

## Part II

# Gene flow in biparental models of inheritance

## Chapter 5

# Introduction

In a population of sexually reproducing organisms, genealogical relatedness is not equivalent to genetic relatedness. An organism will typically inherit only a small proportion of their genes from each of their genealogical ancestors, at least over those portions of the genome subject to recombination. We are interested in patterns both of genealogical and genetic inheritance, which we study in a model of sexually reproducing organisms. Regardless of ploidy, we treat individuals as haploid in the usual manner by calling each gamete in a diploid organism an “individual”. *Genetic inheritance* depends on which part of the genome we are examining, and while knowledge of *genealogical inheritance* determines probabilities of genetic inheritance, genealogically related individuals may not be genetically related at all.

Here is one scheme for keeping track of genetic inheritance within a genealogy. Imagine that we put “marks” at a few loci on the genome of a few individuals, and as time progresses, each copy made of their DNA preserves those marks (regardless of mutation, etcetera), so that every locus directly (genetically) descended from an originally marked locus is also marked. When two individuals mate, each of their offspring inherits a patchwork of marked and unmarked loci, stitched together from the parents’ patterns of marks by recombination.

Now imagine that we want to run time in reverse — put marks on the genomes of a certain set of individuals, and as we move back in time mark those bits of genome that are the ancestors of the present-day marked genome. Suppose that at some time during this process we discover that one particular individual we are following is the offspring of two other individuals (as we observe her shrink back into a seed, and then the pollination event that produced that seed, say). Then, in order to keep our accounting correct, we must distribute her marks over her parents’ genomes depending on where the whimsy of recombination chose to copy each bit of genome from. Some individuals may then have loci with more than one mark — the number of marks indicates the number of modern-day genetic descendants of that individual with a mark at that locus.

These two processes are equivalent, in the sense that we can answer forwards-in-time ques-

tions using the reverse-time process, and vice-versa.

Here is another scheme for tracking inheritance: suppose we mark as above a large number of unlinked<sup>1</sup> loci, and we are only interested in the *proportion* of loci that have marks. Rather than a cut-and-paste patchwork, recombination takes an average of the parents' proportions. If we visualize this proportion as the level in a partially filled bucket, when two parents produce an offspring, the offspring's bucket will be filled to the level midway between the parents' levels. A cartoon of this process is depicted in Figure 5.1.

If we follow this accounting scheme backwards in time, when we discover who the parents of a certain individual were, we empty the offspring's bucket in two equal parts into the parents' buckets (which we assume are large enough to not overflow). An example of this process with the same sequence of reproductive events is depicted in Figure 5.2.

In the “bucket” model of inheritance, what does the level of the bucket mean? In forwards time, the level of an individual's bucket is the probability that at a given locus, she has descended from one of the originally marked individuals. Equivalently, it is the expected number of genetic ancestors she has among the originally marked individuals at a randomly chosen locus. These two interpretations are equivalent, since each individual has exactly one genetic ancestor at each locus at any given time in the past. The levels all stay between 0 and 1 as the dynamics progress, as is required by this interpretation. In reverse time, the level of an individual's bucket is her expected number of genetic descendants at a randomly chosen locus that are marked in the present day.

One consequence of this observation is that the proportion of genes that one individual has inherited from another can be found by looking either backwards or forwards in time. This can be seen in Figures 5.1 and 5.2: the 7<sup>th</sup> proportion at the bottom of Figure 5.1 is equal to the 4<sup>th</sup> proportion at the top of Figure 5.2.

Both modes of inheritance tracking, with arbitrary recombination patterns, fit into a common framework. Call our collection of loci  $U$ , which for definiteness we take to be a compact subset of the real numbers. We associate with each locus  $u \in U$  a nonnegative real number  $x(u)$ , and in the above examples interpret  $x(u)$  as either the number of marks, or the level of the bucket respectively, at  $u$  (an individual could have more than one bucket if we had kept track of different chromosomes separately). If individuals with values  $x(u)$  and  $y(u)$  at  $u$  mate and produce an offspring with value  $z(u)$ , recombination preserves the total amount of genome — “everything in  $z$  had to come from somewhere” — so for each locus  $u$  there is a number  $0 \leq r(u) \leq 1$  such that  $z(u) = r(u)x(u) + (1 - r(u))y(u)$ .

In the “marked genome” example,  $r(u)$  will be either 0 or 1 — the genome at each locus is inherited from either one parent or the other. If we are keeping track of proportions,  $r(u) = \frac{1}{2}$  always, since half the loci come from each parent. This yields a general method for keeping account

---

<sup>1</sup> *Unlinked* means that each is passed on to the offspring with probability  $\frac{1}{2}$ , independently of the others.

of inheritance in offspring in forwards time:

$$z(u) = R(u)x(u) + (1 - R(u))y(u), \quad (5.0.1)$$

where now, since recombination is random, we take  $R$  to be a random function from  $U$  into  $[0, 1]$ , whose distribution is determined by the rules of recombination. This can be stated more simply, as  $z = Rx + (1 - R)y$ , where multiplication of functions is done pointwise. We will assume that  $R$  and  $(1 - R)$  have the same distribution, as we are not distinguishing between the sexes.

In reverse time, if we discover that individuals with states  $x$  and  $y$  are the parents of  $z$ , we update  $x$  to  $x + Rz$  and  $y$  to  $y + (1 - R)z$ , and set  $z = 0$  everywhere.

So far, we have only specified a scheme for tracking inheritance across the genome, once we know all demographic information: who died, and who gave birth to whom, at what times. To complete the model, we need to specify the population dynamics, which we take to be fairly simple. We will work with a continuous-time model of a neutral, randomly-mating population of fixed size  $N$ . We suppose that any two individuals can mate, and that each individual chooses to do so at the times of an independent, homogeneous Poisson process with rate 1, choosing their mate uniformly from all other individuals. Each mating event produces a random number of offspring that kill and replace other uniformly chosen individuals (not their siblings or their parents). The mate choice, number of offspring, the recombination patterns of the offspring, and which individuals are chosen to be replaced are all independent of the inheritance patterns of the individuals involved and the state of the population.

In this model, in either time direction, we will keep track of a collection of  $N$  randomly evolving functions, since each individual's pattern of inheritance is encoded as a function on  $U$ . We will refer to this function as the individual's *type*. There is one more simplification that is sometimes convenient to make, since the model is invariant under relabelings of the individuals — all that matters is the collection of their values. This information can be encoded as a probability measure on the set of functions from  $U$  to  $\mathbb{R}_{\geq 0}$ , that tells us, for instance, what proportion of the individuals have a certain pattern of inheritance.

**Note 1.** It would be straightforward to include a positive probability of selfing, or to allow the recombination patterns of the offspring in each mating event to have an arbitrary joint distribution. In view of the results of Kammerle and Mohle (summarized in Section 5.1) more general exchangeable (Cannings-type) models could be used as well. However, we do not pursue these directions in the current work.

