the new levels. Let (l^*_{sel}, κ^*) denote the element of η which obtains the minimum

$$\min\left\{\frac{l_i - v^*}{\sigma(\kappa_i, \zeta)} : (l_i, \kappa_i) \in \eta, l_i > v^*\right\}. \tag{3.2.2}$$

It is clear that if $\sigma(\kappa_c, \zeta) = \sigma(\kappa_r, \zeta)$ this is identical to neutral events. The individual (l^*_{sel}, κ^*) is chosen as the parent of the event and removed from the population. All new individuals are assigned the parent's type as in neutral events. The levels of old individuals in the population are then changed. If the level of the individual was smaller than v^* it remains unaffected by the birth. If the level of the individual was larger than v^*, it is moved to $\sigma(\kappa, \zeta)(l - l^*_{neu} + v^*)/\sigma(\kappa^*, \zeta)$.

For both neutral and selective events once the parent has been selected, and the old levels moved to their new locations, thinning takes place. Thinning does not affect the new individuals. The thinning takes the level of each individual which is neither a child or the parent of the event present within the population and multiplies it by $1/(1-u)$. The combined effect of this movement is defined explicitly through the function $\mathcal{J}_{neu}$, defined in (3.2.4), for neutral events and $\mathcal{J}_{sel}$, defined in (3.2.6), for selective events. We note that instead of removing the parent from the population we will consider the parent to be identified with the lowest offspring. The movement of the levels is chosen in this way to maintain levels with a conditionally Poisson system after any event.

We now define the main process of interest.

Definition 3.2.6 (Lookdown representation of LFVSFE)**.** *Fix $s \in (0,1)$. Let Π^{neu}, Π^{sel} be a pair of independent Poisson point processes with intensity measures $(1-s)\mathrm{d}t \otimes \nu(\mathrm{d}u)$ and $s\mathrm{d}t \otimes \nu(\mathrm{d}u)$ respectively on $\mathbb{R}^+ \times (0,1)$. Moreover, let Π^{env} be a Poisson process with rate E, independent of Π^{neu}, Π^{sel}. Let $\sigma : \mathcal{K} \times \{-1, 1\} \to \mathbb{R}^+$ be a function.*

The lookdown representation of LFVSRE is the process taking values in the set of $\{-1, 1\}$ times the purely atomic measures on $\mathbb{R} \times \mathcal{K}$ with dynamics described as follows.

1. *If $(t, u) \in \Pi^{neu}$*

   (a) *a group of new individuals with levels $(v_1, v_2, \dots)$ is added to the population. Their levels are distributed according to a Poisson process with intensity u.*

   (b) *Let $v^* = \min\{v_1, v_2, \dots\}$. The type of the new individuals is chosen to be the same as the type of the individual (κ^*, l^*) whose level is the lowest above v^*, that is*

$$l^*_{neu} = \min\{l : (l, \kappa) \in \eta, l > v^*\}. \tag{3.2.3}$$

(c) As a result of an event the levels with position l before an event will have new position given by

$$\mathcal{J}_{neu}(l, l^*_{neu}, v^*) = \begin{cases} \frac{1}{1-u}(l - (l^*_{neu} - v^*)) & \text{if } l > l^*_{neu}, \\ \frac{1}{1-u} l & \text{if } l < l^*_{neu}, \\ v^* & \text{if } l = l^*_{neu}. \end{cases} \tag{3.2.4}$$

2. If $(t, u) \in \Pi^{sel}$

(a) a group of new individuals with levels $(v_1, v_2, \dots)$ is added to the population. Their levels are distributed according to a Poisson process with intensity u.

(b) Let $v^* = \min\{v_1, v_2, \dots\}$. The type of the new individuals is chosen to be the same as the type of the individual (l^*, κ^*) whose level minimizes

$$\left\{ \frac{l_i - v^*}{\sigma(\kappa_i, \zeta)} : (l_i, \kappa_i) \in \eta, l_i > v^* \right\}. \tag{3.2.5}$$

(c) As a result of an event the levels with position l and type κ before an event will maintain their type but will have new position given by

$$\mathcal{J}_{sel}((l, \kappa), (l^*_{sel}, \kappa^*), \zeta, v^*) = \begin{cases} v^* & \text{if } l = l^*_{sel}, \\ \frac{1}{1-u}\left(l - (l^* - v^*)\frac{\sigma(\kappa,\zeta)}{\sigma(\kappa^*,\zeta)}\right) & \text{if } l \neq l^*_{sel}, l > v^*, \\ \frac{1}{1-u} l & \text{if } l < v^*. \end{cases} \tag{3.2.6}$$

3. If $t \in \Pi^{env}$, the environmental variable ζ_t is resampled uniformly from $\{-1, 1\}$.

The way the levels move is largely inspired by Kurtz and Rodrigues (2011).

As in the spatial case our results require a symmetry condition on the values of $\sigma(\kappa, \zeta)$. To be more precise, we require that

$$\mathbb{E}_\pi\left[\frac{\sigma(\kappa_r, \zeta)}{\sigma(\kappa_c, \zeta)} - 1\right] = 0, \tag{3.2.7}$$

where π is the stationary distribution of ζ.

We note that the generator of the process from Definition 3.2.6 is given by

$$Af(\eta,\zeta) = A_{neu}f(\eta,\zeta) + A_{sel}f(\eta,\zeta) + A_{env}f(\eta,\zeta), \tag{3.2.8}$$

where

$$\begin{aligned}
A_{neu}f(\eta,\zeta) &= (1-s)\int_0^\infty \Bigg[ue^{-uv^*}g(\kappa^*,v^*)e^{-u\int_{v^*}^\infty (1-g(\kappa^*,v))\mathrm{d}v} \\
&\qquad \times \prod_{(\kappa,l)\in\eta, l\neq l^*} g(\kappa, \jmath_{neu}(l, l^*_{neu}, v^*))\Bigg]\mathrm{d}v^* - f(\eta), \\
A_{sel}f(\eta,\zeta) &= s\int_0^\infty \Bigg[ue^{-uv^*}g(\kappa^*,v^*)e^{-u\int_{v^*}^\infty (1-g(\kappa^*,v))\mathrm{d}v} \\
&\qquad \times \prod_{(\kappa,l)\in\eta, l\neq l^*_{sel}} g(\kappa, \jmath_{sel}((l,\kappa),(l^*_{sel},\kappa^*),\zeta,v^*))\Bigg]\mathrm{d}v^* - f(\eta), \\
A_{env}f(\eta,\zeta) &= \mathbb{E}_\pi[f(\eta,\zeta)] - f(\eta,\zeta),
\end{aligned}$$

and f is of the form (3.2.1). Observe that the only differences between neutral and selective events are the choice of the parent l^* and the movement of levels.

3.2.1 Scaling

We are interested in the evolution of the subpopulation of a rare type within the population evolving according to LFVSFE. To quantify the rarity, we consider a population with total density K (which is equal to one for the usual Lambda-Fleming-Viot model), and will let K tend to infinity. We wish the rare type to make up about $\mathcal{O}(1/K)$ of the population at each level of the scaling. This is represented in the look-down process by the intensity of levels. The levels of individuals are initially Poisson distributed with intensity K with individuals given the rare type with probability $\mathcal{O}(1/K)$ and given the common type otherwise.

In order to recover the correct scaling limit we need to readjust our parameters. Let w^N denote the proportion of individuals of the rare type at the N^{th} stage of scaling. Let $X^N = Kw^N$ denote the size of the population of the rare type. Our scaling limit describes the behaviour of X^N. The right scaling of the other parameters is suggested by work from Chapter 2 and Biswas et al. (2018) as discussed in Remark 3.1.5. We speed

up the reproduction rate by N and increase the total population size to $K(N)$, but scale down both the impact and the selection coefficient. The impact of an event at the N^{th} stage of the approximation will be given by $u/J(N)$. The selection coefficient in the presence of fluctuations will be $s\widehat{S}(N)/S(N)$ and $s/S(N)$ in the absence of fluctuations. In order to simplify the notation we drop the explicit dependence of scaling parameters on N in what follows. For our results to hold, certain relations between the parameters need to be satisfied. In Theorem 3.2.10, Theorem 3.2.12 and Theorem 3.2.13 we specify scaling limits for the model with fluctuating selection, the model where the direction of selection does not change and the neutral model respectively. In particular, our results require a specific relation between the rate of the changes in the environment and the rate of selection. We therefore assume that the rate of the environmental events is $\widehat{S}^2$.

The generator of the scaled process, where η is a Poisson point process with first marginal given by K times Lebesgue measure, can then be written as

$$A^N f(\eta,\zeta) = A^N_{neu} f(\eta,\zeta) + \widehat{S} A^N_{sel} f(\eta,\zeta) + \widehat{S}^2 A^N_{env} f(\eta,\zeta), \tag{3.2.9}$$

where

$$\begin{aligned}
A^N_{neu} f(\eta,\zeta) = N\Bigg(\int_0^\infty \Bigg[\frac{uK}{J} e^{-\frac{uK}{J}v^*} g(\kappa^*, v^*) e^{-\frac{uK}{J}\int_{v^*}^\infty (1-g(\kappa^*,v))\mathrm{d}v} \\
\times \prod_{(\kappa,l)\in\eta, l\neq l^*} g(\kappa, \beth_{neu}(l,l^*,v^*)) \Bigg] \mathrm{d}v^* - f(\eta) \Bigg), \\
A^N_{sel} f(\eta,\zeta) = \frac{sN}{S}\Bigg(\int_0^\infty \Bigg[\frac{uK}{J} e^{-\frac{uK}{J}v^*} g(\kappa^*, v^*) e^{-\frac{uK}{J}\int_{v^*}^\infty (1-g(\kappa^*,v))\mathrm{d}v} \\
\times \prod_{(\kappa,l)\in\eta, l\neq l^*_{sel}} g(\kappa, \beth_{sel}(l,l^*_{sel},v^*)) \Bigg] \mathrm{d}v^* - f(\eta) \Bigg), \\
A^N_{env} f(\eta,\zeta) = \mathbb{E}_\pi[f(\eta,\zeta)] - f(\eta,\zeta).
\end{aligned}$$

3.2.2 Main result of this section

We now state the main results of this section. We recall the definitions of some of the classical models in terms of their generators, in order to state the results formally.

Definition 3.2.7 (Feller diffusion)**.** *Let $a > 0$, $b \in \mathbb{R}$. The Feller diffusion is the process taking values in $\mathbb{R}$ with generator C_{fd}, defined for every $f \in C^\infty(\mathbb{R})$, given by*

$$C_{fd}f(y) = ayf''(y) + byf'(y).$$

Its lookdown representation is characterised by the process with generator A_{fd} given by

$$A_{fd}f(\eta) = f(\eta)\sum_i 2a\int_{l_i}^{\infty}(g(v)-1)\mathrm{d}v + f(\eta)\sum_i (al_i^2 - bl_i)\frac{g'(l_i)}{g(l_i)}. \tag{3.2.10}$$

Definition 3.2.8 (Feller diffusion in random environment)**.** *Let $a, b > 0$. The Feller diffusion in a random environment is the process taking values in $\mathbb{R}$ with generator C_{fdr}, defined for every $f \in C^\infty(\mathbb{R})$, of the form*

$$C_{fdr}f(y) = (ay + b^2y^2)f''(y) + b^2yf'(y).$$

Its lookdown representation is characterised by the process with generator A given by

$$\begin{aligned} Af(\eta) = f(\eta)\sum_i 2a\int_{l_i}^{\infty}(g(v)-1)\mathrm{d}v + f(\eta)\sum_i (al_i^2 - b^2l_i)\frac{g'(l_i)}{g(l_i)} \\ + b^2 f(\eta)\sum_j\left(\sum_{i\neq j} l_jl_i\frac{g'(l_i)g'(l_j)}{g(l_i)g(l_j)} + l_j^2\frac{g''(l_i)}{g(l_i)}\right). \end{aligned} \tag{3.2.11}$$

For a more detailed discussion of the lookdown constructions for the Feller diffusion and the Feller diffusion in a random environment we refer to Kurtz and Rodrigues (2011), Section 2.

Remark 3.2.9. *We observe that our choice of test functions guarantees that all terms appearing in* (3.2.10) *and* (3.2.11) *are well-defined even if $g(l_i) = 0$, by considering this identity*

$$f(l)\sum_i\left(\frac{g'(l_i)}{g(l_i)}\right) = \sum_i\left(g'(l_i)\prod_{j\neq i}g(l_j)\right).$$

Theorem 3.2.10. *Let $X^N(t)$ denote the total intensity of individuals of the rare type at time t. Suppose that X_0^N converges to X_0. Moreover, suppose that, as N tends to*

infinity,

$$J, K, S, \widehat{S} \to \infty; \quad \frac{K}{J} \to 0; \quad \frac{u^2 NK}{J^2} \to 2a; \quad \frac{N^2}{KJ^2} \to 0;$$
$$\mathbb{E}_\pi\left[\left\{\frac{suN}{SJ}\left(\frac{\sigma(\kappa_r,\zeta)}{\sigma(\kappa_c,\zeta)}-1\right)\right\}^2\right] \to b^2; \quad \frac{\widehat{S}}{K} \to 0; \quad \frac{\widehat{S}}{S} \to 0. \tag{3.2.12}$$

In addition, assume that there exists an m such that, as $N \to \infty$, $NK^m/J^m \to 0$. Then the sequence X_N converges weakly to the Feller diffusion in a random environment with initial condition X_0 and parameters $2a$, b^2.

We have endeavoured to keep the conditions in (3.2.12) as general as possible but examples satisfying the conditions are easy to find as described below.

Example 3.2.11. *Fix $k \in (1/2, 1)$, $\gamma \in (0, min(k, (3-k)/2))$.. The conditions of Theorem 3.2.10 are satisfied if*

$$J = N^{\frac{1+k}{2}} \quad S = N^{\frac{3-k}{2}} \quad K = N^k \quad \widehat{S} = N^\gamma.$$

As a by-product of our technique, we prove an analogous result for the neutral model and the model with selection.

Theorem 3.2.12. *Let $X^N(t)$ denote the total intensity of individuals of the rare type at time t. Suppose that X_0^N converges to X_0 and the intensity of selective events is zero. Moreover, suppose that, as N tends to infinity,*

$$J, K \to \infty; \quad \frac{K}{J} \to 0; \quad \frac{N^2}{KJ^2} \to 0; \quad \frac{u^2 NK}{J^2} \to 2a. \tag{3.2.13}$$

In addition, assume that there exists an m such that, as $N \to \infty$, $NK^m/J^m \to 0$. Then the sequence $X_N(t)$ converges weakly to the critical Feller diffusion ($b = 0$) with initial condition X_0 and variance parameter $2a$.

Theorem 3.2.13. *Let $X^N(t)$ denote the total intensity of individuals of the rare type at time t. Suppose that X_0^N converges to X_0, σ does not depend on the environment*

and $\widehat{S} = 1$. *Moreover, suppose that, as* N *tends to infinity,*

$$J, K, S \to \infty; \quad \frac{K}{J} \to 0; \quad \frac{N^2}{KJ^2} \to 0; \quad \frac{u^2 NK}{J^2} \to 2a; \quad \frac{suN}{SJ}\left(\frac{\sigma(\kappa_r)}{\sigma(\kappa_c)} - 1\right) \to b \tag{3.2.14}$$

In addition, assume that there exists an m *such that, as* $N \to \infty$, $NK^m/J^m \to 0$. *Then the sequence* $X_N(t)$ *converges weakly to the Feller diffusion with initial condition* X_0 *and parameters* $2a$ *and* b.

Remark 3.2.14. *We stress that our proofs do not guarantee convergence of the lookdown representations, but only convergence of the projected models.*

3.2.2.1 Projected model

We follow the work in Etheridge and Kurtz (2018), in particular their "first construction" from Section 4.1.1, to show the connection between our lookdown construction and the standard LFV.

The neutral generator is simply the non-spatial counterpart of the lookdown construction from Etheridge and Kurtz (2018) and so we do not treat it here. However, in order to clarify how our form of selection acts from the perspective of the underlying process we investigate further. The neutral generator can also be seen by taking $\sigma(\kappa_c) = \sigma(\kappa_r)$ in the selective case. This section follows Etheridge and Kurtz (2018) and all the techniques in this section are taken from there, but formulated in our notation. The part of the generator of our process which describes selective events is of the form

$$A_{sel}f(\eta) = s\Bigg(\int_0^\infty \Bigg[ue^{-uv^*}g(v^*,\kappa^*)e^{-u\int_{v^*}^\infty (1-g(v,\kappa^*))\mathrm{d}v} \times \prod_{(l,\kappa)\in\eta, l\neq l^*_{sel}} g(\mathcal{J}_{sel}(l, l^*_{sel}, v^*), \kappa)\Bigg]\mathrm{d}v^* - f(\eta)\Bigg).$$

We define $h(\kappa) = \int_0^\infty (1 - g(l,\kappa))\mathrm{d}l$ and note that integration by parts gives

$$\int_0^\infty ue^{-uv^*}g(v^*,\kappa^*)e^{-\frac{uK}{J}\int_{v^*}^\infty (1-g(v,\kappa^*))\mathrm{d}v}\mathrm{d}v^* = e^{-uh(\kappa^*)}.$$

We also note that if η is formed from a Poisson point process with intensity measure $m_{leb} \otimes \Xi(\mathrm{d}\kappa)$ then

$$\{\mathcal{J}_{sel}(l, l^*_{sel}, v^*) : l \neq l^*_{sel}\},$$

is a Poisson point process with intensity measure $m_{leb} \otimes (1-u)\Xi(\mathrm{d}\kappa)$. Consider the distribution of κ^* conditioned on $\Xi(\mathrm{d}\kappa)$. We note that

$$\mathbb{P}[\kappa^* = \kappa_r] = \frac{\sigma(\kappa_r)\Xi(\{\kappa_c\})}{\sigma(\kappa_r)\Xi(\{\kappa_c\}) + \sigma(\kappa_c)\Xi(\{\kappa_r\})},$$

and so when we average our selective generator we arrive at

$$\begin{aligned}\alpha A^N_{sel} f(\Xi) = s e^{-\int_{\mathbb{K}} h(\kappa)\Xi(\mathrm{d}\kappa)} \\ \times \Bigg(\Bigg[\frac{\sigma(\kappa_r)\Xi(\{\kappa_c\})}{\sigma(\kappa_r)\Xi(\{\kappa_c\}) + \sigma(\kappa_c)\Xi(\{\kappa_r\})} e^{-uh(\kappa_r)} \\ + \frac{\sigma(\kappa_c)\Xi(\{\kappa_r\})}{\sigma(\kappa_r)\Xi(\{\kappa_c\}) + \sigma(\kappa_c)\Xi(\{\kappa_r\})} e^{-uh(\kappa_c)} \Bigg] e^{u\int_{\mathbb{K}} h(\kappa)\Xi(\mathrm{d}\kappa)} - 1 \Bigg).\end{aligned}$$

We finally note that our selection can now be seen as a simple weighted choice of parent in the (non-spatial) Lambda Fleming-Viot process. The same calculation can be done for the spatial case.

Remark 3.2.15. *We notice that our way of modelling selection differs from the approach of Etheridge et al. (2014) and Biswas et al. (2018). However, a quick generator calculation shows that this type of selection should lead to the same diffusion approximation as in Etheridge et al. (2014) and Biswas et al. (2018).*

3.2.3 Convergence of the non-spatial model

We are interested in convergence of the model with selection in a fluctuating environment. Recall that we study the behaviour of an establishing mutation under this model. We take this into account by considering test functions which are unaffected by the individuals of the common type, κ_c.

We recall the generator of the rescaled process takes the form

$$A^N f(\eta,\zeta) = A^N_{neu} f(\eta,\zeta) + \widehat{S} A^N_{sel} f(\eta,\zeta) + \widehat{S}^2 A^N_{env} f(\eta,\zeta), \tag{3.2.15}$$

where A^N_{neu} is the part of the generator which describes neutral events, A^N_{sel} is the part of the generator which describes selective events, and A^N_{env} describes the evolution of the environment.

In Sections 3.2.3.2 and 3.2.3.3, we show that the terms A^N_{neu} and A^N_{sel} converge to well-defined limits. However, since $\widehat{S} \to \infty$ as N tends to infinity, it may seem that as N tends to infinity, (3.2.15) will not converge to a non-trivial limit. However, the naive limiting procedure does not take into account the changes in the direction of selection. In order to identify the correct limit, we apply a 'separation of timescales' trick due to Kurtz (1973).

During the calculations in the proof of Theorem 3.2.13 we will show the operator A^N_{sel} can be written as

$$A^N_{sel} f(\eta) = f(\eta) \frac{suN}{JS} \left(\frac{\sigma(\kappa_r,\zeta)}{\sigma(\kappa_c,\zeta)} - 1 \right) \sum_j l_j \frac{g'(\kappa_j, l_j, \xi)}{g(\kappa_j, l_j, \xi)} + \mathcal{O}\left(\frac{1}{S} + \frac{1}{K} \right) \tag{3.2.16}$$

We consider a test function $\widetilde{f}$ of the form

$$\widetilde{f}(\eta,\zeta) = f(\eta) + \frac{1}{\widehat{S}} f(\eta) \frac{suN}{JS} \left(\frac{\sigma(\kappa_r,\zeta)}{\sigma(\kappa_c,\zeta)} - 1 \right) \sum_j \frac{g'(\kappa_j, l_j)}{g(\kappa_j, l_j)} l_j =: f(\eta) + \frac{1}{\widehat{S}} f_1(\eta,\zeta).$$

Observe that since $\widehat{S} \to \infty$, we can prove the test function $\widetilde{f}$ will tend to $f(\eta)$ as $N \to \infty$. We apply the generator (3.2.15) to $\widetilde{f}$. This leads to

$$\begin{aligned} A^N \widetilde{f}(\eta,\zeta) &= A^N_{neu} f(\eta) + \widehat{S} A^N_{sel} f(\eta) + \frac{1}{\widehat{S}} A^N_{neu} f_1(\eta,\zeta) + A^N_{sel}(f_N(\eta,\xi)) - \widehat{S} f_N(\eta,\xi) \\ &= A^N_{neu} f(\eta) + A^N_{sel}(f_N(\eta,\xi)) + \mathcal{O}\left(\frac{1}{\widehat{S}} + \frac{\widehat{S}}{S} + \frac{\widehat{S}}{K} \right) \end{aligned}$$

where we have used (3.2.16), that f does not depend on ξ (to see $A_{env} f(\eta) = \mathbb{E}_\pi[f(\eta)] - f(\eta) = 0$) and (3.2.7), (to see that, as $\mathbb{E}_\pi[f_N(\eta,\zeta)] = 0$, $A_{env} f_N(\eta,\zeta) = -f_N(\eta,\zeta)$).

Therefore, at least heuristically, the identification of $A^N_{sel}(f_N(\eta,\xi))$ should lead to the correct limit.

To make this argument rigorous, we shall use a theorem due to Kurtz (1992) which we recall here. Let us introduce some notation. For a metric space E, let $l_m(E)$ be the space of measures on $[0,\infty)\times E$ such that $\mu\in l_m(E)$ if and only if $\mu([0,t)\times E)=t$.

Theorem 3.2.16 (Kurtz (1992), Theorem 2.1)**.** *Let E_1, E_2 be complete separable metric spaces, and set $E=E_1\times E_2$. For each n, let $\{(X_n,Y_n)\}$ be a stochastic process with sample paths in $D_E([0,\infty))$ adapted to a filtration $\{\mathcal{F}^n_t\}$. Assume that $\{X_n\}$ satisfies the compact containment condition, that is, for each $\epsilon>0$ and $T>0$, there exists a compact $K\subset E$ such that*

$$\inf_n \mathbb{P}[X_n(t)\in K, t\le T]\ge 1-\epsilon, \tag{3.2.17}$$

and assume that $\{Y_n(t): t\ge 0, n=1,2,\ldots\}$ is relatively compact (as a collection of E_2-valued random variables). Suppose that there is an operator $A:\mathcal{D}(A)\subset\overline{C}(E_1)\to C(E_1\times E_2)$ such that for $f\in\mathcal{D}(A)$ there is a process ϵ^f_n for which

$$f(X_n(t))-\int_0^t Af(X_n(s),Y_n(s))\mathrm{d}s+\epsilon^f_n(t) \tag{3.2.18}$$

is an $\{\mathcal{F}^n_t\}$-martingale. Let $\mathcal{D}(A)$ be dense in $\overline{C}(E_1)$ in the topology of uniform convergence on compact sets. Suppose that for each $f\in\mathcal{D}(A)$ and each $T>0$, there exists $p>1$ such that

$$\sup_n \mathbb{E}\left[\int_0^T |Af(X_n(t),Y_n(t)|^p\mathrm{d}t\right]<\infty \tag{3.2.19}$$

and

$$\lim_{n\to\infty}\mathbb{E}\left[\sup_{t\le T}|\epsilon^f_n(t)|\right]=0. \tag{3.2.20}$$

Let Γ_n be the $l_m(E_2)$-valued random variable given by

$$\Gamma_n([0,t]\times B)=\int_0^t \mathbf{1}_B(Y_n(s))\mathrm{d}s.$$

Then $\{(X_n, \Gamma_n)\}$ *is relatively compact in* $D_{E_1}[0,\infty) \times l_m(E_2)$*, and for any limit point* (X, Γ) *there exists a filration* $\{\mathcal{G}_t\}$ *such that*

$$f(X(t)) - \int_0^t \int_{E_2} Af(X(s), y)\Gamma(\mathrm{d}s \times \mathrm{d}y) \tag{3.2.21}$$

is a $\{\mathcal{G}_t\}$*-martingale for each* $f \in \mathcal{D}(A)$*.*

We will use Theorem 3.2.16 with our environment taking the place of Y_n. We observe that

$$f(\eta_t) - \int_0^t Af(\eta_t, \zeta_t)\mathrm{d}s + (\widehat{f}(\eta_t, \zeta_t) - f(\eta_t)) + \int_0^t Af(\eta_s, \zeta_s) - A^N \widehat{f}(\eta_s, \zeta_s)\mathrm{d}s \\ = \widetilde{f}(\eta_t, \zeta_t) - \int_0^t A^N \widetilde{f}(\eta_s, \zeta_s)\mathrm{d}s,$$

where A is given in (3.2.11). Since $\widetilde{f}_t - \int_0^t A^N \widetilde{f}_s \mathrm{d}s$ is a martingale, we have written our problem in the form (3.2.18) with

$$\epsilon_N^f(t) = (\widetilde{f}(\eta_t, \zeta_t) - f(\eta_t)) + \int_0^t Af(\eta_s, \zeta_s) - A^N \widetilde{f}(\eta_s, \zeta_s)\mathrm{d}s \tag{3.2.22}$$

To check that the assumptions of Theorem 3.2.16 are satisfied, we work with both the lookdown representation and the projected model. The projected model allows us to check the compact containment condition (3.2.17) and prove the L^p estimate (3.2.19). Both are achieved via an intensity estimate given by the following Lemma.

Lemma 3.2.17. *Let* $X^N = Kw^N$ *denote the total intensity of individuals of the rare type. Assume that* $\mathbb{E}[X^N(0)] < \infty$*. Then for any* $T > 0$

$$\sup_{t \le T} \sup_N \mathbb{E}[X^N(t)] < \infty, \tag{3.2.23}$$

$$\lim_{H \to \infty} \sup_N \mathbb{P}\left[\sup_{t \le T} X^N(t) > H\right] = 0. \tag{3.2.24}$$

We discuss the proof in Section 3.2.3.1.

The part of the argument which allows us to identify the correct limit and shows that condition (3.2.20) is satisfied, that is

$$\lim_{n\to\infty} \mathbb{E}\left[\sup_{t\leq T} |\epsilon_N^f(t)|\right] = 0, \tag{3.2.25}$$

is more involved, and requires the use of the lookdown representation. We will first look at the behaviour of

$$\int_0^t A^N f(\eta_s, \zeta_s) - Af(\eta_s, \zeta_s)\mathrm{d}s.$$

The terms involving A^N_{neu} and A^N_{sel} are tackled separately.

Let η_t^r be the process obtained from η_t by only considering the individuals of the rare type, $\eta_t^r := \{l : (l, \kappa_r) \in \eta_t\}$.

Proposition 3.2.18. *Under the conditions of Theorem 3.2.12,*

$$\mathbb{E}\Bigg[\sup_{t\leq T}\Bigg|\int_0^t A^N_{neu} f(\eta_s) - \left(f(\eta_s^r)\sum_{l_i(t)\in\eta_t^N} a l_i^2 \frac{g'(l_i(t))}{g(l_i(t))} + 2af(\eta_s^r)\sum_{l_i(t)\in\eta_s^r}\int_{l_i(t)}^\infty (1-g(\kappa_i, v))\mathrm{d}v\right)\mathrm{d}s\Bigg|\Bigg] \to 0.$$

This proposition will be proved via Taylor's formula through Proposition 3.2.24 and Proposition 3.2.25 in Section 3.2.3.2. An analogous proposition applies to the terms involving A^N_{sel}.

Proposition 3.2.19. *Under the conditions of Theorem 3.2.13 for any $T \in \mathbb{R}$*

$$\mathbb{E}\left[\sup_{t\leq T}\left|\int_0^t A^N_{sel} f(\eta_s^N) - f(\eta_s^N)\left(\sum_{l_i(t)\in\eta_t^N} -bl_i\frac{g'(l_i)}{g(l_i)}\right)\mathrm{d}s\right|\right] \to 0.$$

We discuss the proof of this proposition, along with calculations which allow us to fully justify the separation of timescales procedure in Section 3.2.3.3. These three propositions will allow us to conclude Theorem 3.2.12 and Theorem 3.2.13. We then proceed to use these results to prove Theorem 3.2.10 in Section 3.2.3.4. We notice that if the random perturbation Y appearing in the statement of Theorem 3.2.16 is

trivial, the statement itself reduces to the usual condition for relative compactness of the sequence of stochastic processes, for example in Ethier and Kurtz (1986), Theorem 3.9.1 and Theorem 3.9.4. Therefore as a by-product of our construction we give a proof of Theorem 3.2.12 and Theorem 3.2.13. The proof of Theorem 3.2.12 is the most technical of this section.

Remark 3.2.20. *Our proof guarantees the relative compactness of the sequences of scaled lookdown representations and that limit points must satisfy a martingale problem. However, we do not have a proof of uniqueness of the martingale problem characterizing the limiting equation. We will use the Markov Mapping Theorem to deduce the relative compactness of the sequence of projected models and a martingale problem characterising limit points. Lemma A.13 from Kurtz and Rodrigues (2011) guarantees that every projection given by the Markov Map solves the projected martingale problem. Since this projected martingale problem has unique solutions, the relative compactness is enough to guarantee convergence of the sequence of projected models.*

Remark 3.2.21. *We are interested in the behaviour of a rare subpopulation. As we have discussed earlier, we would like it to form $\mathcal{O}(1/K)$ of the population. For technical reasons, instead of the process $X^N(t)$, it is sometimes convenient to consider a stopped process $X^N(t \wedge \tau^N)$, where*

$$\tau^N := \inf\{t > 0 : X_t^N > Z^N\}.$$

We require the sequence of real numbers Z^N to be finite for each N and to tend to infinity as N tends to infinity. This requirement coupled with Lemma 3.2.17 guarantees that the convergence of the stopped processes translates directly into convergence of the unstopped processes. We shall see that technical assumptions will require that $Z^N \to \infty$ sufficiently slowly. However, these assumptions will not change our proof.

3.2.3.1 Intensity estimate

This subsection is devoted to the proof of our intensity estimate.

Proof[Proof of Lemma 3.2.17] The generator of the projected process can be written as

$$\begin{aligned}\mathcal{L}f(w,\zeta) = &\{wf((1-u)w+u,\zeta) + (1-p)f((1-u)w,\zeta) - f(w,\zeta)\}\\ &+ s\Big[\frac{\sigma(\kappa_r,\zeta)w}{\sigma(\kappa_r,\zeta)w+\sigma(\kappa_s,\zeta)(1-w)}pf((1-u)w+u,\zeta)\\ &+ \frac{\sigma(\kappa_c,\zeta)(1-w)}{\sigma(\kappa_r,\zeta)w+\sigma(\kappa_s,\zeta)(1-w)}wf((1-u)w+u,\zeta) - f(w,\zeta)\Big]\\ &+ \mathcal{L}^{env}f(p,\zeta).\end{aligned}$$

Recall that we are interested in an estimate for the intensity of the process describing the evolution of the rare individuals. We therefore substitute $X = Kw$. Under the assumptions of Theorem 3.2.10 this leads to the generator

$$\begin{aligned}\mathcal{L}^N f(X,\zeta) = &N\Big\{\frac{X}{K}f\Big(\Big(1-\frac{u}{J}\Big)X + K\frac{u}{J},\zeta\Big) + \Big(1-\frac{X}{K}\Big)f\Big(\Big(1-\frac{u}{J}\Big)X,\zeta\Big) - f(X,\zeta)\Big\}\\ &+ N\widehat{S}\frac{s}{S}\Bigg[\frac{\sigma(\kappa_r,\zeta)\frac{X}{K}}{\sigma(\kappa_r,\zeta)\frac{X}{K}+\sigma(\kappa_s,\zeta)\Big(1-\frac{X}{K}\Big)}f\Big(\Big(1-\frac{u}{J}\Big)X + K\frac{u}{J},\zeta\Big)\\ &+ \frac{\sigma(\kappa_c,\zeta)\Big(1-\frac{X}{K}\Big)}{\sigma(\kappa_r,\zeta)\frac{X}{K}+\sigma(\kappa_s,\zeta)\Big(1-\frac{X}{K}\Big)}f\Big(\Big(1-\frac{u}{J}\Big)X,\zeta\Big) - f(X,\zeta)\Bigg]\\ &+ \widehat{S}^2\mathcal{L}^{env}f(X,\zeta).\end{aligned}$$

This means that for any $f \in C^\infty(\mathbb{R})$,

$$f(X^N(T),\zeta_T) = f(X_0,\zeta_0) + \int_0^T \mathcal{L}(X^N(s),\zeta_s)\mathrm{d}s + M(T),$$

where M is a martingale. Substituting $f(x,\zeta) = x$ leads to

$$\begin{aligned}X(T) =&X(0) + N\widehat{S}\frac{s}{S}\frac{u}{J}\int_0^T\Bigg[\frac{\sigma(\kappa_r,\zeta)}{\sigma(\kappa_r,\zeta)\frac{X}{K}+\sigma(\kappa_c,\zeta)\Big(1-\frac{X}{K}\Big)}\frac{X}{K}(K-X)\\ &+\frac{-\sigma(\kappa_c,\zeta)}{\sigma(\kappa_r,\zeta)\frac{X}{K}+\sigma(\kappa_c,\zeta)\Big(1-\frac{X}{K}\Big)}\Big(1-\frac{X}{K}\Big)X\Bigg]\mathrm{d}s + M(T)\\ =&X(0) + N\widehat{S}\frac{s}{S}\frac{u}{J}\int_0^T\Bigg[\frac{\sigma(\kappa_r,\zeta)}{\sigma(\kappa_r,\zeta)\frac{X}{K}+\sigma(\kappa_c,\zeta)\Big(1-\frac{X}{K}\Big)}X\end{aligned}$$

$$-\left(\frac{\sigma(\kappa_r,\zeta)\frac{X}{K}}{\sigma(\kappa_r,\zeta)\frac{X}{K}+\sigma(\kappa_c,\zeta)\Big(1-\frac{X}{K}\Big)}+\frac{\sigma(\kappa_c,\zeta)\Big(1-\frac{X}{K}\Big)}{\sigma(\kappa_r,\zeta)\frac{X}{K}+\sigma(\kappa_c,\zeta)\Big(1-\frac{X}{K}\Big)}\right)X\Bigg]\mathrm{d}s$$
$$+M(T).$$

Since

$$\frac{\sigma(\kappa_r,\zeta)}{\sigma(\kappa_r,\zeta)\frac{X}{K}+\sigma(\kappa_c,\zeta)\Big(1-\frac{X}{K}\Big)}$$
$$=\frac{\sigma(\kappa_r,\zeta)}{\sigma(\kappa_c,\zeta)}+\frac{X}{K}\frac{\sigma(\kappa_c,\zeta)\sigma(\kappa_r,\zeta)-\sigma^2(\kappa_r,\zeta)}{\sigma(\kappa_c,\zeta)\Big(\sigma(\kappa_r,\zeta)\frac{X}{K}+\sigma(\kappa_c,\zeta)\Big(1-\frac{X}{K}\Big)\Big)}, \quad (3.2.26)$$

this expression can be written as

$$X(T)=X(0)+N\widehat{S}\frac{s}{S}\frac{u}{J}\Bigg\{\int_0^T\Bigg[\Big(\frac{\sigma(\kappa_r,\zeta)}{\sigma(\kappa_c,\zeta)}-1\Big)X\Bigg]\mathrm{d}s$$
$$+\int_0^T\left[\frac{X^2}{K}\frac{\sigma(\kappa_c,\zeta)\sigma(\kappa_r,\zeta)-\sigma^2(\kappa_r,\zeta)}{\sigma(\kappa_c,\zeta)\Big(\sigma(\kappa_r,\zeta)\frac{X}{K}+\sigma(\kappa_c,\zeta)\Big(1-\frac{X}{K}\Big)\Big)}\right]\mathrm{d}s\Bigg\}+M(T). \quad (3.2.27)$$

Consider the stopping time $\tau=\inf\{t\geq 0: X^N>H\}$ and the stopped process $\widehat{X}^N(t)=X^N(t\wedge\tau)$. Since (3.2.27) holds at a bounded stopping time, taking the expectation and using the symmetry condition (3.2.7) leads to

$$\mathbb{E}[\widehat{X}^N(T)]=\mathbb{E}[\widehat{X}(0)]+N\widehat{S}\frac{s}{S}\frac{u}{J}\mathbb{E}\Bigg[\int_0^T\Bigg\{\Big(\frac{\sigma(\kappa_c,\zeta)}{\sigma(\kappa_r,\zeta)}-1\Big)\widehat{X}^N(s)\Bigg\}\mathrm{d}s$$
$$+\int_0^T\left\{\frac{(\widehat{X}^N(s))^2}{K}\frac{\sigma^2(\kappa_r,\zeta)+\sigma(\kappa_c,\zeta)\sigma(\kappa_r,\zeta)}{\sigma(\kappa_c,\zeta)\Big(\sigma(\kappa_r,\zeta)\frac{\widehat{X}^N(s)}{K}+\sigma(\kappa_c,\zeta)\Big(1-\frac{\widehat{X}^N(s)}{K}\Big)\Big)}\right\}\mathrm{d}s\Bigg]$$
$$\leq\mathbb{E}[\widehat{X}(0)]+N\frac{H}{K}\frac{s\widehat{S}}{S}\frac{u}{J}C_\sigma\int_0^T\mathbb{E}[\widehat{X}^N(s)]\mathrm{d}s,$$

where C_σ is a constant depending on σ. By Grönwall's inequality

$$\mathbb{E}[\widehat{X}^N(T)]\leq\mathbb{E}[\widehat{X}^N(0)]\exp\Big(N\frac{H}{K}\frac{s\widehat{S}}{S}\frac{u}{J}C_\sigma T\Big).$$

If one takes $H=Z^N$ from Remark 3.2.21, combined with the conditions from Theorem 3.2.10, concludes the proof of (3.2.23). To see that (3.2.24) holds, it is enough to

observe that by Markov's inequality

$$\mathbb{P}\left[\sup_{0\leq t\leq T} X_t \geq H\right] = \mathbb{P}[\widehat{X}_T \geq H] \leq \frac{\mathbb{E}[\widehat{X}_T]}{H}, \tag{3.2.28}$$

and for any T, $\varepsilon > 0$ we can choose H_1, N_1 such that for $H \geq H_1$, $N \geq N_1$ and $t \leq T$ the right hand side of (3.2.28) is less than ε. □

We also take note of a simple corollary which we will use later in our main proof.