We begin the subsequent chapters by formally defining in Section 6.1 the forwards and reverse time measure-valued processes we described above. We then introduce in Section 6.2 the branching processes which are the natural approximations for the trees of ancestors and descendants of a given individual, and use these to describe the deterministic processes which are the infinite population size limit of the finite-population processes. A surprise is that the *reverse-time* branching

process gives the most natural description of the *forwards-time* limiting process, and vice-versa. The convergence is proved in Section 7.2 by showing that the branching processes are Feynman-Kac dual processes to the processes of interest, an approach which is closely related to the *function-valued duals* discussed in [61].

In Section 6.4 we construct the forwards- and reverse-time processes jointly, and use this construction to prove that the branching processes do in fact approximate the genealogies of finite numbers of individuals, as well as pointing out an interesting duality between the two time directions. In Section 7.1 we discuss the generators of each process, and prove approximation results which are useful in other sections.

**Acknowledgements:** This part grew out of a problem discussed with Erick Matsen, Steve Evans, and Monty Slatkin, and took shape from discussions with these collaborators, especially Erick Matsen and Steve Evans, who read several drafts both early and late. Thanks also to Rick Durrett and Shankar Bhamidi, who both provided useful references and insightful comments.

## 5.1 Previous work

There has been work on several models of biparental ancestry in the same spirit as ours, although most have not explicitly incorporated accounting of genetic ancestry, and instead described which individuals are related to one another in the genealogical (not necessarily genetic) sense.

The first seems to have been Kammerle [175, 176], who introduced and studied a general family of exchangeable “bisexual” Moran models, and by analyzing the limiting Markov chains in the infinite population size limit, under certain conditions on the offspring distributions, related the probability of eventual extinction of the offspring of a given individual to the probability of extinction of a certain Galton-Watson process. Kammerle also determined the second-order behavior of this quantity by showing that looking backwards in time, the total number of individuals (suitably renormalized) who left behind descendants in the present day converges to a Ornstein-Uhlenbeck process (viewed only at integer times), and observed a relation between the forward- and reverse-time perspectives.

Mohle later extended Kammerle’s work to a somewhat more general bisexual model in which the two sexes are distinguished [229], and to the case of variable population sizes [230]. Mohle obtained similar results to those of Kammerle for these models, and also derived necessary and sufficient conditions for the almost sure eventual fixation of any allelic type. Furthermore, in an exchangeable model with fixed population size, Mohle and Sagitov [231] described the analogue to the coalescent, and verified that under certain conditions, the limiting process was indeed Kingman’s coalescent as expected.

More biological interest was drawn to biparental models when Joseph Chang made the striking observation in [43] that in a certain discrete-time biparental Wright-Fisher-type model, the number of generations back in time until there existed a most recent common ancestor of the entire

population is asymptotically exactly  $\log_2 N$ , where  $N$  is the total population size. This is in contrast striking to the uniparental situation, where it is well-known by analysis of Kingman's coalescent that the time back to the most recent common ancestor of a randomly-mating population of size  $N$  is of order  $N$ , and furthermore when rescaled by  $N$  converges to a nontrivial distribution [249]. Rohde, Olsen, and Chang [271] extended Chang's results to a finite number of randomly mating populations connected by weak migration, and through computer simulation estimated that there may have been a most recent common ancestor to all living humans around 2,500 years ago.

Chang's results on the extreme recentness of common ancestors in at least one biparental model begged the question of to what extent were such genealogical common ancestors also genetic ancestors — if there existed an individual 2,500 years ago who is an ancestor of everyone alive today, through the whim of recombination, that individual might not share genetic material with very many modern-day humans at all.

Evans & Matsen [224] proved several results in this direction, showing that the expected amount of descendant genetic material grows superlinearly with the number of genealogical descendants, and that once there appears a single common ancestor, there will soon appear many more.

We will expand on both recurrent themes of approximation by branching processes and the relationship between the forwards and reverse time perspectives. We do not in the present work resolve the question of genetic relatedness of population common ancestors, but we do present a framework for its analysis, as well as many other related questions.

The models in this work could be viewed from several different angles. We have chosen to formulate them primarily as measure-valued processes, but they could also be formulated as particle systems, or even as questions about certain random matrices. These connections are made clearer in Section 6.4, and for now we will only mention that the forwards-time model is closely related to the *smoothing* model of Holley and Liggett [149], while the reverse-time processes is closely related to the *potlatch* processes introduced by Liggett and Spitzer [212] and studied further by Holley and Liggett [149]. In particular, the associated particle systems are *monotone* in the sense described in [213].



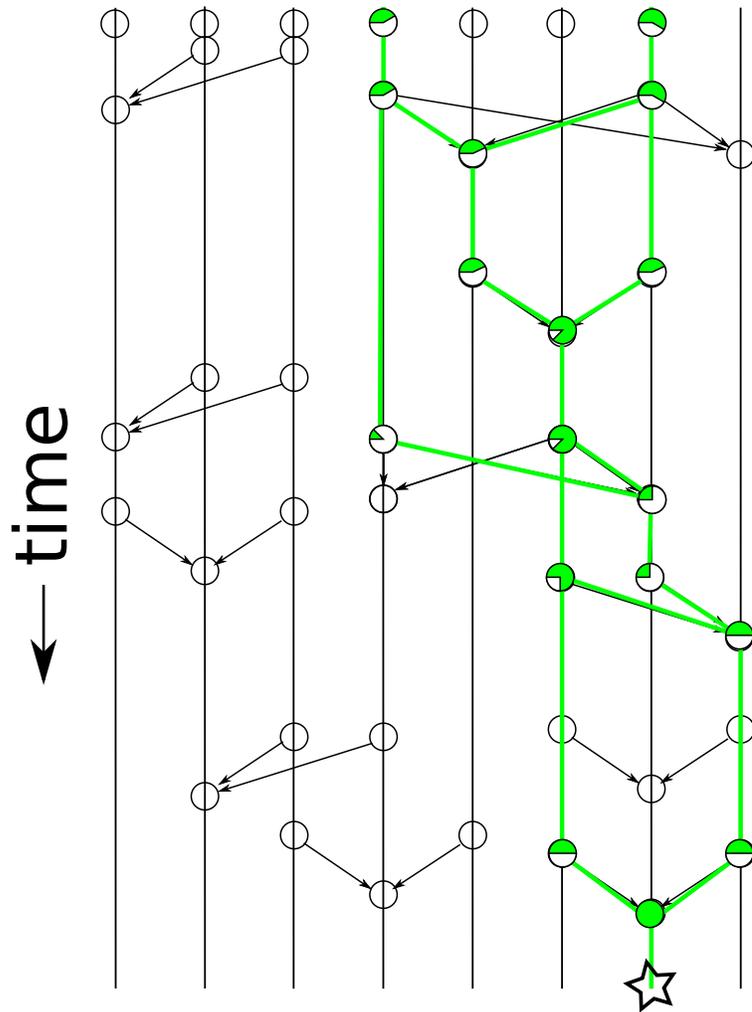


Figure 5.2: A diagram of the reverse-time process, again with the “bucket” model. The process evolves up the page, tracing backwards in time the genes belonging to the individual marked with a star. The events are the same as in Figure 5.1. Note that at each time, the sum of the proportions is constant (in this case, equal to one).