Corollary 3.2.22. *The lookdown process satisfies the compact containment condition,* (3.2.17) *in Theorem 3.2.16.*

Proof We note that the characterisation of convergence given in Theorem 3.5.57 ensures that a sequence of lookdown processes satisfies the compact containment condition if and only if the projected processes also satisfy compact containment condition. □

Remark 3.2.23. *We do not give details but a simpler calculation also proves Lemma 3.2.17 under the assumptions of Theorem 3.2.12 and Theorem 3.2.13.*

3.2.3.2 Neutral model - proof of Theorem 3.2.12

Even though this subsection contains the proof of the result for the least complicated model, the proof itself is the most involved one. The relative simplicity of the proof for the more complicated model demonstrates the power of this lookdown method. Proofs of Theorem 3.2.13 and Theorem 3.2.10 heavily rely on technical observations from this subsection. We recall that the generator of the neutral part of the process takes the form

$$A^N_{neu} f(\eta) = N\Bigg(\int_0^\infty \Bigg[\frac{uK}{J} e^{-\frac{uK}{J}v^*} g(\kappa^*, v^*) e^{-\frac{uK}{J}\int_{v^*}^\infty (1-g(\kappa^*,v))\mathrm{d}v} \times \prod_{(\kappa,l)\in\eta, l\neq l^*} g(\kappa, \mathcal{J}_{neu}(l, l^*, v^*)) \Bigg] \mathrm{d}v^* - f(\eta) \Bigg), \tag{3.2.29}$$

where (κ^*, v^*) denotes the parent and $\mathcal{J}_{neu}$ is defined as in (3.2.4), that is

$$\mathcal{J}_{neu}(l, l^*_{neu}, v^*) = \begin{cases} \frac{1}{1-u}(l - (l^*_{neu} - v^*)) & \text{if } l > l^*_{neu}, \\ \frac{1}{1-u} l & \text{if } l < l^*_{neu}, \\ v^* & \text{if } l = l^*_{neu}. \end{cases}$$

Before stating the propositions and lemmas which prove Theorem 3.2.12, let us rewrite (3.2.29) in a more convenient form.

We start by observing that a Taylor expansion of g gives

$$\begin{aligned} &g(\kappa^*, v^*) \prod_{(\kappa,l)\in\eta, l\neq l^*} g(\kappa, \mathcal{J}(l, l^*, v^*)) - f(\eta) \\ &= f(\eta) \sum_l \frac{g'(\kappa, l)}{g(\kappa, l)} (\mathcal{J}(l, l^*, v^*) - l) + \sum_{l,\widehat{l}} \frac{g'(\widehat{\kappa}, \widehat{l})}{g(\widehat{\kappa}, \widehat{l})} \frac{g'(\kappa, l)}{g(\kappa, l)} \mathcal{O}((\mathcal{J}(l, l^*, v^*) - l)(\mathcal{J}(\widehat{l}, l^*, v^*) - \widehat{l})) \\ &\qquad + \sum_l \frac{g''(\kappa, l)}{g^2(\kappa, l)} \mathcal{O}((\mathcal{J}^2(l, l^*, v^*) - l)), \end{aligned} \tag{3.2.30}$$

and a Taylor expansion of the exponential function about 0 leads to

$$e^{-\frac{uK}{J}\int_{v^*}^{\infty}(1-g(\kappa^*,v))\mathrm{d}v} = 1 - \frac{uK}{J}\int_{v^*}^{\infty}(1 - g(\kappa^*, v))\mathrm{d}v + \mathcal{O}\bigg(\frac{uK}{J}\int_{v^*}^{\infty}(1 - g(\kappa^*, v))\mathrm{d}v\bigg)^2. \tag{3.2.31}$$

Applying (3.2.31) we may rewrite (3.2.29) as

$$A^N_{neu} = A^N_{neu,1} + A^N_{neu,2} + \mathcal{O}\bigg(\frac{K}{J} A^N_{neu,2}\bigg),$$

where

$$A^N_{neu,1} = N\bigg(\int_0^{\infty} \frac{uK}{J} e^{-\frac{uK}{J}v^*} \Big[g(\kappa^*, v^*) \prod_{(\kappa,l)\in\eta, l\neq l^*} g(\kappa, \mathcal{J}(l, l^*, v^*)) - f(\eta)\Big] \mathrm{d}v^*\bigg), \tag{3.2.32}$$

$$\begin{aligned} A^N_{neu,2} = N\bigg(&-\int_0^{\infty} \frac{u^2K^2}{J^2} e^{-\frac{uK}{J}v^*} \int_{v^*}^{\infty}(1 - g(\kappa^*, v))\mathrm{d}v \; g(\kappa^*, v^*) \\ &\times \prod_{(\kappa,l)\in\eta, l\neq l^*} g(\kappa, \mathcal{J}(l, l^*, v^*))\mathrm{d}v^*\bigg). \end{aligned} \tag{3.2.33}$$

We begin with statements of the propositions that identify limits of (3.2.32) and (3.2.33) separately. The proofs appear later in this section.

Proposition 3.2.24. *Under the conditions of Theorem 3.2.12,*

$$\mathbb{E}\left[\sup_{t\leq T}\left|\int_0^t A^N_{neu,1}f(\eta_s) - f(\eta^r_s)\left(\sum_{l_i(t)\in\eta^r_t} al_i^2\frac{g'(l_i(t))}{g(l_i(t))}\right)\mathrm{d}s\right|\right]\to 0.$$

Proposition 3.2.25. *Under the conditions of Theorem 3.2.12,*

$$\mathbb{E}\left[\sup_{t\leq T}\left|\int_0^t\left\{A^N_{neu,2}f(\eta_s) - 2af(\eta^r_s)\sum_{l_i(t)\in\eta^r_s}\int_{l_i(t)}^\infty (1-g(\kappa_i,v))\mathrm{d}v\right\}\mathrm{d}s\right|\right]\to 0.$$

The proof of Proposition 3.2.24 will use the following elementary lemma:

Lemma 3.2.26. *Let Γ be a stochastic process. Assume that the second moment of Γ is bounded uniformly for all times up to time T by ϵ, that is $\mathbb{E}[(\Gamma(s))^2]\leq\epsilon$, for $0\leq s\leq T$. Then*

$$\mathbb{E}\left[\sup_{t<T}\left|\int_0^t\Gamma(s)\mathrm{d}s\right|\right] < T\sqrt{\epsilon}.$$

Proof By Jensen's inequality

$$\mathbb{E}\left[\sup_{t\leq T}\left(\int_0^t\Gamma(s)\mathrm{d}s\right)^2\right]\leq\ \mathbb{E}\left[\sup_{t\leq T}\ t\int_0^t\Gamma^2(s)\mathrm{d}s\right]=\ T\mathbb{E}\left[\int_0^T\Gamma^2(s)\mathrm{d}s\right].$$

The inequality follows by assumption and a final application of Jensen's inequality. □

Proof [Proof of Proposition 3.2.24] By Lemma 3.2.26, it suffices to prove that, conditioned on the position of rare levels,

$$\mathbb{E}\Bigg[\Bigg\{N\Bigg(\int_0^\infty\frac{uK}{J}e^{-\frac{uK}{J}v^*}\Bigg\{g(\kappa^*,v^*)\times\prod_{(\kappa(t),l(t))\in\eta,l(t)\neq l^*(t)}g(\kappa,\mathcal{J}(l(t),l^*(t),v^*))-f(\eta_s)\Bigg\}\mathrm{d}v^*\Bigg)-f(\eta^r_s)\left(\sum_i al_i^2\frac{g'(l_i(t))}{g(l_i(t))}\right)\Bigg\}^2\Bigg]\to 0,$$

uniformly for $0\leq s\leq T$.

For simplicity we use the notation $l_0 = 0$. We will use an ordering l_i by assuming $l_i > l_{i-1}$ and (3.2.30) to observe that $A^N_{neu,1} f(\eta)$ can be approximated by

$$\begin{aligned}
Nf(\eta)\Bigg(\sum_i \int_{l_{i-1}}^{l_i} \frac{uK}{J} e^{-\frac{uK}{J}v^*} \Bigg[\frac{g'(\kappa_i, l_i)}{g(\kappa_i, l_i)}(v^* - l_i) \\
+ \sum_{l \neq l_i} \frac{g'(\kappa, l)}{g(\kappa, l)} \Bigg(\frac{l\frac{u}{J}}{1 - \frac{u}{J}} - \mathbf{1}_{l > l_i} \frac{l_i - v^*}{1 - \frac{u}{J}} \Bigg) \Bigg] \mathrm{d}v^* \Bigg) \\
= Nf(\eta) \sum_i \Bigg[\frac{g'(\kappa_i, l_i)}{g(\kappa_i, l_i)} \Big(-(l_i - l_{i-1}) e^{-\frac{uK}{J} l_{i-1}} + \frac{J}{uK} \Big(e^{-\frac{uK}{J} l_{i-1}} - e^{-\frac{uK}{J} l_i} \Big) \Big) \\
+ \sum_{j \neq i} \Bigg(\frac{g'(\kappa_j, l_j)}{g(\kappa_j, l_j)} \frac{l_j \frac{u}{J}}{1 - \frac{u}{J}} \Big(e^{-\frac{uK}{J} l_{i-1}} - e^{-\frac{uK}{J} l_i} \Big) \Bigg) \\
+ \sum_{j \neq i} \sum_{j < i} \frac{g'(\kappa_j, l_j)}{g(\kappa_j, l_j)} \frac{1}{1 - \frac{u}{J}} \Big(-(l_i - l_{i-1}) e^{-\frac{uK}{J} l_{i-1}} + \frac{J}{uK} \Big(e^{-\frac{uK}{J} l_{i-1}} - e^{-\frac{uK}{J} l_i} \Big) \Big) \Bigg], \quad (3.2.34)
\end{aligned}$$

where the second line follows from integration. At this stage we note that we can drop the factor $1 - u/J$ at the cost of an error of order NK/J^3, which tends to zero as N tends to infinity. We also swap the order of summation to rewrite (3.2.34) as

$$\begin{aligned}
Nf(\eta) \sum_j \frac{g'(\kappa_j, l_j)}{g(\kappa_j, l_j)} \Bigg[\sum_{i \leq j} \Bigg(- (l_i - l_{i-1}) e^{-\frac{uK}{J} l_{i-1}} + \frac{J}{uK} \Big(e^{-\frac{uK}{J} l_{i-1}} - e^{-\frac{uK}{J} l_i} \Big) \Bigg) \\
+ l_j \frac{u}{J} \sum_{i \neq j} \Big(e^{-\frac{uK}{J} l_{i-1}} - e^{-\frac{uK}{J} l_i} \Big) \Bigg].
\end{aligned}$$

We observe that since the second and the third terms in the inner sum are telescoping sums and $l_0 = 0$ we simplify our expression to

$$\begin{aligned}
f(\eta) \sum_j \frac{g'(\kappa_j, l_j)}{g(\kappa_j, l_j)} (P_1 + P_2 + P_3) := Nf(\eta) \sum_j \frac{g'(\kappa_j, l_j)}{g(\kappa_j, l_j)} \Bigg[\frac{J}{uK} \Big(1 - e^{-\frac{uK}{J} l_j} \Big) \\
+ l_j \frac{u}{J} \Big(1 - e^{-\frac{uK}{J} l_{j-1}} + e^{-\frac{uK}{J} l_j} \Big) + \sum_{i \leq j} \Big(-(l_i - l_{i-1}) e^{-\frac{uK}{J} l_{i-1}} \Big) \Bigg], \quad (3.2.35)
\end{aligned}$$

where

$$\begin{aligned}
P_1 &= \frac{NJ}{uK} \Big(1 - e^{-\frac{uK}{J} l_j} \Big), \\
P_2 &= l_j \frac{uN}{J} \Big(1 - e^{-\frac{uK}{J} l_{j-1}} + e^{-\frac{uK}{J} l_j} \Big),
\end{aligned}$$

$$P_3 = N \sum_{i \leq j} \Big(-(l_i - l_{i-1}) e^{-\frac{uK}{J} l_{i-1}} \Big).$$

Now we treat each term in the new sum separately. Using the Taylor expansion of the exponential function about 0 we observe that

$$\begin{aligned}
P_1 &= N \sum_{1 \leq k \leq n+1} \frac{(-1)^{k-1}}{k!} l_j^k \left(\frac{uK}{J} \right)^{k-1} + \mathcal{O}\left(\frac{N(uK)^{n+1}}{J^{n+1}} \right), \\
P_2 &= \frac{uNl_j}{J} + \mathcal{O}\Big(\frac{NK}{J^2} (l_j - l_{j-1}) \Big), \\
P_3 &= N \sum_{1 \leq k \leq n+1} \frac{(-1)^k}{(k-1)!} \left(\frac{uK}{J} \right)^{k-1} \sum_{i \leq j} (l_i - l_{i-1}) l_{i-1}^{k-1} + \mathcal{O}\left(\frac{N(uK)^{n+1}}{J^{n+1}} \right).
\end{aligned}$$

We focus our attention on P_3. We investigate the terms corresponding to different values of k separately. We observe that the first three terms involve

$$\sum_{i \leq j} (l_i - l_{i-1}) = l_j, \tag{3.2.36}$$

$$\begin{aligned}
\sum_{i \leq j} (l_i - l_{i-1}) l_{i-1} &= \frac{1}{2} \left(\sum_{i \leq j} (l_i^2 - l_{i-1}^2) - \sum_{i \leq j} (l_i - l_{i-1})^2 \right) \\
&= \frac{1}{2} \left(l_j^2 - \sum_{i \leq j} (l_i - l_{i-1})^2 \right)
\end{aligned} \tag{3.2.37}$$

$$\begin{aligned}
\sum_{i \leq j} \Big((l_i - l_{i-1}) l_{i-1}^2 \Big) &= \frac{1}{3} \sum_{i \leq j} \Big(l_i^3 - l_{i-1}^3 \Big) - \frac{1}{3} \sum_{i \leq j} (l_i - l_{i-1})^3 - \sum_{i \leq j} (l_i - l_{i-1})^2 l_{i-1} \\
&= \frac{1}{3} l_j^3 - \frac{1}{3} \sum_{i \leq j} (l_i - l_{i-1})^3 - \sum_{i \leq j} (l_i - l_{i-1})^2 l_{i-1},
\end{aligned} \tag{3.2.38}$$

respectively. We are therefore interested in $\sum_i (l_i - l_{i-1})^2$ conditioned on the locations of the rare levels, that is, conditioned on η^r.

Recall that conditioned on the locations of the levels of the rare type, the levels of the common type will be Poisson distributed with intensity $K - Z^N$. Conditioned on the number of levels of the rare type between 0 and l_j these levels will be independent uniformly distributed on $[0, l_j]$. We denote n independent uniformly distributed random variables on $[0, 1]$ by $u_1, \ldots, u_n$ and define the order statistics by letting $u_{(i)} = u_k$ if and only if $\#\{\widehat{k} : u_{\widehat{k}} \leq u_k\} = i$ for $i \in \{1, \ldots n\}$. We note that n points uniformly distributed on $[0, 1]$ can be identified with $n+1$ points uniformly distributed on the unit

circle with one of these points chosen at random to be a reference point corresponding to both 0 and 1. This then leads us to see that $u_{(i)} - u_{(i-1)}$ is equal in distribution to $u_{(1)}$ for $i \in \{1, \dots, n+1\}$, where, by convention, $u_{(n+1)} := 1$ and $u_{(0)} := 0$. From this one can see that

$$\mathbb{E}\left[\sum_{i=1}^{j}(l_i - l_{i-1})^2 | l_0 = 0, j = n+1, l_j\right] = l_j^2 \mathbb{E}\left[\sum_{i=1}^{n+1}(u_{(i)} - u_{(i-1)})^2\right] = l_j^2 (n+1)\frac{2}{(n+1)(n+2)}.$$

We then use that the number of levels of the rare type within $[0, l_j]$ will be Poisson distributed to see

$$\begin{aligned}\mathbb{E}\left[\sum_{i=1}^{j}(l_i - l_{i-1})^2 | l_0 = 0, l_j\right] =& \sum_{i=0}^{\infty}\frac{2l_j^2}{n+2}\frac{\Big(l_j(K-Z^N)\Big)^n \exp(-\Big(l_j(K-Z^N)\Big))}{n!} \\ =& \frac{2l_j}{(K-Z^N)} + \mathcal{O}\Big(\exp\big(-l_j(K-Z^N)\big)\Big).\end{aligned}$$

Identical arguments show

$$\begin{aligned}\mathbb{E}\left[\left(\sum_{i=1}^{j}(l_i - l_{i-1})^2\right)^2 | l_0 = 0, l_j\right] =& \frac{4l_j}{(K-Z^N)^3}(3 + x(K-Z^N)) \\ &+ \mathcal{O}\Big((K-Z^N)\exp\big(-l_j(K-Z^N)\big)\Big), \quad (3.2.39) \\ \mathrm{Var}\left(\sum_{i=1}^{j}(l_i - l_{i-1})^2 | l_0 = 0, l_j\right) =& \mathcal{O}\Big(\frac{l_j}{(K-Z^N)^3}\Big).\end{aligned}$$

We note that we are considering j to be a random variable throughout this calculation, corresponding to the level of a given rare individual.

From this calculation since we multiply (3.2.37) by NK/J in we see that we require $\frac{N^2}{KJ^2} \to 0$. Therefore, we may approximate $\left(\frac{uK}{J}\right)\sum_{i\le j}(l_i - l_{i-1})^2$ by $\frac{ul_j}{J}$. For $k = 3$ we again condition on the number of levels of the rare type beneath l_j and see that

$$\mathbb{E}\left[\sum_{i=1}^{j}(l_i - l_{i-1})^2 l_{i-1} | l_0 = 0, j = n+1, l_j\right] = (n+1)\mathbb{E}\Big[(l_I - l_{I-1})^2 l_{I-1} | l_0 = 0, j = n+1, l_j\Big],$$

where I is chosen uniformly at random from $(1, \cdots, n+1)$. We then again consider the levels as $n+1$ points chosen uniformly at random from a circle to see that this will be $\frac{l_j^3}{n+2}$. This then gives us

$$\mathbb{E}\left[\sum_{i\leq j}(l_i - l_{i-1})^2 l_{i-1} | l_0 = 0, l_j\right] = \frac{l_j^2}{K - Z^N} + \mathcal{O}\Big(\exp(-l_j(K - Z^N))\Big). \tag{3.2.40}$$

We see that for $k \geq 3$

$$\mathbb{E}\left[\sum_{i\leq j}\Big((l_i - l_{i-1})l_{i-1}^k\Big) - \frac{1}{k+1}l_j^{k+1}\right] = \mathcal{O}\Big(\frac{1}{K}\Big), \tag{3.2.41}$$

$$\mathrm{Var}\left(\sum_{i\leq j}\Big((l_i - l_{i-1})l_{i-1}^k\Big) - \frac{1}{k+1}l_j^{k+1}\right) = \mathcal{O}\Big(\frac{1}{K^3}\Big), \tag{3.2.42}$$

which will suffice as each of these terms will be multiplied by $\frac{NK^{k-1}}{J^{k-1}}$ in (3.2.35). We then note that $\frac{NK}{J^2}$ is bounded and $\frac{K}{J} \to 0$. We combine (3.2.36), (3.2.37), (3.2.38), (3.2.41) to see that, conditioned on l_j,

$$\mathbb{E}\left[\sum_{i\leq j}\Big(-(l_i - l_{i-1})e^{-\frac{uK}{J}l_{i-1}}\Big)\middle| l_j\right]$$
$$= \sum_{0\leq k\leq n}\frac{(-1)^{k+1}}{(k+1)!}\left(\frac{uK}{J}\right)^k l_j^{k+1} - \frac{u}{J}l_j + \frac{1}{2}\frac{u^2K}{J^2}l_j^2 + \mathcal{O}\left(\frac{K^2}{J^3}\right) + \mathcal{O}\Big(\frac{K}{J}\Big)^{n+1}.$$

Finally, we observe that all these approximations and cancellations allow us to approximate (3.2.35) by

$$f(\eta)\sum_j \frac{g'(\kappa_j, l_j)}{g(\kappa_j, l_j)}\left(\frac{1}{2}\frac{u^2NK}{J^2}l_j^2 + \mathcal{O}\left(\frac{NK^2}{J^3}\right) + \mathcal{O}\left(\frac{NK^{n+1}}{J^{n+1}}\right) + \mathcal{O}\Big(\frac{NK}{J^2}(l_j - l_{j-1})\Big)\right),$$

plus a random, mean zero, correction term with variance $\mathcal{O}(N^2K^2/(J^2(K-Z)^3))$. □

The proof of Proposition 3.2.25 is more involved than that of Proposition 3.2.24. Once again we begin with an elementary lemma.

Lemma 3.2.27. *Let $\Gamma(s)$ be a stochastic process. For any partition $0 = t_0 < t_1 < \cdots < t_m = T$,*

$$\mathbb{E}\left[\sup_{t\leq T}\left|\int_0^t \Gamma(s)\mathrm{d}s\right|\right] \leq \sum_{j=1}^{m}\sqrt{\mathbb{E}\left[\left(\int_{t_{j-1}}^{t_j}\Gamma(s)\mathrm{d}s\right)^2\right]} + \mathbb{E}\left[\sup_j \int_{t_j-1}^{t_j}|\Gamma(s)|\mathrm{d}s\right].$$

Proof We observe that

$$\begin{aligned}\mathbb{E}\left[\sup_{t\leq T}\left|\int_0^t \Gamma(s)\mathrm{d}s\right|\right] &\leq \mathbb{E}\left[\sum_{j=1}^{m}|\int_{t_{j-1}}^{t_j}\Gamma(s)\mathrm{d}s| + \sup_j\int_{t_j-1}^{t_j}|\Gamma(s)|\mathrm{d}s\right]\\ &\leq \sum_{j=1}^{m}\mathbb{E}\left[\left|\int_{t_{j-1}}^{t_j}\Gamma(s)\mathrm{d}s\right|\right] + \mathbb{E}\left[\sup_j\int_{t_j-1}^{t_j}|\Gamma(s)|\mathrm{d}s\right]\\ &\leq \sum_{j=1}^{m}\sqrt{\mathbb{E}\left[\left(\int_{t_{j-1}}^{t_j}\Gamma(s)\mathrm{d}s\right)^2\right]} + \mathbb{E}\left[\sup_j\int_{t_j-1}^{t_j}|\Gamma(s)|\mathrm{d}s\right].\end{aligned}$$

which concludes the proof. □

We require a few more computations to transform $A^N_{neu,2}$ into a more convenient form. First of all, we observe that as $(1-g)$ is bounded, the following approximation is valid:

$$\begin{aligned}\int_{l_{i-1}}^{l_i}\int_{v^*}^{\infty}(1-g(\kappa_i,v))\mathrm{d}v\,\mathrm{d}v^* &= \int_{l_{i-1}}^{\infty}\int_{l+i-1}^{l_i\wedge v}(1-g(\kappa_i,v))\mathrm{d}v^*\,\mathrm{d}v\\ &= \int_{l_i}^{\infty}\int_{l+i-1}^{l_i}(1-g(\kappa_i,v))\mathrm{d}v^*\,\mathrm{d}v + \int_{l_{i-1}}^{l_i}\int_{l+i-1}^{v}(1-g(\kappa_i,v))\mathrm{d}v^*\,\mathrm{d}v\\ &= (l_i-l_{i-1})\int_{l_i}^{\infty}(1-g(\kappa_i,v))\mathrm{d}v + \mathcal{O}\Big((l_i-l_{i-1})^2\Big). \quad (3.2.43)\end{aligned}$$

For convenience, we order the individuals present in the system according to their level, that is we consider $\eta = \{(\kappa_i, l_i)\}_{i\geq 1}$ where $l_i < l_{i+1}$. We observe that the ordering leads to the following simplification:

$$A^N_{neu,2}f(\eta) = -\frac{u^2NK}{J}\sum_i\left(\int_{l_{i-1}}^{l_i}\frac{uK}{J}e^{-\frac{uK}{J}v^*}f(\eta)\int_{v^*}^{\infty}(1-g(\kappa_i,v))\mathrm{d}v\,\mathrm{d}v^*\right).$$

Since $\frac{K}{J}\to 0$, we may use (3.2.43) to further simplify $A^N_{neu,2}$ to

$A^N_{neu,2}f(\eta)$

$$= -\frac{u^2NK^2}{J^2} f(\eta) \sum_i (l_i - l_{i-1}) \Big(\int_{l_i}^{\infty} (1 - g(\kappa_i, v)) \mathrm{d}v \, \mathrm{d}v^* \Big) \Big(1 + \mathcal{O}\Big(\frac{K}{J} + \frac{1}{K^2} \Big) \Big).$$

By this calculation and Lemma 3.2.27, it is clear that in order to prove Proposition 3.2.25, it is enough to show that for any partition $0 = t_0 < t_1 < \cdots < t_m = T$,

$$\sum_{j=1}^{m} \sqrt{\mathbb{E}\left[\left(\int_{t_{j-1}}^{t_j} \Gamma^N(s) \mathrm{d}s \right)^2 \right]} + \mathbb{E}\left[\sup_j \int_{t_j - 1}^{t_j} |\Gamma^N(s)| \mathrm{d}s \right] \to 0, \tag{3.2.44}$$

where

$$\Gamma^N(s) = \sum_i \left(\frac{u^2NK^2}{J^2} (l_i(t) - l_{i-1}(t)) - 2a \right) f(\eta_t) \int_{l_i(t)}^{\infty} (1 - g(\kappa_i, v)) \mathrm{d}v \, \mathrm{d}v^* \mathrm{d}t. \tag{3.2.45}$$

We first turn our attention to the parts without a supremum which we calculate directly.

Lemma 3.2.28. *Conditioned on the locations of the rare levels,*

$$\mathbb{E}\left[\left(\int_0^T \sum_i \left(\frac{u^2NK^2}{J^2} (l_i(t) - l_{i-1}(t)) - 2a \right) f(\eta_t) \int_{l_i(t)}^{\infty} (1 - g(\kappa_i, v)) \mathrm{d}v \, \mathrm{d}v^* \mathrm{d}t \right)^2 \right] = \mathcal{O}\Big(\frac{LT}{N} \Big) + o(T^2), \tag{3.2.46}$$

for any $L(N)$ such that $\frac{L}{J} \to \infty$.

Proof We recall that there exists λ_g such that for all $v \geq \lambda_g$, $g(v) = 1$. Therefore

$$f(\eta_t) \int_{l_i(t)}^{\infty} (1 - g(\kappa_i, v)) \mathrm{d}v$$

is bounded for each test function. As η_t is constant between events we condition on exactly n reproductive events occurring within $[0, T]$. We denote the time of events by t_i for $i \in \{1, \ldots, n\}$. Observe that t_i are independent random variables uniformly distributed on $[0, T]$. For ease of notation we let $t_0 = 0$ and $t_{n+1} = T$. We will use $\widehat{l}_k$ to denote a rare type individual's level and $\widetilde{l}_k$ to be the highest level across the population (that is, from individuals of both rare and common type) below $\widehat{l}_k$ and consider

$$\mathbb{E}\left[\sum_{i=0}^{n} \sum_{j=0}^{n} (t_{i+1} - t_i)(t_{j+1} - t_j) \right.$$

$$\left(\frac{u^2NK^2}{J^2}(\widehat{l}_k(t_i)-\widetilde{l}_k(t_i))-2a\right)\left(\frac{uNK^2}{J^2}(\widehat{l}_k(t_j)-\widetilde{l}_k(t_j))-2a\right)\Bigg]. \quad (3.2.47)$$

We note that the times between events are independent of the level process and so we use

$$\mathbb{E}[(t_{i+1}-t_i)(t_{j+1}-t_j)] = \begin{cases} \frac{T^2}{(n+1)(n+2)} & \text{if } i \neq j, \\ \frac{2T^2}{(n+1)(n+2)} & \text{if } i = j. \end{cases}$$

We also use the Cauchy–Schwarz inequality to see

$$\mathbb{E}[(\widehat{l}_k(t_i)-\widetilde{l}_k(t_i))(\widehat{l}_k(t_j)-\widetilde{l}_k(t_j))] \leq \frac{2}{(K-Z^N)^2},$$

where we note that we are still looking at the stopped process to see $Z^N/K \to 0$.

We denote the intensity, after n events, of the levels of all individuals which were offspring in one of those n events by I_n. It is then easy to see $I_{n+1} = I_n(1-\frac{u}{J})+\frac{uK}{J}$ and $I_0 = 0$. This recurrence equation has a solution given by

$$I_n = K\left(1-\left(1-\frac{u}{J}\right)^n\right).$$

We now introduce some notation in order to simplify the calculations presented in this section. For $i<j$ we define $\widetilde{l}_k^{new}(t_j)$ as the highest level below $\widehat{l}_k$ of an individual (of either rare or common type) which has been born since t_i. Then we can see that

$$(\widehat{l}_k(t_j)-\widetilde{l}_k(t_j)) \leq (\widehat{l}_k(t_j)-\widetilde{l}_k^{new}(t_j)),$$

and so

$$\mathbb{E}[(\widehat{l}_k(t_i)-\widetilde{l}_k(t_i))(\widehat{l}_k(t_j)-\widetilde{l}_k(t_j))] \leq \frac{1}{K-Z^N}\frac{1}{I_{j-i}}.$$

We then also see that

$$\begin{aligned}\mathbb{E}[(\widehat{l}_k(t_i)-\widetilde{l}_k(t_i))(\widehat{l}_k(t_j)&-\widetilde{l}_k(t_j))] \\ &\geq \mathbb{E}[(\widehat{l}_k(t_i)-\widetilde{l}_k(t_i))(\widehat{l}_k(t_j)-\widetilde{l}_k(t_j))\mathbf{1}_{\widetilde{l}_k(t_j)=\widetilde{l}_k^{new}(t_j)}]\end{aligned}$$

$$
\begin{aligned}
=&\mathbb{E}[(\widehat{l}_k(t_i)-\widetilde{l}_k(t_i))(\widehat{l}_k(t_j)-\widetilde{l}_k^{new}(t_j))]\\
&-\mathbb{E}[(\widehat{l}_k(t_i)-\widetilde{l}_k(t_i))(\widehat{l}_k(t_j)-\widetilde{l}_k^{new}(t_j))\mathbf{1}_{\widetilde{l}_k(t_j)\neq\widetilde{l}_k^{new}(t_j)}]\\
\geq&\frac{1}{I_{j-i}}\frac{1}{K-Z^N}-\sqrt{\frac{2}{(K-Z^N)^2}\frac{2}{I_{j-i}^2}\mathbb{P}[\widetilde{l}_k(t_j)\neq\widetilde{l}_k^{new}(t_j)]},
\end{aligned}
$$

where we have used the Cauchy-Schwartz inequality for the final line.

We note that

$$
\mathbb{P}[\widetilde{l}_k(t_j)\neq\widetilde{l}_k^{new}(t_j)]\leq\frac{K(1-\frac{u}{J})^{j-i}}{K(1-\frac{u}{J})^{j-i}+I_{j-i}}.
$$

We now consider a function $L(N)$. By splitting (3.2.47) into parts with $|i-j|\leq L$ and $|i-j|>L$, we bound it above by

$$
\begin{aligned}
&\frac{4(n+1)LT^2}{(n+1)(n+2)}\left(\frac{N^2K^4}{J^4}\frac{2}{(K-Z^N)^2}+4a^2\right)\\
&+\frac{C(n+1)(n+1-L)T^2}{(n+1)(n+2)}\left\{\frac{N^2K^4}{J^4}\frac{1}{K-Z^N}\frac{1}{I_L}-\left(\frac{NK^2}{J^2}\frac{1}{K-Z^N}+\frac{NK^2}{J^2}\frac{1}{I_L}\right)2a+4a^2\right.\\
&\qquad\left.+\mathcal{O}\left(\frac{N^2K^4}{J^4}\frac{1}{K-Z^N}\frac{1}{I_L}\sqrt{\frac{K(1-\frac{u}{J})^L}{K(1-\frac{u}{J})^L+I_L}}\right)\right\}.
\end{aligned}
$$

We observe that

$$
\begin{aligned}
\sum_{n=1}^{\infty}\frac{n+1}{(n+1)(n+2)}\frac{(NT)^ne^{-NT}}{n!}=&\frac{1}{NT}-\frac{1}{(NT)^2}+\frac{e^{-NT}}{(NT)^2}-\frac{e^{-NT}}{2},\\
\sum_{n=1}^{\infty}\frac{(n+1)^2}{(n+1)(n+2)}\frac{(NT)^ne^{-NT}}{n!}\leq&1.
\end{aligned}
$$

We may combine the two results above to see

$$
\begin{aligned}
&\mathbb{E}\left[\left(\int_0^T\sum_i\left(\frac{u^2NK^2}{J^2}(l_i(t)-l_{i-1}(t))-2a\right)f(\eta_t)\int_{l_i(t)}^{\infty}(1-g(\kappa_i,v))\mathrm{d}v\,\mathrm{d}v^*\mathrm{d}t\right)^2\right]\\
&\qquad=\mathcal{O}\left(\frac{LT}{N}\left(\frac{N^2K^4}{J^4}\frac{2}{(K-Z)^2}+4a^2\right)\right.\\
&\qquad+T^2\left\{\frac{N^2K^4}{J^4}\frac{1}{K-Z}\frac{1}{I_L}-\left(\frac{NK^2}{J^2}\frac{1}{K-Z}+\frac{NK^2}{J^2}\frac{1}{I_L}\right)2a\right.
\end{aligned}
$$

$$+ 4a^2 + \mathcal{O}\left(\frac{N^2K^4}{J^4}\frac{1}{K-Z}\frac{1}{I_L}\sqrt{\frac{K(1-\frac{u}{J})^L}{K(1-\frac{u}{J})^L + I_L}}\right)\Bigg\}\Bigg).$$

Therefore, in order to conclude that (3.2.46) converges to 0, it is enough to find L which satisfies

$$\frac{L}{N} \to 0; \quad \frac{NK^2}{J^2}\frac{1}{I_L} \to 2a; \quad \frac{K(1-\frac{u}{J})^L}{K(1-\frac{u}{J})^L + I_L} \to 0.$$

Since $\frac{N}{J} \to \infty$, this is achieved by any $L(N)$ such that $\frac{L}{J} \to \infty$ and $\frac{L}{N} \to 0$. For example, we may choose $L(N) = \sqrt{NJ}$.

□

We now turn our attention back to (3.2.44). We consider (3.2.44) and $t_j = j\delta(N)$ where $\delta(N)$ is a function of N. We see that to ensure the first part of (3.2.44) converges to zero, we need choices of L and δ such that

$$\frac{L}{N} \to 0; \quad \frac{L}{J} \to \infty; \quad \frac{L}{N\delta} \to 0. \tag{3.2.48}$$

To show that the second term of (3.2.44) converges to zero, it is enough to show boundedness of $\mathbb{E}[\sup_{t\leq T}\Gamma^N(t)]$, provided that there exists a $\delta(N)$ satisfying (3.2.48) such that $\delta(N) \to 0$ as $N \to \infty$. This is satisfied by taking $\delta(N) = \sqrt{L/N}$.

To show $\mathbb{E}[\sup_{t\leq T}\Gamma^N(t)]$ is bounded we consider an auxiliary process β^N, defined by

$$\beta_t^N = \frac{NK^2}{J^2}\sum_{l(t)\in\eta_t^r}\Big(l(t) - \widehat{l}(t)\Big)h_\lambda(l(t)), \tag{3.2.49}$$

where h is a decreasing, positive cut-off function (that is, we assume that there exists a λ such that $h(v) = 0$ for $v > \lambda$) and $\widehat{l}(t) \in \eta_t$ is the first level below $l(t)$. To show the required bound on the expectation of the supremum of the integral of X, (and therefore conclude the proof), we show that β^N is dominated by a bounded submartingale. We note that under the conditions of Theorem 3.2.10 $\frac{NK^2}{J^2} = \mathcal{O}(K)$.

Lemma 3.2.29. *Define β^N as in* (3.2.49). *Then β^N is dominated by a bounded submartingale.*

Proof We recall that the offspring in an event can be ordered $(v_1, v_2, \dots)$ with $v_1 = v^*$. Therefore, as the parent is thought of as moving to v^* we will refer to $(v_i)_{i\geq 2}$ as the 'children' after an event. Recall that within the period of time $[0,t]$ with probability $\mathcal{O}(t^2)$ we will see two or more events. Therefore

$$\begin{aligned}
&\frac{\mathbb{E}\Big[\sum_{l(t)\in\eta_t^r}\Big(l(t)-\widehat{l}(t)\Big)h(l(t)) - \sum_{l(0)\in\eta_0^r}\Big(l(0)-\widehat{l}(0)\Big)h(l(0))\Big]}{t} \\
&\quad = N\int_0^\infty \frac{uK}{J}e^{-\frac{uK}{J}v^*}\Bigg\{\sum_{l(0)\in\eta_t^r}\Bigg\{\Big[\Big(\mathcal{J}(l(0))-\mathcal{J}(\widehat{l}(0))\Big)-\Big(l(0)-\widehat{l}(0)\Big)\Big]h(l(0)) \\
&\qquad\qquad + \Big(\mathcal{J}(\widehat{l}(0))-\widehat{l}(t)\Big)h(l(0)) \\
&\quad + \Big(l(t)-\widehat{l}(0)\Big)(\mathcal{J}(l(0))-l(0))\mathcal{O}(||h'||_\infty)\Bigg\} + \sum_{i=2}^{\infty}(v_i-\widehat{v}_i)h(v_i)\mathbf{1}_{\{v_i \text{ is rare}\}}\Bigg\}\mathrm{d}v^* + \mathcal{O}(t),
\end{aligned}$$

where we are using $\widehat{v}_i$ to denote the highest level below the ith 'child', v_i. We now require that $h(l) = 0$ for all $l \geq \lambda_h$. We note that in the proof of Proposition 3.2.24 we have shown that the parts involving $\mathcal{J}(l(0)) - l(0)$ converge in L^2 to $al(0)^2$ and so we can see that this is approximated by

$$\begin{aligned}
&\sum_{l(0)\in\eta_t^r}\Big\{\Big(al^2(0)-a\widehat{l^2}(0)\Big)h(l(0)) + \Big(l(0)-\widehat{l}(0)\Big)al^2(0)\mathcal{O}(||h'||_\infty)\Big\} \\
&+N\int_0^\infty \frac{uK}{J}e^{-\frac{uK}{J}v^*}\Bigg\{\sum_{l(0)\in\eta_t^r}\Big(\mathcal{J}(\widehat{l}(0))-\widehat{l}(t)\Big)h(l(0)) + \sum_{i=2}^{\infty}(v_i-\widehat{v}_i)h(v_i)\mathbf{1}_{\{v_i \text{ is rare}\}}\Bigg\}\mathrm{d}v^*.
\end{aligned}$$

We note that since the ordering of the $l(t) \in \eta_t$ which are not the parent is unaffected by an event, the only possibilities that will force $\mathcal{J}(\widehat{l}(0)) < \widehat{l}(t)$ is when a 'child' has been born between the new locations of levels or when $\widehat{l}(0)$ is chosen as the parent. However we use the last sum appearing in the previous equation and the fact that h is decreasing to see that the generator applied to β_t^N, for large enough N, is bounded below by

$$\begin{aligned}
-C_h\beta_t^N + NS\frac{NK^2}{J^2}\sum_{l_i\in\eta_t^r}\int_{l_{i-2}}^{\frac{l_{i-2}}{1-\frac{u}{J}}}\frac{uK}{J}e^{-\frac{uK}{J}v^*}\left(v^* - \frac{l_{i-2}}{1-\frac{u}{J}}\right)\mathrm{d}v^* \\
\geq -C_h\beta_t^N - \sum_{l_i\in\eta_t^r}\mathcal{O}\left(\frac{NK^2}{J^3}l_{i-2}^2\right),
\end{aligned}$$

where $C_h = \lambda_h ||h'||_\infty$. We therefore see that

$$\beta_t^N + \int_0^t C_h \beta_s^N + C\frac{NK^2Z}{J^3}\mathrm{d}s,$$

is a sub-martingale. Combined with the conditions of Theorem 3.2.10 this concludes the proof of the lemma. □

To complete the proof of Proposition 3.2.25 we see that, by using Lemma 3.2.28,

$$\mathbb{E}\left[\left(\beta_T^N + \int_0^T C_h \beta_s^N + C\frac{NK^2Z}{J^3}\mathrm{d}s\right)^2\right],$$

is bounded. Therefore, by Jensen's inequality,

$$\mathbb{E}[\sup_{t\leq T}\beta_t^N] \leq \sqrt{\mathbb{E}\left[\left(\sup_{t\leq T}\beta_t^N\right)^2\right]}$$

which is also bounded by Doob's martingale inequality. Proposition 3.2.25 therefore follows by an application of Lemma 3.2.27.