## Chapter 6

# Genetic descendants and ancestral splitting

### 6.1 Definition of the measure-valued processes

We formalize the above models as follows. Let  $U$  be a compact subset of  $\mathbb{R}$ , and let  $\sigma$  be a probability measure on  $\mathbb{R}$  whose support is equal to  $U$ . Let

$$T = \{f \in L^\infty(\sigma) : f \geq 0 \text{ } \sigma\text{-almost everywhere}\},$$

the set of (equivalence classes of)  $\sigma$ -almost everywhere bounded, nonnegative functions on  $U$ . We equip  $T$  with the weak-\* topology, so that  $f_n \rightarrow f \in T$  iff  $\int_U f_n(x)g(x)\sigma(dx) \rightarrow \int_U f(x)g(x)\sigma(dx)$  for all  $g \in L^1(\sigma)$ . Let  $\mathcal{B}(T)$  denote the Borel subsets of  $T$ . Addition and multiplication of elements of  $T$  is defined pointwise. We write  $\|f\|_\infty$  for the  $L^\infty(\sigma)$  norm on  $T$ . When appropriate, we will identify constants with constant functions on  $U$  — for instance, writing 1 for the function which is 1 everywhere on  $U$ . Let  $S \subset T$  be the set of  $f \in T$  that take values in  $[0, 1]$   $\sigma$ -almost everywhere.

Since we have assumed recombination patterns in different offspring to be independent, we need only specify the distribution of a single recombination pattern. This will be described by a random variable  $R$  which we assume takes values in  $S$  and satisfies  $R \stackrel{d}{=} (1 - R)$ . Note that for any  $x, y \in T$ ,  $\mathbb{E}[Rx + (1 - R)y] = (x + y)/2$ , by symmetry of  $R$ , and so  $\mathbb{E}[R]$  is the constant function  $1/2$ .

*Example 6.1.1.* Here are some natural examples of  $U$  and  $R$ .

1. Let  $U$  consist of a single point, and let  $R = 1/2$  (deterministically). This is the introductory “bucket” example, where we keep track only of proportions of a large number of unlinked sites.
2. Let  $U = [0, 1]$ , let  $\theta$  be a random variable taking values in  $[0, 1]$ , and given  $\theta$  with probability  $1/2$  let  $R$  take the value  $\mathbf{1}_{[0, \theta]}$ , the function which takes the value 1 on the interval  $[0, \theta]$  and

is zero otherwise; and let  $R = \mathbf{1}_{[\theta,1]}$  otherwise. This models a segment of the genome as the unit interval, which recombination breaks in a single location distributed according to  $\theta$ .

3. Let  $U = \{0, 1, 2, \dots, n\}$ , let  $R(0) = \frac{1}{2}$ , and let  $\{R(k) : 1 \leq k \leq n\}$  be a collection of independent random variables that take the value 0 or 1 with probability 1/2 each. This models the evolution of  $n$  unlinked sites *jointly* with the overall proportion of inherited genetic material.

Our processes will move in  $M_1(T)$ , the space of probability measures on  $T$ . For finite  $N$ , they will be restricted to the subset of  $M_1(T)$  that corresponds to the empirical distributions for a population of size  $N$ , and is defined as follows. Let  $\delta_x$  denote the probability measure that assigns unit mass to the singleton  $\{x\}$ , and define

$$E_N = \left\{ \mu \in M_1(T) : \exists \sigma(i) \in T \text{ s.t. } \mu = \frac{1}{N} \sum_{i=1}^N \delta_{\sigma(i)} \right\}, \quad (6.1.1)$$

the set of probability measures on  $T$  that can be written as the sum of  $N$  equally-sized atoms. Note that for  $\mu \in E_N$  and  $k \leq N$  we may unambiguously define the distribution of a *uniform sample of size  $k$  without replacement* from  $\mu$ , for instance, by ordering the atoms and choosing a uniformly distributed subset of size  $k$  of their indices.

The *offspring distribution* is a random variable  $L$  that has the distribution of the number of offspring borne to a single individual at one reproductive event. To avoid tiresome caveats, we will assume that there exists an  $L_{\max}$  such that  $1 \leq L \leq L_{\max}$  almost surely, and we will always take  $N$  to be at least  $L_{\max} + 2$ . We will in general need some sort of boundedness of  $L$ , and our results apply with little change to the proofs for at least any  $L$  such that  $\mathbb{E}[e^{\alpha L}] < \infty$  for all  $\alpha$  in some neighborhood of zero.

**Definition 6.1.1.** The *genetic descent process on  $T$  for a population of size  $N$*  is a  $E_N$ -valued process  $(D_t^N)_{t \geq 0}$  that evolves as follows. Let  $L_1, L_2, \dots$  be an infinite sequence of independent copies of  $L$ , and for each  $k \geq 1$ , conditional on  $L_k$  but independent of everything else, let  $(R_1^k, \dots, R_{L_k}^k)$  be independent copies of  $R$ .

Let  $\tau_1, \tau_2, \dots$  be the arrival times of an independent Poisson process with rate  $N$ . Let  $D_0^N \in E_N$  be given, and let  $D_t^N$  be constant except at the arrival times of the Poisson process. At arrival time  $\tau_k$ , conditioned on  $(D_t^N)_{t < \tau_k}$  and  $L_k$ , let  $(\xi_1^k, \xi_2^k, \chi_1^k, \dots, \chi_{L_k}^k)$  be a uniform sample of size  $(L_k + 2)$  drawn without replacement from  $D_{\tau_k-}^N$ . Then let

$$D_{\tau_k}^N = D_{\tau_k-}^N + \frac{1}{N} \sum_{i=1}^{L_k} \left( \delta_{R_i^k \xi_1^k + (1-R_i^k) \xi_2^k} - \delta_{\chi_i^k} \right).$$

**Definition 6.1.2.** The *genetic ancestry process on  $T$  for a population of size  $N$*  is a  $E_N$ -valued process  $(A_t^N)_{t \geq 0}$  that evolves as follows. Let  $L_1, L_2, \dots$  be an infinite sequence of independent copies of  $L$ , and for each  $k \geq 1$ , conditional on  $L_k$  but independent of everything else, let  $(R_1^k, \dots, R_{L_k}^k)$  be independent copies of  $R$ .

Let  $\tau_1, \tau_2, \dots$  be the arrival times of an independent Poisson process with rate  $N$ . Let  $A_0^N \in E_N$  be given, and let  $A_t^N$  be constant except at the arrival times of the Poisson process. At arrival time  $\tau_k$ , conditioned on  $(A_t^N)_{t < \tau_k}$  and  $L_k$ , let  $(\xi_1^k, \xi_2^k, \chi_1^k, \dots, \chi_{L_k}^k)$  be a uniform sample of size  $(L_k + 2)$  drawn without replacement from  $A_{\tau_k-}^N$ . Then let

$$A_{\tau_k}^N = A_{\tau_k-}^N + \frac{1}{N} \left( \delta_{\xi_1 + \sum_i R_i^k \chi_i} + \delta_{\xi_2 + \sum_i (1-R_i^k) \chi_i} + L_k \delta_0 - \delta_{\xi_1} - \delta_{\xi_2} - \sum_{i=1}^{L_k} \delta_{\chi_i} \right).$$

If  $Q$  is a random probability measure on  $T$ , the *mean* of  $Q$  is the quantity

$$\int_T x Q(dx).$$

Note that mean of  $A_t^N$  is invariant, while the mean of  $D_t^N$  is merely a martingale.

These processes are closely related. In Section 6.4 we make their relation more explicit through a joint construction of  $D^N$  and  $A^N$  on a bounded time interval.

## 6.2 Trees and root systems

In a randomly-mating population of size  $N$ , it is easy to see that with high probability the descendants of a single individual (or any relatively small number of individuals) will not interbreed with each other for a period of time of order  $\log N$  — so over this period of time the genealogy of those descendants will be well-approximated by the genealogy of a continuous-time Markov branching process. A similar observation holds looking backwards in time — the ancestral tree of any individual will be well-approximated by a (different) continuous-time Markov branching process, up until some time of order  $\log N$ . These observations, in their respective discrete-generation models, were a common thread in previous work on biparental models.

Given these observations, it is not surprising that these branching processes give a description of the finite-time evolution of each process in the large-population-size limit. However, it is not immediate, since while for each  $t$ , the family trees of a *finite* collection of individuals are asymptotically disjoint and independent on  $[0, t]$ , the family trees of the entire population are intermeshed.

More surprisingly, it turns out that the description of the *forwards-time* genetic descendant process is most easily given in terms of the *reverse-time* branching process of ancestors, and that correspondingly the description of the *reverse-time* ancestral recombination process is given in terms of the *forwards-time* branching process of descendants.

To describe the limiting processes, we must first define these branching processes.

**Definition 6.2.1.** The (*reverse-time*) *ancestral branching process* on  $T$  is a finite sequence of particles  $(W_1(t), \dots, W_{M_W(t)}(t))$  with types  $W_i(t) \in T$  for each  $1 \leq i \leq M_W(t)$  that evolves as follows. Each particle lives for an exponentially distributed amount of time with rate parameter  $\mathbb{E}[L]$ , after which it splits in two. At each split, an independent copy of  $R$  is produced, and if the type of the

particle that split was  $z$ , then a particle with type  $Rz$  replaces the parent, while a particle of type  $(1 - R)z$  is appended to the end of the sequence.

**Definition 6.2.2.** The (*forwards-time*) *descendant branching process on  $T$*  is a finite sequence of particles  $(Z_1(t), \dots, Z_{M_Z(t)}(t))$  with types  $Z_i(t) \in T$  for each  $1 \leq i \leq M_Z(t)$  that evolves as follows. Each particle lives for an exponentially distributed amount of time with rate parameter  $\mathbb{E}[L]$ , and upon its death it is removed from the list and the remaining particles are renumbered. Each living particle carries an independent rate 2 Poisson process, and at each arrival time of this process, the particle produces a random number of offspring with distribution  $L$  (and does not die); the offspring are appended to the end of the list. If the type of the parental particle was  $z$ , the distribution of the offspring is that of  $(R_1z, \dots, R_Lz)$ , where  $(R_1, R_2, \dots, R_L)$  are independent copies of  $R$ , independent of everything else.

**Note 2.** Both are standard continuous-time multitype branching processes. The genealogy of the ancestral branching process is the same as a Yule process with branching rate  $\mathbb{E}[L]$  — see [10] for much information about this process. The descendant branching process falls into the usual framework if we redefine it so that each individual reproduces and dies at rate  $2 + \mathbb{E}[L]$ , and has the following offspring distribution: If the parental type is  $z$ , and the offspring are  $Z'_1, \dots, Z'_K$ , then

$$\mathbb{P}\{K = 0\} = \frac{\mathbb{E}[L]}{2 + \mathbb{E}[L]}, \text{ and} \tag{6.2.1}$$

$$\begin{aligned} \mathbb{P}\{K = k + 1 \text{ and } Z'_1 = z, Z'_2 = r_1z, \dots, Z'_{k+1} = r_kz\} \\ = \frac{2}{2 + \mathbb{E}[L]} \mathbb{P}\{L = k\} \prod_{i=1}^k \mathbb{P}\{R = r_i\} \quad \text{for } k \geq 1, \end{aligned} \tag{6.2.2}$$

where the probabilities are interpreted as densities where appropriate, depending on the distribution of  $R$ . If we ignore types, both branching process have growth rate  $\mathbb{E}[L]$  — that is,  $\mathbb{E}[M_Z(t)] = \mathbb{E}[M_W(t)] = \exp(t\mathbb{E}[L])$ .

We postpone our claim that the ancestral (respectively, descendant) processes of a finite number of individuals is asymptotically given by independent copies of the ancestral (respectively, descendant) branching processes until Section 6.4.

### 6.3 Limiting flows

We are now ready to describe the processes in the large-population-size limit, which are *deterministic* flows on  $M_1(T)$ . First, a note about their relation to the intrinsically stochastic process of descendants (or ancestors) originating with a *single* individual. To obtain a nontrivial limit process  $(D_t)_{t \geq 0}$ , we need to have the initial measures  $D_0^N$  converge to some measure  $D_0 \in M_1(T)$  that assigns positive mass to something besides  $\{0\}$ . However, if we are interested in the how the genes

originating from a single individual move through the population, the natural initial conditions are  $D_0^N = 1/N\delta_1 + (1 - 1/N)\delta_0$ , which converge to  $\delta_0$  as  $N \rightarrow \infty$ .

If we begin with, say,  $D_0 = p\delta_1 + (1 - p)\delta_0$ , this corresponds to marking a *proportion*  $p$  of the population at time 0 and tracking their descendant genes through time — and it is this process which we show behaves deterministically. Indeed, this is what we might expect from the additivity property — for finite  $N$ ,  $D_t^N$  is the sum of a large number of “single-individual” processes. These “single-individual” processes are certainly not independent, but it turns out that in some sense, there is enough independence to allow a law of large number effect (although this is not how our results are obtained).

The limiting processes defined here turn out to be *dual* (see Section 7.2 for discussion of the duality) to the branching processes of Section 6.2 through the following function. Given a measure  $\mu \in M_1(T)$  and a sequence of values  $(w_1, w_2, \dots, w_n) \in T^n$  (possibly of length zero) define

$$Q(w, \mu) = \int_T \cdots \int_T \left( \sum_{i=1}^n w_i y_i \right) \prod_{i=1}^n \mu(dy_i) \in M_1(T), \quad (6.3.1)$$

where if  $n = 0$ , then  $Q(w, \mu) = 0$ .

**Definition 6.3.1.** Fix a distribution  $D_0 \in M_1(T)$ . The *descendant flow on  $T$*  begun at  $D_0$  is a deterministic  $M_1(T)$ -valued process  $(D_t)_{t \geq 0}$  defined as follows: Let  $(W_t)_{t \geq 0}$  be a copy of the ancestral branching process on  $T$  begun with a single particle of type 1 (the constant function). Then for all  $t \geq 0$ ,

$$D_t(dx) = \mathbb{P}\{Q(W(t), D_0) \in dx\}, \quad (6.3.2)$$

or equivalently, for all bounded, continuous functions  $f$  on  $T$ ,

$$\int_T f(x) D_t(dx) = \mathbb{E}[f(Q(W_t, D_0))], \quad (6.3.3)$$

where the symbols  $\mathbb{P}$  and  $\mathbb{E}$  refer to the distribution of the branching process  $W(t)$ .