3.2.3.3 Selective model - proof of Theorem 3.2.13

The generator of the model presented in Theorem 3.2.13 is of the form

$$A^N f(\eta) = A^N_{neu} f(\eta) + A^N_{sel} f(\eta).$$

The analysis for the neutral part of the generator, A^N_{neu}, is the same as in Section 3.2.3.2, with the exception of a slight modification of Lemma 3.2.29, which we discuss in Lemma 3.2.32. We therefore turn our attention to A^N_{sel}, which is the part of the generator describing the selection. It can be written as

$$A^N_{sel} f(\eta) = \frac{sN}{S}\Bigg(\int_0^\infty \Bigg[\frac{uK}{J} e^{-\frac{uK}{J}v^*} g(\kappa^*, v^*) e^{-\frac{uK}{J}\int_{v^*}^\infty (1-g(\kappa^*,v))\mathrm{d}v} \times \prod_{(\kappa,l)\in\eta, l\neq l^*_{sel}} g(\kappa, \beta_{sel}((l,\kappa),(l^*_{sel},\kappa^*),v^*))\Bigg]\mathrm{d}v^* - f(\eta)\Bigg), \quad (3.2.50)$$

where we recall that the movement of the levels $\mathcal{J}_{sel}$ is specified by

$$\mathcal{J}_{sel}((l,\kappa),(l^*_{sel},\kappa^*),v^*) = \begin{cases} v^* & \text{if } l = l^*_{sel}, \\ \frac{1}{1-\frac{u}{J}}\Big(l - (l^*_{sel} - v^*)\frac{\sigma(\kappa)}{\sigma(\kappa^*)}\Big) & \text{if } l \neq l^*_{sel}, l > v^*, \\ \frac{1}{1-\frac{u}{J}} l & \text{if } l < v^*. \end{cases}$$

As in Section 3.2.3.2, we transform A^N_{sel} into a more convenient form. We use the Taylor approximation (3.2.31) to approximate the generator by

$$A^N_{sel} = A^N_{sel,1} + A^N_{sel,2} + \mathcal{O}\Big(\frac{K}{J}A^N_{sel,2}\Big),$$

where

$$\begin{aligned} &A^N_{sel,1}f(\eta) \\ &= \frac{sN}{S}\Bigg(\int_0^\infty \frac{uK}{J}e^{-\frac{uK}{J}v^*}\Big[g(\kappa^*,v^*)\prod_{(\kappa,l)\in\eta, l\neq l^*_{sel}} g(\kappa,\mathcal{J}_{sel}(l,l^*_{sel},v^*)) - f(\eta)\Big]\mathrm{d}v^*\Bigg), \end{aligned} \tag{3.2.51}$$

$$\begin{aligned} A^N_{sel,2}f(\eta) = \frac{N}{S}\Bigg(&- \int_0^\infty \frac{u^2K^2}{J^2}e^{-\frac{uK}{J}v^*}\int_{v^*}^\infty (1-g(\kappa^*,v))\mathrm{d}v\; g(\kappa^*,v^*) \\ &\times \prod_{(\kappa,l)\in\eta, l\neq l^*_{sel}} g(\kappa,\mathcal{J}_{sel}(l,l^*_{sel},v^*))\mathrm{d}v^*\Bigg). \end{aligned} \tag{3.2.52}$$

Once again, we treat $A^N_{sel,1}$ and $A^N_{sel,2}$ separately. We order the levels $\eta = \{(\kappa_i,l_i)\}_{i\geq 1}$ present in the system by requiring that for each i, $l_i < l_{i+1}$. Arguing in the same way as for the neutral case, we may conclude that the term involving $A^N_{sel,2}$ tends to 0 as $N\to\infty$, as the requirements $NK/J^2 \to C$ combined with $1/S \to 0$ imply that $NK/J^2S \to 0$. We turn our attention to $A^N_{sel,1}$. Recalling the difference between l^*_{neu} and l^*_{sel}, we note that for large N we will very rarely see $l^*_{sel} \neq l^*_{neu}$ and so we consider

$$\begin{aligned} &A^N_{sel,3}f(\eta) \\ &\quad = \frac{sN}{S}\Bigg(\int_0^\infty \frac{uK}{J}e^{-\frac{uK}{J}v^*}\Big[g(\kappa^*,v^*)\prod_{(\kappa,l)\in\eta, l\neq l^*_{neu}} g(\kappa,\mathcal{J}_{sel}(l,l^*_{neu},v^*)) - f(\eta)\Big]\mathrm{d}v^*\Bigg). \end{aligned} \tag{3.2.53}$$

We will prove Proposition 3.2.19 by showing that $A^N_{sel,3}$ satisfies a suitable estimate and then showing that $A^N_{sel,1} - A^N_{sel,3}$ converges to 0 by virtue of Lemma 3.2.31.

Lemma 3.2.30. *Under the conditions of Theorem 3.2.13,*

$$\mathbb{E}\Bigg[\Bigg|\frac{sN}{S}\Bigg(\int_0^\infty \frac{uK}{J}e^{-\frac{uK}{J}v^*}\Bigg[g(\kappa^*,v^*)\prod_{(\kappa,l)\in\eta_t,l\neq l^*_{neu}} g(\kappa,\beth_{sel}(l,l^*_{neu},v^*)) - f(\eta)\Bigg]\mathrm{d}v^*\Bigg) \\ - b\Bigg(f(\eta^r_t)\sum_j \frac{g'(\kappa_i,l_i)}{g(\kappa_i,l_i)}\frac{suN}{JS}l_i\Bigg)\Bigg|\Bigg] \to 0,$$

Proof Just as in Proposition 3.2.24, we approximate (3.2.53) using (3.2.30) by

$$\begin{aligned}
&\frac{sN}{S}f(\eta)\Bigg(\sum_i\int_{l_{i-1}}^{l_i}\frac{uK}{J}e^{-\frac{uK}{J}v^*}\Bigg[\frac{g'(\kappa_i,l_i)}{g(\kappa_i,l_i)}(v^*-l_i) \\
&\qquad + \sum_{l\neq l_i}\frac{g'(\kappa,l)}{g(\kappa,l)}\Bigg(\frac{l\frac{u}{J}}{1-\frac{u}{J}} - \mathbf{1}_{l>l_i}\frac{\sigma(\kappa)}{\sigma(\kappa^*)}\frac{l_i-v^*}{1-\frac{u}{J}}\Bigg)\Bigg]\mathrm{d}v^*\Bigg) \\
&= Nf(\eta)\sum_i\Bigg[\frac{g'(\kappa_i,l_i)}{g(\kappa_i,l_i)}\Big(-(l_i-l_{i-1})e^{-\frac{uK}{J}l_{i-1}} + \frac{J}{uK}\Big(e^{-\frac{uK}{J}l_{i-1}} - e^{-\frac{uK}{J}l_i}\Big)\Big) \\
&\qquad + \sum_{j\neq i}\Bigg(\frac{g'(\kappa_j,l_j)}{g(\kappa_j,l_j)}\frac{l_j\frac{u}{J}}{1-\frac{u}{J}}\Big(e^{-\frac{uK}{J}l_{i-1}} - e^{-\frac{uK}{J}l_i}\Big)\Bigg) \\
&+ \sum_{j\neq i}\sum_{j<i}\frac{\sigma(\kappa_j)}{\sigma(\kappa_i)}\frac{g'(\kappa_j,l_j)}{g(\kappa_j,l_j)}\frac{1}{1-\frac{u}{J}}\Big(-(l_i-l_{i-1})e^{-\frac{uK}{J}l_{i-1}} + \frac{J}{uK}\Big(e^{-\frac{uK}{J}l_{i-1}} - e^{-\frac{uK}{J}l_i}\Big)\Big)\Bigg] \\
&=: \frac{\mathcal{S}_1}{S} + \frac{sN}{S}f(\eta)\sum_i\sum_{j<i}\Big(\frac{\sigma(\kappa_j)}{\sigma(\kappa_i)}-1\Big)\frac{g'(\kappa_j,l_j)}{g(\kappa_j,l_j)}\frac{1}{1-\frac{u}{J}}\Big(-(l_i-l_{i-1})e^{-\frac{uK}{J}l_{i-1}} \\
&\qquad + \frac{J}{uK}\Big(e^{-\frac{uK}{J}l_{i-1}} - e^{-\frac{uK}{J}l_i}\Big)\Big),
\end{aligned}$$

where we have used integration by parts and $\mathcal{S}_1$ is equal to the terms appearing in (3.2.34). Now we notice that $\mathcal{S}_1$ is multiplied by $1/S$ and since $S \to \infty$ it can be neglected. Consider then the second term. We can neglect the factor $1 - u/J$ at the cost of an error of order NK/J^3, which tends to zero. We then change the order of summation and approximate the second term by

$$\frac{sN}{S}f(\eta)\Big(\frac{\sigma(\kappa_r)}{\sigma(\kappa_c)}-1\Big)\sum_j\frac{g'(\kappa_j,l_j)}{g(\kappa_j,l_j)}\sum_{i<j}\Bigg(-(l_i-l_{i-1})e^{-\frac{uK}{J}l_{i-1}} + \frac{J}{uK}\Big(e^{-\frac{uK}{J}l_{i-1}} - e^{-\frac{uK}{J}l_i}\Big)\Bigg).$$

Now we proceed precisely as in the proof of Proposition 3.2.24. The final approximation of $A^N_{sel,3}$ then takes the form

$$f(\eta)\Big(\frac{\sigma(\kappa_r)}{\sigma(\kappa_c)}-1\Big)\sum_j \frac{g'(\kappa_j,l_j)}{g(\kappa_j,l_j)}\Big(\frac{suN}{JS}l_j+\mathcal{O}\Big(\frac{1}{S}\Big)\Big).$$

□

We now turn our attention to $A^N_{sel,1}-A^N_{sel,3}$. We must treat the cases when the rare type is favoured or unfavoured separately. However, we show in both cases that $\mathbb{E}[\Big|A^N_{sel,1}-A^N_{sel,3}\Big|]\to 0$ and conclude with Jensen's inequality.

Lemma 3.2.31. *Let l^*_{neu} be the lowest level above v^*. Let l^*_{sel} be the level above v^* which minimises $\frac{l-v^*}{\sigma}$. Then*

$$\mathbb{E}\left[\left|\frac{sN}{S}f(\eta)\int_0^\infty \frac{uK}{J}e^{-\frac{uK}{J}v^*}g(v^*,\kappa^*)\sum_l \frac{g'(l)}{g(l)}(\mathcal{J}_{sel}(l,l^*_{sel},v^*)-\mathcal{J}_{sel}(l,l^*_{neu},v^*))\mathrm{d}v^*\right|\right]\to 0.$$

Proof We note that for $l<v^*$, $\mathcal{J}_{sel}(l,l^*_{sel},v^*)=\mathcal{J}_{sel}(l,l^*_{neu},v^*)$ as

$$\mathcal{J}_{sel}(l,l^*,v^*)=\begin{cases}\frac{1}{1-\frac{u}{J}}\Big(l-(l^*-v^*)\frac{\sigma(\kappa)}{\sigma(\kappa^*)}\Big) & l>v^*,\ l\neq l^*,\\ v^* & l=l^*,\\ \frac{1}{1-\frac{u}{J}}l & l<v^*,\end{cases}$$

and so we need only consider

$$\frac{sN}{S}f(\eta)\int_0^\infty \frac{uK}{J}e^{-\frac{uK}{J}v^*}g(v^*,\kappa^*)\sum_{l>v^*}\frac{g'(l)}{g(l)}(l^*_{neu}-l^*_{sel})\frac{\sigma(\kappa)}{\sigma(\kappa^*)}\mathrm{d}v^*. \tag{3.2.54}$$

We proceed as in Section 3.2.3.2 and consider v^* in the interval $[l_{i-1},l_i]$. We note that $l^*_{neu}=l_i$ and if l_i is of the favoured type then $l^*_{sel}=l_i$ also. If l_i is unfavoured we denote the lowest favoured type above l_i by l^s_i then $l^*_{sel}\neq l_i$ if and only if

$$l_{i-1}\leq v^*\leq\Big(l_i-\frac{\sigma_w}{\sigma_s-\sigma_w}(l^s_i-l_i)\Big)\wedge l_{i-1},$$

where $\sigma_w = \sigma(\kappa_w)$, $\sigma_s = \sigma(\kappa_s)$, κ_w is the unfavoured type and κ_s is the favoured type. Therefore we see that (3.2.54) can be written as

$$\frac{sN}{S} f(\eta) \sum_{(l_i,\kappa_i)\in\eta,\kappa_i=\kappa_w} \int_{l_{i-1}}^{(l_i - \frac{\sigma_w}{\sigma_s-\sigma_w}(l_i^s - l_i))\wedge l_{i-1}} \frac{uK}{J} e^{-\frac{uK}{J}v^*} g(v^*,\kappa^*) \sum_{l>v^*} \frac{g'(l)}{g(l)} (l_n^* - l_s^*) \frac{\sigma(\kappa)}{\sigma(\kappa^*)} \mathrm{d}v^*.$$

We can then integrate and use a Taylor expansion to see that we need only show

$$\mathbb{E}\left[\frac{sN}{S} f(\eta) \sum_j \frac{g'(l_j)}{g(l_j)} \sum_{(l_i,\kappa_i)\in\eta, l_i<l_j, \kappa_i=\kappa_w} \frac{uK}{J} (l_i^s - l_i) \Big(l_i - l_{i-1} - \frac{\sigma_w}{\sigma_s - \sigma_w}(l_i^s - l_i) \Big) \wedge 0 \right],$$

tends to 0 as $N \to \infty$. If the rare type is unfavoured we can then conclude by noting that $\mathbb{E}\Big[\frac{suNK}{SJ}(l_i^s - l_i)(l_i - l_{i-1})\Big] = \mathcal{O}\Big(\frac{NK}{SJ}\frac{1}{(K-Z)^2}\Big)$.

If the rare type is favoured we may bound it by

$$\sum_{\widehat{l}_{k-1}<l_i<\widehat{l}_k} \frac{suNK}{JS} (l_i - l_{i-1})^2 \mathbf{1}_{\{l_i - l_{i-1} - \frac{\sigma_w}{\sigma_s-\sigma_w}(\widehat{l}_k - l_i) > 0\}}, \tag{3.2.55}$$

where we recall that $\widehat{l}_k$ is the kth highest rare level. As before, we condition on there being n individuals of the common type with levels between $\widehat{l}_k$ and $\widehat{l}_{k-1}$. We can then consider the n levels as independent uniformly distributed rather than the ordered levels l_i. Suppose that $z, z_1, \ldots, z_{n-1}$ are independent uniformly distributed random variables on $[0,1]$, $Y = 1 - z$ and $X = \min(\{z\}, \{(z - z_i)\}_{z_i<z})$. We then see that the expectation of(3.2.55) is bounded by

$$n\frac{suKN}{JS}\mathbb{E}\Big[X^2 \mathbf{1}_{X>cY}\Big].$$

We now look at

$$\begin{aligned}
\mathbb{E}[X^2\mathbf{1}_{X>cY}] &= \int_0^1 \mathbb{E}[X^2\mathbf{1}_{X>c(1-u)}|z=u]\mathrm{d}u \\
&= \int_{\frac{c}{1+c}}^1 \mathbb{E}[X^2\mathbf{1}_{X>c(1-u)}|z=u]\mathrm{d}u \\
&= \int_{\frac{c}{1+c}}^1 \int_0^u x^2 \mathbf{1}_{x>c(1-u)} \mathrm{d}\mathbb{P}[X=x|z=u]\mathrm{d}u \\
&= \int_{\frac{c}{1+c}}^1 \int_{c(1-u)}^u x^2 \mathrm{d}\mathbb{P}[X=x|z=u]\mathrm{d}u.
\end{aligned} \tag{3.2.56}$$

Since

$$\begin{aligned}\mathbb{P}[X > x | z = u] =& \mathbb{P}\big[[u-x, u] \cap \{z_i\} = \emptyset\big] = (1-x)^{n-1},\\ \mathrm{d}\mathbb{P}[X = x | z = u] =& (n-1)(1-x)^{n-2},\end{aligned}$$

we may bound (3.2.56) by

$$\begin{aligned}\mathbb{E}[X^2 \mathbf{1}_{X>cY}] \leq& \int_{\frac{c}{1+c}}^1 \int_{c(1-u)}^1 x^2 (n-1)(1-x)^{n-2} \mathrm{d}x \, \mathrm{d}u\\ =& \int_{\frac{c}{1+c}}^1 \int_0^{1-c+cu} (n-1)(1-v)^2 v^{n-2} \mathrm{d}v \, \mathrm{d}u\\ \leq& \frac{n-1}{c}\left[\frac{v^n}{n(n-1)} - 2\frac{v^{n+1}}{n(n+1)} + \frac{v^{n+2}}{(n+1)(n+2)}\right]_0^1\\ =& \frac{6}{cn(n+1)(n+2)}.\end{aligned}$$

We can then conclude by noting that

$$\sum_{n=1}^{\infty} \frac{1}{(n+1)(n+2)} \frac{(K-Z^N)^n e^{-(K-Z^N)}}{n!} = \mathcal{O}\Big(\frac{1}{(K-Z^N)^2} + \exp(-(K-Z^N))\Big),$$

which again leads to an error of order $\mathcal{O}\Big(\frac{NK}{SJ}\frac{1}{(K-Z)^2}\Big)$, completing the proof of the Lemma. □

Proof[Proof of Proposition 3.2.19] By an application of Lemma 3.2.26, Proposition 3.2.19 follows from a combination of Lemma 3.2.30 and Lemma 3.2.31. □

Lemma 3.2.32. *If we define β^N as in* (3.2.49) *but with the addition of selection. Then β^N is a dominated by a bounded submartingale.*

Remark 3.2.33. *Although the statements of Lemma 3.2.29 and Lemma 3.2.32 are the same, the behaviour of the process β^N is subtly different, as in Lemma 3.2.32 the movement of the levels is affected by selective events. This difference does not effect the proof however we include them as seperate lemmas for completeness.*

Proof As in the proof of Lemma 3.2.29 (and using the notation introduced there)

$$\frac{\mathbb{E}\Big[\sum_{l(t)\in\eta_t^r}\Big(l(t) - \widehat{l}(t)\Big)h(l(t)) - \sum_{l(0)\in\eta_0^r}\Big(l(0) - \widehat{l}(0)\Big)h(l(0))\Big]}{t}$$

$$= N\int_0^\infty \Big(1+\frac{s}{S}\Big)\frac{uK}{J}e^{-\frac{uK}{J}v^*}\Bigg\{\sum_{l(0)\in\eta_t^r}\Bigg\{\Big[\Big(\mathcal{J}(l(0))-\mathcal{J}(\widehat{l}(0))\Big)-\Big(l(0)-\widehat{l}(0)\Big)\Big]h(l(0))$$
$$+\Big(\mathcal{J}(\widehat{l}(0))-\widehat{l}(t)\Big)h(l(0))$$
$$+\Big(l(t)-\widehat{l}(0)\Big)(\mathcal{J}(l(0))-l(0))\mathcal{O}(||h'||_\infty)\Bigg\}+\sum_{i=2}^{\infty}(v_i-\widehat{v}_i)h(v_i)\mathbf{1}_{\{v_i \text{ is rare}\}}\Bigg\}\mathrm{d}v^*+\mathcal{O}(t).$$

Considering combination of Proposition 3.2.24 and Proposition 3.2.19 we know that the parts involving $\mathcal{J}(l(0))-l(0)$ converge to $al(0)-bl^2(0)$. Since $\frac{s}{S}\to 0$ and observing Lemma 3.2.31, we may proceed as in the proof of Lemma 3.2.29 to conclude. □

3.2.3.4 Model with selection in fluctuating environment: proof of Theorem 3.2.10

Proof We begin by identifying the limit and recall that the rescaled generator takes the form

$$A^N f(\eta,\zeta)=A^N_{neu}f(\eta,\zeta)+\widehat{S}A^N_{sel}f(\eta,\zeta)+\widehat{S}^2A_{env}f(\eta,\zeta), \tag{3.2.57}$$

where A^N_{neu} is defined as in (3.2.29), A^N_{sel} is defined as in (3.2.50), and

$$A_{env}f(\eta,\xi)=\mathbb{E}_\pi[f(\eta,\xi)]-f(\eta,\xi).$$

We also use (3.2.16) and that $A_{neu}f_1$ is of order 1 by previous calculations. To identify the correct limit it is therefore enough to evaluate $A^N_{sel}(f_N(\eta,\xi))$, which can be approximated as

$$A^N_{sel}(f_N(\eta,\xi))=A^N_{sel}\Bigg(f(\eta)\frac{suN}{JS}\Big(\frac{\sigma(\kappa_r,\zeta)}{\sigma(\kappa_c,\zeta)}-1\Big)\sum_j l_j\frac{g'(\kappa_j,l_j)}{g(\kappa_j,l_j)}\Bigg)$$
$$=\Big[\frac{suN}{JS}\Big(\frac{\sigma(\kappa_r,\zeta)}{\sigma(\kappa_c,\zeta)}-1\Big)\Big]^2 f(\eta)$$
$$\times\Bigg\{\Bigg(\sum_{i\neq j}l_il_j\frac{g'(\kappa_i,l_i,)g'(\kappa_j,l_j)}{g(\kappa_i,l_i)g(\kappa_j,l_j)}\Bigg)+\Bigg(\sum_j l_j\frac{g'(\kappa_j,l_j)+l_jg''(\kappa_j,l_j)}{g(\kappa_j,l_j)}\Bigg)\Bigg\}+\mathcal{O}\Big(\frac{1}{S}+\frac{1}{K^{1/2}}\Big),$$

The bound (3.2.20) follows directly from proofs of Propositions 3.2.19, 3.2.24, 3.2.25, with Lemma 3.2.29 modified in a way analogous to Lemma 3.2.32. We can then conclude by noting that our class of test functions are point separating with dense span. □

3.3 SuperBrownian motion in a random environment

In this section we present a precise definition of superBrownian motion in a random environment. We begin by defining Branching Brownian Motion in a random environment and recalling the original definition of the corresponding superprocess from Mytnik (1996). Then we describe a lookdown construction for both of these models based on the ideas in Kurtz and Rodrigues (2011).

3.3.1 Definitions

Branching Brownian Motion in a random environment (BBMRE) can be described as follows. A collection of particles on $\mathbb{R}^d$ where each moves according to independent standard Brownian motions. Each particle, if alive, gives birth to one new particle at a time, at rate a. The initial offspring location is the same as that of the parent. After the birth, the offspring moves and reproduces independently of all other particles. The particles die at instantaneous rate $a - \zeta_t(x)b$, where $\zeta_t(x)$, taking values in $\{-1, 1\}$ as before, models the random environment and x is the current location of the particle. We assume that $a > b > 0$. If $\zeta_t(x)$ is positive, the particle is less likely to die, if it is negative, it is more likely to do so. The evolution of the environment and particles are independent. We give a more formal definition below.

We begin by recalling the description of our environment, which is used for all models in this section. Our environment is modelled through a simple random field.

Definition 3.3.34. *Let Π^{env} be a Poisson process with intensity E, dictating the times of the changes in the environment. Let $q(x, y) \in C_0\Big(\mathbb{R}^d \times \mathbb{R}^d\Big)$ (continuous functions vanishing at infinity) and let $\{\xi^{(m)}(\cdot)\}_{m\geq 0}$ be a family of identically distributed random fields on $\mathbb{R}^d$ such that*

$$\mathbb{P}\Big[\xi^{(m)}(x) = -1\Big] = \frac{1}{2} = \mathbb{P}\Big[\xi^{(m)}(x) = +1\Big],$$

$$\mathbb{E}\Big[\xi^{(m)}(x)\xi^{(m)}(y)\Big] = q(x,y).$$

Set $\tau_0 = 0$ *and write* $\{\tau_m\}_{m\geq 1}$ *for the points in* Π^{env} *and define*

$$\zeta(t,\cdot) := \sum_{m=0}^{\infty} \xi^{(m)}(\cdot)\mathbf{1}_{[\tau_m,\tau_{m+1})}(t).$$

Since the exact labelling of our particles is not important, we identify the particle component of the process with a counting measure, that is for a vector $\overline{x} = (x_1, \dots x_n)$

$$\mu_{\overline{x}} = \sum_i \delta_{x_i} \quad x_i \in \mathbb{R}^d.$$

We are now ready to state the definition of Branching Brownian motion in a random environment.

Definition 3.3.35 (Branching Brownian motion in a random environment (BBMRE))**.** *Branching Brownian motion in the random environment* ζ *is the stochastic process taking values in* $\{-1,1\} \times \mathbb{R}^d$ *times purely atomic measures on* $\mathbb{R}^d$ *whose evolution consists of four ingredients.*

1. ***Spatial motion*** *The location of each particle,* x_i*, evolves according to a standard Brownian motion, independently of all other particles.*
2. ***Birth events*** *At exponential rate a (independent for each particle), a particle gives birth to a new particle (a new particle is added to the system). The location of the offspring is the same as the location of their parent. The behaviour of the new particle after the birth is independent of all other particles.*
3. ***Death events*** *Each particle dies (is removed from the system) at instantaneous rate* $a - \zeta_t(x_i)b$*, where* x_i *is its location.*
4. ***Environment changes*** *The environment evolves as described in Definition 3.3.34.*

Alternatively, we may define the BBMRE by the means of the generator. For a counting measure $\mu = \sum_i \delta_{x_i}$, define

$$f(\mu) = \sum_i f(x_i).$$

Let $f(\zeta,\mu) = f_0(\zeta)f_1(\mu) = f_0(\zeta)\Pi_i h(x_i)$ be a function such that $h \in C_c(\mathbb{R}^d)$, that is h is a continuous function with a compact support. We define the generator of the BBMRE as

$$\begin{aligned}\mathcal{L}f(\zeta,\mu)&\\ &= f_1(\mu)A^{env}f_0(\zeta) + f_0(\zeta)\Bigg(\sum_{i=1}^{n} Bf_1(\mu) + \sum_i a(f_1(\mu_{b(x|x_i)}) - f_1(\mu))\\ &\qquad + \sum_i (a - \zeta(x_i)b(f_1(\mu_{d(x|x_i)}) - f_1(\mu))\Bigg),\end{aligned}$$

where $\mu_{b(x|x_i)}$ denotes the addition of a particle at location x_i, and $\mu_{d(x|x_i)}$ denotes a removal of the particle at location x_i.

It is well known that the high density limit for the Branching Brownian Motion gives rise to a SuperBrownian motion, for example see Etheridge (2000). An analogous result for branching random walk in a random environment, under certain scaling of the environment, was established in Mytnik (1996). The limiting object is SuperBrownian motion in a random environment (SBMRE).

Definition 3.3.36 (SuperBrownian motion in a random environment (SBMRE))**.** *Let $q(x,y)$ be a covariance function which belongs to $C_0\Big(\mathbb{R}^d \times \mathbb{R}^d\Big)$. The superBrownian motion in a random environment is the (unique) process for which, for all $\phi \in \mathcal{D}(\Delta)$,*

$$X_t(\phi) = X_t(\phi) - X_0(\phi) - \int_0^t \frac{1}{2} X_s(\Delta\phi)\mathrm{d}s \tag{3.3.1}$$

is a square-integrable martingale with quadratic variation given by

$$\langle X(\phi)\rangle_t = \int_0^t X_s(\phi^2)\mathrm{d}s + \int_0^t \int_{\mathbb{R}^d\times\mathbb{R}^d} q(x,y)\phi(x)\phi(y)X_s(\mathrm{d}x)X_s(\mathrm{d}y)\mathrm{d}s.$$

An equivalent characterisation of the SBMRE can be given in terms of the generator, see Theorem 4.8 in Mytnik (1996). Namely, for $f \in \overline{C}^2(R_+), \phi \in \mathcal{D}(\Delta)$ and $q(x,y)$ as in Definition 3.3.36, the generator is given by

$$\mathcal{L}f(\mu(\phi))$$

$$= f'(\mu(\phi))\mu(\Delta\phi) + \frac{1}{2}f''(\mu(\phi))\Big(\mu(\phi^2) + \int_{\mathbb{R}^d\times\mathbb{R}^d} q(x,y)\phi(x)\phi(y)\mu(\mathrm{d}x)\mu(\mathrm{d}y)\mathrm{d}s\Big).$$

We would like to point out that the process of Definition 3.1.3 which is obtained as a limiting behaviour of the scaled SLFV in a random environment differs from the one from Definition 3.3.36 only by a presence of a drift term.

Remark 3.3.37. *Uniqueness of solutions to the martingale problem of Definition 3.3.36 is not immediately clear. It was established by Mytnik (1996) using a novel approximate duality technique, and later re-proved by the means of the log-Laplace transform by Crisan (2004). A uniqueness result for the process of Definition 3.1.3 is a simple consequence of Dawson's Girsanov Theorem, see Dawson (1978) and Etheridge (2000), Chapter 7.*

Remark 3.3.38. *A model similar to that in Mytnik (1996) was studied in Sturm (2003). The main difference between the two is in the behaviour of the environment. For the limiting model in Sturm (2003),* (3.3.1) *is again a martingale, but with quadratic variation of the form*

$$\langle X(\phi)\rangle_t = \int_0^t \int_{\mathbb{R}^d\times\mathbb{R}^d} q(x,y)\phi(x)\phi(y)X_s(\mathrm{d}x)X_s(\mathrm{d}y)\mathrm{d}s.$$

In this case, the density of the process can be described as a solution to an SPDE in all dimensions, whereas the SBMRE has a density only in dimension one and the analogous SPDE has no solution in dimensions $d \geq 2$.

Remark 3.3.39. *An alternative construction of the SBMRE has been suggested by Nakashima (2015) based on the model introduced by Birkner et al. (2005b).*

3.3.2 Lookdown representation for BBMRE and SBMRE

In this section we describe a new construction of the SBMRE, inspired by the constructions of Kurtz and Rodrigues (2011). As a by-product of this construction, we provide a lookdown construction for the BBMRE. A precise statement of the results is given in Theorem 3.3.41 and Theorem 3.3.42. We will use the lookdown representation of SBMRE in Section 3.4 to describe the behaviour of the 'rare' type (by which we mean

a new mutation establishing in the population) in the Spatial Lambda-Fleming-Viot model with fluctuating selection.

In order to motivate what follows, let us informally describe the construction of SBMRE from Mytnik (1996). The SBMRE is constructed via a series of approximations. Consider a sequence of mean 0 random fields ζ_k, taking values in $\{-1, 1\}$, with correlation as in Definition 3.3.34, althoughin Mytnik (1996) they consider a more general class of random fields. At stage n of the approximation, we start with a population with size of order n. Over each time interval $(k/n, (k+1)/n), k \in \mathbb{N}$, each individual, if alive, moves independently of all the others, according to a standard Brownian motion. At times k/n, each individual either splits into two with probability $1/2 + \zeta_k(x_i)/\sqrt{n}$, or dies with probability $1/2 - \zeta_k(x_i)/\sqrt{n}$. The state of the environment is resampled with each reproduction event. If for each set B we define

$$X_t^n(B) = \frac{\text{number of particles in } B \text{ alive at time } t}{n},$$

and assume that X_0^n converges to some X_0, then passing to the limit as n tends to infinity the process X^n converges to a SBMRE with initial condition X_0. Intuitively, this procedure corresponds to increasing the rate of the branching events in BBMRE by n, while scaling down the impact of the environment by $\sqrt{n}$. Observe that Mytnik (1996) considers a model with non-overlapping generations, for purely technical reasons.

We now move to a description of a lookdown construction for BBMRE (with overlapping generations). Fix $\lambda \in \mathbb{R}$. Consider a system of particles in the geographical space $\mathbb{R}^d$. Each particle moves independently, according to a standard Brownian motion. Each particle i is assigned a level, l_i. The level takes values in $[0, \lambda]$.

In order to take environmental fluctuations into account we use the process $\zeta_t(x) \in \{-1, 1\}$, introduced in Definition 3.3.34, to model the environment. The branching rate of the particles depends on both their level and the state of the environment. We assume that every particle gives birth at instantaneous rate $2a(\lambda - l_i(t))$, with the location of offspring being the same as the location of the parent. The initial level of the offspring is distributed uniformly on the interval $[l_i, \lambda]$. The levels of the particles evolve according

to an ODE with a random coefficient

$$\frac{\mathrm{d}l_i}{\mathrm{d}t} = al_i^2 - \zeta(x_i)\sqrt{\lambda}bl_i.$$

The particle dies when its level reaches λ. We are now ready to define the process of interest.

Definition 3.3.40 (Lookdown representation of Branching Brownian motion in a random environment)**.** *The lookdown representation of BBMRE is the process taking values in* $(\mathbb{R}^d \times \mathbb{R})^\infty \times \{-1, 1\}$ *with dynamics specified by four components.*

1. ***Spatial motion*** *The spatial location of each particle,* $x_i \in \mathbb{R}^d$*, evolves according to a standard Brownian motion with generator* $B = \frac{1}{2}\Delta$*.*

2. ***Birth events*** *Each particle gives birth at instantaneous rate* $2a(\lambda - l_i(t))$*, where* l_i *is the level of the particle. The spatial location of the offspring is the same as the location of their parent,* x_i*. The level of the offspring is chosen uniformly at random from* $[l_i, \lambda]$*.*

3. ***Level movement*** *The level of each of the particles evolves according to the equation*

$$\frac{\mathrm{d}l_i}{\mathrm{d}t} = al_i^2 - \zeta(x_i)\sqrt{\lambda}bl_i.$$

4. ***Environment changes*** *The environment evolves as described in Definition 3.3.34.*

A formal definition, via the generator of the process, is given in (3.3.4). It is convenient to identify the process with the counting measure

$$\mu_{\overline{x},\overline{l}} = \sum_i \delta_{x_i,l_i}.$$

We now state the first result of this section.

Theorem 3.3.41 (Lookdown construction for Branching Brownian Motion in random environment)**.** *Let* μ_0 *be the measure associated with the initial state of the population.*

Let $\gamma : \mathcal{N}(\mathbb{R}^d \times [0,\lambda)) \to \mathcal{M}(\mathbb{R}^d)$ *be given by*

$$\gamma\left(\sum_i \delta_{x_i,l_i}\right) = \frac{1}{\lambda}\sum_i \delta_{x_i}.$$

The lookdown process of Definition 3.3.40 corresponds to BBMRE, in the sense that if $\eta(\overline{x},\overline{l})$ *is a solution to the martingale problem for the process defined in Definition 3.3.40, then* $\gamma\Big(\eta(\overline{x},\overline{l})\Big)$ *is the solution to the martingale problem for the process of Definition 3.3.35.*

Our next aim is to write down the generator of the process obtained by passing to the limit as $\lambda \to \infty$, which would correspond to passing to the limit with n passing to infinity in the sequence of approximations described at the beginning of this section. In order to formulate the result for the limiting process, we need to consider a special class of test functions of the form

$$f(\zeta,x,l) = f_0(\zeta)f_1(x,l) = f_0(\zeta)\prod_i g(x_i,l_i), \tag{3.3.2}$$

with the additional requirement there exists $\lambda_g \in \mathbb{R}$ such that

$$g(x_i,l_i) = 1 \text{ for } l_i > \lambda_g.$$

This ensures that we ignore all individuals with levels above λ_g. We are now in a position to state the main result of this section.

Theorem 3.3.42 (Lookdown construction for superBrownian motion in random environment)**.** *Consider the process with generator given by*

$$\begin{aligned} A_\infty(f) = {} & f(x,l)\sum_i \frac{Bg(x_i,l_i)}{g(x_i,l_i)} \\ & + f(x,l)\sum_i 2a\int_{l_i}^{\lambda_g}(g(x_i,v)-1)\mathrm{d}v + f(x,l)\sum_i (al_i^2 - b^2 l_i)\frac{\partial_l g(x_i,l_i)}{g(x_i,l_i)} \\ & - f(x,l)\sum_i b^2 l_i\left(\sum_{j\neq i} q(x_i,x_j)l_j\frac{\partial_l g(x_i,l_i)\partial_l g(x_j,l_j)}{g(x_i,l_j)g(x_i,l_i)} + \frac{l_i\partial_l^2 g(x_i,l_i)}{g(x_i,l_i)}\right). \end{aligned} \tag{3.3.3}$$

Let μ_0 be the measure associated with the initial state of the population in the limit. Define the map $\gamma : \mathcal{N}(\mathbb{R}^d \times [0,\infty)) \to \mathcal{M}(\mathbb{R}^d)$ by

$$\gamma\left(\sum_i \delta_{x_i,l_i}\right) = \begin{cases} \lim_{\lambda\to\infty} \frac{1}{\lambda} \sum_{l_i \le \lambda} \delta_{x_i} & \text{if the limit exists,} \\ \mu_0 & \text{otherwise .} \end{cases}$$

The process described by the limiting generator (3.3.3) *corresponds to the SBMRE of Definition 3.1.3, in the sense that if $\eta(\overline{x}, \overline{l})$ is a solution to the martingale problem for the process described by the limiting generator* (3.3.3)*, then $\gamma(\eta(\overline{x}, \overline{l}))$ is a solution to the martingale problem for the process in Definition 3.1.3.*

We now wish to specify the generator for the process in Definition 3.3.40. In general, we consider the set of test functions of the form (3.3.2). To make the functional setup more precise, we need to borrow the following condition from Kurtz and Rodrigues (2011), which guarantees that the assumptions of the Markov Mapping Theorem are satisfied.

Condition 3.3.43 (Based on Kurtz and Rodrigues (2011), Condition 3.1)**.** *We assume that the following conditions on the operator B, the coefficients a, b and the test functions g are satisfied.*

1. *The operator B is defined on a subset of the space of bounded continuous functions, and its domain is closed under multiplication and separating.*

2. *The test functions are of the form* (3.3.2)*, where*

$$g(x,l) = \prod_{j=1}^{m} (1 - g_1^j(x) g_2^j(l))$$

and $g_1^j \in \mathcal{D}(B)$ and g_2^j are twice differentiable with support in $[0,\lambda]$. Moreover, there exists ρ_g such that

$$0 \le g_1^j g_2^j < \rho_g < 1.$$

3. *There exists a continuous, non-negative function ψ_B such that for every test function g, and for every $x \in \mathbb{R}^d$*

$$\sup_l |Bg(x,l)| \leq c_g \psi_B(x)$$

for some constant c_g which depends only on the function g.

4. *The following bound holds for the test functions:*

$$\int_0^\infty |g(x,l) - 1| \mathrm{d}l + \sup_l \{l + l^2\} \partial_l g(x,l) \leq c'_g \psi_B(x)$$

for some constant c'_g which depends only on the function g.

5. $a > 0$, $\lambda a - \sqrt{\lambda} b > 0$.

Remark 3.3.44. *We notice that the conditions for the generator of the motion B are satisfied by the Laplacian, the generator of the Brownian motion. This is the only generator that we consider. We provide the general construction for the sake of completeness and to highlight potential extensions of this work.*

The generator A_λ of the process of Definition 3.3.40 can be written as

$$\begin{aligned} A_\lambda f(\zeta,x,l) =& f_1(x,l) A_\lambda^{env} f_0(\zeta) + f(\zeta,x,l) \sum_i \frac{Bg(x_i,l_i)}{g(x_i,l_i)} \\ &+ f(\zeta,x,l) \sum_i 2a \int_{l_i}^{\lambda} (g(x_i,v) - 1) \mathrm{d}v \\ &+ f(\zeta,x,l) \sum_i (a l_i^2 - \sqrt{\lambda} \zeta(x_i) b l_i) \frac{\partial_{l_i} g(x_i,l_i)}{g(x_i,l_i)}. \end{aligned} \tag{3.3.4}$$

A naive limiting procedure does not lead to a well defined object, since the form of the generator does not take into account the cancellations coming from fluctuations in the environment. In order to identify the correct limit, we once again use the separation of timescales trick. Proofs of Theorem 3.3.41 and Theorem 3.3.42 are presented in Section 3.5.4.

Remark 3.3.45. *Our computations do not lead to any surprising results - we show that the SBMRE can be obtained as a scaling limit of BBMRE. Our arguments, combined*

with tightness of the sequence of projected processes, guarantees convergence of the sequence of projected models. We therefore provide a new construction for the SBMRE. However, we still refer to results described in Remark 3.3.37 to guarantee the uniqueness of solutions to the projected martingale problem. The question of uniqueness of solutions to the martingale problem for the limiting process with levels is left as future work.

3.4 Scaling limits of the SLFV - dynamics of the rare type

In this section we are interested in a spatial analogue of the results from Section 3.2. We show that, under a certain scaling, the dynamics of the subpopulation with a rare mutation, which is a part of a population evolving according to a version of the Spatial-Lambda-Fleming-Viot model with selection in a fluctuating environment, is given by a superBrownian motion in a random environment. Since under our scaling the proof of the result does not differ significantly from the one discussed in Section 3.2, our discussion will be rather brief and focus on highlighting the differences and required modifications.

We begin with a description of the model. As before, we consider a population with two genetic types, rare and common, which we denote by κ_r and κ_c, respectively. We also consider the random field ζ, specified by Definition 3.3.34. Since in our model we consider a population with a countable number of individuals, the state of the population can be represented as

$$\eta = \sum_{(x,\kappa,l)} \delta_{x,\kappa,l}.$$

We assume that η is a conditionally Poisson system with Cox measure $\Xi(\mathrm{d}x, \mathrm{d}\kappa) \times m_{leb}(\mathrm{d}l)$.

The evolution of the population is determined by reproduction events of two types, neutral and selective, driven by independent Poisson point processes Π^{neu} and Π^{sel}, which specify the time, location, impact and radius of the events. They are analogous to those in Section 3.2, but now we assume that an event has a location and radius,

only individuals within the ball of given radius centred at the location of the event are affected. A rigorous definition of the model follows.

Definition 3.4.46 (Lookdown representation of SLFVSRE)**.** *Let μ be a measure on $(0,\infty)$ and for each $r \in (0,\infty)$, let ν_r be a probability measure on $[0,1]$, such that the mapping $r \mapsto \nu_r$ is measurable and*

$$\int_{(0,\infty)} r^d \int_{[0,1]} u\, \nu_r(\mathrm{d}u)\mu(\mathrm{d}r) < \infty. \tag{3.4.1}$$

Fix $s \in [0,1]$. Let Π^{neu}, Π^{sel} be a pair of independent Poisson processes with intensity measures $(1-s)\mathrm{d}t \otimes \mathrm{d}y \otimes \mu(\mathrm{d}r)\nu_r(\mathrm{d}u)$ and $s\mathrm{d}t \otimes \mathrm{d}y \otimes \mu(\mathrm{d}r)\nu_r(\mathrm{d}u)$ respectively. Let Π^{env} be a Poisson process independent of Π^{neu}, Π^{sel}.

The lookdown representation of SLFVSRE is a process taking values in purely atomic measures on $\mathbb{R}^d \times \mathbb{R} \times \{\kappa_r, \kappa_c\} \times \{-1, 1\}$ with dynamics described as follows.

1. *If $(t, y, r, u) \in \Pi^{neu}$*

   (a) *a group of new individuals with levels $(v_1, v_2, \dots,)$ is added to the population within the ball $B_r(y)$. Their spatial positions are chosen uniformly at random from the ball centered at x with radius r. Their levels are distributed according to a Poisson process with intensity $u|B_r|$.*

   (b) *Let $v^* = \min\{v_1, v_2, \dots, \}$. The type of each new individual is chosen to be the same as the type of the individual with the lowest level above v^* within the ball $B_r(y)$.*

   (c) *As a result of an event the individual originally with level, l, and position, x, has a new level given by*

$$\mathcal{J}_{neu}(l, l^*, v^*, x, (y, r)) = \begin{cases} l & \text{if } x \notin B_r(y), \\ \frac{1}{1-\frac{u}{J}}(l - (l^* - v^*)) & \text{if } l > l^*, x \in B_r(y), \\ \frac{1}{1-\frac{u}{J}} l & \text{if } l < l^*, x \in B_r(x), \\ v^* & \text{if } l = l^*, x \in B_r(y). \end{cases}$$

2. *If $(t, y, r, u) \in \Pi^{sel}$*

(a) a group of new individuals with levels $(v_1, v_2, \dots,)$ *is added to the population. Their levels distributed according to a Poisson process with intensity* $u|B_r|$.

(b) Let $v^* = \min\{v_1, v_2, \dots, \}$. *The type of each new individual is chosen to be the same as the type of the individual with level above* $v*$ *minimizing* $(l_i - v^*)/\sigma(\kappa_i, \zeta)$ *within the ball* $B(x, r)$.

(c) As a result of an event the individual originally with level, l, *and position,* x, *has a new level given by*

$$\mathcal{J}_{sel}(l, l^*, v^*, x, (y, r)) = \begin{cases} l & \text{if } x \notin B_r(y), \\ \frac{1}{1-\frac{u}{J}}(l - (l^* - v^*)\frac{\sigma(\kappa,\zeta)}{\sigma(\kappa^*,\zeta)}) & \text{if } l > l^*, x \in B_{r/M}(y), \\ \frac{1}{1-\frac{u}{J}} l & \text{if } l < l^*, x \in B_r(y), \\ v^* & \text{if } l = l^*, x \in B_r(y). \end{cases}$$

3. The dynamics of Π^{env} *are specified by Definition 3.3.34.*

Definition 3.4.46 is more general than we require, however, we include it to underline the possibility of extending our results. However, in the interest of keeping our notation as simple as possible, from now on we shall specialise to fix the radius and impact of reproduction events.

Assumption 3.4.47. *From now on, we fix* $r \in (0, \infty)$ *and* $u \in (0, 1)$ *and take*

$$\mu = \delta_r, \qquad \nu_r = \delta_u.$$

The integrability condition (3.4.1) is trivially satisfied for our model with fixed radius and impact.

3.4.1 Scaling and statement of main results

As in Section 3.2 we record two theorems which are a by-product of our technique. We recall that we have defined $V_r := |B_r(0)|$ and $C_d := \int_{|x|\leq 1} |x|^2 \mathrm{d}x$ in dimension d.

Theorem 3.4.48. *Suppose that* X_0^N *is absolutely continuous with respect to Lebesgue measure, that the support* $\mathrm{supp}(X_0^N) \subseteq D$, *where* D *is a compact subset of* $\mathbb{R}^d$ *(inde-*

pendent of N), and that X_0^N converges weakly to X_0. Furthermore suppose that the intensity of selective events is 0 *and as N tends to infinity,*

$$\frac{C_d u r^{d+2} N}{JM^2} \to C_1; \quad J, K, M \to \infty; \quad \frac{K}{JM^d} \to 0; \quad \frac{N^2}{M^d K J^2} \to 0; \quad \frac{u^2 V_r N K}{J^2 M^d} \to a,$$

In addition, assume that there exists an n such that, as N tends to infinity, $N(K/J)^n \to 0$. *Then the sequence* $X_N(t)$ *converges weakly to superBrownian motion without drift, initial condition* X_0, *diffusion parameter* C_1 *and quadratic variation parameter* $2a$.

Theorem 3.4.49. *Suppose that X_0^N is absolutely continuous with respect to Lebesgue measure, that the support* $\mathrm{supp}(X_0^N) \subseteq D$, *where D is a compact subset of* $\mathbb{R}^d$ *(independent of N), and that X_0^N converges weakly to X_0. Furthermore, suppose that* $\widehat{S} = 1$, $\sigma(\kappa, \zeta) = \sigma(\kappa)$ *and as N tends to infinity,*

$$\begin{aligned} &\frac{C_d u r^{d+2} N}{JM^2} \to C_1 \quad J, K, M \to \infty; \quad \frac{K}{JM^d} \to 0; \\ &\frac{N^2}{M^d K J^2} \to 0; \quad \frac{u^2 V_r N K}{J^2 M^d} \to a; \frac{suNV_r}{JS}\left(\frac{\sigma(\kappa_r)}{\sigma(\kappa_c)} - 1\right) \to b, \end{aligned}$$

In addition, assume that there exists an n such that, as N tends to infinity, $N(K/J)^n \to 0$. *Then the sequence* $X_N(t)$ *converges weakly to a superBrownian motion with initial condition* X_0, *diffusion parameter* C_1, *growth rate b and quadratic variation parameter* $2a$.

The strategy of the proof is analogous to that laid out in Section 3.2.3. We therefore focus on describing the differences. In contrast to the situation described in the Remark 3.2.20 we do not claim to show any results on the limits of sequences of lookdown representations. However, we would expect similar results to follow with an extension of the techniques used here. We focus on the result for the projected version of the model. We divide the generator of the lookdown representation into two separate parts: one part describes the spatial movement of the particles, the other describes the evolution of the levels. We show that the selective events (and therefore the fluctuations in the direction of selection) do not affect the movement of the particles in the limit. For the convergence of the neutral model we refer to results in Chapter 2. We then use the lookdown representation to deduce the form of the limiting generator and to

justify the separation of timescales trick for the selective part of the generator. The part of the proof which deals with the evolution of the levels is analogous to that of Section 3.2, and we do not repeat it here. For the reader's convenience we include the intensity estimate and discuss the splitting of the generator.

Once again we consider a stopped process, with $Z^N = \langle X^N, 1\rangle$, see Remark 3.2.21.

3.4.2 Intensity estimate

As in Section 3.2.3.1 we observe that the generator of the projected process is given by

$$\begin{aligned}
\mathcal{L}^N f(\langle\phi, X_0^N\rangle) = NM^d\Bigg[&\int_{\mathbb{R}^d}\int_{B_r^M(x)} \frac{1}{|B_r^M|} w_0^N(z) f\Bigg(K\frac{u}{J}\int_{B_r^M(x)}\phi(y)\mathrm{d}y \\
&+ K\Big(1-\frac{u}{J}\Big)\int_{B_r^M(x)}\phi(y)w_0^N(y)\mathrm{d}y\Bigg) - f\Big(K\int_{B_r^M(x)}\phi(y)w_0^N(y)\mathrm{d}y\Big)\mathrm{d}z\mathrm{d}x \\
&+ \int_{\mathbb{R}^d}\int_{B_r^M(x)}(1-w_0^N(z)) f\Bigg(K\Big(1-\frac{u}{J}\Big)\int_{B_r^M(x)}\phi(y)w_0^N(y)\mathrm{d}y\Bigg) \\
&- f\Big(K\int_{\mathbb{R}^d}\phi(y)w_0^N(y)\mathrm{d}y\Big)\mathrm{d}z\ \mathrm{d}x\Bigg] \\
+ N\frac{s\widehat{S}}{S}M^d\Bigg[&\int_{\mathbb{R}^d}\int_{B_r^M(x)}\frac{1}{|B_r^M|}\frac{\sigma(\kappa_r,\zeta)w_0^N(z)}{\sigma(\kappa_r,\zeta)w_0^N(z)+\sigma(\kappa_s,\zeta)(1-w_0^N(z))} \\
&f\Bigg(K\frac{u}{J}\int_{B_r^M(x)}\phi(y)\mathrm{d}y + K\Big(1-\frac{u}{J}\Big)\int_{B_r^M(x)}\phi(y)w_0^N(y)\mathrm{d}y\Bigg) \\
&- f\Big(K\int_{B_r^M(x)}\phi(y)w_0^N(y)\mathrm{d}y\Big)\mathrm{d}z\mathrm{d}x \\
+\int_{\mathbb{R}^d}\int_{B_r^M(x)}&\frac{\sigma(\kappa_c,\zeta)(1-w_0^N(z))}{\sigma(\kappa_r,\zeta)w_0^N(z)+\sigma(\kappa_s,\zeta)(1-w_0^N(z))} f\Bigg(K\Big(1-\frac{u}{J}\Big)\int_{B_r^M(x)}\phi(y)w_0^N(y)\mathrm{d}y\Bigg) \\
&- f\Big(K\int_{\mathbb{R}^d}\phi(y)w_0^N(y)\mathrm{d}y\Big)\mathrm{d}z\ \mathrm{d}x\Bigg]. \qquad (3.4.2)
\end{aligned}$$

The first two terms represent the part of the generator describing the effect of the neutral events and is the same as in Chapter 2. The last two terms represent the effects of selective events. Substituting $Kw_0^N = X_0^N$ this becomes

$$\mathcal{L}^N f(\langle\phi, X_0^N\rangle) = \mathcal{L}^N_{neu} f(\langle\phi, X_0^N\rangle) + \mathcal{L}^N_{sel} f(\langle\phi, X_0^N\rangle)$$

$$
\begin{aligned}
&NM^d\Bigg[\int_{\mathbb{R}^d}\int_{B_r^M(x)}\frac{1}{|B_r^M|}\frac{X_0^N(z)}{K}f\Bigg(K\frac{u}{J}\int_{B_r^M(x)}\phi(y)\mathrm{d}y\\
&+\Big(1-\frac{u}{J}\Big)\int_{B_r^M(x)}\phi(y)X_0^N(y)\mathrm{d}y\Bigg)-f\Big(\int_{B_r^M(x)}\phi(y)X_0^N(y)\mathrm{d}y\Big)\mathrm{d}z\mathrm{d}x\\
&+\int_{\mathbb{R}^d}\int_{B_r^M(x)}\Big(1-\frac{X_0^N(z)}{K}\Big)f\Bigg(\Big(1-\frac{u}{J}\Big)\int_{B_r^M(x)}\phi(y)X_0^N(y)\mathrm{d}y\Bigg)\\
&-f\Big(\int_{\mathbb{R}^d}\phi(y)X_0^N(y)\mathrm{d}y\Big)\mathrm{d}z\,\mathrm{d}x\Bigg]\\
&+N\frac{s\widehat{S}}{S}M^d\Bigg[\int_{\mathbb{R}^d}\int_{B_r^M(x)}\frac{1}{|B_r^M|}\frac{\sigma(\kappa_r,\zeta)\frac{X_0^N(z)}{K}}{\sigma(\kappa_r,\zeta)\frac{X_0^N(z)}{K}+\sigma(\kappa_s,\zeta)\Big(1-\frac{X_0^N(z)}{K}\Big)}\\
&f\Bigg(K\frac{u}{J}\int_{B_r^M(x)}\phi(y)\mathrm{d}y+\Big(1-\frac{u}{J}\Big)\int_{B_r^M(x)}\phi(y)X_0^N(y)\mathrm{d}y\Bigg)\\
&-f\Big(\int_{B_r^M(x)}\phi(y)X_0^N(y)\mathrm{d}y\Big)\mathrm{d}z\mathrm{d}x\\
&+\int_{\mathbb{R}^d}\int_{B_r^M(x)}\frac{\sigma(\kappa_c,\zeta)\Big(1-\frac{X_0^N(z)}{K}\Big)}{\sigma(\kappa_r,\zeta)\frac{X_0^N(z)}{K}+\sigma(\kappa_s,\zeta)\Big(1-\frac{X_0^N(z)}{K}\Big)}f\Bigg(\Big(1-\frac{u}{J}\Big)\int_{B_r^M(x)}\phi(y)X_0^N(y)\mathrm{d}y\Bigg)\\
&-f\Big(\int_{\mathbb{R}^d}\phi(y)X_0^N(y)\mathrm{d}y\Big)\mathrm{d}z\,\mathrm{d}x\Bigg]. \qquad (3.4.3)
\end{aligned}
$$

We restate useful lemmas from Chapter 2, which describe the form and estimates for the neutral part of the generator.

Definition 3.4.50. *We denote by $\mathcal{A}^N$ the operator*

$$\mathcal{A}^N(\phi) := \frac{C(d)Nur^{d+2}}{JM^2}\Delta\phi,$$

with $C(d) := \int_{\mathcal{B}_1(0)} x^2\mathrm{d}x$.

Lemma 3.4.51. *For $f(x,\zeta)=x$ and $\phi_s(x):\mathbb{R}\times\mathbb{R}^d\to\mathbb{R}\in C_0^{2,3}$,*

$$\int_0^t \mathcal{L}_{neu}^N(\langle\phi,X^N(s)\rangle)\mathrm{d}s=\int_0^t\Big\langle X_s^N,\dot{\phi}_s\Big\rangle+\Big\langle X_s^N,\mathcal{A}^N(\phi_s)\Big\rangle\mathrm{d}s+N_t^N(\phi), \qquad (3.4.4)$$

where

$$|N_t^N(\phi)|\le\mathcal{O}\Bigg(\frac{N\sup_{0\le s\le t}\|\phi_s\|_{C^3}}{JM^3}V_r\Bigg)\int_0^t\Big\langle X_s^N,\mathbf{1}\Big\rangle\mathrm{d}s. \qquad (3.4.5)$$

We proceed to the proof of Lemma 3.4.52. We notice that the proof is a simple combination of our proof of Lemma 3.2.17 and the proof of, Lemma 2.5.20.

Lemma 3.4.52. *Let $X = Kw$ denote the total intensity of individuals of the rare type. Assume that $\mathbb{E}[X^N(0)] < \infty$. Then for any $T > 0$*

$$\sup_{t\leq T}\sup_N \mathbb{E}[\langle X^N, 1\rangle] <\infty, \tag{3.4.6}$$

$$\lim_{H\to\infty}\sup_N \mathbb{P}\left[\sup_{t\leq T}\langle X^N, 1\rangle \geq H\right] =0. \tag{3.4.7}$$

Proof We observe that with the test functions chosen as in Lemma 3.4.51, the part of the generator describing the change in the population resulting from a selective event can be written as

$$\begin{aligned}
&\mathcal{L}^N_{sel}(\langle\phi, X^N(s)\rangle) = N\frac{us\widehat{S}}{SJ}M^d\Bigg[\\
&\int_{\mathbb{R}^d}\frac{1}{|B^M_r|}\int_{B^M_r(x)}\frac{\sigma(\kappa_r,\zeta)}{\sigma(\kappa_r,\zeta)\frac{X}{K}+\sigma(\kappa_c,\zeta)\Big(1-\frac{X}{K}\Big)}\frac{X(z)}{K}\int_{B^M_r(x)}(K-X(y))\phi(y)\mathrm{d}y\mathrm{d}z\mathrm{d}x\\
&-\int_{\mathbb{R}^d}\frac{1}{|B^M_r|}\int_{B^M_r(x)}\frac{\sigma(\kappa_r,\zeta)}{\sigma(\kappa_r,\zeta)\frac{X}{K}+\sigma(\kappa_c,\zeta)\Big(1-\frac{X}{K}\Big)}\Big(1-\frac{X(z)}{K}\Big)\int_{B^M_r(x)}X(y)\phi(y)\mathrm{d}y\mathrm{d}z\mathrm{d}x\Bigg]\\
&= N\frac{us\widehat{S}}{SJ}M^d\Bigg[\int_{\mathbb{R}^d}\frac{1}{|B^M_r|}\int_{B^M_r(x)}\frac{\sigma(\kappa_r,\zeta)}{\sigma(\kappa_r,\zeta)\frac{X}{K}+\sigma(\kappa_c,\zeta)\Big(1-\frac{X}{K}\Big)}X(z)\int_{B^M_r(x)}\phi(y)\mathrm{d}y\mathrm{d}z\mathrm{d}x\\
&\qquad-\int_{\mathbb{R}^d}\frac{1}{|B^M_r|}\int_{B^M_r(x)}\int_{B^M_r(x)}X(y)\phi(y)\mathrm{d}y\mathrm{d}z\mathrm{d}x\Bigg].
\end{aligned}\tag{3.4.8}$$

We observe that by Taylor's Theorem $\phi(y)$ can be locally approximated as

$$\phi(y) = \phi(z) + \nabla\phi(z)(y-z) + \|\phi\|_{C^2(\mathbb{R}^d)}\mathcal{O}\Big(|y-z|^2\Big).$$

Therefore by using a calculation analogous to (3.2.26) and the fact that $|y-z| < R/M$ within a ball $B_r(x)$ we may approximate (3.4.8) by

$$\begin{aligned}
&\mathcal{L}^N_{sel}(\langle\phi, X^N(s)\rangle) =\\
&N\frac{us\widehat{S}}{SJM^{2d}}\Big(\frac{\sigma(\kappa_r,\zeta)}{\sigma(\kappa_c,\zeta)}-1\Big)\langle X,\phi\rangle + \frac{1}{K}\langle X,\phi\rangle^2\mathcal{O}\Big(C_\sigma N\frac{us\widehat{S}}{SJM^{2d}}\Big)+\mathcal{O}\Big(\frac{N\widehat{S}}{SJM}\Big),
\end{aligned}\tag{3.4.9}$$

where C_σ is a constant depending on σ.

As in Chapter 2, let $\chi^N(t) = \mathbb{E}[\langle X_t^N, 1\rangle]$ and let h_R denote a sequence of smooth functions such that h_R is supported on the ball of radius $2R$ centred at zero and is equal to 1 on the ball of radius R centred at zero.

Assume in addition that the sequence h_R satisfies

$$\Delta h_R \leq \epsilon; \quad h_R \leq h_{R+1}.$$

We combine (3.4.4) and (3.4.9) and take expectation in (3.4.3) to obtain

$$\begin{aligned}
\mathbb{E}\Big[\langle X_t^N, h_R\rangle\Big] &= \mathbb{E}\Big[\langle X_0^N, h_R\rangle\Big] + \mathbb{E}\bigg[\int_0^t \Big\langle X_s^N, \mathcal{A}^N(h_R)\Big\rangle \mathrm{d}s\bigg] \\
&\quad + N\frac{us\widehat{S}}{SJM^{2d}}\mathbb{E}\bigg[\bigg(\frac{\sigma(\kappa_r,\zeta)}{\sigma(\kappa_c,\zeta)} - 1\bigg)\int_0^t \langle X, h_R\rangle + \frac{C_\sigma}{K}\langle X, h_R\rangle^2 \mathrm{d}s\bigg] + \mathbb{E}\Big[N_t^N(h_R)\Big] \\
&\leq \chi^N(0) + C\|\Delta h_R\| + \mathcal{O}\bigg(\frac{N\sup_{0\leq s\leq t}\|h_R\|_{C^3}}{JM^3}V_r + N\frac{us\widehat{S}C_\sigma}{SJM^{2d}}\bigg)\chi^N(0), \quad (3.4.10)
\end{aligned}$$

where we have used the properties of h_R. Letting R tend to infinity and using the Monotone Convergence Theorem, we arrive at

$$\chi^N(t) \leq \chi^N(0) + C\epsilon + \mathcal{O}\bigg(\frac{N\sup_{0\leq s\leq t}\|h_R\|_{C^3}}{JM^3}V_r + N\frac{us\widehat{S}C_\sigma}{SJM^{2d}}\bigg)\chi^N(0).$$

We can then conclude by Grönwall's inequality. The second part of the statement follows exactly as the second part of the proof of Lemma 3.2.17. □

3.4.3 Sketch of the proof of Theorem 3.1.4

The generator of the SLFVRE process is given by

$$\begin{aligned}
&A^N f(\eta) \\
&= NM^d \int_{\mathbb{R}^d}\Bigg(\int_0^\infty \Bigg[\frac{uK|B_r|}{JM^d}e^{-\frac{uK|B_r|}{JM^d}v^*}\widehat{g}_{y,r/M}(\kappa^*,v^*)e^{-\frac{uK|B_r|}{JM^d}\int_{v^*}^\infty(1-\widehat{g}_{y,r/M}(\kappa^*,v))\mathrm{d}v} \\
&\qquad \prod_{(x,\kappa,l)\in\eta, l\neq l^*} g\Big(x,\kappa,\mathfrak{J}_{neu}(l,l^*,v^*,x,(y,r/M))\Big)\Bigg]\mathrm{d}v^* - f(\eta)\Bigg)\mathrm{d}y
\end{aligned}$$

$$+ N\frac{s\widehat{S}}{S}M^d \int_{\mathbb{R}^d}\Bigg(\int_0^\infty \Bigg[\frac{uK|B_r|}{JM^d}e^{-\frac{uK|B_r|}{JM^d}v^*}\widehat{g}_{y,r/M}(\kappa^*,v^*)e^{-\frac{uK|B_r|}{JM^d}\int_{v^*}^\infty(1-\widehat{g}_{y,r/M}(\kappa^*,v))\mathrm{d}v}$$
$$\prod_{(x,\kappa,l)\in\eta, l\neq l^*} g\Big(x,\kappa,\mathfrak{J}_{sel}(l,l^*,v^*,x,(y,r/M))\Big)\Bigg]\mathrm{d}v^* - f(\eta)\Bigg)\mathrm{d}y + A^{env}, \quad (3.4.11)$$

where A^{env} is specified as in (3.1.1) and $\widehat{g}_{y,r/M}(\kappa,l) := \frac{1}{|B_{r/M}|}\int_{B_{r/M}(y)} g(z,\kappa,l)\mathrm{d}z$.

We split the generator into three parts, by adding and subtracting $g(y,\kappa^*,v)$ inside the integral. The first two parts describes the movement of the levels.

$$A^N_{full,neu}f(\eta)$$
$$= NM^d\int_{\mathbb{R}^d}\Bigg(\int_0^\infty\Bigg[\frac{uK|B_r|}{JM^d}e^{-\frac{uK|B_r|}{JM^d}v^*}g(x^*,\kappa^*,v^*)e^{-\frac{uK|B_r|}{JM^d}\int_{v^*}^\infty(1-\widehat{g}_{y,r/M}(\kappa^*,v))\mathrm{d}v}$$
$$\prod_{(x,\kappa,l)\in\eta_{B_{r/M}(y)},l\neq l^*} g\Big(x^*,\kappa,\mathfrak{J}_{neu}(l,l^*,v^*,x,(y,r/M))\Big)\prod_{(x,\kappa,l)\notin\eta_{B_{r/M}(y)},l\neq l^*} g\Big(x,\kappa,l\Big)\Bigg]\mathrm{d}v^*$$
$$- f(\eta)\Bigg)\mathrm{d}y, \quad (3.4.12)$$

where $\eta_{B_{r/M}(y)} := \{(x,\kappa,l)\in\eta : x\in B_{r/M}(y)\}$ and

$$A^N_{full,sel}f(\eta)$$
$$= N\frac{s\widehat{S}}{S}M^d\int_{\mathbb{R}^d}\Bigg(\int_0^\infty\Bigg[\frac{uK|B_r|}{JM^d}e^{-\frac{uK|B_r|}{JM^d}v^*}g(x^*,\kappa^*,v^*)e^{-\frac{uK|B_r|}{JM^d}\int_{v^*}^\infty(1-\widehat{g}_{y,r/M}(\kappa^*,v))\mathrm{d}v}$$
$$\prod_{(x,\kappa,l)\in\eta_{B_{r/M}(y)},l\neq l^*} g\Big(x^*,\kappa,\mathfrak{J}_{sel}(l,l^*,v^*,x,(y,r/M))\Big)\prod_{(x,\kappa,l)\notin\eta_{B_{r/M}(y)},l\neq l^*} g\Big(x,\kappa,l\Big)\Bigg]\mathrm{d}v^*$$
$$- f(\eta)\Bigg)\mathrm{d}y, \quad (3.4.13)$$

Observe that $\widehat{g}_{y,R/M}$ is symmetrical with respect to y. Therefore (3.4.12) and (3.4.13) can be treated using arguments which are analogous to those which we applied to $A_{neu,1}, A_{neu,2}, A_{sel,1}$ and $A_{sel,2}$ in Section 3.2. The second two parts are given by

$$A^N_{full,neu,2}f(\eta) = NM^d\int_{\mathbb{R}^d}\Bigg(\int_0^\infty\Bigg[\frac{uK|B_r|}{JM^d}e^{-\frac{uK|B_r|}{JM^d}v^*}\Big(\widehat{g}_{y,r/M}(\kappa^*,v^*) - g(y,\kappa^*,v^*)\Big)$$
$$\times e^{-\frac{uK|B_r|}{JM^d}\int_{v^*}^\infty(1-\widehat{g}_{y,r/M}(\kappa^*,v))\mathrm{d}v}$$

$$\times \prod_{(x,\kappa,l)\in\eta, l\neq l^*} g\Big(x,\kappa,\mathcal{J}_{neu}(l,l^*,v^*,x,(y,r/M))\Big)\Bigg]\mathrm{d}v^*\Bigg)\mathrm{d}y \quad (3.4.14)$$

and

$$A^N_{full,sel,2}f(\eta) = N\frac{s\widehat{S}}{S}M^d\int_{\mathbb{R}^d}\Bigg(\int_0^\infty\Bigg[\frac{uK|B_r|}{JM^d}e^{-\frac{uK|B_r|}{JM^d}v^*}\Big(\widehat{g}_{y,r/M}(\kappa^*,v^*)-g(y,\kappa^*,v^*)\Big)$$
$$\times e^{-\frac{uK|B_r|}{JM^d}\int_{v^*}^\infty(1-\widehat{g}_{y,r/M}(\kappa^*,v))\mathrm{d}v}$$
$$\times \prod_{(x,\kappa,l)\in\eta, l\neq l^*} g\Big(x,\kappa,\mathcal{J}_{sel}(l,l^*,v^*,x,(y,r/M))\Big)\Bigg]\mathrm{d}v^*\Bigg)\mathrm{d}y. \quad (3.4.15)$$

We observe that our scaling implies that if the contribution from (3.4.14) is non-negligible, the contribution from (3.4.15) vanishes in the limit. For that reason the results from Chapter 2 are sufficient to deduce the behaviour of the spatial movement of the generator of all cases of interest. From

$$A^N_{full,sel,2}f(\eta) = NM^d\int_{\mathbb{R}^d}\int_0^\infty\frac{uK|B_r|}{JM^d}e^{-\frac{uK|B_r|}{JM^d}v^*}\Big(\widehat{g}_{y,r/M}(\kappa^*,v^*)-g(x^*,\kappa^*,v^*)\Big)$$
$$\times e^{-\frac{uK|B_r|}{JM^d}\int_{v^*}^\infty(1-\widehat{g}_{y,r/m}(\kappa^*,v))\mathrm{d}v}\prod_{(x,\kappa,l)\in\eta, l\neq l^*} g(x,\kappa,\mathcal{J})\mathrm{d}v^*\mathrm{d}y,$$

we can see as in the calculations in Section 3.2 that this leads to

$$\alpha A^N_{full,sel,2}h(\Xi)$$
$$= NM^d\int_{\mathbb{R}^d\times\mathcal{K}}\frac{M^d}{|B_r|}\int_{B_{r/M}(x^*)}e^{-\int_{\mathbb{R}/B_{r/M}(y)}h(z,\kappa)\Xi(\mathrm{d}z,\mathrm{d}\kappa)}e^{-\left(1-\frac{u}{J}\right)\int_{B_{r/M}(y)}h(z,\kappa)\Xi(\mathrm{d}z,\mathrm{d}\kappa)}$$
$$\times\Big[e^{-\frac{uK|B_r|}{JM^d}\widehat{h}_{y,r/M}(\kappa^*)}-e^{-\frac{uK|B_r|}{JM^d}h(x^*,\kappa^*)}\Big]\mathrm{d}y\,\Xi(\mathrm{d}x^*,\mathrm{d}\kappa^*).$$

Performing a simple Taylor expansion and noting this is identical to calculations appearing in Section 2.4 we see that this will give us

$$\alpha A^N_{full,sel,2}f(\Xi) = \exp(-\langle h,\Xi\rangle)\left\langle\frac{C(d)ur^dN}{JM^2}\Delta h+\mathcal{O}(1/M),\Xi\right\rangle,$$

where we recall that $C(d) := \int_{B_1}x^2\mathrm{d}x$ and $\langle h,\Xi\rangle := \int_{\mathbb{R}^d\times\mathcal{K}}h(x,\kappa)\Xi(\mathrm{d}x,\mathrm{d}\kappa)$.

For the sake of completeness we state the propositions describing the evolution of the levels. We observe that both (3.4.12) and (3.4.13) are integrals over compact sets (recall that $g(x,l)$ is equal to 1 outside of a compact set) of their non-spatial counterparts studied in Section 3.2.3. The task of analysing this generator is then a simple expansion of arguments in that section. In particular, the sparsity condition $K/M^d \to \infty$ leads to following analogous of Proposition 3.2.18 and Proposition 3.2.19.

Proposition 3.4.53. *Under the conditions of Theorem 3.4.48,*

$$\mathbb{E}\Bigg[\sup_{t\leq T}\Bigg|\int_0^t A^N_{full,neu}f(\eta_s) - \left(f(\eta^r_s)\sum_{l_i(s)\in\eta^N_s} al_i^2\frac{\partial_l g(x_i,l_i(s))}{g(x_i,l_i(s))} + 2af(\eta^r_s)\sum_{l_i(s)\in\eta^r_s}\int_{l_i(s)}^{\infty}(1-g(x_i,v))\mathrm{d}v\right)\mathrm{d}s\Bigg|\Bigg] \to 0.$$

Proposition 3.4.54. *Under the conditions of Theorem 3.4.49 for any $T \in \mathbb{R}$*

$$\mathbb{E}\left[\sup_{t\leq T}\left|\int_0^t A^N_{full,sel}f(\eta^N_s) - f(\eta^N_s)\left(\sum_{l_i(t)\in\eta^N_t} -bl_i\frac{g'(x_i,l_i)}{g(x_i,l_i)}\right)\mathrm{d}s\right|\right] \to 0.$$

The proofs of Theorem 3.1.4, Theorem 3.4.48 and Theorem 3.4.49 now follow by applying Theorem 3.2.16 to the process characterised by the generator of the projected version of $A^N_{full,neu}$, $A^N_{full,sel}$ and A_{env}.

3.5 Supplementary results

We present in this section some key lemma's required for the proofs present in this chapter. We also present the proofs of Theorem 3.3.41 and Theorem 3.3.42 as an aside to the main body of work.

3.5.1 Poisson random measures

We present some facts about Poisson random measures. Most of the facts presented in this section have been stated in the papers on lookdown constructions, see, for example Kurtz and Rodrigues (2011), Etheridge and Kurtz (2018). We present them again as they are useful for many calculations involving lookdown constructions.

Lemma 3.5.55. *Let ξ be a Poisson random measure with mean measure ν. Let $f \in L^1(\mathbb{R}^d, \nu)$. Then*

$$\mathbb{E}\Big[\exp\Big(\int f(z)\xi(\mathrm{d}z)\Big)\Big] = \exp\Big(\int (e^{f(x)} - 1)\nu(\mathrm{d}x)\Big)$$

Similarly, the expected value and variance of the integral with respect to a Poisson random measure is given by

$$\mathbb{E}\Big[\int f(z)\xi(\mathrm{d}z)\Big] = \int f(x)\nu(\mathrm{d}x) \quad Var\Big[\int f(z)\xi(\mathrm{d}z)\Big] = \int f^2(x)\nu(\mathrm{d}x)$$

Definition 3.5.56 (Conditionally Poisson system)**.** *Consider a counting measure ξ on $\mathbb{R}^d$. Let Ξ be a locally finite random measure on $\mathbb{R}^d$. We say that ξ is conditionally Poisson with Cox measure Ξ if, conditioned on Ξ, ξ is a Poisson random measure with mean measure Ξ.*

Conditionally Poisson systems are sometimes referred to as Cox processes. We notice that to check that a Poisson random measure ξ is actually a Cox process, it is enough to check that

$$\mathbb{E}\Big[\exp\Big(-\int_{\mathbb{R}^d} f\mathrm{d}\xi\Big)\Big] = \mathbb{E}\Big[\exp\Big(-\int_{\mathbb{R}^d}(1 - e^f)\mathrm{d}\Xi\Big)\Big]$$

for all positive Borel-measurable functions f. Our main application of the presented theory is to show convergence of the particles systems to their high intensity limits. We need some more definitions, as the convergence of sequences of conditionally Poisson systems requires a rather exotic topology.

Consider a family of continuous functions $h_k : \mathbb{R}^d \to [0, 1]$ such that

$$\bigcup_k S_{h_k} = \mathbb{R}^d,$$

where S_f denotes the support of f. Let $\mathcal{M}_{h_k}(\mathbb{R}^d)$ be the collection of Borel measures on $\mathbb{R}^d$ such that

$$\int_{\mathbb{R}^d} h_k f \mathrm{d}\nu < \infty.$$

Let $\mathrm{d}\nu^k = h_k \mathrm{d}\nu$. The space $\mathcal{M}_{h_k}(\mathbb{R}^d)$ endowed with the topology of weak convergence of $\mathrm{d}\nu^k$ is metrizable. We observe that checking convergence in $\mathcal{M}_{h_k}(\mathbb{R}^d)$ is equivalent to checking convergence of $\int_{\mathbb{R}^d} f \mathrm{d}\nu_n$ for all bounded and continuous functions which satisfy

$$\int_{\mathbb{R}^d} f \mathrm{d}\nu^k < \infty \text{ for } f \leq c h_k$$

for some constant $c > 0$. The space $\mathcal{M}_{h_k}(\mathbb{R}^d \times [0, \infty))$ can be defined in a similar way.

Theorem 3.5.57 (Kurtz and Rodrigues (2011), Theorem A.9)**.** *Let ξ_n be a sequence of conditionally Poisson random measures on $\mathbb{R}^d \times [0, \infty)$ with Cox measures $\{\Xi_n \times \Lambda\}$. Then $\xi_n \Rightarrow \xi$ in $\mathcal{M}_{h_k}(\mathbb{R}^d \times [0, \infty))$ if and only if $\Xi_n \Rightarrow \Xi$ in $\mathcal{M}_{h_k}(\mathbb{R}^d)$. If the limit exists, ξ is a conditionally Poisson random measure with Cox measure $\Xi \times \Lambda$.*

3.5.2 Markov Mapping Theorem

We recall some basic definitions and introduce the necessary notation. For a detailed account of this introductory material we refer to Ethier and Kurtz (1986), Chapter 1, and Lunardi (2012).

Let (E, d), (E_0, d_0) be a pair of complete, separable metric spaces (with metrics d and d_0, respectively). Let $B(E)$ be the space of bounded measurable functions on E. We notice that equipped with the usual supremum norm $\|\cdot\|_\infty$ $B(E)$ forms a Banach space. Let $C(E) \subset B(E)$ denote the subspace of continuous functions on E. A subspace A of $B(E) \times B(E)$ is a multivalued linear operator. It domain is given by $\mathcal{D} = \{f : (f, g) \in A\}$ and its range by $\mathcal{R} = \{g : (f, g) \in A\}$.

Definition 3.5.58 (Dissipative operator)**.** *We say that the operator A is dissipative if for each $(f, g) \in A$ and $\lambda > 0$*

$$\|\lambda f - g\| \geq \lambda \|f\|.$$

Definition 3.5.59 (Graph separable pre-generator)**.** *We say that an operator $A \subset B(E) \times B(E)$ is a pre-generator if it is dissipative and there exists a sequence of functions μ_n mapping E to the set of probability measures over $\mathcal{P}(E)$, and a sequence of $\lambda_n \in E$,*

such that for each $(f, g) \in A$

$$g(x) = \lim_{n\to\infty} \lambda_n \int_E (f(y) - f(x))\mu_n(x, \mathrm{d}y).$$

If in addition there exists a countable subset $\{f_k\} \subset \mathcal{D}(A) \bigcup C(E)$ *such that the graph of* A *is contained in the closure of the linear span of* (f_n, Af_n)*, we say that it is graph-separable.*

We notice that the generators of Markov process are graph-separable pre-generators. Let $D_E[0, \infty)$ denote the space of càdlàg functions and $M_E[0, \infty)$ denote the space of Borel measurable functions on $[0, \infty)$ taking values in E.

Theorem 3.5.60 (Kurtz and Rodrigues (2011), Theorem A.15)**.** *Let* $A \subset \overline{C}(E) \times C(E)$ *and let* ψ *be a continuous function taking values in* $\mathbb{R}$ *such that* $\psi \geq 1$*. Suppose that for each* $f \in \mathcal{D}(A)$ *there exists a constant* $c_f > 0$ *such that*

$$|Af(x)| \leq c_f \psi(x).$$

Let A_0 *be defined as*

$$A_0 f(x) = \frac{Af(x)}{\psi(x)}.$$

Suppose that A_0 *is graph-separable pre-generator and suppose that* $\mathcal{D}(A) = \mathcal{D}(A_0)$ *is closed under multiplication and separating. Let* $\gamma : E \to E_0$ *be Borel measurable, and let* α *be a transition function from* E_0 *into* E *satisfying* $\alpha(y, \gamma^{-1}(y)) = 1$*. Assume that for each* $y \in E_0$

$$\widetilde{\psi} = \int_E \psi(y, z)\alpha(y, \mathrm{d}z) < \infty.$$

and define

$$C = \left\{ \left(\int_E f(z)\alpha(\cdot, \mathrm{d}z), \int_E Bf(z)\alpha(\cdot, \mathrm{d}z) \right) : f \in \mathcal{D}(B) \right\}.$$

Let $\mu_0 \in \mathcal{P}(E_0)$*, and define* $\nu_0(y) = \int_{\mathbb{R}} \alpha(y, \cdot)\mu_0(\mathrm{d}y)$*.*

1. *Let $\widetilde{Y}$ be a solution of a martingale problem for (C, μ_0). Assume that it satisfies the moment condition*

$$\int_0^t \mathbb{E}\Big[\widetilde{\psi}(\widetilde{Y}(s))\Big]\mathrm{d}s < \infty \quad \forall t \geq 0. \tag{3.5.1}$$

Then there exists a solution X of the martingale problem for (A, ν_0) such that $\widetilde{Y}$ has the same distribution on $\mathcal{M}_{E_0}[0, \infty)$ as $\widetilde{Y} = \gamma \circ Y$.