**Definition 6.3.2.** Fix a distribution  $A_0 \in M_1(T)$ . The *ancestral flow on  $T$*  begun at  $A_0$  is a deterministic  $M_1(T)$ -valued process  $(D_t)_{t \geq 0}$  defined as follows: Let  $(Z_t)_{t \geq 0}$  be a copy of the descendant branching process on  $T$  begun with a single particle of type 1. Then for all  $t \geq 0$ ,

$$A_t(dx) = \mathbb{P}\{Q(Z(t), A_0) \in dx\}. \quad (6.3.4)$$

or equivalently, for all bounded, continuous functions  $f$  on  $T$ ,

$$\int_T f(x) A_t(dx) = \mathbb{E}[f(Q(Z_t, A_0))], \quad (6.3.5)$$

where the symbols  $\mathbb{P}$  and  $\mathbb{E}$  refer to the distribution of the branching process  $Z(t)$ .

The operation relating each deterministic flow to its dual branching process can be understood as follows. In the case of the ancestral flow, we create a tree with the branching process  $Z$  run

for time  $t$ , and interpret the evolution of types in  $Z$  as a labeling of each edge of the tree leading from parent to child by an independent copy of  $R$ . Then we label the leaves with independent  $T$ -valued random variables with distribution  $A_0$ , and recursively label each internal node with a value in  $T$  by multiplying each of its children by the value on the intervening edge, and sum them up. Eventually we will label the root, which is therefore a random variable with distribution in  $M_1(T)$ . Then  $A_t$  is the *distribution* of the value at the root. A similar description holds for  $D$  with  $Z$  replaced by  $W$ . This procedure is schematically depicted in Figure 6.1.

The branching property implies that  $(A_t)_{t \geq 0}$  and  $(D_t)_{t \geq 0}$  are in fact flows, e.g. that for each  $s \geq 0$ ,

$$A_{t+s}(dx) = \mathbb{P}\{Q(Z(s), A_t) \in dx\}. \quad (6.3.6)$$

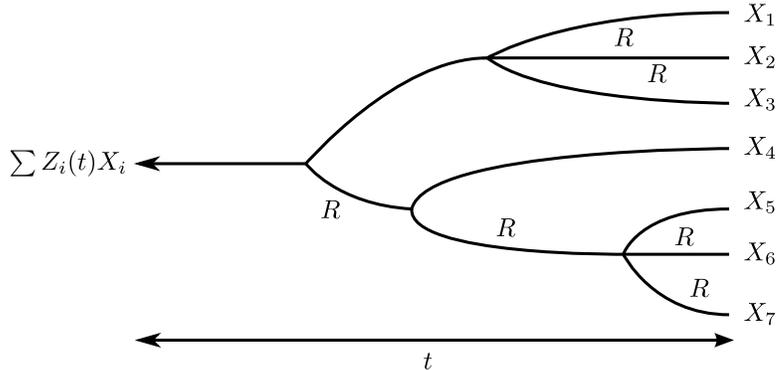


Figure 6.1: The value of  $A_t$  is obtained from the distribution of  $Z_t$ . Here  $X_1, X_2, \dots$  are independent samples from  $A_0$ , and the various copies of  $R$  are also i.i.d.. Values are propagated up the tree by multiplying along edges and adding at vertices. The *distribution* of the value at the root is the *value* of  $A_t$ .

The following theorem is proved in Section 7.2, after some preliminary work with generators in Section 7.1. Convergence of the processes will be in  $D(R_+, M_1(T))$ , the space of càdlàg  $M_1(T)$ -valued functions on  $R_+$ , equipped with the Skorokhod topology associated with the topology of weak convergence on  $M_1(T)$ .

**Theorem 6.3.1.** *If  $A_0^N$  converges to  $A_0$ , then the genetic ancestry processes  $A^N$  converge in distribution to the ancestral flow  $A$  as  $N \rightarrow \infty$ . If  $D_0^N$  converges to  $D_0$ , then the genetic descendant processes  $D^N$  converge in distribution to the descendant flow  $D$  as  $N \rightarrow \infty$ .*

## 6.4 Joint definition

In this section we provide the promised joint construction, which elucidates the model and provides connections to other work, particularly previous work on biparental models [176, 230] and

the literature on particle systems [213]. We also use this construction to prove the assertion that the ancestral and descendant processes begun with a finite number of individuals are asymptotically identical to the branching processes introduced in Section 6.2.

It is clear that the forwards and reverse time processes can be coupled, in the sense that we can use the same collection of random variables  $\tau$ ,  $R$ ,  $L$ ,  $I$ ,  $J$ , and  $K$  to construct each process, but use them in “reverse order” for one. It should also be clear that both  $A^N$  and  $D^N$  are *monotone* [213], in the sense that their semigroups preserve the natural partial order on  $M_1(T)$ . Both these statements are demonstrated with the following “particle system” construction.

We denote by  $[N]$  the collection of integers  $\{1, 2, \dots, N\}$ , and by  $(N)_k$  the falling factorial,  $(N)_k = N(N-1)\cdots(N-k+1)$ .

**Definition 6.4.1.** For each  $i \in [N]$ , let  $\Pi_i$  be a Poisson point process on the set

$$\begin{aligned} & \{(t, j, (k_1, x_1), \dots, (k_\ell, x_\ell)) : t \in \mathbb{R}, \ell \geq 1, \{x_1, \dots, x_\ell\} \subset T^\ell, \\ & \text{and } \{j, k_1, \dots, k_\ell\} \subset [N]^\ell, \text{ with } \{i, j, k_1, \dots, k_\ell\} \text{ distinct}\} \end{aligned}$$

with intensity measure

$$\pi_i(dt, j, (k_1, dr_1), \dots, (k_\ell, dr_\ell)) = \frac{1}{(N)_{\ell+1}} p_\ell \nu(dr_1) \cdots \nu(dr_\ell) dt$$

where  $p_\ell = \mathbb{P}\{L = \ell\}$  and  $\nu$  is the distribution of  $R$ . In other words, if a point in  $\Pi_i$  has the form  $(\tau, J, (K_1, R_1), \dots, (K_L, R_L))$  then  $\tau$  is the arrival time of a unit Poisson process,  $L$  is a copy of our familiar offspring distribution, and given  $L$ , the indices  $\{J, K_1, \dots, K_L\}$  are a uniformly distributed subset of  $[N] \setminus \{i\}$  and  $(R_1, \dots, R_L)$  are independent copies of  $R$ .

For each  $i, j \in [N]$ , and  $t \in \mathbb{R}$ , let  $\Theta(i, i; t, t) = 1$  and let  $\Theta(i, j; t, t) = 0$  if  $i \neq j$ . Update  $\Theta$  at each arrival time of the point process, so that if  $(\tau, J, (K_1, R_1), \dots, (K_L, R_L))$  is a point in  $\Pi_i$  with  $\tau > t$ , then for each  $m, n \in [N]$ ,

$$\Theta(m, n; t, \tau) = \begin{cases} R_\ell \Theta(m, i; t, \tau-) + (1 - R_\ell) \Theta(m, J; t, \tau-) & \text{if } n = K_\ell \text{ for some } \ell, \text{ and} \\ \Theta(m, n; t, \tau-) & \text{otherwise.} \end{cases}$$

We can define  $\Theta$  recursively backwards in time as well, so it is defined for all time coordinates in  $\mathbb{R}$ : if  $(\tau, J, (K_1, R_1), \dots, (K_L, R_L))$  is a point in  $\Pi_i$  with  $\tau < t$ , then  $\Theta$  satisfies

$$\Theta(m, n; \tau-, t) = \begin{cases} \Theta(i, n; \tau, t) + \sum_{\ell=1}^L R_\ell \Theta(K_\ell, n; \tau, t) & \text{if } m = i \\ \Theta(J, n; \tau, t) + \sum_{\ell=1}^L (1 - R_\ell) \Theta(K_\ell, n; \tau, t) & \text{if } m = J \\ 0 & \text{if } m = K_\ell \text{ for some } \ell \\ \Theta(m, n; \tau, t) & \text{otherwise.} \end{cases}$$

If we construct the full ancestral network by connecting parents to offspring with a directed edge, weight each directed reproductive edge with the recombination function (the value of  $R$ )

associated with that reproductive event, and weight each longer path in the network with the product of weights on constituent edges, then  $\Theta(i, j; t, t + s)$  is the sum of the weights of all directed paths leading from  $i$  at time  $t$  to  $j$  at time  $t + s$ , chosen to be right-continuous in  $t$  and in  $s$ .

Fix  $t \in \mathbb{R}$  and  $(\eta(1), \dots, \eta(N)) \in T^N$ , and define for  $s \geq 0$

$$\eta_s(j) = \sum_{i=1}^N \eta(i) \Theta(i, j; t, t + s).$$

If we define

$$A_s^N = \frac{1}{N} \sum_{j=1}^N \delta_{\eta_s(j)},$$

then  $(A_s^N)_{s \geq 0}$  is a version of the genetic ancestry process begun at  $\frac{1}{N} \sum_{i=1}^N \delta_{\eta(i)}$ .

Similarly, if we fix  $s \in \mathbb{R}$  and  $(\sigma(1), \dots, \sigma(N)) \in T^N$ , and define for  $t \geq 0$

$$\sigma_t(j) = \sum_{i=1}^N \sigma(i) \Theta(i, j; (s - t)^-, s),$$

and

$$D_t^N = \frac{1}{N} \sum_{j=1}^N \delta_{\sigma_t(j)},$$

then  $(D_t^N)_{t \geq 0}$  is a version of the genetic descendant process begun at  $\frac{1}{N} \sum_{i=1}^N \delta_{\sigma(i)}$ .

The following corollary says that a single sample from the random probability measure  $D_t^N$  and a single sample from the random probability measure  $A_t^N$  have the same distribution. This does not imply that  $D_t^N$  and  $A_t^N$  have the same distribution — for instance, if  $D_t^N$  was supported on a single (random) value, while  $A_t^N$  was a nonrandom distribution, then this would say that the distribution of the value that  $D_t^N$  is supported on is given by  $A_t^N$ .

**Corollary 6.4.1.** *Let  $D_0^N = (1/N)\delta_1 + (1 - 1/N)\delta_0$  and let  $X$  have the distribution of a single sample from  $D_t^N$ , e.g.  $\mathbb{E}[g(X)] = \mathbb{E}[\int_T g(x) D_t^N(dx)]$ , for all  $g \in \bar{C}(T)$ . Similarly, let  $A_0^N = 1/N\delta_1 + (1 - 1/N)\delta_0$  and let  $Y$  have the distribution of a single sample from  $A_t^N$ . Then  $X \stackrel{d}{=} Y$ .*

*Proof.* Construct  $D^N$  as above with  $s = 0$ . Then, using the exchangeability of the initial conditions first for  $D^N$  and then for  $A^N$ ,

$$\begin{aligned} \mathbb{E}[g(X)] &= \frac{1}{N} \mathbb{E} \left[ \sum_{j=1}^N g(\sigma_t(j)) \right] \\ &= \frac{1}{N} \mathbb{E} \left[ \sum_{j=1}^N g(\Theta(1, j; 0, t)) \right] \\ &= \frac{1}{N^2} \mathbb{E} \left[ \sum_{i=1}^N \sum_{j=1}^N g(\Theta(i, j; 0, t)) \right] \end{aligned}$$

We can also construct  $A^N$  as above, and again using exchangeability,

$$\begin{aligned}\mathbb{E}[g(Y)] &= \frac{1}{N} \mathbb{E} \left[ \sum_{i=1}^N g(\eta_t(i)) \right] \\ &= \frac{1}{N} \mathbb{E} \left[ \sum_{i=1}^N g(\Theta(i, 1; 0, t)) \right] \\ &= \frac{1}{N^2} \mathbb{E} \left[ \sum_{i=1}^N \sum_{j=1}^N g(\Theta(i, j; 0, t)) \right]\end{aligned}$$

□

We now prove our claim that the branching processes described in Section 6.2 asymptotically describe the genealogical trees (in either time direction) of a finite number of individuals.

First we define using the above construction what we mean by the offspring process of each individual alive at time 0. For each  $1 \leq i \leq N$ , and each  $t \geq 0$ , define indices  $1 \leq I_1(i, t) \leq I_2(i, t) \leq \dots \leq I_{M_I(i, t)}(i, t) \leq N$  to be the indices of those individuals at time  $t$  that are genealogically descended from individual  $i$  at time 0. Without loss of generality, we may assume that for all  $x \neq 0$ ,  $\mathbb{P}\{Rx = 0\} = 0$ , for instance by attaching an extra point  $u_0$  to  $U$  and setting  $R(x_0) = 1/2$ . If we do this, then we may define the indices  $I_k(i, t)$  to be exactly those indices such that  $\Theta(i, I_k(i, t); 0, t) \neq 0$ . Define the  $M(T)$ -valued process  $(\xi_i(t))_{t \geq 0}$  by

$$\xi_i(t) = \sum_{k=1}^{M_i(t)} \delta_{\Theta(i, I_k(i, t))}.$$

Similarly, define indices  $1 \leq J_1(i, t) \leq J_2(i, t) \leq \dots \leq J_{M_J(i, t)}(i, t) \leq N$  to be the indices of those individuals at time  $-t$  ( $t$  units of time in the past) from which individual  $i$  at time 0 is genealogically descended. Define  $(\zeta_i(t))_{t \geq 0}$  by

$$\zeta_i(t) = \sum_{k=1}^{M_J(i, t)} \delta_{\Theta(J_k(i, t), i)}.$$

Fix an  $n \geq 1$ , let  $W^1, W^2, \dots, W^n$  be independent copies of the ancestral branching process of Definition 6.2.1, and let  $Z^1, Z^2, \dots, Z^n$  be independent copies of the descendant branching process of Definition 6.2.2. We denote, for instance, the state of  $W^k$  at time  $t$  by  $(W_1^k(t), W_2^k(t), \dots, W_{M_W^k(t)}^k(t))$ , where  $W_i^k(t) \in T$ , and  $M_W^k(t)$  is the total size of  $W^k$  at time  $t$ . For each  $1 \leq k \leq n$ , define the measure-valued versions

$$\begin{aligned}\widetilde{W}_t^k &= \sum_{i=1}^{M_W^k(t)} \delta_{W_i^k(t)}, \\ \text{and } \widetilde{Z}_t^k &= \sum_{i=1}^{M_Z^k(t)} \delta_{Z_i^k(t)}.\end{aligned}$$