2. *If, in addition, uniqueness holds for the martingale problem for (A, ν_0), then uniqueness holds for the martingale problem for (C, μ_0).*

The original proof of the theorem was inspired by the proofs of generalisations of Burke's Output Theorem appearing in Kliemann et al. (1990) and the proof of equivalence of martingale problems for the Moran model and its lookdown representation in Donnelly and Kurtz (1996).

Since then the Markov Mapping Theorem has been a useful tool in mathematical population genetics (Etheridge and Kurtz (2018), Kurtz and Rodrigues (2011)), mathematical biology (Gupta (2012)), mathematical finance (Stockbridge (2002)) and analysis of infinite dimensional stochastic differential equations (Kurtz (2010)).

The main power of the Markov Mapping Theorem comes in simplifications of proofs of equivalence of seemingly different martingale problems. The main source of the power is in exploitation of properties of exchangeable process and conditionally Poisson systems.

3.5.3 Kurtz-Rodrigues' Martingale Lemma

The following Lemma plays an important role in our applications of the Markov Mapping Theorem. Intuitively, it clarifies why the averaged process is a solution to a projected martingale problem.

Lemma 3.5.61 (Kurtz and Rodrigues (2011), Lemma A.13)**.** *Let $\{\mathcal{F}_t\}$ and $\{\mathcal{G}_t\}$ be filtrations with $\mathcal{G}_t \subset \mathcal{F}_t$. Suppose that for each $t \geq 0$*

$$\mathbb{E}[|X_t| + \int_0^t |Y_s|\mathrm{d}s] < \infty.$$

and that

$$M_t = X_t - \int_0^t Y_s \mathrm{d}s$$

is an $\mathcal{F}_t$-martingale. Then

$$\widehat{M}_t = \mathbb{E}[X_t|\mathcal{G}_t] - \int_0^t \mathbb{E}[Y_s|\mathcal{G}_s]\mathrm{d}s$$

is a $\{\mathcal{G}_t\}$martingale.

3.5.4 Proofs of Theorem 3.3.41 and Theorem 3.3.42

Proof[Proof of Theorem 3.3.41] Before we can proceed, additional objects need to be defined. Let $\alpha_\lambda(n,l)$ be the joint distribution of n i.i.d. uniformly distributed random variables on $[0,\lambda]$. Recall that μ denotes a point measure representing positions of individuals. For a test function $f_1(x_i,l_i)$, we define the projection onto type space $\widehat{f}$ as

$$\widehat{f}(\mu) = \prod_i \widehat{g}(x_i) = e^{-\sum_i \mathcal{I}(g(x_i))},$$

where the average for a single level is defined as

$$e^{-\mathcal{I}(g(x_i))} = \widehat{g}(x_i) = \frac{1}{\lambda}\int_0^\lambda g(x_i,z)\mathrm{d}z.$$

To calculate the generator of the projected model (the generator averaged over the distribution of the levels), we need to evaluate

$$\int A_\lambda f(\zeta,x,l)\alpha_\lambda(\mathrm{d}l).$$

Let us integrate the four terms appearing in A_λ separately. We begin with the two terms which are least involved - the movement of particles and the environment. Since both of those terms do not depend on the levels, integrals with respect to them do not

alter our projections, namely

$$\int f(\zeta, x, l) \sum_i \frac{Bg(x_i, l_i)}{g(x_i, l_i)} \alpha_\lambda(\mathrm{d}l) = \sum_{i=1}^{n} B\widehat{f}(\zeta, \mu) = nB\widehat{f}(\zeta, \mu). \tag{3.5.2}$$

Analogously,

$$\int \lambda f_1(x, l) A_\lambda^{env} f_0(\zeta, x) \alpha_\lambda(n, \mathrm{d}l) = \lambda \widehat{f_1}(\mu) A_\lambda^{env} f_0(\zeta, \mu). \tag{3.5.3}$$

In order to evaluate terms describing births and movement of the levels, which both do depend on the exact value of the level, it is convenient to note that (here we follow the calculation on p. 492 in Kurtz and Rodrigues (2011))

$$\lambda^{-1} 2a \int_0^\lambda g(x, z) \int_z^\lambda (g(x, v) - 1) \mathrm{d}v \mathrm{d}z = a\lambda e^{-\jmath_g} - 2a\lambda^{-1} \int_0^\lambda g(x, z)(\lambda - z) \mathrm{d}z, \tag{3.5.4}$$

where we have used Fubini's Theorem, and

$$\begin{aligned} \lambda^{-1} \int_0^\lambda (az^2 - \zeta bz) g'(x, z) \mathrm{d}z &= -\lambda^{-1} \int_0^\lambda (2az - \zeta b)(g(z) - 1) \mathrm{d}z \\ &= \lambda^{-1} 2a \int_0^\lambda z g(x, z) \mathrm{d}z + a\lambda + b(e^{-\jmath_g} - 1), \end{aligned} \tag{3.5.5}$$

where we have integrated by parts. It will also be useful to describe the changes in our system due to births and deaths. Whenever a birth event occurs, the new individual is located at the same place as the parent. If a death occurs, the individual is just removed from the system. Therefore, if we denote the new collection of particles after a birth at location y by $(b(\overline{x}|y))$ and the new collection of particles after a death at location x_j by $d(\overline{x}|x_j)$, we see that

$$\mu_{b(\overline{x}|y)} = \delta_y + \sum_{i=1}^{n} \delta_{x_i}, \quad \mu_{d(\overline{x}|x_j)} = -\delta_{x_j} + \sum_{i=1}^{n} \delta_{x_i}.$$

Armed with these observations and identities (3.5.4), (3.5.5) we proceed to evaluate the remaining terms. A simple calculation shows that

$$\begin{aligned}
\int f(\zeta,x,l)\Big\{\sum_i 2a\int_{l_i}^{\lambda}(g(x_i,v)-1)\mathrm{d}v+\sum_i(al_i^2-\sqrt{\lambda}\zeta(x_i)bl_i)\frac{\partial_{l_i}g(x_i,l_i)}{g(x_i,l_i)}\Big\}\alpha_\lambda(\mathrm{d}l)\\
=\sum_j e^{\sum_{i\neq j}\jmath_{g(x_i)}}\Big\{2a\lambda e^{-\jmath_{g(x_j)}}+2a\lambda^{-1}\int_0^{\lambda}g(x_j,z)\mathrm{d}z\\
-2a\lambda^{-1}\int_0^{\lambda}g(x_j,z)\mathrm{d}z+a\lambda+\sqrt{\lambda}b\zeta(x_j)(e^{\jmath_{g(x_j)}}-1)\Big\}\\
=a\lambda e^{\sum_i\jmath_{g(x_i)}}\sum_j(e^{-\jmath_{g(x_j)}}-1)+\sum_j(\lambda a-\sqrt{\lambda}b\zeta(x_j))e^{\sum_{i\neq j}\jmath_{g(x_i)}}(1-e^{\jmath_{g(x_j)}})\\
=\sum_i\lambda a(\widehat{f}(\mu_{b(x|x_i)})-\widehat{f}(\mu))+\sum_i(\lambda a-\sqrt{\lambda}\zeta(x_i)b)(\widehat{f}(\mu_{d(x|x_i)})-\widehat{f}(\mu)). \quad (3.5.6)
\end{aligned}$$

Combining (3.5.2), (3.5.3) and (3.5.6) we have established that the projected generator can be written as

$$\begin{aligned}
\mathcal{L}_\lambda\widehat{f}(\zeta,\mu)=\lambda\widehat{f}_1(\mu)A_\lambda^{env}f_0(\xi)+\sum_i B_{x_i}\widehat{f}(\mu)+\sum_i\lambda a(\widehat{f}(\mu_{b(x|x_i)})-\widehat{f}(\mu))\\
+\sum_i(\lambda a-\zeta(x_i)\sqrt{\lambda}b)(\widehat{f}(\mu_{d(x|x_i)})-\widehat{f}(\mu)),
\end{aligned}$$

which is the generator of the BBMRE with birth rate λa and death rate $(\lambda a-\sqrt{\lambda}\zeta b)$, as claimed.

Now we only need to check that all assumptions of the Markov Mapping Theorem are satisfied. Fortunately our Condition 3.3.43 has been imposed to guarantee just that. The map $\gamma:\mathcal{N}(\mathbb{R}^d\times[0,\lambda))\to\mathcal{M}(\mathbb{R}^d)$ (mapping counting measures to measures on $\mathbb{R}^d$) is given by

$$\gamma\left(\sum_i\delta_{x_i,l_i}\right)=\frac{1}{\lambda}\sum_{l_i}\delta_{x_i}$$

The moment condition (3.5.1) is satisfied if we consider ψ of the form

$$\psi(x,l)=1+\sum_i\psi_B(x_i)(1+a+b)e^{-l_i}, \quad (3.5.7)$$

so that the averaged $\widetilde{\psi}$ is of the form

$$\widetilde{\psi}(x)=1+\sum_i\psi_B(x_i)(1+a+b)(1-e^{-\lambda}). \quad (3.5.8)$$

We note that the 1 appearing in the definitions of ψ and $\widetilde{\psi}$ has been added only to ensure that both of these functions are greater than or equal to 1. □

Proof[Proof of Theorem 3.3.42] We define a test function of the form

$$h_1(\zeta, x, l) = -f_1(x, l) b \sum_i \zeta(x_i) l_i \frac{\partial_l g(x_i, l_i)}{g(x_i, l_i)}$$

and apply the generator (3.3.4) to a test function of the form $G = f_1 + \frac{1}{\sqrt{\lambda}} h_1$. This leads to

$$\begin{aligned}
A_\lambda\Big(f_1(x,l) + \frac{1}{\sqrt{\lambda}} h_1\Big) &= f_1(x,l) \sum_i \frac{B g(x_i, l_i)}{g(x_i, l_i)} \\
&+ f_1(x,l) \sum_i 2a \int_{l_i}^{\lambda} (g(x_i, v) - 1) \mathrm{d}v + f(x,l) \sum_i (a l_i^2 - \sqrt{\lambda} \zeta(x_i) b l_i) \frac{\partial_l g(x_i, l_i)}{g(x_i, l_i)} \\
&+ \sqrt{\lambda} f_1(x,l) \sum_i \zeta(x_i) l_i \frac{\partial_l g(x_i, l_i)}{g(x_i, l_i)} \\
&- \frac{1}{\sqrt{\lambda}} \Bigg\{ f_1(x,l) b \zeta(x_i) \sum_i \Big[\frac{B g(x_i, l_i)}{g(x_i, l_i)} + \frac{B \partial_l g(x_i, l_i)}{g(x_i, l_i)} \Big] \\
&+ \Big[\sum_i \zeta(x_i) l_i \frac{\partial_l g(x_i, l_i)}{g(x_i, l_i)} \Big] f_1(x,l) \sum_i 2a \int_{l_i}^{\lambda} (g(x_i, v) - 1) \\
&+ f_1(x,l) \sum_i [a l_i^2 - \sqrt{\lambda} b \zeta(x_i) l_i] \Bigg(\sum_{j \neq i} l_j \zeta(x_j) \frac{\partial_l g(x_i, l_i) \partial_l g(x_j, l_j)}{g(x_i, l_j) g(x_i, l_i)} \\
&+ \frac{\zeta(x_i) \partial_l g(x_i, l_i) + \zeta(x_i) l_i \partial_{l^2}^2 g(x_i, l_i)}{g(x_i, l_i)} \Bigg) \Bigg\}, \quad (3.5.9)
\end{aligned}$$

where we have used the fact that f_1 does not depend on the environment and that $\mathbb{E}_\pi[h_1] = 0$, (where $\mathbb{E}_\pi$ is the expected value over the stationary distribution for the environment) since $\mathbb{E}_\pi[\zeta] = 0$. Passing to the limit in (3.5.9) as λ tends to infinity we obtain

$$\begin{aligned}
A_\infty(f_1) &= f_1(x,l) \sum_i \frac{B g(x_i, l_i)}{g(x_i, l_i)} \\
&+ f_1(x,l) \sum_i 2a \int_{l_i}^{\infty} (g(x_i, v) - 1) \mathrm{d}v + f_1(x,l) \sum_i (a l_i^2 - b^2 l_i) \frac{\partial_{l_i} g(x_i, l_i)}{g(x_i, l_i)} \\
&- f_1(x,l,n) \sum_i b^2 \zeta(x_i) l_i \Bigg(\sum_{j \neq i} \zeta(x_j) l_j \frac{\partial_{l_i} g(x_i, l_i) \partial_{l_j} g(x_j, l_j)}{g(x_i, l_j) g(x_i, l_i)} + \frac{\zeta(x_i) l_i \partial_{l_i^2}^2 g(x_i, l_i)}{g(x_i, l_i)} \Bigg).
\end{aligned}$$

Conditions of Theorem 3.2.16 are satisfied if we consider $A = A_\infty(f_1)$ which would lead too $\varepsilon_n^f = \mathcal{O}\left(\frac{1}{\lambda}\right)$. We now show that if we average the levels of the limiting generator, we obtain the generator of the SBMRE of Definition 3.3.36. The general principle is the same as for the proof of Theorem 3.3.41 - we average out the levels and refer to the Markov Mapping Theorem to show that the distribution of the projected process is the distribution of the SBMRE.

We consider a Poisson random measure with distribution $\alpha(\mu, \mathrm{d}x \times \mathrm{d}l)$ on $\mathbb{R}^d \times \mathbb{R}_+$ with mean measure $\mu \times m_{leb}$, where m_{leb} is Lebesgue measure. Just as in the Branching Brownian motion case, we consider a special set of test functions of the form

$$h(x) = \int_0^\infty (1 - g(x, l))\mathrm{d}l.$$

In this setup, for a test function f, the projected (averaged) test function, $\widehat{f}$, takes the form

$$\widehat{f}(\mu) = \alpha f(\mu) = \int f(x, v)\alpha(\mu, \mathrm{d}x \times \mathrm{d}v) = e^{\int_{\mathbb{R}^d} \int_0^\infty (1-g(x,v))\mathrm{d}v\mu(\mathrm{d}x)} = e^{-\langle h, \mu\rangle},$$

which is a simple consequence of properties of Poisson random measures. Once again, we integrate the groups of terms that behave similarly separately. Also, to make the calculations easier to read, we write the averaging 'level by level' - performing the computation for a single level wherever possible.

Since the part of the generator which describes the movement of the particles does not depend on the value of the level l the averaging is simply

$$\alpha\left(f_1(x, l)\sum_i \frac{Bg(x_i, l_i)}{g(x_i, l_i)}\right) = \int_{\mathbb{R}^d} -Bh(y)\mu(\mathrm{d}y)e^{-\langle h,\mu\rangle}, \tag{3.5.10}$$

We now turn our attention to the terms which behave in a very similar fashion to those in (3.5.6). The computation is analogous.

$$\begin{aligned}&\int_{\mathbb{R}^d} \int_0^\infty avg(y, z) \int_v^\infty (1 - g(y, z))\mathrm{d}z\mathrm{d}v\mu(\mathrm{d}y)e^{-\langle h,\mu\rangle} \\ &\quad + \int_{\mathbb{R}^d} \int_0^\infty (av^2 - b^2 v)\partial_v g(y, v)\mathrm{d}v\mu(\mathrm{d}y)e^{-\langle h,\mu\rangle}\end{aligned}$$

$$
\begin{aligned}
&= \int_{\mathbb{R}^d} \int_0^\infty avg(y,z) \int_v^\infty (1-g(y,z))\mathrm{d}z\mathrm{d}v\mu(\mathrm{d}y)e^{-\langle h,\mu\rangle} \\
&\quad - \int_{\mathbb{R}^d} \int_0^\infty (2av-b^2)g(y,v)\mathrm{d}v\mu(\mathrm{d}y)e^{-\langle h,\mu\rangle} \\
&= \int_{\mathbb{R}^d} \int_0^\infty avg(y,z) \int_v^\infty (1-g(y,z))\mathrm{d}z\mathrm{d}v\mu(\mathrm{d}y)e^{-\langle h,\mu\rangle} \\
&\quad - \int_{\mathbb{R}^d} \int_0^\infty 2a \int_v^\infty g(z,v)\mathrm{d}z\mathrm{d}v\mu(\mathrm{d}y)e^{-\langle h,\mu\rangle} + \int_{\mathbb{R}^d} \int_0^\infty b^2 g(y,v)\mathrm{d}v\mu(\mathrm{d}y)e^{-\langle h,\mu\rangle} \\
&= \int_{\mathbb{R}^d} a\bigg(\int_0^\infty [g(y,v)-1]\mathrm{d}v\bigg)^2 \mu(\mathrm{d}y)e^{-\langle h,\mu\rangle} + \int_{\mathbb{R}^d} \int_0^\infty b^2 g(y,v)\mathrm{d}v\mu(\mathrm{d}y)e^{-\langle h,\mu\rangle} \\
&= \int_{\mathbb{R}^d} \{ah^2(y) + b^2 h(y)\}\mu(\mathrm{d}y)e^{-\langle h,\mu\rangle}, \qquad (3.5.11)
\end{aligned}
$$

where we have integrated by parts and used the analogues of identities (3.5.4), (3.5.5).

Finally, the projections of the terms which are a direct consequence of separation of timescales lead to

$$
\begin{aligned}
&\int_{\mathbb{R}^d\times\mathbb{R}^d} b^2 \int_0^\infty \int_0^\infty \zeta(y_1)\zeta(y_2) v\partial_v g(y,v) z\partial_z g(y,z)\mathrm{d}v\mathrm{d}z\mu(\mathrm{d}y_1)\mu(\mathrm{d}y_2)e^{-\langle h,\mu\rangle} \\
&\quad = -\int_{\mathbb{R}^d\times\mathbb{R}^d} b^2\zeta(y_1)\zeta(y_2) \int_0^\infty (g(y,v)-1)\mathrm{d}v \int_0^\infty (g(y,z)-1)\mathrm{d}v\mathrm{d}z\mu(\mathrm{d}y_1)\mu(\mathrm{d}y_2)e^{-\langle h,\mu\rangle} \\
&\quad = \int_{\mathbb{R}^d\times\mathbb{R}^d} b^2\zeta(y_1)\zeta(y_2)h(y_1)h(y_2)\mu(\mathrm{d}y_1)\mu(\mathrm{d}y_2)e^{-\langle h,\mu\rangle}, \qquad (3.5.12)
\end{aligned}
$$

where we have integrated by parts, and

$$
\begin{aligned}
&\int_{\mathbb{R}^d} b^2 \int_0^\infty v^2\partial_v^2 g(y,v)\mathrm{d}v\mu(\mathrm{d}y)e^{-\langle h,\mu\rangle} = -\int_{\mathbb{R}^d} 2b^2 \int_0^\infty v\partial_v g(y,v)\mathrm{d}v\mu(\mathrm{d}y)e^{-\langle h,\mu\rangle} \\
&\quad = -\int_{\mathbb{R}^d} 2b^2 \int_0^\infty (1-g(y,v))\mathrm{d}v\mu(\mathrm{d}y)e^{-\langle h,\mu\rangle} = -\int_{\mathbb{R}^d} 2b^2 h(y)\mu(\mathrm{d}y)e^{-\langle h,\mu\rangle}, \qquad (3.5.13)
\end{aligned}
$$

where we have integrated by parts twice.

Combining the calculations (3.5.10), (3.5.11), (3.5.12), (3.5.13) and appealing to Theorem 3.2.16 to average over the environment we arrive at

$$
\begin{aligned}
\mathcal{L}\widehat{f}(\mu) = \mathcal{L}e^{\langle f,\mu\rangle} \int_{\mathbb{R}^d} \Big\{-Bh(y) + ah^2(y) - b^2 h(y)\Big\}\mu(\mathrm{d}y)e^{-\langle h,\mu\rangle} \\
+ \int_{\mathbb{R}^d\times\mathbb{R}^d} b^2 q(y_1,y_2)h(y_1)h(y_2)\mu(\mathrm{d}y_1)\mu(\mathrm{d}y_2)e^{-\langle h,\mu\rangle},
\end{aligned}
$$

which is the generator of the SBMRE.

As before, to ensure that the solution of the martingale problem for the lookdown process gives us information about the solution of the martingale problem for the projected process, we need to specify the Markov map γ and check that the conditions of the Markov Mapping Theorem are satisfied. Once again we appeal to Condition 3.3.43.

The Markov map γ is given by

$$\gamma\left(\sum_i \delta_{x_i,l_i}\right) = \begin{cases} \lim_{\lambda\to\infty} \frac{1}{\lambda} \sum_{l_i\leq\lambda} \delta_{x_i} & \text{if measures converge,} \\ \mu_0 & \text{otherwise .} \end{cases}$$

Our class of test functions is separating over the counting measures, and closed under multiplication. The moment condition 3.5.1 is satisfied if we consider ψ of the form (3.5.7) and the averaged $\widetilde{\psi}$ of the form (3.5.8). □

Chapter 4

Multiple merger coalescents: truncated offspring distributions and large sample sizes

4.1 Introduction

The passing of genetic information from parents to offspring is a fundamental characteristic of living organisms. In population genetics, mathematical models of how genetic copies are passed on between generations usually focus on the number of offspring, or gene copies, produced by each individual at any given time. From these models one develops appropriate tools for statistical inference. Commonly, data takes the form of the DNA sequence of a given gene (or at a given genetic locus) for each individual in a sample from the population. The differences between these sequences are then used to infer information about the genealogical tree relating the genes in the sample. In particular, if we know the distribution of the genealogical trees under our population models, we can aim to use DNA sequence data to distinguish between different evolutionary scenarios. The statistic we will focus on is the sight frequency spectrum which we will discuss further in Section 4.2.1.

The celebrated Kingman coalescent provides a reference point introduced in Kingman (1982a), Kingman (1982b) and Kingman (1982c). It arises as an approximation to the

genealogical trees (viewed on an appropriate timescale, determined by the 'effective population size') relating individuals in a sample from a large population. As discussed in Section 1.1.2 the Kingman coalescent approximation holds for a whole raft of population models, and it is this stability under small perturbations of the underlying population model that makes it such a powerful tool.

A key feature of Kingman's coalescent is that at most two ancestral lineages merge in a given coalescence event. This reflects two underlying assumptions: first, the sample size is a negligible fraction of the total effective population size; second, in the corresponding population models, any given individual has a negligibly small number of offspring relative to the total effective population size. Regarding the first assumption, in many cases the effective population size is much smaller than census population size as discussed in Árnason (2004) and Tenesa et al. (2007). With rapid developments in DNA sequencing technology, sample sizes are growing rapidly, so that the common assumption about sample size can increasingly be called into question. In addition, developments in algorithm design have enabled efficient computer simulations of ancestral trees of very large samples for example in Kelleher et al. (2016). Regarding the second assumption, observations from ecology and molecular population genetics suggest that in certain marine species family sizes may be highly variable. In particular, individuals may have very many offspring. For instance, May (1967), Árnason (2004) and Árnason and Halldórsdóttir (2015) discuss Atlantic cod (*Gadus morhua*), Li and Hedgecock (1998) discusses Pacific oysters (*Crassostrea gigas*), Turner et al. (2002) looks at red drum (*Sciaenops ocellatus*), Niwa et al. (2016) analyses Japanese sardines (*Sardinella zunasi*) and Hedgecock (1994) looks at more general marine life with analysis of Pacific oysters.

These organisms are characterised by broadcast spawning, external fertilisation, extremely high fecundity and a so-called Type III survivorship curve (high initial mortality followed by a slow decrease in survival probability). This is often referred to as sweepstakes reproduction in the litererature for instance discussed in Beckenbach (1994) and Hedgecock (1994). Similar considerations can apply to some plant populations, which distribute pollen, and some insect populations, where individuals of one gender far outnumber those of the other.

In *multiple-merger* coalescent processes, three or more lineages may merge in a single coalescence event. These are analysed in Sagitov (1999), Pitman (1999) and Donnelly and Kurtz (1999). Multiple-merger coalescents are referred to as Λ-coalescent if at most one group of ancestral lines merge each time, and Ξ-coalescents if at least two distinct groups of lines can merge at the same time. It has been suggested that multiple-merger coalescents may provide a more appropriate description, compared to the Kingman coalescent, of the genealogy of a sample from highly fecund populations with sweepstakes mode of reproduction for instance Eldon and Wakeley (2006), Árnason (2004) and Sargsyan and Wakeley (2008) have analysed genetic data. A subfamily of multiple-merger coalescents, known as Beta-coalescents, has been a focus of mathematical analysis in Schweinsberg (2003) and Birkner et al. (2005a). Beta-coalescents are parametrised by a quantity, α, which captures the tail behaviour of the distribution of the random number of juveniles contributed by a single individual in sweepstakes reproduction. For $\alpha \in [1, 2)$ the Beta-coalescent arises naturally from a population model that we shall refer to as the Schweinsberg model from Schweinsberg (2003). In this model, which is a Cannings model that shares certain features with the Wright-Fisher model, each of the current N individuals, independently produces a random number, X, of juveniles. A new set of (surviving) offspring is then formed by sampling N individuals to survive to the next generation uniformly at random from the resultant pool of juveniles. It is assumed that $\mathbb{P}[X > x] \sim Cx^{-\alpha}$ as $x \to \infty$. The value of α will determine the timescale of coalescence. If c_N is the probability that two individuals, sampled at random from the population, have a common parent in the previous generation, then as $N \to \infty$, $N^{\alpha-1}c_N \to C\alpha B(2-\alpha, \alpha)$ (where the function $B(x, y)$ is the beta function to be made explicit later). Measuring time in units of $1/c_N$ generations, as $N \to \infty$, the genealogy of a sample from the population converges to a Beta-coalescent with parameter α.

An obvious objection to the Schweinsburg model as a justification of the Beta-coalescent model is that the parameter, α, is determined by the extreme tail behaviour of the distribution of the number of juveniles, and while some organisms can produce vast numbers of juveniles, no organism can produce arbitrarily large numbers of juveniles. Moreover, Beta-coalescents would be expected to cease to be a good approximation to genealogies of very large samples from a population of finite size under the Schweinsberg

model, just as the Kingman coalescent breaks down as an approximation to the genealogy of a large sample from a population of finite size and Wright-Fisher sampling, for example see Bhaskar et al. (2014), Wakeley and Takahashi (2003), Fu (2006), Melfi and Viswanath (2018a) and Melfi and Viswanath (2018b).

The aims in this chapter are threefold. First we address the problems of arbitrary numbers of juveniles by considering a Schweinsberg type model modified to incorporate a cutoff, dependent on total population size, of the number of juveniles that can be produced by a single individual. We will then identify the corresponding coalescent. Depending on the cut-off, this will be a Beta-coalescent, a 'truncated' Beta-coalescent or a Kingman coalescent. Second, we shall investigate the variability in c_N for large N. This provides some insight into how variable estimates of effective population size, defined to be $1/c_N$, will be for a population evolving under this model. Third, we shall investigate the breakdown in the coalescent approximation for large sample sizes. In this context, we focus on when the breakdown will be evident in data. We shall argue that, in practice, this will occur at a larger sample size than that at which anomalies first appear in the full coalescent tree, but nonetheless at sample sizes that are certainly not out of reach of modern datasets. To our knowledge this is the first study of the impact of large sample size on gene genealogies in the context of skewed offspring distribution and multiple-merger coalescents.

The rest of this chapter is laid out as follows. In Section 4.2, we provide some background on the main statistical tool we analyse, the site frequency spectrum, Schweinsburg's model and the Beta-coalescent. In Section 4.3 we introduce our modification of the Schweinsberg model. We then give a precise statement of our main results in Section 4.4, presenting the timescaling in Proposition 4.4.1, and the resulting coalescent models in Proposition 4.4.2 and Proposition 4.4.3. The heuristics on quantifying the impact of large sample size are in Section 4.5, with the calculations behind them in Section 4.7. We then gather the details of the proofs into Section 4.6. Many of the ideas follow Schweinsberg (2003), although we have taken care to keep track of the error bounds in contrast to Schweinsberg (2003). We note that the error bounds for Schweinsberg (2003) can be recovered formally by taking $\Psi(N) = \infty$ in our results. In particular, in Section 4.7 we consider the impact of large sample size. In Section 4.7.1 we

consider convergence to the Kingman coalescent under our modified Schweinsberg model and in Section 4.7.2 we consider convergence to the Beta-coalescent. In Section 4.7 we give heuristics for the sample size required to see its effect on the site-frequency spectrum in the case of the Kingman coalescent and the Beta-coalescent.

4.2 Background

4.2.1 Mutation and the site frequency spectrum

In practice, one does not observe the coalescent process directly, but rather a projection of it, determined by the mutations that fall on the coalescent tree. We will assume that the infinitely many sites model of mutations, as discussed in Kimura (1969) and Watterson (1975), applies. In the infinitely many sites model an individual retains all the mutations that fall on its ancestral line. For a sample of size n, we define the 'unfolded site frequency spectrum' to be the vector $(M_1(n), M_2(n), \ldots, M_{n-1}(n))$, where $M_j(n)$ is the number of new mutations carried by exactly j of the n individuals in the sample. The expected value of $M_j(n)$ is just the expected number of mutations to fall on branches of the genealogical tree that are ancestral to exactly j individuals in the sample. This shows that, although the site-frequency spectrum is a simple summary statistic of the full DNA sequence data, it contains valuable information about the pattern of observed genetic variation due to the structure of the genealogical tree. Under the Kingman coalescent, it is well known that if the number of new mutations per individual per generation is $\widetilde{\mu}$, the 'effective population size' is N_e, and $\theta := 2N_e\mu$, then $\mathbb{E}[M_j(n)] = \theta/j$, Fu (1995).

The amount of information available to us about the genealogy of the sample is strictly limited by the mutation rate, it would have to be infinite for us to be able to resolve the tree entirely. The effective population size appearing in the scaling of θ above, is sometimes called the coalescent effective population size in the literature, for instance in Wakeley and Sargsyan (2008). It is the amount by which time must be rescaled in order for the Kingman coalescent to approximate the genealogy of a sample from the population. For the Schweinsberg model Schweinsberg (2003) shows if the limiting coalescent is the Beta-coalescent with parameter $\alpha \in (1, 2)$, the corresponding scaling of

time will be of order $N^{\alpha-1}$, where N is population size. We write $\widetilde{\mu}$ for the mutation rate per individual per generation. In the Beta-coalescent, the expected time until the first coalescence event among a sample of n lineages is of order $1/n^{\alpha}$. Then if we are to see any mutations at all before the first merger, which is a minimum requirement but certainly not enough to resolve the coalescent tree, then we require $\widetilde{\mu}N^{\alpha-1}n/n^{\alpha}$ to be order one; that is $\widetilde{\mu}(N/n)^{\alpha-1}$ must be order one. In the following discussion of mutation rates we are considering an individual to be a single genetic locus rather than a member of the species. In Kong et al. (2012) the mutation rate in humans has been estimated to be around 1.2×10^{-8} per site per generation. In Kumar and Subramanian (2002) looking at mammals, the point mutation rate has been estimated to be around 2.2×10^{-9}. In Pacific cod, a mutation rate estimate around 8.3×10^{-9} has been reported by Canino et al. (2010). A population genetic study, Carr and Marshall (2008), of the mitochondrial DNA of 32 Atlantic cod found 298 polymorphic sites over 15655 sites, and gave an estimated mutation rate of 1.1×10^{-8}. From the mitochondrial DNA study on Atlantic cod, $\widetilde{\mu}$ is estimated to be on the order of 10^{-4}. Given these mutation rate estimates, we see that the closer α is to 1, that is the more extreme the sweepstakes reproduction is, the harder it becomes to obtain the relation $\widetilde{\mu}(N/n)^{\alpha-1} = \mathcal{O}(1)$.

In reality, rather than using the full site frequency spectrum, it is common to use summary statistics. In order to distinguish between population growth and sweepstakes reproduction, Koskela (2018) uses the two-dimensional statistic $(M_1(n), \sum_{j\geq k} M_j(n))$. The statistic $M_1(n)$ is referred to as the singletons in the sample in the literature. The method is not particularly sensitive to the choice of k. The 'lumped tail', $\sum_{j\geq k} M_j(n))$, is a proxy for the internal branch length of the coalescent tree, with the number of singletons reflecting the external branch length. Often we consider a lumped tail for $j \geq k$ rather than $j \geq 2$ as this has proved a more robust analysis. This statistic also provides a means of estimating the parameter α that provides the best fit of data to a Beta-coalescent model.

Previous studies of the impact of large sample size on gene genealogies have been focussed on models with Wright-Fisher sampling as in Wakeley and Takahashi (2003), Fu (2006), Bhaskar et al. (2014), Melfi and Viswanath (2018a) and Melfi and Viswanath (2018b). In order to allow the sample size to exceed the effective population size,

Wakeley and Takahashi (2003) assume that in each generation $N - N_e > 0$ individuals die without reproducing, and then assign N offspring to N_e parents by multinomial sampling with equal weights. Through both analysis and simulation, they demonstrate that a sign of sample size exceeding the effective population size is an excess of singletons in the site frequency spectrum, compared to the predictions of the Kingman coalescent.

In Section 4.5, we provide a heuristic argument to explain why in the model of Wakeley and Takahashi (2003) the genealogy will, with high probability, deviate from the Kingman coalescent as soon as the sample size is of the order of $N_e^{1/3}$. However, in Wakeley and Takahashi (2003), the deviation of their statistic from the predictions of the Kingman coalescent, only becomes evident when the sample size is on the order of the effective population size, at which point it revealed itself as an excess of singletons. We also provide an explanation for why the deviation only becomes evident at such large sample sizes.

There are many possible explanations for an excess of singletons in a site frequency spectrum, including large sample size, population growth, and sweepstakes reproduction. An interesting aspect of the approach of Wakeley and Takahashi (2003) is that for sufficiently large sample size, they are able to estimate the mutation rate per generation ($\widetilde{\mu}$) and the effective population size separately (as opposed to only their product).

In Wakeley and Takahashi (2003) they re-analysed mitochondrial DNA data of introduced populations of Pacific oysters from Boom et al. (1994) and show that under their model, in order to explain the excess of singletons simply through large sample size relative to effective population size, one requires an implausibly large mutation rate. However, this does not necessarily imply that sweepstakes reproduction is the true explanation for the excess. Indeed, multiple merger coalescents reflect strong genetic drift, and a consequent loss of diversity, and it seems very plausible that these models too will require an implausibly high mutation rate to explain observed data.

Elegant asymptotic results, as sample size $n \to \infty$, of branch lengths, and the site-frequency spectrum, have been obtained for various coalescent models. The results can be broadly partitioned into studies of the random coalescent tree sampled at a fixed point in time, the 'static' case, and studies of the evolving coalescent, or the 'dynamic'

case. All these asymptotic results are obtained under the assumption that the relevant coalescent model holds for arbitrarily large sample size.

We write T_k for the random time during which the coalescent process, restricted to $\{1,\dots,n\}$, has exactly k blocks. Under the Kingman-coalescent, the T_k are independent exponentials with rate $k(k-1)/2$. The total tree length $L_n := 2T_2 + \cdots + nT_n$ has clear biological significance, through its relation to the total mutations which can have occurred since the most recent common ancestor and are present in the population. Using Chebyshev's inequality one can show that $L_n/\log(n) \to 2$ in probability, as sample size $n \to \infty$.

If E_n denotes the length of a single external branch in a Kingman coalescent, Caliebe et al. (2007) show that $nE_n \xrightarrow{\mathrm{d}} E$, where E has density $x \mapsto 8/(2+x)^3$ for $x \geq 0$. Let B_k^n denote the random total length of branches supporting $k \in \{1,\dots,n-1\}$ leaves within the coalescent restricted to a sample size of n. As discussed before these are closely related to $M_k(n)$. In Janson and Kersting (2011) they obtain the asymptotic distribution of B_1^n

$$\frac{1}{2}\sqrt{\frac{n}{\log(n)}}(B_1^n - 2) \xrightarrow{\mathrm{d}} N(0,1)$$

where $N(0,1)$ is the standard normal. This is extended to the joint distribution of $(B_1,\dots,B_k)$ under the Kingman coalescent in Dahmer and Kersting (2015)

$$\frac{1}{2}\sqrt{\frac{n}{\log(n)}}\Big(B_1^n - \tfrac{2}{1},\dots,B_k^n - \tfrac{2}{k}\Big) \xrightarrow{\mathrm{d}} N(0,I_k) \tag{4.2.1}$$

where I_k is the $k \times k$ identity matrix. Furthermore, they use the result in (4.2.1) to derive a convergence in distribution result for the site-frequency spectrum:

$$(M_1(n),\dots,M_k(n)) \xrightarrow{\mathrm{d}} (M_1,\dots,M_k)$$

where the M_i are independent Poisson with mean $2\theta/i$. The scaling $\sqrt{n/(4\log n)}$ carries over to the evolving Kingman coalescent analysed in Pfaffelhuber and Wakolbinger (2006) and Pfaffelhuber et al. (2010). Later in Dahmer and Kersting (2017) the total length of the external branches in the evolving Kingman coalescent is considered and

they show that

$$\left(\frac{1}{2}\sqrt{\frac{n}{\log(n)}}(L_{t/n}-2)\right)_{t\in\mathbb{R}}$$

converges in finite-dimensional distributions to a stationary Gaussian process.

Asymptotics for the lengths of branches in a genealogy under multiple-merger coalescents in the static case have been extensively studied, for instance Kersting et al. (2014), Dahmer et al. (2014), Möhle (2010), Berestycki et al. (2007) Berestycki et al. (2008) and references therein. In Berestycki et al. (2008) they consider a Λ-coalescent in particular the Beta-coalescent with $1<\alpha<2$ as a special case, and show that

$$\frac{L_n}{n^{2-\alpha}} \xrightarrow{\text{p}} \frac{\alpha(\alpha-1)}{A(2-\alpha)\Gamma(2-\alpha)},$$

where A is a constant. Furthermore, Berestycki et al. (2007) obtain asymptotic results on the number of blocks of the Beta-coalescent as $t\to 0$. They also obtain convergence results for the site-frequency spectrum under the Beta-coalescent

$$\frac{M_k(n)}{n^{2-\alpha}} \to \theta B\frac{(2-\alpha)\Gamma(k+\alpha-2)}{\Gamma(\alpha-1)k!} \quad \text{almost surely,} \tag{4.2.2}$$

where θ is the scaled mutation rate and $B=\alpha(\alpha-1)/(A\Gamma(2-\alpha)(2-\alpha))$ for some constant A. Then, in Berestycki et al. (2014), they strengthen this to an almost sure convergence. When α is close to 1 Birkner et al. (2013a) show a very large sample size is needed for a good agreement between exact computations of the expected site-frequency spectrum, and the asymptotic results in (4.2.2).

4.3 Beta coalescents and a modified Schweinsberg model

The Beta-coalescents and the coalescents considered by Eldon and Wakeley (2006) are all special cases of a much more general class known as Λ-coalescents, which were constructed independently by Donnelly and Kurtz (1999), Pitman (1999) and Sagitov (1999) and which have since been studied by many authors, see Berestycki (2009) for a survey. Like the Kingman coalescent, they are Markov processes, taking their values among partitions of $\mathbb{N}$. If $\{\pi(t), t\geq 0\}$ is a Λ-coalescent, we write $\{\pi_n(t), t\geq 0\}$ for

its restriction to $[n] = \{1, 2, \ldots, n\}$ for each $n \in \mathbb{N}$. The elements of $[n]$ are used to label individuals in a sample from the population. Each block of the partition $\pi_n(t)$ corresponds to an individual ancestral to the sample at time t before the present, and the elements of the block are the labels of the descendants of that ancestor in the sample. In what follows, we shall abuse terminology and refer to the pure death process that records the total number of lineages ancestral to the sample at each time t before the present as the coalescent. The restriction $\{\pi_n(t), t \geq 0\}$ is itself a Markov process. If there are currently k blocks in $\pi_n(t)$ then each transition involving j of the blocks merging into one happens at rate $b_{k,j}$ (which is *independent* of n) and these are the only possible transitions. The rate of such mergers is not arbitrary; Pitman (1999) shows a one-to-one correspondence between these coalescents and finite measures on $[0, 1]$. For a given finite measure Λ on $[0, 1]$, in the corresponding coalescent, if there are currently k lineages ancestral to the sample, then the rate at which a particular subset of j of them coalesces into a single lineage is

$$b_{k,j} = \int_0^1 x^{j-2}(1-x)^{k-j}\Lambda(\mathrm{d}x), \quad j \in \{2, 3, \ldots, k\}. \tag{4.3.1}$$

A much-studied subfamily of Λ-coalescents is the Beta-coalescent.