**Theorem 6.4.1.** *Let  $\zeta$ ,  $\xi$ ,  $\widetilde{W}$ , and  $\widetilde{Z}$  be defined as above, and  $t_N = \frac{x}{\mathbb{E}[L]} \log N$  for some  $0 \leq x < 1/2$ .*

1. *Let  $\alpha$  denote the distribution of  $(\zeta_1(s), \dots, \zeta_n(s))_{0 \leq s \leq t_N}$ , and let  $\beta$  denote the distribution of  $(\widetilde{W}^1(s), \dots, \widetilde{W}^n(s))_{0 \leq s \leq t_N}$ . Then  $d_{TV}(\alpha, \beta) \rightarrow 0$  as  $N \rightarrow \infty$ , where  $d_{TV}$  denotes total variation distance.*
2. *Let  $\alpha'$  denote the distribution of  $(\xi_1(s), \dots, \xi_n(s))_{0 \leq s \leq t_N}$ , and let  $\beta'$  denote the distribution of  $(\widetilde{Z}^1(s), \dots, \widetilde{Z}^n(s))_{0 \leq s \leq t_N}$ . Then  $d_{TV}(\alpha', \beta') \rightarrow 0$  as  $N \rightarrow \infty$ .*

*Proof.* We state the proof for  $\zeta$  and  $\widetilde{W}$ ; the proof works for both cases.

Let  $\tau$  be the first time that at least two of the individuals in any of the first  $n$  ancestral trees are involved in the same reproductive event. In terms of the point processes  $\Pi_i$  we used to define  $\Theta$ , and the indices  $J$  we used to define  $\zeta$ ,

$$\tau = \inf\{t \geq 0 : (-t, j, (k_1, r_1), \dots, (k_\ell, r_\ell)) \in \Pi_i \text{ with } \{J_a(b, t), J_c(d, t)\} \subset \{i, j, k_1, \dots, k_\ell\} \\ \text{for some } 1 \leq a \leq M_J(b, t), 1 \leq c \leq M_J(d, t), 1 \leq b \leq n \text{ and } 1 \leq d \leq n\}.$$

Then it is clear from the definitions that  $(\zeta_1(s), \dots, \zeta_n(s))_{0 \leq s < \tau}$  has the same distribution as  $(\widetilde{W}_1(s), \dots, \widetilde{W}_n(s))_{0 \leq s < \tau}$ , and that we can couple the two processes by defining  $(\widetilde{W}_1(s), \dots, \widetilde{W}_n(s)) = (\zeta_1(s), \dots, \zeta_n(s))$  for  $0 \leq s < \tau$ , and allowing the processes to evolve independently after  $\tau$ . By the characterization of total variation distance in terms of a maximal coupling, this implies that  $d_{TV}(\alpha', \beta') \leq \mathbb{P}\{\tau \leq t\}$ .

Since reproductive events occur at rate  $N$  and involve a uniformly chosen set of size  $L + 2$ , for any fixed pair of individuals, the reproductive events in which they are both involved form a Poisson process with rate  $\mathbb{E}[(L + 2)(L + 1)]/(N - 1)$ . It follows that the set of reproductive events that involve at least two of a given set of  $M$  individuals forms a Poisson process with rate bounded above by  $\mathbb{E}[(L + 2)(L + 1)]M(M - 1)/(N - 1)$ . Therefore, if we write  $M(s) = \sum_{i=1}^n M_J(i, s)$  then

$$\mathbb{P}\{\tau \in ds \mid \tau \geq s\} \leq \mathbb{E}[(L + 2)(L + 1)] \frac{M(s)(M(s) - 1)}{N - 1},$$

and so by the conditional Gronwall's inequality, and the fact that  $1 - e^{-x} \leq x$  for  $x \geq 0$ ,

$$\begin{aligned} \mathbb{P}\{\tau \leq t\} &\leq \mathbb{E} \left[ 1 - \exp \left( - \frac{\mathbb{E}[(L + 2)(L + 1)]}{N - 1} \int_0^t M(s)(M(s) - 1) ds \right) \right] \\ &\leq \mathbb{E} \left[ \frac{\mathbb{E}[(L + 2)(L + 1)]}{N - 1} \int_0^t M(s)(M(s) - 1) ds \right]. \end{aligned} \tag{6.4.1}$$

Note also that this is an upper bound for the expected number of reproductive events involving at least two of the individuals we are following over the time period  $[0, t]$ .

Let  $\widetilde{M}_W(t) = \sum_{k=1}^n M_W^k(t)$ , and note that  $M(s) = \widetilde{M}_W(s)$  for all  $0 \leq s < \tau$ , and that  $M_W(s)$  stochastically dominates  $M(s)$  for times  $s > \tau$ . Furthermore, we can easily compute the moments of  $\widetilde{M}_W(s)$ , following Athreya & Ney (III.4) [10] (but avoiding their error at the end of that section). Indeed, as we have already noted,  $\widetilde{M}_W(s)$  is a supercritical branching process whose

generating function is given by  $u(\gamma) = \mathbb{E}[L]\gamma(\gamma - 1)$ . Since this is a Yule process, it is possible to find the full moment-generating function of the process, but we proceed so that the computations below will apply equally well to  $M_Z$ . Recall that  $M_W^1$  is the total size of a single copy of  $W$ . If we define

$$F(\gamma, t) = \mathbb{E}[\gamma^{M_W^1(t)}]$$

Kolmogorov's backward equation says that

$$\partial_t F(\gamma, t) = u(F(\gamma, t)),$$

and so differentiating with respect to  $\gamma$  (first once, then twice) and letting  $\gamma$  approach 1 from below gives us that

$$\begin{aligned} \partial_t \mathbb{E}[M_W^1(t)] &= u'(1)\mathbb{E}[M_W^1(t)], \\ \text{and } \partial_t \mathbb{E}[M_W^1(t)(M_W^1(t) - 1)] &= u''(1)\mathbb{E}[M_W^1(t)]^2 + u'(1)\mathbb{E}[M_W^1(t)(M_W^1(t) - 1)]. \end{aligned}$$