Definition 4.3.1 (Beta-coalescent)**.** *The Beta-coalescent with parameter α is the Λ-coalescent in which, if there are currently k lineages ancestral to the sample, the rate at which a particular subset of j of them coalesces into a single lineage is obtained from (4.3.1) by setting*

$$\Lambda(\mathrm{d}x) = \frac{1}{\Gamma(2-\alpha)\Gamma(\alpha)} x^{1-\alpha}(1-x)^{\alpha-1}\mathrm{d}x, \quad 1 \leq \alpha < 2. \tag{4.3.2}$$

In Birkner et al. (2005a) they discuss the connections Beta-coalescents have to continuous-state branching processes. Even though one can take any positive $\alpha < 2$ in (4.3.2), the Beta-coalescent with $\alpha < 1$ does not correspond to any known population model inspired by natural reproduction dynamics. As before, we shall restrict our attention to haploid populations. However an analogue of Schweinsberg's model for diploid populations has been studied in Birkner et al. (2018).

Definition 4.3.2 (Schweinsberg type models)**.** *Consider a haploid population of fixed size N which evolves in discrete generations. In each generation, each individual, independently, produces a random number of juveniles (potential offspring). The N individuals that survive to maturity (and form the next generation) are sampled uniformly at random from the combined pool of all juveniles. If there are fewer than N juveniles from which to sample, we suppose that the population is unchanged over the generation.*

Since we are only interested in models in which the total number of juveniles produced each time will almost always exceed the size of the adult population, N, the specification of what happens when there are fewer than N juveniles will be completely irrelevant. However, we assume the population remains fixed for simplicity.

The coalescence probability will be used to rescale time in population models as is standard in the literature, for instance see Möhle (1998) and Sagitov (1999).

Definition 4.3.3 (The coalescence probability)**.** *Define c_N as the probability that two individuals, chosen uniformly at random without replacement and at the same time from a given population, have a common parent in the previous generation.*

The effective population size, N_e, is defined in different ways within the literature. We will be following Schweinsberg (2003) and take as our definition of N_e, $N_e := 1/c_N$. In a Wright-Fisher population of size N is $1/c_N = N$, and the ancestral process is the Kingman coalescent with time measured in units of N generations. With Schweinsberg's model Schweinsberg (2003) shows in a population of size N, c_N can be much larger than $1/N$.

Theorem 4.3.4 (Schweinsberg (2003))**.** *Suppose that a haploid population of size N evolves according to Definition 4.3.2. Let X^N be the random number of juveniles produced by an individual. Assume that $Ck^\alpha \mathbb{P}[X^N \geq k] \to 1$ as $k \to \infty$ for some $\alpha > 0$ and a normalising constant $C > 0$. Write $\{\pi_{n,N}(t), t \geq 0\}$ for the ancestral process of a sample of size n. Then the scaled ancestral process $\{\pi_{n,N}(\lfloor t/c_N \rfloor), t \geq 0\}$ converges as $N \to \infty$ to a process whose law if $\alpha \geq 2$ is that of the Kingman coalescent restricted to $\{1, \ldots, n\}$ and if $\alpha \in [1, 2)$ is that of the Beta-coalescent (see Definition 4.3.1) restricted to $\{1, \ldots, n\}$.*

Under the assumptions of Theorem 4.3.4, Schweinsberg (2003) shows if $\alpha < 1$ the ancestral process converges to a discrete time analogue of a Ξ-coalescent. The measure Ξ is determined by α and there is no scaling of time. We do not consider this scenario in this work in large part as, given the mutation rates cited above, it is highly non-trivial to explain any observed genetic variation without a rescaling of time.

4.3.1 Model definition

We shall consider a variant of Schweinsberg's model in which the distribution of the number of juvenile offspring produced by each individual can depend upon the total population size. More precisely, writing X^N for the random number of juvenile offspring of an individual in a population of size N, we suppose that $1 < \alpha < 2$ and there exist functions $f(k)$ and $g(k)$ such that

$$g(k)k^{-\alpha}\mathbf{1}_{k\leq\psi(N)} \leq \mathbb{P}[X^N \geq k] \leq f(k)k^{-\alpha}\mathbf{1}_{k\leq\psi(N)}, \tag{4.3.3}$$

and

$$\begin{aligned} g(k) &\to g_\infty > 0, \\ f(k) &\to f_\infty > 0, \end{aligned} \tag{4.3.4}$$

as $k \to \infty$, where g_∞, f_∞ are finite constants. We also assume

$$\mathbb{E}[X^N] =: m_N \to m_\infty > 1, \tag{4.3.5}$$

as $N \to \infty$. For example, one can choose f and g in (4.3.3) such that

$$\mathbb{P}[X^N = k] = \frac{(\psi(N)+1)^\alpha}{(\psi(N)+1)^\alpha - 1}\big(k^{-\alpha} - (k+1)^{-\alpha}\big), \quad 1 \leq k \leq \psi(N). \tag{4.3.6}$$

We shall typically suppress the superscript N and write X_i for the number of juvenile offspring of the ith individual in the population. We write

$$S_N = \sum_{i=1}^{N} X_i$$

for the total number of juveniles produced. Recall that if $S_N < N$, then we simply maintain the original population. As in Lemma 5 in Schweinsberg (2003), $\mathbb{P}[S_N < N]$ is exponentially small in N. We also define

$$\overline{f}(k) = \sup_{i \geq k} f(i),$$
$$\underline{f}(k) = \inf_{i \geq k} f(i),$$

with $\overline{g}$ and $\underline{g}$ similarly defined. We note that, as $g_\infty > 0$, g can be chosen to ensure $\underline{g}(i) > 0$ for all i and we shall assume this throughout.

In order to state the timescaling associated with our model in (4.3.3) it will be convenient to define (as in Lemma 6 from Schweinsberg (2003))

$$\mathcal{C}_N = N\mathbb{E}\left[\frac{X_1(X_1-1)}{S_N^2}\mathbf{1}_{S_N \geq N}\right]. \tag{4.3.7}$$

Later in Lemma 4.6.8 we prove that

$$c_N = \mathcal{C}_N\left(1 + \mathcal{O}\left(\frac{1}{N}\right)\right),$$

where, as before, c_N denotes the probability that two individuals chosen at random from the current generation are derived from the same parent in the previous generation as defined in Definition 4.3.3.

4.4 Results

Our first result identifies $\mathcal{C}_N$ when (4.3.3) holds. Most work in population genetics concerned with deriving coalescent processes ignores the error in the approximation. We carefully control the order of the error in all our approximations, in contrast to Schweinsberg (2003). This provides some insight into the variability that we can expect in estimates of the effective population size under such a model. Our results highlight the importance of how the truncation $\psi(N)$ depends on N.

Proposition 4.4.1. *Suppose that the population evolves according to the Schweinsberg model with juvenile offspring distribution satisfying* (4.3.3)*. In each of the statements*

below, L is a function of N, over which one can optimise for any specific choice of model.

1. *Suppose that $f_\infty = g_\infty$, $\psi(N)/N \to 0$, and $L(N)/\psi(N) \to 0$ as $N \to \infty$, then*

$$\frac{N}{\psi(N)^{2-\alpha}}\mathcal{C}_N = \frac{2f_\infty}{(2-\alpha)m_N^2} + \mathcal{O}(|\overline{f}(L) - \underline{g}(L)|) + \mathcal{O}\left(\frac{L^{2-\alpha}}{\psi(N)^{2-\alpha}}\right) + \mathcal{O}\left(\frac{\psi(N)}{N}\right) + \mathcal{O}\left(\frac{\psi(N)^{\frac{2-\alpha}{3}}}{N^{\frac{1}{3}}}\right). \quad (4.4.1)$$

2. *Suppose that $f_\infty = g_\infty$, $\psi(N)/N \to K$, and $L(N)/N \to 0$ as $N \to \infty$, then*

$$N^{\alpha-1}\mathcal{C}_N = f_\infty \frac{2}{m_N^\alpha}\int_{\frac{m_N}{K+m_N}}^1 y^\alpha(1-y)^{1-\alpha}\mathrm{d}y + +\mathcal{O}(|\overline{f}(L) - \underline{g}(L)|) + \mathcal{O}\left(\left|\frac{\psi(N)}{N} - K\right|\right) + \mathcal{O}((L/N)^{2-\alpha}) + \mathcal{O}\left(\frac{1}{N^{\frac{\alpha-1}{3}}}\right). \quad (4.4.2)$$

3. *Suppose that $f_\infty = g_\infty$, $\psi(N)/N \to \infty$ and $L(N)/N \to 0$ as $N \to \infty$, then*

$$N^{\alpha-1}\mathcal{C}_N = f_\infty \frac{2}{m_N^\alpha}\int_0^1 y^\alpha(1-y)^{1-\alpha}\mathrm{d}y + \mathcal{O}(|\overline{f}(L) - \underline{g}(L)|) + \mathcal{O}\Big((N/\psi(N))^{1+\alpha}\Big) + \mathcal{O}((L/N)^{2-\alpha}) + \mathcal{O}\Big(N^{\frac{1-\alpha}{1+\alpha}}\Big). \quad (4.4.3)$$

These results are kept in a general form, however, the parts involving L occur due to the flexibility of our model. In particular the relationship between the functions f and g. Similarly, parts involving Ψ also appear due to the flexibility in our cut off. The remaining factor will tend to be key when considering specific models. We will prove Proposition 4.4.1 in Section 4.6.2. We will first introduce Proposition 4.6.3 which includes an extra function, which can be optimised over to give Proposition 4.4.1 as a simple corollary.

The results in Proposition 4.4.1 highlight the importance of keeping track of the errors. Although in case 1 the limiting process is the Kingman coalescent (4.4.1) reveals that the timescale depends on the truncation. In case 2, (4.4.2) shows the effect of the truncation on the leading term, which is a truncated beta function. If we let $\psi(N) = \mathcal{O}(N \log N)$, (4.4.3) shows that convergence can be extremely slow.

We next identify the limiting coalescent for fixed sample sizes, again maintaining tight control over the errors.

Remark 4.4.1. *In Schweinsberg (2003), when $\alpha \geq 2$ and without a cutoff, they show convergence to the Kingman coalescent depends on the small families as the large families are sufficiently rare. Therefore, for any cutoff, $\Psi(N) \geq 2$, when $\alpha \geq 2$ we will simply have convergence to the Kingman coalescent.*

Proposition 4.4.2. *In Proposition 4.4.1 Case 1, the scaled ancestral process $\{\pi_{n,N}(\lfloor t/c_N \rfloor), t \geq 0\}$ corresponding to a sample of size n converges to a process whose law is that of the Kingman coalescent restricted to $\{1, \ldots, n\}$. Moreover*

$$\frac{N}{c_N}\mathbb{E}\left[\frac{(X_1)_3}{S_N^3}\mathbf{1}_{S_N \geq N}\right] = \mathcal{O}\left(\frac{\psi(N)}{N}\right), \tag{4.4.4}$$

$$\frac{N^2}{c_N}\mathbb{E}\left[\frac{(X_1)_2(X_2)_2}{S_N^4}\mathbf{1}_{S_N \geq N}\right] = \mathcal{O}\left(\frac{\psi(N)^{2-\alpha}}{N}\right). \tag{4.4.5}$$

If you take $\psi(N) = N/\log(N)$ then $\psi(N)/N \to 0$ very slowly, and Proposition 4.4.2 implies that for almost any real population the probability of multiple mergers is not negligible even though the ancestral process converges to the Kingman coalescent. We will prove Proposition 4.4.2 in Section 4.6.3. We would like to emphasise at this point that the convergence to Kingman's coalescent is driven by the cutoff function. In particular we see that this cutoff ensures families are sufficiently small so we will not see multiple mergers.

We now specify the limiting coalescent processes for the cases in which Beta-coalescents are obtained.

Proposition 4.4.3. *In Proposition 4.4.1 Case 2, the scaled ancestral process $\{\Psi_{n,N}(\lfloor t/c_N \rfloor), t \geq 0\}$ corresponding to a sample of size n converges to a process whose law is given by a Λ-coalescent restricted to $\{1, \ldots, n\}$, with*

$$\Lambda(\mathrm{d}x) = \frac{1}{B(\frac{K}{m_\infty + K}, 2-\alpha, \alpha)} x^{1-\alpha}(1-x)^{\alpha-1}\mathbf{1}_{x \leq \frac{K}{m_\infty + K}}\mathrm{d}x,$$

where we have used the notation

$$B\Big(\frac{K}{m_\infty + K}, 2-\alpha, \alpha\Big) = \int_0^{\frac{K}{m_\infty+K}} x^{1-\alpha}(1-x)^{\alpha-1}\mathrm{d}x,$$

and m_∞ was defined in (4.3.5)*.*

In Proposition 4.4.1 Case 3, the scaled ancestral process $\{\Psi_{n,N}(\lfloor t/c_N \rfloor), t \geq 0\}$ corresponding to a sample of size n converges to a process whose law is given by a Λ-coalescent restricted to $\{1, \ldots, n\}$, with

$$\Lambda(\mathrm{d}x) = \frac{1}{B(2-\alpha, \alpha)} x^{1-\alpha}(1-x)^{\alpha-1}\mathrm{d}x,$$

where

$$B(2-\alpha, \alpha) = \int_0^1 x^{1-\alpha}(1-x)^{\alpha-1}\mathrm{d}x = \Gamma(\alpha)\Gamma(2\alpha-1).$$

Moreover, in both cases,

$$\frac{N^2}{c_N}\mathbb{E}\left[\frac{(X_1)_2(X_2)_2}{S_N^4}\mathbf{1}_{S_N \geq N}\right] = \mathcal{O}\Big(N^{1-\alpha}\Big). \tag{4.4.6}$$

We will prove Proposition 4.4.3 in Section 4.6.4. We note that as $\alpha \downarrow 1$ Proposition 4.4.3 implies that the error in ignoring simultaneous mergers increases, when the limit is a Beta-coalescent. This is natural as, formally, by taking $\alpha = 1$ Schweinsberg (2003) shows that the limit will be a Ξ-coalescent rather than a Beta-coalescent.

The results in Proposition 4.4.1, 4.4.2, and 4.4.3 do not relate to sample size; they only pertain to the 'bare-bones' population model of genetic reproduction as given in (4.3.3). In particular the convergence results hold for a fixed sample size as $N \to \infty$. In Section 4.7 we give arguments for the sample size needed to impact convergence or inference. We present the key observations here.

Conjecture 4.4.2. *Suppose model* (4.3.3) *holds. In the case $f_\infty = g_\infty$ and $\psi(N)/N \to 0$, if the sample size $n \geq \mathcal{O}(N/\psi(N))$, the total probability of 3-mergers is non-negligible compared to the total probability of 2-mergers.*

In Section 4.7.1 we will give arguments for Conjecture 4.4.2. Then in Section 4.7.2 we consider the sample size required to invalidate convergence to a Beta-coalescent. A proof of Lemma 4.7.10 is given in Section 4.7.2.

Lemma 4.4.3. *Suppose model* (4.3.3) *holds,* $n/N \to 0$, $f_\infty = g_\infty$, $n \geq k$, *let* $K > 0$ *be fixed, and write*

$$r_N(k) := \mathbb{E}\left[\mathbf{1}_{S_N \geq N}\binom{X_1}{k}\binom{S_N - X_1}{n-k}/\binom{S_N}{n}\right].$$

1. *When* $\psi(N)/N \to K$ *then*

$$\begin{aligned} Nr_N(k) = N^{1-\alpha} f_\infty m_N^{-\alpha}\binom{n}{k}\Bigg(k\int_0^{\frac{K}{K+m_N}} u^{k-1-\alpha}(1-u)^{n-k+\alpha}\mathrm{d}u \\ -(n-k)\int_0^{\frac{K}{K+m_N}} u^{k-\alpha}(1-u)^{n-k+\alpha-1}\mathrm{d}u\Bigg)\Bigg(1 + \mathcal{O}(|\overline{f}(L) - \underline{g}(L)|) + \mathcal{O}(\beta_N/N) \\ + \mathcal{O}\Big(\Big|\frac{\psi(N)}{N} - K\Big|\Big) + \mathcal{O}\Big((L/N)^{k-\alpha}\Big) + \mathcal{O}(k^2/L) + \mathcal{O}\Big(N\psi(N)^{2-\alpha}/\beta_N^2\Big)\Bigg), \end{aligned} \tag{4.4.7}$$

   when $L/M \to 0$.

2. *When* $\psi(N)/N \to \infty$ *then*

$$\begin{aligned} Nr_N(k) = N^{1-\alpha} f_\infty m_N^{-\alpha}\binom{n}{k}\Bigg(k\int_0^{1} u^{k-1-\alpha}(1-u)^{n-k+\alpha}\mathrm{d}u \\ -(n-k)\int_0^{1} u^{k-\alpha}(1-u)^{n-k+\alpha-1}\mathrm{d}u\Bigg)\Bigg(1 + \mathcal{O}(|\overline{f}(L) - \underline{g}(L)|) + \mathcal{O}(\beta_N/N) \\ + \mathcal{O}\Big((N/\psi(N))^{n-k+\alpha+1}\Big) + \mathcal{O}\Big((L/N)^{k-\alpha}\Big) + \mathcal{O}(k^2/L) + \mathcal{O}(N/\beta_N^\alpha)\Bigg). \end{aligned} \tag{4.4.8}$$

This agrees with our results in Proposition 4.6.3 and we can see that if we cannot choose L such that both $k^2/L \to 0$ and $L/N \to 0$ we will not expect to see convergence to the limiting rates suggested by Proposition 4.4.3. Recall that L is an arbitrary function of N (see Proposition 4.4.1). Therefore, the requirement to have convergence of rates becomes $k^2/N \to 0$. In particular the error term $\mathcal{O}(k^2/L)$ is shown in the proof to always be negative. Therefore showing that when the limiting Lambda coalescent sees

mergers of at least $\sqrt{N}$ lineages we will be seeing a reduced number of these mergers and so a different topology in the coalescent tree.

4.5 Large sample size

We wish to establish how large a sample size we can take before the errors that we make in using the limiting coalescent processes that we have identified to estimate the distribution of the site frequency spectrum become significant. Of particular interest is the comparison between the proportion of singletons and the 'lumped tail' of the spectrum used by Koskela (2018). Since mutation rates are not infinite these statistics will not, in particular, be able to differentiate between mergers that happen instantaneously and those which occur in quick succession, if both lead to the same topology for the genealogical tree. In what follows we will will refer to events where two or more separate mergers happen instantaneously as simultaneous mergers.

4.5.1 Large Sample Size in a modified Wright-Fisher Model

We begin with a brief discussion of the modified Wright-Fisher Model considered by Wakeley and Takahashi (2003), which is closely related to the Wright-Fisher model discussed in Section 1.1.1. The population is assumed to be of constant size, N, and evolve in discrete generations. In each generation, first sample N_e 'potential' parents, uniformly at random, from the population; and then determine the family sizes of those individuals by multinomial sampling with equal weights. In other words, each of the N individuals in the offspring generation chooses a parent uniformly at random from among the subset of N_e available to reproduce.

When $N_e = N$, the model reduces to the Wright-Fisher model, and a detailed analysis of the coalescent for large samples is then provided by Melfi and Viswanath (2017). Here we satisfy ourselves with some heuristics. Two types of events can cause deviations from the predictions of the Kingman coalescent: simultaneous mergers and multiple mergers. The probability of seeing two pairs of lineages within a sample of size k merge (into two separate parents) in a single generation is approximately $2\binom{k}{4}\frac{N_e-1}{N_e^3}$. Similarly, the probability of seeing a 3-merger in a single generation within a sample

size of k is approximately $\binom{k}{3}\frac{1}{N_e^2}$. If neither of these events happens, the number of generations that it takes for the number of ancestral lineages to change is about $N_e/\binom{k}{2}$, so the chance that we see a simultaneous merger before we see our first coalescence event is of order $\frac{N_e}{k^2}\frac{k^4}{N_e^2} = \frac{k^2}{N_e}$. To estimate the chance that we see a simultaneous merger before the most recent common ancestor of a sample of size n, we sum over $k \leq n$ to find that the probability is of order $\mathcal{O}(\frac{n^3}{N_e})$. In other words, if n is at least on the order of $N_e^{1/3}$ there is an appreciable probability that we will see the simultaneous merger of two pairs of lineages somewhere in the genealogical tree.

We can use the same heuristics to find the order of sample size at which we expect to start seeing 3-mergers in the genealogy. The chance of a 3-merger before either a pairwise merger (or two or more simultaneous pairwise mergers) when the sample size is k is of the order $C\frac{N_e}{k^2}\frac{k^3}{N_e^2} = \frac{k}{N_e}$. Summing over k as before, we expect to start seeing multiple mergers somewhere in the genealogy, once n is at least $\mathcal{O}(\sqrt{N_e})$.

In fact, following Melfi and Viswanath (2017), one can ask how big a sample is allowed if we only want to see pairwise mergers in the genealogy with at most c pairs merging at once, and entirely analogous heuristics yield that this breaks down when the sample exceeds $\mathcal{O}(n^{c/(2c+1)})$.

Although these arguments suggest that we can expect the genealogy of a sample to differ from that predicted by the Kingman coalescent once the sample is at least $\mathcal{O}(N_e^{1/3})$, the results of Wakeley and Takahashi (2003) illustrate that this is not evident in the summary statistic that compares singletons to total number of segregating sites in the site frequency spectrum. Indeed, this only shows significant deviation from the predictions of the Kingman coalescent when the sample size is of the same order as the effective population size, N_e. It has already been observed by Bhaskar et al. (2014) that simultaneous pairwise mergers may not be too disruptive. To see why, notice that when the number of ancestral lineages is large, and so coalescence events are frequent, we can only hope to detect the difference between two mergers happening at the same time, and two mergers in quick succession, if those mergers lead to different tree topologies; that is if, in the second scenario, the parental lineage of the first merger is also involved in the subsequent merger. If there are currently k ancestral lineages, this happens with probability on the order of $1/k$, and our heuristic above says that we will see such

events with a frequency of the same order as triple mergers. Even if this happens, our rather coarse summary statistic is only likely to be changed if such anomalous events make up a significant proportion of coalescence events; that is they must occur on a comparable timescale to simple pairwise mergers. This requires $k^3/N_e^2 \sim k^2/N_e$, which in turn requires sample sizes on the order of N_e.

Even when the limiting coalescent is a Kingman coalescent, the modified Schweinsberg model that we consider here behaves rather differently from that of Wakeley and Takahashi (2003). Under the conditions of Proposition 4.4.2, for sample size k, we see simultaneous mergers at rate

$$\frac{N^2\mathbb{E}\left[\frac{(X_1)_2(X_2)_2}{S_N^4}\mathbf{1}_{S_N\geq N}\right]\frac{1}{2}\binom{k}{4}}{N\mathbb{E}\left[\frac{(X_1)_2}{S_N^2}\mathbf{1}_{S_N\geq N}\right]\binom{k}{2}} \sim \frac{k^4\psi(N)^{2-\alpha}/N}{k^2}.$$

If we define $N_e = 1/c_N$, this is of order k^2/N_e, just as in the modified Wright-Fisher example above. On the other hand, we see 3-mergers at rate

$$\frac{N\mathbb{E}\left[\frac{(X_1)_3}{S_N^3}\mathbf{1}_{S_N\geq N}\right]\binom{k}{3}}{N\mathbb{E}\left[\frac{(X_1)_2}{S_N^2}\mathbf{1}_{S_N\geq N}\right]\binom{k}{2}} \sim \frac{k^3\psi(N)/N}{k^2} = k\frac{\psi(N)}{N}. \tag{4.5.1}$$

If $\psi(N)$ is unbounded, this is significantly bigger than the k/N_e that we saw in the modified Wright-Fisher model. This is not surprising perhaps, but then suggests that the Kingman approximation for this model will be poor at sample sizes closer to $N/\psi(N)$, which could be substantially smaller than N_e.

4.5.2 Large Sample Size in the Lambda Coalescent

We now turn our attention to the cases in which the limiting coalescent is a Λ-coalescent, that is under the conditions of Proposition 4.4.3. Once again there are two possible ways in which our coalescent may not be well approximated by the limiting coalescent: first it may contain simultaneous mergers; second the rate of multiple mergers may be poorly approximated.

Consider first the rate of simultaneous mergers. As before, we wish to compare to the time-scale of all events which for a Beta coalescent with k lineages is of order $c_N k^\alpha$

(instead of $c_N k^2$ in the Kingman case). Our heuristic argument then suggests that the probability that the next event will involve a simultaneous merger, if the current sample size is k, is approximately

$$\frac{N^2\mathbb{E}\left[\frac{(X_1)_2(X_2)_2}{S_N^4}\mathbf{1}_{S_N\geq N}\right]\frac{1}{2}\binom{k}{4}}{N\mathbb{E}\left[\frac{(X_1)_2}{S_N^2}\mathbf{1}_{S_N\geq N}\right]k^\alpha} \sim k^{4-\alpha}N^{1-\alpha}. \tag{4.5.2}$$

The denominator in the ratio on the left is just the rate at which we expect to see jumps in the limiting coalescent, while the numerator is the probability that we see a simultaneous merger of two pairs of lineages in a single generation. Unlike in the Kingman coalescent, in which the number of ancestral lineages simply reduces by one at each merger, we don't have an explicit expression for the jump chain of a Beta-coalescent. However, we know from our calculations of Section 4.7.2, that it requires a sample of size $n = \mathcal{O}(N^{\alpha/2})$ before we see a merger of size $\sqrt{N}$, and the same argument gives that, in our sample of size n, we cannot expect to see mergers of more than $n^{1/\alpha}$ ancestral lineages. This tells us that an event consisting of two simultaneous mergers, one of k lineages and one of l lineages, say, will, with high probability, not be distinguished in the site frequency spectrum from a k merger rapidly followed by an l merger, as (just as in the Kingman case) the lineages involved are very unlikely to overlap. Moreover, our calculations in Section 4.7.2 show for sample sizes less than $\mathcal{O}(N^{\alpha/2})$, the rate of k mergers is correctly predicted by the limiting coalescent for all possible k, and since our coalescent must be consistent if we subsample, there cannot be unusual clusters of mergers of particular sizes. In conclusion, arguing as for the Kingman coalescent, we do not expect to see any change in the tree topology for sample sizes less than $\mathcal{O}(N^{\alpha/2})$, and nor do we expect to see measurable differences in our statistic of the site frequency spectrum.

In order to see when the topology of the tree will be affected we must look at the rates of k mergers in a sample of n lineages. Therefore, we consider the probability that a particular set of k ancestral lineages out of a sample of n converges in a single generation, and compare to the predictions of the Λ-coalescent limit. In Section 4.7.2, we find that the limiting coalescent consistently overestimates the rate of k-mergers

and that this becomes significant once sample sizes are $\mathcal{O}(N^{\alpha/2})$. If we follow Koskela (2018) in using the statistic $\left(M_1(n), \sum_{j\geq k} M_j(n)\right)$ to estimate the parameter α for such sample sizes, then we expect to underestimate α. The explanation is that the ratio of external to total branches in the beta coalescent converges to $2-\alpha$ as the sample size tends to infinity. Substituting actual rates instead of those predicted by the limiting coalescent has the effect of preventing coalescence of lineages, which only serves to extend the external branches. However, we are using the same c_N which depends only on pairwise mergers. Consequently we will find a higher value for the ratio of external branch length to total branch length, which will lead us to predict a smaller α.

4.6 Convergence proofs

4.6.1 Key Lemmas

First we recall some key lemmas. Our model is easily recast as a Cannings model, enabling us to make use of the work of Möhle and Sagitov (2001). For completeness, but following Schweinsberg (2003), we record the results that we need here. First we fix notation by describing the Cannings model for reproduction in a haploid population of constant size N. Again the population evolves in discrete generations. In each generation, writing ν_k for the number of *adult* offspring of the kth individual in the parental generation, $\underline{\nu} \equiv (\nu_1, \nu_2, \ldots, \nu_N)$ is an exchangeable random vector with $\sum_{k=1}^{N} \nu_k = N$. We assume that these vectors are i.i.d. across generations.

Recalling our notation c_N for the probability that two individuals chosen at random from the current generation are derived from the same parent in the previous generation, we see that for the Cannings model

$$c_N = \frac{\mathbb{E}[\nu_1(\nu_1 - 1)]}{N-1}. \tag{4.6.1}$$

We now record general criteria for the convergence of the corresponding rescaled ancestral processes. The first, due to Möhle (2000), guarantees convergence to a Kingman coalescent. Recall the standard notation for descending factorials: $(\theta)_k = \theta(\theta-1)(\theta-2)\cdots(\theta-k+1)$.

Proposition 4.6.1 (Möhle (2000), section 4)**.** *Suppose* $c_N \to 0$ *and*

$$\lim_{N\to\infty} \frac{\mathbb{E}[(\nu_1)_3]}{N^2 c_N} = 0. \tag{4.6.2}$$

Then, for each finite sample size $n \in \mathbb{N}$*, the rescaled ancestral processes* $(\pi_{n,N}(\lfloor t/c_N \rfloor), t \geq 0)$ *converge as* $N \to \infty$ *to the Kingman coalescent restricted to* $\{1, \ldots, n\}$*.*

If the family-size distribution is such that on the timescale determined by c_N we see 'large families' (comprising a non-trivial proportion of the total population) then we recover a coalescent with multiple mergers. The proposition below follows from Theorem 3.1 of Sagitov (1999).

Proposition 4.6.2 (Sagitov (1999))**.** *Suppose*

$$\lim_{N\to\infty} c_N = 0,$$

and

$$\lim_{N\to\infty} \frac{\mathbb{E}[(\nu_1)_2(\nu_2)_2]}{N^2 c_N} = 0. \tag{4.6.3}$$

If there exists a probability measure Λ *on* $[0,1]$ *such that*

$$\lim_{N\to\infty} \frac{N}{c_N} \mathbb{P}[\nu_1 > Nx] = \int_x^1 y^{-2} \Lambda(dy), \tag{4.6.4}$$

for all $x \in (0,1)$*, then, for each finite sample size* $n \in \mathbb{N}$*, the rescaled ancestral processes* $\left(\pi_{n,N}(\lfloor \frac{t}{c_N} \rfloor), t \geq 0\right)$ *converge as* $N \to \infty$ *to the* Λ*-coalescent restricted to* $[n]$*.*

Let us pause to consider what the conditions of Propositions 4.6.1 and 4.6.2 guarantee. First observe that

$$\frac{N\mathbb{E}[(\nu_1)_3]}{(N)_3} = \frac{\mathbb{E}[(\nu_1)_3]}{N^2}\left(1 + \mathcal{O}(\frac{1}{N})\right), \tag{4.6.5}$$

is the probability that a given sample of three ancestral lineages merge in a single generation, whereas

$$c_N = \frac{N\mathbb{E}[(\nu_1)_2]}{(N)_2},$$

is the probability that a given pair of lineages merge. Thus, at least for small samples, equation (4.6.2), which says that the ratio of these two probabilities tends to zero, guarantees that we will not see any mergers of three or more lineages before all lineages have reached a common ancestor through pairwise mergers. However, if the sample size is n, then initially the rate of pairwise mergers is $\binom{n}{2}c_N$ and of triple mergers is $\binom{n}{3}$ times the expression in (4.6.5). Thus if n is large enough that

$$\frac{n\mathbb{E}[(\nu_1)_3]}{N^2 c_N} = \mathcal{O}(1),$$

this conclusion is no longer valid.

The expression

$$\frac{(N)_2\mathbb{E}[(\nu_1)_2(\nu_2)_2]}{(N)_4},$$

is the probability that two (given) distinct pairs of lineages merge (pairwise) in a single generation and so, in the same way, condition (4.6.3) guarantees that for small samples we see no *simultaneous* mergers on the timescale of the coalescent. Möhle (2000) checks that equation (4.6.2) implies (4.6.3) and so under the conditions of Proposition 4.6.1, we expect convergence to the Kingman coalescent. Similarly, under the conditions of Proposition 4.6.2, although we will see multiple mergers, we do not expect to see simultaneous multiple mergers and so the limiting ancestral process is expected to be a Λ-coalescent. Once again, these conclusions will fail as soon as

$$\frac{n^2\mathbb{E}[(\nu_1)_2(\nu_2)_2]}{N^2 c_N} = \mathcal{O}(1).$$

4.6.2 Asymptotic rate of the coalescent: calculating c_N

We recall the definition of

$$\mathcal{C}_N = N\mathbb{E}\left[\frac{X_1^N(X_1^N-1)}{S_N^2}\mathbf{1}_{S_N\geq N}\right]. \tag{4.6.6}$$

4.6.2.1 Preliminary lemmas

Following the ideas from Schweinsberg (2003), Lemma 12, we present a supplementary lemma.

Lemma 4.6.4. *For $1 < \alpha < 2$, $M \in \mathbb{R}$ and any $L(N) \leq \psi(N)$*

$$\begin{aligned}
\mathbb{E}\left[\frac{X_1^N(X_1^N-1)}{(X_1^N+M)^2}\right] \leq & \overline{f}(1)\frac{2M+1}{M^{1+\alpha}}\int_{\frac{M}{L-1+M}}^{1} y^\alpha(1-y)^{1-\alpha}\mathrm{d}y \\
& + \overline{f}(L)\frac{2M+1}{M^{1+\alpha}}\int_{\frac{M}{\psi(N)-1+M}}^{\frac{M}{L-2+M}} y^\alpha(1-y)^{1-\alpha}\mathrm{d}y, \\
\mathbb{E}\left[\frac{X_1^N(X_1^N-1)}{(X_1^N+M)^2}\right] \geq & \underline{g}(1)\frac{2M}{(M+1)^{1+\alpha}}\int_{\frac{M+1}{L+M+1}}^{\frac{M+1}{2+M}} y^\alpha(1-y)^{1-\alpha}\mathrm{d}y \\
& + \underline{g}(L)\frac{2M}{(M+1)^{1+\alpha}}\int_{\frac{M+1}{\psi(N)+M+1}}^{\frac{M+1}{L-1+M+1}} y^\alpha(1-y)^{1-\alpha}\mathrm{d}y.
\end{aligned}$$

Proof First we note

$$\begin{aligned}
\mathbb{E}\left[\frac{X_1^N(X_1^N-1)}{(X_1^N+M)^2}\right] &= \sum_{k=1}^{\infty}\left(\frac{k(k-1)}{(k+M)^2} - \frac{(k-1)(k-2)}{(k-1+M)^2}\right)\mathbb{P}[X_1^N \geq k] \\
&= \sum_{k=1}^{\psi(N)}(k-1)\left(\frac{2kM+k+2M^2}{(k+M)^2(k-1+M)^2}\right)\mathbb{P}[X_1^N \geq k]. \quad (4.6.7)
\end{aligned}$$

We note that

$$\begin{aligned}
(k-1)\left(\frac{2kM+k+2M^2}{(k+M)^2(k-1+M)^2}\right) &\leq (k-1)\left(\frac{(k+M)(2M+1)}{(k+M)^2(k-1+M)^2}\right) \\
&\leq \left(\frac{(k-1)(2M+1)}{(k-1+M)^3}\right),
\end{aligned}$$

and

$$\begin{aligned}
(k-1)\left(\frac{2kM+k+2M^2}{(k+M)^2(k-1+M)^2}\right) &\geq (k-1)\left(\frac{(k+M)2M}{(k+M)^2(k-1+M)^2}\right) \\
&\geq \frac{2M(k-1)}{(k-1+M+1)^3}.
\end{aligned}$$

We observe that, when $\alpha \geq 1$,

$$\frac{k}{(k+M)^3}\frac{1}{k^\alpha},$$

is decreasing for positive k. We then see that

$$\int_a^{b+1}\frac{x^{1-\alpha}}{(x+M)^3}\mathrm{d}x \leq \sum_{k=a}^{b}\frac{k}{(k+M)^3}\frac{1}{k^\alpha} \leq \int_{a-1}^{b}\frac{x^{1-\alpha}}{(x+M)^3}\mathrm{d}x.$$

We finally note

$$\int_a^b \frac{x^{1-\alpha}}{(x+M)^3}\mathrm{d}x = \int_{\frac{M}{b+M}}^{\frac{M}{a+M}} \left(\frac{M(1-y)}{y}\right)^{1-\alpha}\left(\frac{y}{M}\right)^3 My^{-2}\mathrm{d}y$$
$$= \frac{1}{M^{1+\alpha}}\int_{\frac{M}{b+M}}^{\frac{M}{a+M}} y^\alpha(1-y)^{1-\alpha}\mathrm{d}y.$$

Splitting the sum appearing in (4.6.7) into two parts, $k < L$ and $L < k < \psi(N)$ and exploiting these observations completes the proof. □

Proposition 4.4.1 will follow from the proposition below as an easy corollary by optimising the choice of β_N below.

Proposition 4.6.3. *Suppose that the population evolves according to the Schweinsberg model with juvenile offspring distribution satisfying* (4.3.3)*. Let β_N be any increasing sequence of real numbers with $\beta_N \to \infty$ as $N \to \infty$.*

1. *When $f_\infty = g_\infty$ and $\psi(N)/N \to 0$ then*

$$\frac{N}{\psi(N)^{2-\alpha}}\mathcal{C}_N = \frac{2f_\infty}{(2-\alpha)m_N^2} + \mathcal{O}(|\overline{f}(L) - \underline{g}(L)|) + \mathcal{O}\left(\frac{L^{2-\alpha}}{\psi(N)^{2-\alpha}}\right)$$
$$+ \mathcal{O}\left(\frac{\beta_N}{N}\right) + \mathcal{O}\left(\frac{\psi(N)}{N}\right) + \mathcal{O}\left(\frac{N\psi(N)^{2-\alpha}}{\beta_N^2}\right),$$

   where L is any function of N with $L(N)/\psi(N) \to 0$ as $N \to \infty$.

2. *When $f_\infty = g_\infty$ and $\psi(N)/N \to K$ then*

$$N^{\alpha-1}\mathcal{C}_N = f_\infty\frac{2}{m_N^\alpha}\int_{\frac{m_N}{K+m_N}}^1 y^\alpha(1-y)^{1-\alpha}\mathrm{d}y + +\mathcal{O}(|\overline{f}(L) - \underline{g}(L)|) + \mathcal{O}(\beta_N/N)$$
$$+ \mathcal{O}\left(\left|\frac{\psi(N)}{N} - K\right|\right) + \mathcal{O}((L/N)^{2-\alpha}) + \mathcal{O}\Big(N\psi(N)^{2-\alpha}/\beta_N^2\Big),$$

   where L is any function of N with $L(N)/N \to 0$ as $N \to \infty$.

3. *When $f_\infty = g_\infty$ and $\psi(N)/N \to \infty$ then*

$$N^{\alpha-1}\mathcal{C}_N = f_\infty\frac{2}{m_N^\alpha}\int_0^1 y^\alpha(1-y)^{1-\alpha}\mathrm{d}y + \mathcal{O}(|\overline{f}(L) - \underline{g}(L)|) + \mathcal{O}(\beta_N/N)$$
$$+ \mathcal{O}\Big((N/\psi(N))^{1+\alpha}\Big) + \mathcal{O}((L/N)^{2-\alpha}) + \mathcal{O}(N/\beta_N^\alpha),$$

where L is any function of N with $L(N)/N \to 0$ as $N \to \infty$.

Remark 4.6.5. *To reduce the constant order term in Case 3 to a more familiar form, note that*

$$2\int_0^1 y^\alpha (1-y)^{1-\alpha} \mathrm{d}y = 2\frac{\Gamma(\alpha+1)\Gamma(2-\alpha)}{\Gamma(3)} = \alpha\Gamma(\alpha)\Gamma(2-\alpha).$$

Proof[Proof of Proposition 4.4.1]

This now follows immediately from Proposition 4.6.3 on setting $\beta_N = \psi(N)^{\frac{2-\alpha}{3}}/N^{\frac{1}{3}}$ in Case 1, $\beta_N = N^{\frac{4-\alpha}{3}}$ in Case 2, and $\beta_N = N^{\frac{2}{1+\alpha}}$ in Case 3. □

Notation 4.6.4. *As we are interested in the behaviour of $N\mathbb{E}\left[\frac{X_1(X_1-1)}{S_N^2}\mathbf{1}_{S_N \geq N}\right]$, it is convenient to define $\widetilde{S}_N = \sum_{i=2}^N X_i$. We shall be conditioning on events*

$$A_N := \{|\widetilde{S}_N - (N-1)m_N| \leq \beta_N\}, \tag{4.6.8}$$

where β_N is an increasing sequence and we recall from (4.3.5) that $m_N = \mathbb{E}[X_i^N]$. We can see, conditioned on A_N, that

$$M_- \leq \widetilde{S}_N \leq M_+,$$

where

$$\begin{aligned} M_- =& (N-1)m_N - \beta_N, \\ M_+ =& (N-1)m_N + \beta_N. \end{aligned}$$

We shall use the notation M_- and M_+ throughout our calculations. In several lemmas below, we use M to denote an arbitrary constant. When we do this, we are intending to use the result with M replaced by either M_+ or M_-.

To prove Proposition 4.6.3 we shall partition over the events A_N, A_N^c. Our main focus will be the value, for large N, of

$$\mathbb{E}\left[\frac{X_1(X_1-1)}{S_N^2}\mathbf{1}_{S_N \geq N}\middle| A_N\right] = \mathbb{E}\left[\frac{X_1(X_1-1)}{(X_1+\widetilde{S}_N)^2}\middle| A_N\right],$$

under the assumption that $\beta_N/N \to 0$ as $N \to \infty$. The result will then follow if we prove that

$$N\mathbb{E}\left[\frac{X_1(X_1-1)}{S_N^2}\mathbf{1}_{S_N \geq N}\middle| A_N^c\right]\mathbb{P}[A_N^c],$$

is sufficiently small.

Lemma 4.6.6. *In the notation of Proposition 4.6.3 and Notation 4.6.4,*

$$\mathbb{P}[|\widetilde{S}_N - (N-1)m_N| \geq \beta_N] = \mathcal{O}\left(\frac{N}{\beta_N^\alpha}\right).$$

Remark 4.6.7. *Formally, if we take $\psi(N) = \infty$ the rate in Lemma 4.6.6 is optimal as X will then lie in the domain of attraction of a stable law of index α and so we expect*

$$\mathbb{P}[|S_N - Nm_N| \geq \beta_N] = \mathbb{P}\left[\left|\frac{S_N - Nm_N}{N^{1/\alpha}}\right| \geq \frac{\beta_N}{N^{1/\alpha}}\right] \sim C\frac{N}{\beta_N^\alpha},$$

for large N and some constant C.

Proof[Proof of Lemma 4.6.6] Define

$$B_N := \{X_i \leq \beta_N \ \forall i\},$$

and observe that

$$\begin{aligned}
&\mathbb{P}[|\widetilde{S}_N - (N-1)m_N| \geq \beta_N] \\
&\quad = \mathbb{P}\Big[\Big(\mathbf{1}_{B_N}|\widetilde{S}_N - (N-1)m_N| + \mathbf{1}_{B_N^c}|\widetilde{S}_N - (N-1)m_N|\Big) \geq \beta_N\Big] \\
&\quad \leq \mathbb{P}\left[\mathbf{1}_{B_N}|\widetilde{S}_N - (N-1)m_N| \geq \frac{\beta_N}{2}\right] + \mathbb{P}\left[\mathbf{1}_{B_N^c}|\widetilde{S}_N - (N-1)m_N| \geq \frac{\beta_N}{2}\right] \\
&\quad \leq \mathbb{P}\left[\mathbf{1}_{B_N}|\widetilde{S}_N - (N-1)m_N| \geq \frac{\beta_N}{2}\right] + \mathbb{P}[B_N^c].
\end{aligned}$$

We can see that

$$\mathbb{P}[B_N^c] \leq N\mathbb{P}[X_1 > \beta_N] \leq \frac{Nf(\beta_N)}{\beta_N^\alpha} \leq \frac{N\overline{f}(1)}{\beta_N^\alpha}.$$

Now observe that defining $\widetilde{m}_N := \mathbb{E}[X_1 \mathbf{1}_{X_1 \leq \beta_N}]$, and recalling that $m_N := \mathbb{E}[X_1]$,

$$\begin{aligned}\mathbb{E}[\mathbf{1}_{B_N}(\widetilde{S}_N - (N-1)m_N)^2] =& \mathbb{E}[\mathbf{1}_{B_N}\big(\widetilde{S}_N - (N-1)\widetilde{m}_N + (N-1)\widetilde{m}_N - (N-1)m_N\big)^2] \\ =& \mathbb{E}[\mathbf{1}_{B_N}(\widetilde{S}_N - (N-1)\widetilde{m}_N)^2] + (N-1)^2(\widetilde{m}_N - m_N)^2 \\ =& \mathcal{O}(N\beta_N^{2-\alpha}) + \mathcal{O}(N^2/\beta^{2+2\alpha}).\end{aligned}$$

Chebyshev's inequality then implies

$$\begin{aligned}\mathbb{P}\Big[\mathbf{1}_{B_N}|\widetilde{S}_N - (N-1)m_N| \geq \frac{\beta_N}{2}\Big] \leq& 4\frac{\mathbb{E}[\mathbf{1}_{B_N}(\widetilde{S}_N - (N-1)m_N)^2]}{\beta_N^2} \\ =& \mathcal{O}\left(\frac{N\beta_N^{2-\alpha}}{\beta_N^2} + \frac{N^2}{\beta_N^{4+2\alpha}}\right) = \mathcal{O}\left(\frac{N}{\beta_N^{\alpha}}\right),\end{aligned}$$

for $N/\beta_N^{\alpha} \to 0$. □

4.6.2.2 Proof of Proposition 4.6.3

We are now in a position to prove Proposition 4.6.3.

Proof[Proof of Proposition 4.6.3, Case 1]

Consider

$$\mathbb{E}[X_1(X_1-1)] \leq \int_0^{\psi(N)} 2\overline{f}(x)x^{1-\alpha}\mathrm{d}x \leq \overline{f}(1)\frac{2L^{2-\alpha}}{2-\alpha} + \overline{f}(L)\left(\frac{2\psi(N)^{2-\alpha}}{2-\alpha} - \frac{2L^{2-\alpha}}{2-\alpha}\right),$$

for any L. We now see that

$$\begin{aligned}\mathbb{E}\left[\frac{X_1(X_1-1)}{(X_1+\widetilde{S}_N)^2}\middle| A_N\right] \leq& \frac{\mathbb{E}[X_1(X_1-1)]}{((N-1)m_N - \beta_N)^2} \leq \frac{\frac{2\overline{f}(L)\psi(N)^{2-\alpha}}{2-\alpha} + \mathcal{O}(L^{2-\alpha})}{((N-1)m_N - \beta_N)^2} \\ =& \frac{\psi(N)^{2-\alpha}}{N^2}\left(\frac{2\overline{f}(L)}{(2-\alpha)m_N^2} + \mathcal{O}\left(\frac{L^{2-\alpha}}{\psi(N)^{2-\alpha}}\right) + \mathcal{O}\left(\frac{\beta_N}{N}\right)\right).\end{aligned}$$

Similarly we can see that

$$\begin{aligned}\mathbb{E}\left[\frac{X_1(X_1-1)}{(X_1+\widetilde{S}_N)^2}\middle| A_N\right] \geq& \frac{\mathbb{E}[X_1(X_1-1)]}{(\psi(N) + (N-1)m_N + \beta_N)^2} \\ =& \frac{\psi(N)^{2-\alpha}}{N^2}\left(\frac{2\underline{g}(L)}{(2-\alpha)m_N^2} + \mathcal{O}\left(\frac{L^{2-\alpha}}{\psi(N)^{2-\alpha}}\right) + \mathcal{O}\left(\frac{\beta_N}{N}\right) + \mathcal{O}\left(\frac{\psi(N)}{N}\right)\right).\end{aligned}$$

Finally,

$$\frac{N^2}{\psi(N)^{2-\alpha}}\mathbb{E}\left[\frac{X_1(X_1-1)}{S_N^2}\mathbf{1}_{S_N\geq N}\middle|A_N^c\right]\mathbb{P}[A_N^c]$$
$$\leq \frac{N^2}{\psi(N)^{2-\alpha}}\frac{\mathbb{E}[X_1(X_1-1)]\mathbb{P}[A_N^c]}{N^2} = \mathcal{O}\left(\frac{N\psi(N)^{2-\alpha}}{\beta_N^2}\right),$$

where we have used Chebyshev's inequality to bound $\mathbb{P}[A_N^c]$.

□

Proof[Proof of Proposition 4.6.3, Case 2]

Recall, the assumption $\psi(N)/N \to K$ as $N \to \infty$, and the notation $M_- = (N-1)m_N - \beta_N$ from Notation 4.6.4. We use Lemma 4.6.4 to see that if $L(N)/N$ is bounded, then

$$\begin{aligned}
N^\alpha\mathbb{E}\left[\left|\frac{X_1(X_1-1)}{(X_1+\widetilde{S}_N)^2}\right|A_N\right] \leq & f_\infty\frac{2}{m_N^\alpha}\int_{\frac{m_N}{K+m_N}}^1 y^\alpha(1-y)^{1-\alpha}\mathrm{d}y + \mathcal{O}(|\overline{f}(L)-f_\infty|)\\
&+\mathcal{O}\left(\left|\frac{M_-}{\psi(N)-1+M_-}-\frac{m_N}{K+m_N}\right|\right)\\
&+\mathcal{O}\left(\int_{\frac{M_-}{L-1+M_-}}^1 y^\alpha(1-y)^{1-\alpha}\right)\\
&+\mathcal{O}\left(\left|\frac{2M_-+1}{M_-^{1+\alpha}}N^\alpha-\frac{2}{m_N^\alpha}\right|\right)\\
\leq & f_\infty\frac{2}{m_N^\alpha}\int_{\frac{m_N}{K+m_N}}^1 y^\alpha(1-y)^{1-\alpha}\mathrm{d}y + \mathcal{O}(|\overline{f}(L)-f_\infty|)\\
&+\mathcal{O}(\beta_N/N)+\mathcal{O}\left(\left|\frac{\psi(N)}{N}-K\right|\right)+\mathcal{O}((L/N)^{2-\alpha}).
\end{aligned}$$

We also see, recalling $M_+ = (N-1)m_N + \beta_N$,

$$\begin{aligned}
N^\alpha\mathbb{E}\left[\left|\frac{X_1(X_1-1)}{(X_1+\widetilde{S}_N)^2}\right|A_N\right] \geq & g_\infty\frac{2}{m_N^\alpha}\int_{\frac{m_N}{K+m_N}}^1 y^\alpha(1-y)^{1-\alpha}\mathrm{d}y + \mathcal{O}(|\underline{g}(L)-g_\infty|)\\
&+\mathcal{O}\left(\left|\frac{M_+}{\psi(N)+1+M_+}-\frac{m_N}{K+m_N}\right|\right)\\
&+\mathcal{O}\left(\int_{\frac{M_++1}{L+M_++1}}^1 y^\alpha(1-y)^{1-\alpha}\right)\\
&+\mathcal{O}\left(\left|\frac{2M_+}{(M_++1)^{1+\alpha}}N^\alpha-\frac{2}{m_N^\alpha}\right|\right)
\end{aligned}$$

$$\geq g_\infty \frac{2}{m_N^\alpha} \int_{\frac{m_N}{K+m_N}}^1 y^\alpha (1-y)^{1-\alpha} \mathrm{d}y + \mathcal{O}(|\underline{g}(L) - g_\infty|)$$
$$+ \mathcal{O}(\beta_N/N) + \mathcal{O}\left(\left|\frac{\psi(N)}{N} - K\right|\right) + \mathcal{O}((L/N)^{2-\alpha}).$$

We finally turn our attention to

$$N^\alpha \mathbb{E}\left[\frac{X_1(X_1-1)}{S_N^2}\mathbf{1}_{S_N \geq N}\middle| A_N^c\right]\mathbb{P}[A_N^c]$$
$$\leq \frac{\mathbb{E}[X_1(X_1-1)]\mathbb{P}[A_N^c]}{N^{2-\alpha}} = \mathcal{O}\left(\frac{\psi(N)^{2-\alpha}}{N^{2-\alpha}}\frac{N\psi(N)^{2-\alpha}}{\beta_N^2}\right), \quad (4.6.9)$$

where we have used Chebyshev's inequality to bound $\mathbb{P}[A_N^c]$. □

Proof[Proof of Proposition 4.6.3, Case 3] Now assuming that $\psi(N)/N \to \infty$ and $L(N)/N$ bounded, we again turn to Lemma 4.6.4 to see

$$N^\alpha \mathbb{E}\left[\left|\frac{X_1(X_1-1)}{(X_1+\widetilde{S}_N)^2}\right| A_N\right] \leq f_\infty \frac{2}{m_N^\alpha}\int_0^1 y^\alpha(1-y)^{1-\alpha}\mathrm{d}y + \mathcal{O}(|\overline{f}(L) - f_\infty|)$$
$$+ \mathcal{O}\left(\left|\frac{2M_- + 1}{M_-^{1+\alpha}}N^\alpha - \frac{2}{m_N^\alpha}\right|\right)$$
$$+ \mathcal{O}\left(\int_0^{\frac{M_-}{\psi(N)-1+M_-}} y^\alpha(1-y)^{1-\alpha}\mathrm{d}y\right)$$
$$+ \mathcal{O}\left(\int_{\frac{M_-}{L-1+M_-}}^1 y^\alpha(1-y)^{1-\alpha}\right)$$
$$\leq f_\infty \frac{2}{m_N^\alpha}\int_0^1 y^\alpha(1-y)^{1-\alpha}\mathrm{d}y + \mathcal{O}(|\overline{f}(L) - f_\infty|)$$
$$+ \mathcal{O}(\beta_N/N) + \mathcal{O}\Big((N/\psi(N))^{1+\alpha}\Big) + \mathcal{O}((L/N)^{2-\alpha}).$$

Similarly, recalling that $M_+ = (N-1)m_N + \beta_N$,

$$N^\alpha \mathbb{E}\left[\left|\frac{X_1(X_1-1)}{(X_1+\widetilde{S}_N)^2}\right| A_N\right] \geq g_\infty \frac{2}{m_N^\alpha}\int_0^1 y^\alpha(1-y)^{1-\alpha}\mathrm{d}y + \mathcal{O}(|\underline{g}(L) - g_\infty|)$$
$$+ \mathcal{O}\left(\left|\frac{2M_+}{(M_+ + 1)^{1+\alpha}}N^\alpha - \frac{2}{m_N^\alpha}\right|\right)$$
$$+ \mathcal{O}\left(\int_0^{\frac{M_+ + 1}{\psi(N)+M_+}} y^\alpha(1-y)^{1-\alpha}\mathrm{d}y\right)$$
$$+ \mathcal{O}\left(\int_{\frac{M_+ + 1}{L+M_+ + 1}}^1 y^\alpha(1-y)^{1-\alpha}\right)$$

$$\geq g_\infty \frac{2}{m_N^\alpha} \int_0^1 y^\alpha (1-y)^{1-\alpha} \mathrm{d}y + \mathcal{O}(|\underline{g}(L) - g_\infty|) + \mathcal{O}(\beta_N/N) + \mathcal{O}\Big((N/\psi(N))^{1+\alpha}\Big) + \mathcal{O}((L/N)^{2-\alpha}).$$

We finally turn our attention to

$$N^\alpha \mathbb{E}\left[\frac{X_1(X_1-1)}{S_N^2} \mathbf{1}_{S_N \geq N} \middle| A_N^c \right] \mathbb{P}[A_N^c] \leq N^\alpha \mathbb{E}\left[\frac{X_1(X_1-1)}{\max\{X_1^2, N^2\}} \right] \mathbb{P}[A_N^c] \leq N^\alpha 4 \mathbb{E}\left[\frac{X_1(X_1-1)}{(X_1+N)^2} \right] \mathbb{P}[A_N^c] = \mathcal{O}(\mathbb{P}[A_N^c]) = \mathcal{O}\left(\frac{N}{\beta_N^\alpha} \right), \quad (4.6.10)$$

where we have used Lemma 4.6.6 to bound $\mathbb{P}[A_N^c]$ and Lemma 4.6.4 to control all other terms.

□

We conclude this section with a slight modification of Schweinsberg (2003), Lemma 6.

Lemma 4.6.8 (Lemma 6, Schweinsberg (2003))**.** *For all $r \geq 1$ and $k_1, \ldots, k_r \geq 2$, we have*

$$\frac{\mathbb{E}[(v_{1,N})_{k_1} \ldots (v_{r,N})_{k_r}]}{N^{k_1+\cdots+k_r-r} c_N} = \frac{N^r}{c_N} \mathbb{E}\left[\frac{(X_1)_{k_1} \ldots (X_r)_{k_r}}{S_N^{k_1+\cdots+k_r}} \mathbf{1}_{S_N \geq N} \right] \left(1 + \mathcal{O}\left(\frac{(k_1+\cdots+k_r)^2}{N} \right) \right), \quad (4.6.11)$$

and

$$c_N = \mathcal{C}_N \left(1 + \mathcal{O}\left(\frac{1}{N} \right) \right). \quad (4.6.12)$$

Our proof follows that presented in Schweinsberg (2003). However, we include it in order to emphasize that the source of the error term is the distinction between sampling with and without replacement from the juvenile population.

Proof

Label individuals in the current generation by $\{1, 2, \ldots, N\}$. For each $1 \leq r \leq N$, set $k_0 = 0$ and write $B_{k_1,\ldots,k_r}$ for the event that for each $i = 1, \ldots, r$, the individuals labelled $\{k_0 + \cdots + k_{i-1} + 1, \ldots, k_0 + \cdots + k_{i-1} + k_i\}$ are all descended from the individual with label i in the previous generation.

We then see that

$$\mathbb{P}[B_{k_1,\dots,k_r}] = \mathbb{E}[\mathbb{P}[B_{k_1,\dots,k_r}|v_{1,N},\dots,v_{N,N}]] = \mathbb{E}\Big[\frac{(v_{1,N})_{k_1}\dots(v_{r,N})_{k_r}}{(N)_{k_1+\dots+k_r}}\Big],$$

and

$$\begin{aligned}\mathbb{P}[B_{k_1,\dots,k_r}] =& \mathbb{P}[B_{k_1,\dots,k_r}\cap\{S_N\geq N\}] + \mathbb{P}[B_{k_1,\dots,k_r}\cap\{S_N< N\}]\\ =& \mathbb{E}[\mathbb{P}[B_{k_1,\dots,k_r}\cap\{S_N\geq N\}|X_1,\dots,X_N]] + \mathbb{P}[B_{k_1,\dots,k_r}\cap\{S_N< N\}]\\ =& \mathbb{E}\Big[\frac{(X_1)_{k_1}\dots(X_r)_{k_r}}{(S_N)_{k_1+\dots+k_r}}\mathbf{1}_{S_N\geq N}\Big],\end{aligned}$$

where we have used that when $S_N < N$ we will maintain the previous population, and so will see exactly one individual in the ith generation descended from each individual in the previous generation. The result follows on replacing $(S_N)_{k_1+\dots+k_r}$ by $S_N^{k_1+\dots+k_r}$.

We recall the definition of $\mathcal{C}_N$ from (4.6.6). We may deduce (4.6.12) from (4.6.11) with $r = 1$ and $k_1 = 2$, by using that $c_N = \mathbb{E}[(v_{1,N})_2]/(N-1)$ and that, in this case, the left hand side of (4.6.11) is equal to $(N-1)/N$. □

4.6.3 Convergence to the Kingman Coalescent

In this section, we use Proposition 4.6.1 to prove Proposition 4.4.2.

Proof[Proof of Proposition 4.4.2.]

From Lemma 4.6.8, it will suffice to replace c_N by $\mathcal{C}_N$, and from Proposition 4.4.1, $\mathcal{C}_N = \mathcal{O}(\psi(N)^{2-\alpha}/N)$. Recall also that we are assuming that $\psi(N)/N \to 0$. We then see

$$\frac{N}{c_N}\mathbb{E}\Big[\frac{(X_1)_3}{S_N^3}\mathbf{1}_{S_N\geq N}\Big] \leq \frac{N}{c_N}\frac{\mathbb{E}[(X_1)_3]}{N^3} = \mathcal{O}\Big(\frac{N^2}{\psi(N)^{2-\alpha}}\frac{\psi(N)^{3-\alpha}}{N^3}\Big) = \mathcal{O}\Big(\frac{\psi(N)}{N}\Big).$$

For the converse of (4.4.4), recalling the definition of A_N and M_+ from Notation 4.6.4, and setting $\beta_N = N$, we see

$$\begin{aligned}\frac{N}{c_N}\mathbb{E}\Big[\frac{(X_1)_3}{S_N^3}\mathbf{1}_{S_N\geq N}\Big] \geq& \frac{N}{c_N}\mathbb{E}\Big[\frac{(X_1)_3}{(\psi(N)+\widetilde{S}_N)^3}\mathbf{1}_{S_N\geq N}\mathbf{1}_{A_N}\Big]\\ \geq& \frac{N}{c_N}\frac{\mathbb{E}[(X_1)_3]}{(\psi(N)+M_+)^3}\mathbb{P}[A_N]\end{aligned}$$

$$=\mathcal{O}\left(\frac{N^2}{\psi(N)^{2-\alpha}}\frac{\psi(N)^{3-\alpha}}{N^3}\right)=\mathcal{O}\left(\frac{\psi(N)}{N}\right),$$

where we used Lemma 4.6.6 to deduce that $\mathbb{P}[A_N]$ is $\mathcal{O}(1)$.

Now consider (4.4.5).

$$\begin{aligned}\frac{N^2}{c_N}\mathbb{E}\left[\frac{(X_1)_2(X_2)_2}{S_N^4}\mathbf{1}_{S_N\geq N}\right]\leq&\frac{N^2}{c_N}\frac{\mathbb{E}[(X_1)_2(X_2)_2]}{N^4}=\frac{\mathbb{E}[(X_1)_2]^2}{c_N N^2}\\=&\mathcal{O}\left(\frac{\psi(N)^{4-2\alpha}}{\psi(N)^{2-\alpha}N}\right)=\mathcal{O}\left(\frac{\psi(N)^{2-\alpha}}{N}\right).\end{aligned}$$

For the opposite inequality, we define

$$\widehat{A}_N:=\left\{\left|\sum_{i=3}^{N}X_i^N-(N-2)m_N\right|\leq\beta_N\right\},$$

and proceed as before to see

$$\begin{aligned}\frac{N^2}{c_N}\mathbb{E}\left[\frac{(X_1)_2(X_2)_2}{S_N^4}\mathbf{1}_{S_N\geq N}\right]\geq&\frac{N^2}{c_N}\mathbb{E}\left[\frac{(X_1)_2(X_2)_2}{(2\psi(N)+\sum_{i=3}^{N}X_i^N)^4}\mathbf{1}_{S_N\geq N}\mathbf{1}_{\widehat{A}_N}\right]\\\geq&\frac{N^2}{c_N}\frac{\mathbb{E}[(X_1)_2(X_2)_2]}{(2\psi(N)+(N-2)m_N+\beta_N)^4}\mathbb{P}[\widehat{A}_N]\\=&\mathcal{O}\left(\frac{N^3}{\psi(N)^{2-\alpha}}\frac{\psi(N)^{4-2\alpha}}{N^4}\right)=\mathcal{O}\left(\frac{\psi(N)^{2-\alpha}}{N}\right),\end{aligned}$$

where again we chose $\beta_N=N$, and used Lemma 4.6.6. □

4.6.4 Convergence to the Lambda coalescent

In this section we use Proposition 4.6.2 to prove Proposition 4.4.3.

We divide the proof of Proposition 4.4.3. into two parts, first controlling the probability of simultaneous multiple mergers, and then identifying the measure Λ in the limiting coalescent.

Proof[Proof of (4.4.6)]

From Lemma 4.6.8, we can replace c_N by $\mathcal{C}_N$, and from Proposition 4.4.1, $\mathcal{C}_N=\mathcal{O}(1/N^{\alpha-1})$ and (assuming that N is large enough that $m_N>1$), following Schweinsberg (2003) we see

$$\frac{N^2}{c_N}\mathbb{E}\Big[\frac{(X_1)_2(X_2)_2}{S_N^4}\mathbf{1}_{S_N\geq N}\Big] \leq \frac{N^2}{c_N}\mathbb{E}\Big[\frac{(X_1)_2(X_2)_2}{(\max\{X_1^2,N^2\})(\max\{X_2^2,N^2\})}\Big]$$
$$= \frac{N^2}{c_N}\Big(\mathbb{E}\Big[\frac{(X_1)_2}{\max\{X_1^2,N^2\}}\Big]\Big)^2 \leq 16\frac{N^2}{c_N}\Big(\mathbb{E}\Big[\frac{(X_1)_2}{(X_1+N)^2}\Big]\Big)^2 = \mathcal{O}\Big(N^{1-\alpha}\Big),$$

where we have used Lemma 4.6.4 for the final estimate. For the lower bound, using the Notation 4.6.4, we have

$$\begin{aligned}
\frac{N^2}{c_N}\mathbb{E}\Big[\frac{(X_1)_2(X_2)_2}{S_N^4}\mathbf{1}_{S_N\geq N}\Big] \geq& \frac{N^2}{c_N}\mathbb{E}\Big[\frac{(X_1)_2(X_2)_2}{(X_1+M_+)^4}\mathbf{1}_{A_N}\Big]\\
=& \frac{N^2}{c_N}\mathbb{E}\Big[\frac{(X_1)_2}{(X_1+M_+)^4}\Big]\mathbb{E}[(X_2)_2\mathbf{1}_{A_N}]\\
\geq& \frac{N^2}{c_N}\mathbb{E}\Big[\frac{(X_1)_2}{(X_1+M_+)^4}\mathbf{1}_{X_1\leq M_+}\Big]\\
&\times\mathbb{E}\Big[(X_2)_2\mathbf{1}_{|\sum_{i=3}^N X_1-(N-2)m_n|\leq\frac{\beta_N}{2}}\mathbf{1}_{|X_2-m_N|\leq\frac{\beta_N}{2}}\Big]\\
\geq& \frac{N^2}{c_N}\mathbb{E}\Big[\frac{(X_1)_2}{16M_+^4}\mathbf{1}_{X_1\leq M_+}\Big]\\
&\times\mathbb{E}\Big[(X_2)_2\mathbf{1}_{|X_2-m_N|\leq\frac{\beta_N}{2}}\Big]\mathbb{P}\Big[\Big|\sum_{i=3}^N X_1-(N-2)m_n\Big|\leq\frac{\beta_N}{2}\Big]\\
=& \mathcal{O}\bigg(\frac{N^2}{N^{1-\alpha}}\frac{M_+^{2-\alpha}}{M_+^4}\beta_N^{2-\alpha}\bigg) = \mathcal{O}\Big(N^{1-\alpha}\Big),
\end{aligned}$$

where we choose $\beta_N = N$ to optimise on the final line and Lemma 4.6.6 to deduce that the probability on the penultimate line is $\mathcal{O}(1)$.

□

Proof[Proof of Proposition 4.4.3: identifying the measure Λ]

We first note that (4.4.6) and Lemma 4.6.8 reduce the proof of Proposition 4.4.3 to examining

$$\lim_{N\to\infty}\frac{N}{c_N}\mathbb{P}[v_{1,N}>Nx] = \lim_{N\to\infty}\frac{N}{c_N}\mathbb{P}\Big[\frac{X_1}{S_N}\mathbf{1}_{S_N\geq N}>x\Big],$$

for $x>0$. We essentially follow Schweinsberg (2003), pp 129–131. We start in a familiar way by conditioning on $\{\widetilde{S}_N\geq(N-1)m_N-\beta_N\}=\{\widetilde{S}_N\geq M_-\}$. We assume that $\beta_N/N\to 0$ and recall the definition of A_N from Notation 4.6.4. We see

$$\frac{N}{c_N}\mathbb{P}\Big[\frac{X_1}{S_N}\mathbf{1}_{S_N\geq N}>x\Big] \leq \frac{N}{c_N}\mathbb{P}\Big[\frac{X_1}{X_1+M_-}>x\Big] + \frac{N}{c_N}\mathbb{P}\Big[\{\widetilde{S}_N<M_-\}\Big]\mathbb{P}\Big[\frac{X_1}{N}\geq x\Big]$$

$$\begin{aligned}
&\leq \frac{N}{c_N}\mathbb{P}\left[X_1 > \frac{xM_-}{1-x}\right] + \frac{N}{c_N}\mathbb{P}[A_N^c]\mathbb{P}[X_1 \geq xN] \\
&\leq \frac{N}{c_N M_-^\alpha}\frac{(1-x)^\alpha}{x^\alpha} f\left(\frac{xM_-}{1-x}\right)\mathbf{1}_{\frac{xM_-}{1-x}\leq\psi(N)} + \mathcal{O}\left(\frac{N}{\beta_N^\alpha} f(xN)\right).
\end{aligned}$$

In the opposite direction, still requiring $\beta_N/N \to 0$, we have

$$\begin{aligned}
\frac{N}{c_N}\mathbb{P}\left[\frac{X_1}{S_N}\mathbf{1}_{S_N\geq N} > x\right] &\geq \frac{N}{c_N}\mathbb{P}[A_N]\mathbb{P}\left[\frac{X_1}{X_1 + M_+} > x\right] \\
&= \frac{N}{c_N}\mathbb{P}[A_N]\mathbb{P}\left[X_1 > \frac{xM_+}{1-x}\right] \\
&\geq \frac{N}{c_N M_+^\alpha}\frac{(1-x)^\alpha}{x^\alpha} g\left(\frac{xM_+}{1-x}\right)\mathbf{1}_{\frac{xM_+}{1-x}\leq\psi(N)}\left(1 - \mathcal{O}\left(\frac{N}{\beta_N^\alpha}\right)\right),
\end{aligned}$$

where we recall $M_+ = (N-1)m_N + \beta_N$.

Using Lemma 4.6.8 to replace c_N by $\mathcal{C}_N$ from Proposition 4.4.1, the proof of Proposition 4.4.3 is complete. □

4.7 Large sample size calculations

We wish to establish how large a sample size we can take before the errors that we make in using the limiting coalescent processes that we have identified to estimate the distribution of the site frequency spectrum for data become significant. Of particular interest is understanding how large sample size affects the site-frequency spectrum, since the site-frequency spectrum is frequently used in inference, in both the single locus and multi-locus case, for instance see Eldon et al. (2015) and Koskela (2018) respectively. Since mutation rates are not infinite most statistics will not be able to differentiate between mergers that happen instantaneously and those which occur in quick succession if both lead to the same topology for the genealogical tree.

4.7.1 Large Sample Size and the Kingman coalescent

We consider what sample size will show divergence from Kingman in a haploid population of size N with Wright-Fisher sampling, and $c_N = 1/N$. The probability of seeing a 3-merger in a single generation within a sample size of n is $\binom{n}{3}/N^2$. Similarly the probability of seeing two simultaneous pairwise mergers within a sample size of n is

$\binom{n}{4}(N-1)/N^3$. We see that 2-mergers have a rate $\binom{n}{2}/N$. Therefore, in order to see a 3-merger or simultaneous merger they must happen at a rate at least comparable to 2-mergers. A straightforward calculation then leads us to expect to see simultaneous mergers when $n = \mathcal{O}(\sqrt{N})$ and 3-mergers when $n = \mathcal{O}(N)$. One would, therefore, expect the Kingman approximation to break down due to occurrence of simultaneous mergers. However it can be seen, for instance in Wakeley and Takahashi (2003), that one only starts to see the impact of large sample size on the site-frequency spectrum when n is of the same order as the effective population size $(1/c_N)$, which in our example above is N. This leads us to look for when the sample size is sufficiently large to change the topology of the coalescent tree.

For example with sample size n in the case covered by Proposition 4.6.3, (1), we will see 3-mergers (recall (4.6.2)) when

$$\frac{\mathbb{P}[\text{a 3-merger given } n \text{ lines}]}{\mathbb{P}[\text{a 2-merger given } n \text{ lines}]} = \frac{N\mathbb{E}\left[\frac{(X_1)_3}{S_N^3}\mathbf{1}_{S_N \geq N}\right]\binom{n}{3}}{N\mathbb{E}\left[\frac{(X_1)_2}{S_N^2}\mathbf{1}_{S_N \geq N}\right]\binom{n}{2}} \sim \frac{n\psi(N)}{N} \not\to 0, \tag{4.7.1}$$

and we will see simultaneous mergers (recall (4.6.3)) when

$$\frac{\mathbb{P}[\text{two simultaneous 2-mergers given } n \text{ lines}]}{\mathbb{P}[\text{a 2-merger given } n \text{ lines}]} = \frac{N^2\mathbb{E}\left[\frac{(X_1)_2(X_2)_2}{S_N^4}\mathbf{1}_{S_N \geq N}\right]\frac{1}{2}\binom{n}{4}}{N\mathbb{E}\left[\frac{(X_1)_2}{S_N^2}\mathbf{1}_{S_N \geq N}\right]\binom{n}{2}} \sim \frac{n^2\psi(N)^{2-\alpha}}{N} \not\to 0. \tag{4.7.2}$$

We then see that simultaneous mergers happen when n is of order $\sqrt{1/c_N}$ as in the simple Wright-Fisher example above. Whereas 3-mergers only happen when n is of order $N/\psi(N)$ which is larger than $\sqrt{1/c_N}$ for $1 < \alpha < 2$. As such in Proposition 4.6.3, case (1), we would expect the Kingman coalescent to be a valid approximation for sample sizes up to order $N/\psi(N)$. The assumption here is that $\psi(N)/N \to 0$. If $\psi(N) = N/\log(N)$, then our results show that sample size of order $\mathcal{O}(\log N)$ can be sufficient to invalidate the Kingman-coalescent approximation.

4.7.2 Large Sample Size and the Lambda Coalescent

We now turn our attention to proving our claims about the effects of large sample size when the limit is a Λ-coalescent. In particular, we show that as a result of approximating sampling without replacement, by sampling with replacement, we consistently overestimate the rate of k-mergers and that the error is on the same order as the rate of mergers for $k = \mathcal{O}(\sqrt{N})$. We shall then provide a heuristic argument, based on known results for the number of blocks in the limiting coalescent, that suggests that this will significantly distort our prediction of things like the ratio of external branch length to total tree length, and hence the site frequency spectrum, for sample sizes of order $N^{\alpha/2}$.

Recall the notation $\widetilde{S}_N = \sum_{k=2}^{N} X_k$.

$$
\begin{aligned}
&\mathbb{P}[\text{Exactly } k \text{ from } n \text{ lineages merge in 1 generation}] \\
&= \mathbb{E}[\mathbb{P}[\text{Exactly } k \text{ from } n \text{ lineages merge in 1 generation}|X_1, \dots X_N]] \\
&= N\mathbb{E}\left[\frac{\binom{X_1}{k}\binom{S_N - X_1}{n-k}}{\binom{S_N}{n}}\mathbf{1}_{S_N \geq N}\right] \\
&= N\binom{n}{k}\mathbb{E}\left[\frac{X_1(X_1-1)\dots(X_1-k+1)(\widetilde{S}_N)\dots(\widetilde{S}_N-n+k+1)}{(X_1+\widetilde{S}_N)\dots(X_1+\widetilde{S}_N-n+1)}\mathbf{1}_{S_N \geq N}\right].
\end{aligned}
$$

Note that we have not prohibited simultaneous mergers. Now we wish to calculate

$$
\mathbb{E}\left[\frac{X_1(X_1-1)\dots(X_1-k+1)}{(X_1+M)\dots(X_1+M-n+1)}\right].
$$

As usual we consider the set A_N defined in Notation 4.6.4. Conditioned on A_N, for any n with $n/S_N \to 0$, we approximate $(\widetilde{S}_N)\dots(\widetilde{S}_N-n-k+1)$ and $(X_1+\widetilde{S}_N)\dots(X+\widetilde{S}_N-n+1)$ by $\widetilde{S}_N^{n-k}$ and $(X_1+\widetilde{S}_N)^n$ respectively.

We shall make use of the following observation:

$$
\begin{aligned}
\int_a^b \frac{x^p}{(x+M)^q}\mathrm{d}x &= \int_{\frac{M}{b+M}}^{\frac{M}{a+M}} M^m \frac{(1-y)^p}{y^p}\frac{y^q}{M^q}\frac{M}{y^2}\mathrm{d}y \\
&= M^{1+p-q}\int_{\frac{M}{b+M}}^{\frac{M}{a+M}}(1-y)^p y^{q-p-2}\mathrm{d}y = M^{1+p-q}\int_{1-\frac{M}{a+M}}^{1-\frac{M}{b+M}} u^p(1-u)^{q-p-2}\mathrm{d}u. \quad (4.7.3)
\end{aligned}
$$

We also recall from the calculations in Section 4.6.4 that, for $\psi(N)/N \to \infty$,

$$\lim_{N\to\infty} \frac{N}{c_N}\mathbb{P}[\nu_1 > Nx] = \int_x^1 y^{-2}\Lambda(\mathrm{d}y)$$

with $\Lambda(\mathrm{d}y) = C(\alpha)(1-y)^{\alpha-1}/y^{\alpha-1}\mathrm{d}y$, where the constant $C(\alpha) = 1/\Gamma(\alpha)\Gamma(2-\alpha)$.

Lemma 4.7.9. *For $M \in \mathbb{R}$ we have*

$$M^{1+p-q}\int_{1-\frac{M}{a+M}}^{1-\frac{M}{b+M}} u^p(1-u)^{q-p-2}\mathrm{d}u \le \sum_{i=a}^{b}\frac{i^p}{(i+M)^q} \le M^{1+p-q}\int_{1-\frac{M}{a-1+M}}^{1-\frac{M}{b+1+M}} u^p(1-u)^{q-p-2}\mathrm{d}u + \mathcal{O}(M^{p-q})$$

Proof We begin with

$$\sum_{i=a}^{b}\frac{i^p}{(i+M)^q} \le \int_{a-1}^{b+1}\frac{x^p}{(x+M)^q}\mathrm{d}x + \mathcal{O}(M^{p-q}) = M^{1+p-q}\int_{1-\frac{M}{a+M}}^{1-\frac{M}{b+1+M}} u^p(1-u)^{q-p-2}\mathrm{d}u + \mathcal{O}(M^{p-q}),$$

where the $\mathcal{O}(M^{p-q})$ correction appears since for $p \neq q$, $i^p/(i+M)^q$ has a maximum at $i = pM/(q-p)$. A similar calculation provides the opposite inequality. □

We now present a lemma which is similar in nature to Proposition 4.6.3.

Lemma 4.7.10. *Consider a sample of n individuals from a population evolving according to model* (4.3.3)*. Suppose that $n/N \to 0$ and $n \ge k$.*

1. *When $f_\infty = g_\infty$ and $\psi(N)/N \to K$, for any $L = L(N)$ with $L/M \to 0$,*

$$N\mathbb{E}\left[\frac{\binom{X_1}{k}\binom{S_N-X_1}{n-k}}{\binom{S_N}{n}}\mathbf{1}_{S_N\ge N}\right] = N^{1-\alpha}f_\infty m_N^{-\alpha}\binom{n}{k}\left(k\int_0^{\frac{K}{K+m_N}} u^{k-1-\alpha}(1-u)^{n-k+\alpha}\mathrm{d}u - (n-k)\int_0^{\frac{K}{K+m_N}} u^{k-\alpha}(1-u)^{n-k+\alpha-1}\mathrm{d}u\right)\left(1 + \mathcal{O}(|\overline{f}(L) - \underline{g}(L)|) + \mathcal{O}(\beta_N/N) + \mathcal{O}\Big(\Big|\frac{\psi(N)}{N} - K\Big|\Big) + \mathcal{O}\Big((L/N)^{k-\alpha}\Big) + \mathcal{O}(k^2/L) + \mathcal{O}\Big(N\psi(N)^{2-\alpha}/\beta_N^2\Big)\right). \tag{4.7.4}$$

2. When $f_\infty = g_\infty$ and $\psi(N)/N \to \infty$ then

$$N\mathbb{E}\left[\frac{\binom{X_1}{k}\binom{S_N-X_1}{n-k}}{\binom{S_N}{n}}\mathbf{1}_{S_N\geq N}\right] = N^{1-\alpha}f_\infty m_N^{-\alpha}\binom{n}{k}\Big(k\int_0^1 u^{k-1-\alpha}(1-u)^{n-k+\alpha}\mathrm{d}u$$
$$-(n-k)\int_0^1 u^{k-\alpha}(1-u)^{n-k+\alpha-1}\mathrm{d}u\Big)\Big(1+\mathcal{O}(|\overline{f}(L)-\underline{g}(L)|)+\mathcal{O}(\beta_N/N)$$
$$+\mathcal{O}\Big((N/\psi(N))^{n-k+\alpha+1}\Big)+\mathcal{O}(\mathbf{1}_{n\neq k}(N/\psi(N))^{n-k+\alpha+1})$$
$$+\mathcal{O}\Big((L/N)^{k-\alpha}\Big)+\mathcal{O}(k^2/L)+\mathcal{O}(N/\beta_N^\alpha)\Big). \quad (4.7.5)$$

Equation (4.7.5) arises naturally in our proof and so we have left it in this form. However, observe that

$$k\int_0^1 u^{k-1-\alpha}(1-u)^{n-k-\alpha}\mathrm{d}u-(n-k)\int_0^1 u^{k-\alpha}(1-u^{n-k+\alpha-1}\mathrm{d}u$$
$$=\frac{k(n-k+\alpha)\Gamma(k-\alpha)\Gamma(n-k+\alpha)}{\Gamma(n+1)}-\frac{(k-\alpha)(n-k)\Gamma(k-\alpha)\Gamma(n-k-\alpha)}{\Gamma(n+1)}$$
$$=\alpha n\frac{\Gamma(k-\alpha)\Gamma(n-k+\alpha)}{\Gamma(n+1)}=\alpha\int_0^1 u^{k-\alpha-1}(1-u)^{n-k-\alpha-1}\mathrm{d}u,$$

as expected from Proposition 4.4.3.

Proof[Proof of Lemma 4.7.10]

As usual we partition over A_N, A_N^c, with A_N from (4.6.8). First note that

$$(M_- - n + k + 1)^{n-k}\mathbb{E}\Big[\frac{X\dots(X-k+1)}{(X+M_+)\dots(X+M_+-n+1)}\Big]$$
$$\leq \mathbb{E}\left[\frac{X_1(X_1-1)\dots(X_1-k+1)\widetilde{S}_N\dots(\widetilde{S}_N-n+k+1)}{(X_1+\widetilde{S}_N)\dots(X_1+\widetilde{S}_N-n+1)}\mathbf{1}_{S_N\geq N}\middle|A_N\right]$$
$$\leq M_+^{n-k}\mathbb{E}\Big[\frac{X\dots(X-k+1)}{(X+M_-)\dots(X+M_--n+1)}\Big]. \quad (4.7.6)$$

We now look at

$$\mathbb{E}\Big[\frac{X\dots(X-k+1)}{(X+M_-)\dots(X+M_--n+1)}\Big]$$
$$=\sum_{i=k}^{\psi(N)}\Big(\frac{i\dots(i-k+1)}{(i+M_-)\dots(i+M_--n+1)}-\frac{(i-1)\dots(i-k)}{(i+M_--1)\dots(i+M_--n)}\Big)\mathbb{P}[X\geq i]$$
$$=\sum_{i=k}^{\psi(N)}\Big(\frac{(i-1)\dots(i-k+1)}{(i+M_--1)\dots(i+M_--n+1)}\frac{kM-(n-k)i}{(i+M_-)(i+M_--n)}\Big)\mathbb{P}[X\geq i]$$

$$= \sum_{i=k}^{\psi(N)} \Big(\frac{(i-1)\dots(i-k+1)}{(i+M_-)\dots(i+M_--n)} kM \Big) \mathbb{P}[X \geq i] \\ - \sum_{i=k}^{\psi(N)} \Big(\frac{(i-1)\dots(i-k+1)}{(i+M_-)\dots(i+M_--n)} (n-k)i \Big) \mathbb{P}[X \geq i].$$

Our next task is to control the first term in this expression. For any $L \leq \psi(N)$, using Lemma 4.7.9 to bound the sum for $i \geq L$

$$\sum_{i=k}^{\psi(N)} \Big(\frac{(i-1)\dots(i-k+1)}{(i+M_-)\dots(i+M_--n)} kM \Big) \mathbb{P}[X \geq i] \\ \leq \sum_{i=k}^{\psi(N)} \left(\frac{i^{k-1}}{(i+M_--n)^{n+1}} kM \right) \frac{f(i)}{i^\alpha} \\ \leq \overline{f}(L) kM_-(M_--n)^{k-n-1-\alpha} \int_{1-\frac{M_--n}{L+M_--n-1}}^{1-\frac{M_--n}{\psi(N)+M_--n+1}} u^{k-1-\alpha}(1-u)^{n-k+\alpha} \mathrm{d}u \\ + \mathcal{O}\Big((L/M)^{k-\alpha} \Big),$$

where the error arises as $\mathcal{O}(kM_-(M_--n)^{k-n-1-\alpha}(L/M)^{k-\alpha}$.

To find a corresponding lower bound, first observe that

$$\sum_{i=k}^{\psi(N)} \Big(\frac{(i-1)\dots(i-k+1)}{(i+M_-)\dots(i+M_--n)} kM \Big) \mathbb{P}[X \geq i] \\ \geq \sum_{i=k}^{\psi(N)} \left(\frac{(i-k+1)^{k-1}}{(i+M_-)^{n+1}} kM \right) \frac{g(i)}{i^\alpha} \\ \geq \underline{g}(L) \left(kM_-M_-^{k-n-1-\alpha} \int_{1-\frac{M_-}{L+M_-}}^{1-\frac{M_-}{\psi(N)+M_-}} u^{k-1-\alpha}(1-u)^{n-k+\alpha} \mathrm{d}u \right) \Big(1 + \mathcal{O}\Big(k^2/L \Big) \Big),$$

where the $\mathcal{O}(k^2/L)$ error term (which is always negative) appears since, for $i \geq L$, $(i-1)\dots(i-k+1)/i^{k-1} = 1 - \mathcal{O}(k^2/L)$.

Similarly,

$$\sum_{i=k}^{\psi(N)} \Big(\frac{(i-1)\dots(i-k+1)}{(i+M_-)\dots(i+M_--n)} (n-k)i \Big) \mathbb{P}[X \geq i] \\ \leq \sum_{i=k}^{\psi(N)} \left(\frac{i^k}{(i+M_--n)^{n+1}} (n-k) \right) \frac{f(i)}{i^\alpha} \\ \leq \overline{f}(L) \left((n-k)(M_--n)^{k-n-\alpha} \int_{1-\frac{M_--n}{L+M_--n-1}}^{1-\frac{M_--n}{\psi(N)+M_--n+1}} u^{k-\alpha}(1-u)^{n-k+\alpha-1} \mathrm{d}u \right)$$

$$+ \mathcal{O}\Big((L/M)^{k-\alpha+1}\Big),$$

and

$$\begin{aligned}
&\sum_{i=k}^{\psi(N)} \left(\frac{(i-1)\dots(i-k+1)}{(i+M_-)\dots(i+M_--n)}(n-k)i \right) \mathbb{P}[X \geq i] \\
&\quad \geq \sum_{i=k}^{\psi(N)} \left(\frac{(i-k+1)^{k-1}}{(i+M_-)^{n+1}}(n-k) \right) \frac{g(i)}{i^{\alpha-1}} \\
&\quad \geq \underline{g}(L) \left((n-k) M_-^{k-n-\alpha} \int_{1-\frac{M_-}{L+M_-}}^{1-\frac{M_-}{\psi(N)+M_-}} u^{k-\alpha}(1-u)^{n-k+\alpha-1} \mathrm{d}u \right) \Big(1 + \mathcal{O}\Big(k^2/L\Big)\Big).
\end{aligned}$$

Bounds involving M_+ follow from entirely analogous arguments.

We must now estimate the terms conditioned on A_N^c. When $n/N \to 0$, for sufficiently large N,

$$\begin{aligned}
&\mathbb{E}\left[\frac{X_1(X_1-1)\dots(X_1-k+1)\widetilde{S}_N\dots(\widetilde{S}_N-n+k+1)}{(X_1+\widetilde{S}_N)\dots(X_1+\widetilde{S}_N-n+1)} \mathbf{1}_{S_N \geq N} \middle| A_N^c \right] \\
&\qquad \leq \mathbb{E}\left[\frac{X_1(X_1-1)\dots(X_1-k+1)}{(X_1+\widetilde{S}_N-n+k)\dots(X_1+\widetilde{S}_N-n+1)} \mathbf{1}_{S_N \geq N} \middle| A_N^c \right] \\
&\qquad\qquad \leq \mathbb{E}\left[\frac{X_1(X_1-1)}{S_N^2} \mathbf{1}_{S_N \geq N} \middle| A_N^c \right] \left(1 + \mathcal{O}\left(\frac{(n-k)k}{N} \right) \right).
\end{aligned}$$

The error term is calculated by partitioning over $X_1 \leq N$ and $X_1 \geq N$. In the first case, each of the k 'spare terms' $(X_1-j)/(X_1+\widetilde{S}_N-n+j)$, is bounded by $N/(N-(n-k))$; if $X_1 > N$ each such term is bounded by $X_1/(X_1+n-k)$. Now using the elementary facts that $(i+b)^k/i^k = 1 + \mathcal{O}(kb/i)$ and $(i-b)^k/i^k = 1 - \mathcal{O}(kb/i)$, we recover the error term above.

We can now see that the final two error terms in (4.7.4) and (4.7.5) come from (4.6.9) in the case $\psi(N)/N \to K$ and from (4.6.10) when $\psi(N)/N \to \infty$.

□

Corollary 4.7.11. *If $k^2/N \not\to 0$ we will not see convergence to the limiting rates suggested by Proposition 4.4.3.*

In particular, we see from the proof above that the $\mathcal{O}(k^2/L)$ error term is always be negative. Therefore, when the limiting Λ-coalescent has mergers of at least $\sqrt{N}$ lineages

we will be see a reduced number of such mergers, and so a different topology, in the coalescent tree corresponding to the prelimiting model.

We now present some heuristic calculations to establish how large the sample size must be in order for there to be an $\mathcal{O}(1)$ probability of seeing a merger of at least $\sqrt{N}$ lineages in the limiting Λ-coalescent. Since

$$\lambda_{m,k} = \int_0^1 \frac{x^{k-\alpha-1}(1-x)^{m-k+\alpha-1}}{\Gamma(\alpha)\Gamma(2-\alpha)}\mathrm{d}x = \frac{\Gamma(k-\alpha)\Gamma(m-k+\alpha)}{\Gamma(m)\Gamma(\alpha)\Gamma(2-\alpha)},$$

the total rate of k mergers within m lineages is given by

$$\begin{aligned}
\binom{m}{k}\lambda_{m,k} =& \binom{m}{k}\frac{\Gamma(k-\alpha)\Gamma(m-k+\alpha)}{\Gamma(m)\Gamma(\alpha)\Gamma(2-\alpha)} \\
=& \frac{m!}{k!(m-k)!}\frac{\Gamma(k-\alpha)\Gamma(m-k+\alpha)}{(m-1)!\Gamma(\alpha)\Gamma(2-\alpha)} \\
>& \frac{m!}{k!(m-k)!}\frac{\Gamma(k-1)\Gamma(m-k+2)}{(m-1)!\Gamma(\alpha)\Gamma(2-\alpha)}\frac{1}{(k-1)^{\alpha-1}(m-k+2)^{2-\alpha}} \\
=& \frac{m!}{k!(m-k)!}\frac{(k-2)!(m-k+1)!}{(m-1)!\Gamma(\alpha)\Gamma(2-\alpha)}\frac{1}{(k-1)^{\alpha-1}(m-k+2)^{2-\alpha}} \\
=& \frac{m(m-k+1)}{k(k-1)(k-1)^{\alpha-1}(m-k+2)^{2-\alpha}}\frac{1}{\Gamma(\alpha)\Gamma(2-\alpha)},
\end{aligned}$$

where we have used Gautschi's inequality, Gautschi (1959/60),

$$x^{1-s} < \frac{\Gamma(x+1)}{\Gamma(x+s)} < (x+1)^{1-s},$$

for $x > 0$ and $s \in (0,1)$. Summing over $k = 1, \ldots, m$, and approximating the sum by an integral, since $k = \sqrt{N} \to \infty$, the total rate of mergers involving more than k individuals can be approximated (up to a constant) by

$$\int_k^m \frac{m(m-x)^{\alpha-1}}{x^{\alpha+1}}\mathrm{d}x = \int_0^{\frac{m-k}{k}} u^{\alpha-1}\mathrm{d}u = \frac{1}{\alpha}\left(\frac{m-k}{k}\right)^{\alpha},$$

where we have used the substitution $u = \frac{m-x}{x}$.

Now consider the number of ancestral lineages present in the Lambda coalescent. This has been well studied. For example, we recall the following result from Berestycki et al. (2008).

Theorem 4.7.1 (Theorem 1.1, Berestycki et al. (2008))**.** *Let N_t be the remaining number of blocks in the Beta-coalescent, then*

$$\lim_{t\to 0} t^{\frac{1}{\alpha-1}} N_t = (\alpha\Gamma(\alpha))^{\frac{1}{\alpha-1}},$$

almost surely.

We now present some simple heuristic arguments. We will be looking at small time results and so, with an abuse of notation, we take

$$N_t = \left(\frac{\alpha\Gamma(\alpha)}{t}\right)^{\frac{1}{\alpha-1}}.$$

We then see that the expected number of mergers involving more than k individuals when started from m individuals is approximated by

$$\int_{\frac{\alpha\Gamma(\alpha)}{m^{\alpha-1}}}^{\frac{\alpha\Gamma(\alpha)}{k^{\alpha-1}}} \frac{1}{\alpha}\left(\frac{\left(\frac{\alpha\Gamma(\alpha)}{t}\right)^{\frac{1}{\alpha-1}} - k}{k}\right)^{\alpha} \mathrm{d}t = \mathcal{O}\left(\frac{m}{k^{\alpha}}\right).$$

The integral is dominated by the contribution for times arbitrarily close to the starting time, so as soon as $m = \mathcal{O}(N^{\alpha/2})$ we expect to see mergers of at least $\sqrt{N}$ lineages at arbitrarily small times.

Since the external branches of a Beta-coalescent make up a positive proportion of the total branch length even in the limit (unlike in the Kingman case in which the ratio of external to internal branched tends to zero), such mergers can be expected to influence our statistic $(M_1(n), \sum_{j\geq k} M_j(n))$ for sample sizes of $\mathcal{O}(N^{\alpha/2})$, as claimed.

Bibliography

D. Aldous. Stopping times and tightness. *Ann. Probab.*, 6, 1978.

E. Árnason. Mitochondrial cytochrome *b* variation in the high-fecundity Atlantic cod: trans-Atlantic clines and shallow gene genealogy. *Genetics*, 166:1871–1885, 2004.

E. Árnason and K. Halldórsdóttir. Nucleotide variation and balancing selection at the *Ckma* gene in Atlantic cod: analysis with multiple merger coalescent models. *PeerJ*, 3:e786, 2015. doi: 10.7717/peerj.786. URL http://dx.doi.org/10.7717/peerj.786.

A. Barnosky. Distinguishing the effects of the red queen and court jester on miocene mammal evolution in the northern rocky mountains. *Journal of Vertebrate Paleontology*, 21(1):172–185, 2001.

N. H. Barton, A. M. Etheridge, and A. Véber. A new model for evolution in a spatial continuum. *Electron. J. Probab.*, 15:162–216, 2010.

N. H. Barton, A. M. Etheridge, and A. Véber. Modelling evolution in a spatial continuum. *J. Stat. Mech.*, page PO1002, 2013.

A. T. Beckenbach. Mitochondrial haplotype frequencies in oysters: neutral alternatives to selection models. In B Golding, editor, *Non-neutral Evolution*, pages 188–198. Chapman & Hall, New York, 1994.

J. Benton. The red queen and the court jester: Species diversity and the role of biotic and abiotic factors through time. *Science*, 323(5915):728–732, 2009.

J. Berestycki, N. Berestycki, and J. Schweinsberg. Beta-coalescents and continuous stable random trees. *Ann Probab*, 35:1835–1887, 2007.

J. Berestycki, N. Berestycki, and J. Schweinsberg. Small-time behavior of beta coalescents. *Ann. Inst. Henri Poincaré Probab. Stat.*, 44(2):214–238, 2008. ISSN 0246-0203. doi: 10.1214/07-AIHP103. URL https://doi.org/10.1214/07-AIHP103.

J. Berestycki, N. Berestycki, and V. Limic. Asymptotic sampling formulae for Λ-coalescents. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 50(3): 715–731, August 2014. doi: 10.1214/13-aihp546.

N. Berestycki. Recent progress in coalescent theory. *Ensaios Mathématicos*, 16:1–193, 2009.

N. Berestycki, A. M. Etheridge, and A. Véber. Large-scale behaviour of the spatial Λ-Fleming-Viot process. *Ann. Inst. H. Poincaré*, 49(2):374–401, 2013.

J. Bertoin and J-F. Le Gall. Stochastic flows associated to a coalescent process. *Prob. Theor. Rel. Fields*, 126:261–288, 2003.

J. Bertoin and J-F. Le Gall. Stochastic flows associated to coalescent processes III: Limit theorems. *Illinois J Math*, 50(1):147–181, 2006.

A. Bhaskar, A. G. Clark, and Y. S. Song. Distortion of genealogical properties when the sample size is very large. *PNAS*, 111:2385–2390, 2014.

M. Birkner, J. Blath, M. Capaldo, A. M. Etheridge, M. Möhle, J. Schweinsberg, and A. Wakolbinger. Alpha-stable branching and Beta-coalescents. *Elect. J. Probab.*, 10: 303–325, 2005a.

M. Birkner, J. Geiger, and G. Kersting. Branching processes in random environment—a view on critical and subcritical cases. In *Interacting stochastic systems*, pages 269–291. Springer, 2005b.

M. Birkner, J. Blath, and B. Eldon. Statistical properties of the site-frequency spectrum associated with Λ-coalescents. *Genetics*, 195:1037–1053, 2013a.

M. Birkner, J. Blath, and M. Steinrücken. Analysis of DNA sequence variation within marine species using beta-coalescents. *Theoretical Population Biology*, 87:15 – 24, 2013b. ISSN 0040-5809. doi: http://dx.doi.org/10.1016/j.tpb.2013.01.007. URL http://www.sciencedirect.com/science/article/pii/S0040580913000087. Coalescent Theory.

M. Birkner, H. Liu, and A. Sturm. Coalescent results for diploid exchangeable population models. *ArXiv e-prints*, 2018.

N. Biswas, A. Etheridge, and A. Klimek. The spatial Lambda-Fleming-Viot process with fluctuating selection. *ArXiv e-prints*, February 2018.

J. D. G. Boom, E. G. Boulding, and A. T. Beckenbach. Mitochondrial DNA variation in introduced populations of Pacific oyster, *Crassostrea gigas*, in British Columbia. *Canadian Journal of Fisheries and Aquatic Sciences*, 51(7):1608–1614, 1994.

A. Caliebe, R. Neininger, M. Krawczak, and U. Rösler. On the length distribution of external branches in coalescence trees: Genetic diversity within species. *Theoretical Population Biology*, 72(2):245–252, September 2007. doi: 10.1016/j.tpb.2007.05.003.

M. F. Canino, I. B. Spies, K. M. Cunningham, L. Hauser, and W. S. Grant. Multiple ice-age refugia in pacific cod, gadus macrocephalus. *Molecular Ecology*, 19(19):4339–4351, September 2010. doi: 10.1111/j.1365-294x.2010.04815.x.

C. Cannings. The latent roots of certain Markov chains arising in genetics: A new approach, I. Haploid models. *Adv Appl Probab*, 6:260–290, 1974.

C. Cannings. The latent roots of certain markov chains arising in genetics: A new approach, II. further haploid models. *Advances in Applied Probability*, 7(2):264, jun 1975. doi: 10.2307/1426077. URL https://doi.org/10.2307/1426077.

S. M. Carr and H. D. Marshall. Intraspecific phylogeographic genomics from multiple complete mtDNA genomes in atlantic cod (gadus morhua): Origins of the "codmother, " transatlantic vicariance and midglacial population expansion. *Genetics*, 180(1): 381–389, August 2008. doi: 10.1534/genetics.108.089730.

J. Chetwynd-Diggle and A. M. Etheridge. SuperBrownian motion and the spatial Lambda-Fleming-Viot process. *Electron. J. Probab.*, 23:36 pp., 2018.

J. Chetwynd-Diggle and A. Klimek. Rare mutations in the spatial Lambda-Fleming-Viot model in a fluctuating environment and SuperBrownian Motion. *arXiv e-prints*, January 2019.

J. T. Cox and E. A. Perkins. Rescaled Lotka-Volterra models converge to superBrownian motion. *Ann. Probab.*, 33:904–947, 2005.

J. T. Cox, R. Durrett, and E. A. Perkins. Rescaled particle systems converging to superBrownian motion. In *Perplexing problems in probability: Progress in probability 44*, pages 269–284. Birkhäuser, Boston, MA, 1999.

J. T. Cox, R. Durrett, and E. A. Perkins. Rescaled voter models converge to super-Brownian motion. *Ann. Probab.*, 28:185–234, 2000.

D. Crisan. Superprocesses in a Brownian environment. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 460(2041):243–270, 2004. ISSN 1364-5021.

I. Dahmer and G. Kersting. The internal branch lengths of the Kingman coalescent. *The Annals of Applied Probability*, 25(3):1325–1348, June 2015. doi: 10.1214/14-aap1024.

I. Dahmer and G. Kersting. The total external length of the evolving Kingman coalescent. *Probability Theory and Related Fields*, 167(3):1165–1214, Apr 2017. ISSN 1432-2064. doi: 10.1007/s00440-016-0703-7.

I. Dahmer, G. Kersting, and A. Wakolbinger. The total external length of Beta-coalescents. *Comb Prob Comp*, 23:1010–1027, 2014.

D. A. Darling and M. Kac. On occupation times for Markoff processes . *Trans. Amer. Math. Soc.*, 84:444–458, 1957.

D. A. Dawson. Stochastic evolution equations and related measure processes. *J. Multivar. Anal.*, 5:1–52, 1975.

D. A. Dawson. Geostochastic calculus. *Can. J. Statistics*, 6:143–168, 1978.

D. A. Dawson. Measure-valued Markov processes. In *École d'été de probabilités de Saint Flour*, volume 1541. Springer-Verlag, 1993.

P. Donnelly and T. G. Kurtz. A countable representation of the Fleming-Viot measure-valued diffusion. *Ann. Probab.*, 24:698–742, 1996.

P. Donnelly and T. G. Kurtz. Particle representations for measure-valued population models. *Ann. Probab.*, 27:166–205, 1999.

N. El Karoui and S. Roelly. Proprietés de martingales, explosion et representation de Lévy-Khinchine d'une classe du processus de branchement à valeurs mesures. *Stoch. Proc. Appl.*, 38(2):239–266, 1991.

B. Eldon and J. Wakeley. Coalescent processes when the distribution of offspring number among individuals is highly skewed. *Genetics*, 172(4):2621–2633, 2006. ISSN 0016-6731. doi: 10.1534/genetics.105.052175. URL http://www.genetics.org/content/172/4/2621.

B. Eldon, M. Birkner, J. Blath, and F. Freund. Can the site-frequency spectrum distinguish exponential population growth from multiple-merger coalescents. *Genetics*, 199:841–856, 2015.

A. M. Etheridge. *An introduction to superprocesses*, volume 20 of *University lecture notes*. Amer. Math. Soc., 2000.

A. M. Etheridge. Drift, draft and structure: some mathematical models of evolution. *Banach Center Publ.*, 80:121–144, 2008.

A. M. Etheridge. *Some mathematical models from population genetics*, volume 2012 of *Lecture Notes in Mathematics*. Springer, Heidelberg, 2011. ISBN 978-3-642-16631-0. doi: 10.1007/978-3-642-16632-7. URL http://dx.doi.org/10.1007/978-3-642-16632-7. Lectures from the 39th Probability Summer School held in Saint-Flour, 2009, École d'Été de Probabilités de Saint-Flour. [Saint-Flour Probability Summer School].

A. M. Etheridge and T. G. Kurtz. Genealogical constructions of population models. *Annals of Probability*, 2018.

A. M. Etheridge, A. Véber, and F. Yu. Rescaling limits of the spatial Lambda-Fleming-Viot process with selection. *arXiv:1406.5884*, 2014.

A. M. Etheridge, N. Freeman, S. Penington, and D. Straulino. Branching Brownian motion and selection in the Spatial Lambda-Fleming-Viot process. *Ann. Appl. Probab. arXiv:1512.03766*, 2017.

S. N. Ethier and T. G. Kurtz. *Markov processes: characterization and convergence.* Wiley, 1986.

S. N. Evans. Coalescing markov labelled partitions and a continuous sites genetics model with infinitely many types. *Ann. Inst. Henri Poincaré*, 33:339–358, 1997.

N. Freeman. *Super-Brownian motion is a scaling limit of bursting processes.* Chapter 4 of Transfer of Status Thesis, University of Oxford, 2010.

Y-X. Fu. Statistical properties of segregating sites. *Theor Popul Biol*, 48:172–197, 1995.

Y-X. Fu. Exact coalescent for the Wright-Fisher model. *Theor Popul Biol*, 69:385–394, 2006.

F. Galton and H. W. Watson. On the probability of the extinction of families. *The Journal of the Anthropological Institute of Great Britain and Ireland*, 4:138–144, 1875. ISSN 09595295. URL http://www.jstor.org/stable/2841222.

W. Gautschi. Some elementary inequalities relating to the gamma and incomplete gamma function. *J. Math. and Phys.*, 38:77–81, 1959/60. ISSN 0097-1421. doi: 10.1002/sapm195938177. URL https://doi.org/10.1002/sapm195938177.

I. M. Gel'fand and G. E. Shilov. *Generalised Functions: Volume 1. Properties and Operations. Translated by E Saletan.* Academic Press, 1964.

A. Gupta. Stochastic model for cell polarity. *The Annals of Applied Probability*, 22(2): 827–859, 2012.

T. E. Harris. Contact interactions on a lattice. *Ann. Probab.*, 2(6):969–988, 12 1974. doi: 10.1214/aop/1176996493. URL https://doi.org/10.1214/aop/1176996493.

D. Hedgecock. Does variance in reproductive success limit effective population sizes of marine organisms? In A R Beaumont, editor, *Genetics and evolution of marine organisms*, pages 122–134. Chapman and Hall, London, 1994.

E. Hille. *Functional Analysis and Semi-Groups.* American Mathematical Society Colloquium Publications, vol. 31. American Mathematical Society, New York, 1948.

R. Van der Hofstad and A. Sakai. Convergence of the critical finite range contact process to superbrownian motion above the upper critical dimension: The higher point functions. *Electron. J. Probab.*, 15:801–894, 2010.

R. Van der Hofstad and G. Slade. Convergence of critical oriented percolation to superbrownian motion above 4+1 dimensions. *Ann. Inst. Henri Poincaré Probab. Stat.*, 39:413–485, 2003.

R. Van der Hofstad, M. Holmes, and E. A. Perkins. A criterion for convergence to superBrownian motion on path space. *Ann. Probab.*, 45:278–376, 2017.

M. Holmes. Convergence of lattice trees to superBrownian motion above the critical dimension. *Elect. J. Prob.*, 13:671–755, 2008.

J. Jacod and A. N. Shiryaev. *Limit theorems for stochastic processes*, volume 288 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, second edition, 2003. ISBN 3-540-43932-3. doi: 10.1007/978-3-662-05265-5. URL http://dx.doi.org/10.1007/978-3-662-05265-5.

S. Janson and G. Kersting. On the total external length of the Kingman coalescent. *Electronic Journal of Probability*, 16(0):2203–2218, 2011. doi: 10.1214/ejp.v16-955.

J. Kelleher, A. M. Etheridge, and G. McVean. Efficient coalescent simulation and genealogical analysis for large sample sizes. *PLOS Computational Biology*, 12(5): e1004842, may 2016. doi: 10.1371/journal.pcbi.1004842. URL http://dx.doi.org/10.1371/journal.pcbi.1004842.

G. Kersting, J. C. Pardo, and A. Siri-Jégousse. Total internal and external lengths of the Bolthausen-Sznitman coalescent. *Journal of Applied Probability*, 51(A):73–86, December 2014. doi: 10.1239/jap/1417528468.

H. Kesten and B. P. Stigum. A limit theorem for multidimensional galton-watson processes. *Ann. Math. Statist.*, 37(5):1211–1223, 10 1966. doi: 10.1214/aoms/1177699266. URL http://dx.doi.org/10.1214/aoms/1177699266.

M. Kimura. The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. *Genetics*, 61(4):893–903, 1969. ISSN 0016-6731.

M. Kimura and G. H. Weiss. The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics*, 49(4):561, 1964.

J. F. C. Kingman. The coalescent. *Stoch Proc Appl*, 13:235–248, 1982a.

J. F. C. Kingman. Exchangeability and the evolution of large populations. In G Koch and F Spizzichino, editors, *Exchangeability in Probability and Statistics*, pages 97–112, Amsterdam, 1982b. North-Holland.

J. F. C. Kingman. On the genealogy of large populations. *Journal of Applied Probability*, 19:27–43, 1982c. ISSN 00219002. URL http://www.jstor.org/stable/3213548.

W. H. Kliemann, G. Koch, and F. Marchetti. On the unnormalized solution of the filtering problem with counting process observations. *IEEE Transactions on Information Theory*, 36(6):1415–1425, 1990.

A. Kong, M. L. Frigge, G. Masson, S. Besenbacher, P. Sulem, G. Magnusson, S. A. Gudjonsson, A. Sigurdsson, A. Jonasdottir, A. Jonasdottir, W. S. W. Wong, G. Sigurdsson, G. B. Walters, S. Steinberg, H. Helgason, G. Thorleifsson, D. F. Gudbjartsson, A. Helgason, O. Th. Magnusson, U. Thorsteinsdottir, and K. Stefansson. Rate of de novo mutations and the importance of father's age to disease risk. *Nature*, 488(7412): 471–475, August 2012. doi: 10.1038/nature11396.

J. Koskela. Multi-locus data distinguishes between population growth and multiple merger coalescents. *Stat. Appl. Genet. Mol. Biol.*, 17(3):20170011, 21, 2018. ISSN 2194-6302. doi: 10.1515/sagmb-2017-0011. URL https://doi.org/10.1515/sagmb-2017-0011.

S. Kumar and S. Subramanian. Mutation rates in mammalian genomes. *Proceedings of the National Academy of Sciences*, 99(2):803–808, January 2002. doi: 10.1073/pnas.022629899.

T. G. Kurtz. A limit theorem for perturbed operator semigroups with applications to random evolutions. *Journal of Functional Analysis*, 12(1):55–67, 1973.

T. G. Kurtz. Averaging for martingale problems and stochastic approximation. In *Applied Stochastic Analysis*, volume 177. Springer, Berlin, 1992.

T. G. Kurtz. Martingale problems for conditional distributions of Markov processes. *Elect. J. Prob.*, 3, paper 9, 1998.

T. G. Kurtz. Particle representations for measure-valued population processes with spatially varying birth rates. In *Stochastic models (Ottawa, ON, 1998)*, volume 26 of *CMS Conf. Proc.*, pages 299–317. Amer. Math. Soc., Providence, RI, 2000.

T. G. Kurtz. Equivalence of stochastic equations and martingale problems. In *Stochastic analysis 2010*, pages 113–130. Springer, 2010.

T. G. Kurtz and E. Rodrigues. Poisson representations of branching markov and measure-valued branching processes. *Ann. Probab.*, 39(3):939–984, 05 2011. doi: 10.1214/10-AOP574. URL https://doi.org/10.1214/10-AOP574.

T. G. Kurtz, R. Lyons, R. Pemantle, and Y. Peres. *A Conceptual Proof of the Kesten-Stigum Theorem for Multi-Type Branching Processes*, pages 181–185. Springer New York, New York, NY, 1997. ISBN 978-1-4612-1862-3. doi: 10.1007/978-1-4612-1862-3_14. URL http://dx.doi.org/10.1007/978-1-4612-1862-3_14.

A. Lambert and E. Schertzer. Recovering the Brownian coalescent point process from the Kingman coalescent by conditional sampling. *arXiv:1611.01323[math.PR]*, 2016.

G. Li and D. Hedgecock. Genetic heterogeneity, detected by PCR-SSCP, among samples of larval Pacific oysters (*Crassostrea gigas*) supports the hypothesis of large variance in reproductive success. *Can. J. Fish. Aquat. Sci.*, 55(4):1025–1033, apr 1998. doi: 10.1139/f97-312. URL http://dx.doi.org/10.1139/f97-312.

A. J. Lotka. The growth of mixed populations: Two species competing for a common food supply. *Journal of the Washington Academy of Sciences*, 22(16/17):461–469, 1932. ISSN 00430439. URL http://www.jstor.org/stable/24530449.

A. Lunardi. *Analytic semigroups and optimal regularity in parabolic problems*. Springer Science & Business Media, 2012.

C. Ma. Construction of non-gaussian random fields with any given correlation structure. *Journal of Statistical Planning and Inference*, 139(3):780–787, 2009.

A. W. May. Fecundity of Atlantic cod. *J Fish Res Brd Can*, 24:1531–1551, 1967.

A. Melfi and D. Viswanath. The validity of the coalescent approximation for large samples. *bioRxiv*, 2017. doi: 10.1101/170928. URL https://www.biorxiv.org/content/early/2017/08/01/170928.

A. Melfi and D. Viswanath. Single and simultaneous binary mergers in Wright-Fisher genealogies. *Theoretical Population Biology*, 121:60–71, May 2018a. doi: 10.1016/j.tpb.2018.04.001.

A. Melfi and D. Viswanath. The Wright–Fisher site frequency spectrum as a perturbation of the coalescent's. *Theoretical Population Biology*, 124:81–92, December 2018b. doi: 10.1016/j.tpb.2018.09.005.

L. R. Miller. *Evolution of highly fecund organisms*. PhD thesis, University of Oxford, 2015.

M. Möhle. Robustness results for the coalescent. *Journal of Applied Probability*, 35(02): 438–447, June 1998. doi: 10.1017/s0021900200015060.

M. Möhle. Ancestral processes in population genetics—the coalescent. *Journal of Theoretical Biology*, 204(4):629–638, jun 2000. doi: 10.1006/jtbi.2000.2032.

M. Möhle. Asymptotic results for coalescent processes without proper frequencies and applications to the two-parameter Poisson–Dirichlet coalescent. *Stochastic Processes and their Applications*, 120(11):2159–2173, November 2010. doi: 10.1016/j.spa.2010.07.004.

M. Möhle and S. Sagitov. A classification of coalescent processes for haploid exchangeable population models. *Ann. Probab.*, 29(4):1547–1562, 2001. ISSN 0091-1798. doi: 10.1214/aop/1015345761. URL https://doi.org/10.1214/aop/1015345761.

L Mytnik. Superprocesses in random environments. *Ann. Probab.*, 24:1953–1978, 1996.

M. Nakashima. Branching random walks in random environment and superbrownian motion in random environment. *Ann. Inst. H. Poincaré Probab. Statist.*, 51(4): 1251–1289, 11 2015.

C. Neuhauser and S. W. Pacala. An explicitly spatial version of the Lotka-Volterra model with interspecific competition. *Ann. Appl. Probab.*, 9(4):1226–1259, 1999. ISSN 1050-5164. doi: 10.1214/aoap/1029962871. URL https://doi.org/10.1214/aoap/1029962871.

H-S. Niwa, K. Nashida, and T. Yanagimoto. Reproductive skew in japanese sardine inferred from DNA sequences. *ICES Journal of Marine Science: Journal du Conseil*, 73(9):2181–2189, may 2016. doi: 10.1093/icesjms/fsw070. URL http://dx.doi.org/10.1093/icesjms/fsw070.

E. A. Perkins. Dawson-Watanabe Superprocesses and Measure-valued diffusions. In *École d'été de probabilités de Saint Flour XXIX-1999*. Springer-Verlag, 2002.

P. Pfaffelhuber and A. Wakolbinger. The process of most recent common ancestors in an evolving coalescent. *Stochastic Processes and their Applications*, 116(12):1836–1859, December 2006. doi: 10.1016/j.spa.2006.04.015.

P. Pfaffelhuber, A. Wakolbinger, and H. Weisshaupt. The tree length of an evolving coalescent. *Probability Theory and Related Fields*, 151(3-4):529–557, June 2010. doi: 10.1007/s00440-010-0307-6.

J. Pitman. Coalescents with multiple collisions. *Ann. Probab.*, 27(4):1870–1902, 10 1999. doi: 10.1214/aop/1022874819. URL http://dx.doi.org/10.1214/aop/1022874819.

S. Sagitov. The general coalescent with asynchronous mergers of ancestral lines. *J Appl Probab*, 36:1116–1125, 1999.

O. Sargsyan and J. Wakeley. A coalescent process with simultaneous multiple mergers for approximating the gene genealogies of many marine organisms. *Theor Pop Biol*, 74:104–114, 2008.

J. Schweinsberg. Coalescents with simultaneous multiple collisions. *Electron. J. Probab.*, 5:1–50, 2000.

J. Schweinsberg. Coalescent processes obtained from supercritical Galton-Watson processes. *Stochastic Process. Appl.*, 106(1):107–139, 2003. ISSN 0304-4149. URL http://www.sciencedirect.com/science?_ob=GatewayURL&_origin=MR&_method=citationSearch&_piikey=s0304414903000280&_version=1&md5=438c179f2400fbe4f5f43d781809d0f6.

R. Stockbridge. Portfolio optimization in markets having stochastic rates. In Bozenna Pasik-Duncan, editor, *Stochastic Theory and Control*, pages 447–458, Berlin, Heidelberg, 2002. Springer Berlin Heidelberg.

A. Sturm. On convergence of population processes in random environments to the stochastic heat equation with colored noise. *Electron. J. Probab.*, 8:39 p., 2003.

J. E. Taylor. The genealogical consequences of fecundity variance polymorphism. *Genetics*, 182(3):813–837, 2009.

A. Tenesa, P. Navarro, B. J. Hayes, D. L. Duffy, G. M. Clarke, M. E. Goddard, and P. M. Visscher. Recent human effective population size estimated from linkage disequilibrium. *Genome Research*, 17(4):520–526, March 2007. doi: 10.1101/gr.6023607.

T. F. Turner, J. P. Wares, and J. R. Gold. Genetic effective size is three orders of magnitude smaller than adult census size in an abundant, estuarine-dependent marine fish (sciaenops ocellatus). *Genetics*, 162(3):1329–1339, 2002. ISSN 0016-6731.

A. Véber and A. Wakolbinger. The spatial Lambda-Fleming-Viot process: an event based construction and a lookdown representation. *Ann. Inst. H. Poincaré*, 51:570–598, 2015.

V. Volterra. Variations and fluctuations of the numbers of individuals in animal species living together. *Journal du Conseil. Conseil Permanent International pour l'Exploration de la Mer*, 3:3–51, 01 1928. doi: 10.1093/icesjms/3.1.3.

J. Wakeley and O. Sargsyan. Extensions of the coalescent effective population size. *Genetics*, 181(1):341–345, November 2008. doi: 10.1534/genetics.108.092460.

J. Wakeley and T. Takahashi. Gene Genealogies When the Sample Size Exceeds the Effective Size of the Population. *Molecular Biology and Evolution*, 20(2):208–213, 02 2003. ISSN 0737-4038. doi: 10.1093/molbev/msg024. URL https://doi.org/10.1093/molbev/msg024.

S. Watanabe. A limit theorem of branching processes and continuous state branching processes. *J. Math. Kyoto Univ.*, 8:141–167, 1968.

G. A. Watterson. On the number of segregating sites in genetical models without recombination. *Theor Pop Biol*, 7:256–276, 1975.

K. Yosida. On the differentiability and the representation of one-parameter semi-group of linear operators. *J. Math. Soc. Japan*, 1:15–21, 1948. ISSN 0025-5645. doi: 10.2969/jmsj/00110015. URL http://dx.doi.org/10.2969/jmsj/00110015.