The boundary conditions are that  $\mathbb{E}[M_W^1(0)] = 1$  and  $\mathbb{E}[M_W^1(0)(M_W^1(0) - 1)] = 0$ , so if we note that  $\mathbb{E}[L] = u'(1)$ , the solutions are given by

$$\begin{aligned} \mathbb{E}[M_W^1(t)] &= \exp(\mathbb{E}[L]t), \quad \text{and} \\ \mathbb{E}[M_W^1(t)(M_W^1(t) - 1)] &= u''(1) \left( e^{2\mathbb{E}[L]t} - e^{\mathbb{E}[L]t} \right) / \mathbb{E}[L]. \end{aligned}$$

In this case,  $u''(1) = 2\mathbb{E}[L]$ , but we continue to argue more generally. Therefore, since  $\widetilde{M}_W$  is the sum of  $n$  independent copies of  $M_W^1$ , there exists an appropriate constant  $C > 0$  that does not depend on  $t$  such that

$$\mathbb{E} \left[ \int_0^t \widetilde{M}_W(s)(\widetilde{M}_W(s) - 1) ds \right] \leq C e^{2\mathbb{E}[L]t}.$$

Therefore, we have if  $t_N = (x/\mathbb{E}[L]) \log N$  for some  $x < 1/2$ , using this in (6.4.1) and consolidating our constants into  $C'$ ,

$$\mathbb{P}\{\tau \leq t_N\} \leq C' N^{2x-1}$$

for each  $N$ , and hence  $d_{TV}(\alpha', \beta') \leq C' N^{2x-1} \rightarrow 0$  as  $N \rightarrow \infty$ . □

**Note 3.**  $\{\Theta(i, j; 0, t) : t \geq 0\}$  is a right-continuous Markov process with values in the set of  $N \times N$  matrices with entries in  $T$ , and the jumps are effected by multiplication by independent, identically distributed rank  $L + 2$  perturbations of the identity. As such, we could translate many of our questions to questions about this random matrix, to which the methods of Högnäs and Mukherjea [147] may apply. We do not pursue this further.

## Chapter 7

# Convergence results

### 7.1 Generators

In this section we define the generators of both finite-population processes and their limits as  $N \rightarrow \infty$ , and prove various bounds that we will use in later sections. First we treat the generator for the ancestral process  $A^N$ , followed by similar material for the descendant process  $D^N$ .

The generators will have particularly nice forms on the *monomial functions*, defined on  $M_1(T)$  as follows. We denote by  $C(T^m)$  the set of continuous functions on  $T^m$ , and by  $\bar{C}(T^m)$  the set of bounded, continuous functions on  $T^m$ . Recall that  $M_1(T)$  is the set of probability measures on  $T$ , and that  $C(M_1(T))$  is the set of continuous functions on  $M_1(T)$ . For  $f \in \bar{C}(T^m)$  and  $\mu \in M_1(T)$ , define

$$F_f(\mu) = \int_T \cdots \int_T f(x_1, \dots, x_m) \mu(dx_1) \cdots \mu(dx_m). \quad (7.1.1)$$

The set of degree- $m$  monomials is denoted by

$$\mathcal{C}^m = \{F_f(\cdot) : f \in \bar{C}(T^m)\},$$

and we let  $\mathcal{C} = \bigcup_{m \geq 0} \mathcal{C}^m$ .

Recall that  $E_N \subset M_1(T)$  is the set of measures of the form  $\frac{1}{N} \sum_{i=1}^N \delta_{\eta(i)}$  for some  $(\eta(1), \dots, \eta(N))$ . Given a measure  $\mu \in E_N$ , let  $S_N(\mu)$  be a random signed measure on  $T$  whose distribution is described as follows. Let  $R_1, R_2, \dots$  be a sequence of independent copies of  $R$ . Choose an independent copy of  $L$ , and given  $L$  let  $(\xi_1, \xi_2, \chi_1, \dots, \chi_L)$  be a sample without replacement from  $\mu$ . Then let

$$S_N(\mu) = \delta_{\xi_1 + \sum_{i=1}^L R_i \chi_i} + \delta_{\xi_2 + \sum_{i=1}^L R_i \chi_i} + L\delta_0 - \delta_{\xi_1} - \delta_{\xi_2} - \sum_{i=1}^L \delta_{\chi_i}. \quad (7.1.2)$$

Using this notation, we may concisely state the form of the generator of  $A^N$ .

**Definition 7.1.1.** For each  $F \in \bar{C}(M_1(T))$  and  $\mu \in E_N$  such that the following is finite, define

$$G^N F(\mu) = N\mathbb{E}[F(\mu + S_N(\mu)/N) - F(\mu)]. \quad (7.1.3)$$

Returning to the Poisson process that defined  $A^N$ , it is easy to see that  $G^N$  is a generator of  $A^N$ , with domain

$$\{F \in \bar{C}(M_1(T)) : G^N F(\mu) < \infty \forall \mu \in E_N\},$$

in the sense that

$$F(A_t^N) - F(A_0^N) - \int_0^t G^N F(A_s^N) ds$$

is a locally bounded martingale for every  $F \in \bar{C}(M_1(T))$ .

We also define for  $F \in \bar{C}(M_1(T))$  and  $\mu_N \rightarrow \mu$  as  $N \rightarrow \infty$  with  $\mu_N \in E_N$

$$GF(\mu) = \lim_{N \rightarrow \infty} G^N F(\mu_N), \quad (7.1.4)$$

as long as the limit exists and is independent of the approximating sequence  $\mu_N$ .

It is apparent that  $G$  is a first-order operator in some sense. For  $F \in \bar{C}(M_1(T))$ , the *variational derivative* of  $F$  at  $\mu$  [61] is defined for  $x \in T$  as

$$\delta F(\mu; x) = \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} (F(\mu + \epsilon \delta_x) - F(\mu)), \quad (7.1.5)$$

if the limit exists. If  $\delta F(\mu; x)$  exists and belongs to  $\bar{C}(M_1(T))$  for all  $x$ , then we say that  $F$  is *differentiable*.

It at least seems reasonable that we can rewrite  $G$  as a variational derivative. Let  $\xi, \xi_1, \xi_2, \dots$  be iid with distribution  $\mu$ , let  $L$  and  $R_1, R_2, \dots$  be defined as above, and define  $\zeta$  to be a random  $T$ -valued random variable with distribution

$$\begin{aligned} \mathbb{P}\{\zeta = 0\} &= \mathbb{E}[L]/(2 + \mathbb{E}[L]) \\ \mathbb{P}\{\zeta = \xi_1 + \sum_{i=2}^{\ell+1} R_i \xi_i\} &= 2\mathbb{P}\{L = \ell\}/(2 + \mathbb{E}[L]) \end{aligned} \quad (7.1.6)$$

Then, as long as several exchanges of limits and integrals are justified, we would have that

$$GF(\mu) = \mathbb{E}[\delta F(\mu; \zeta) - \delta F(\mu; \xi)]. \quad (7.1.7)$$

This in fact holds at least for  $F \in \mathcal{C}^m$ , as we will now prove. Let  $\nu$  be the distribution of  $R$ , and let  $p_k = \mathbb{P}\{L = k\}$ . Then it is easy to see that (7.1.7) can be written as

$$\begin{aligned} GF_f(\mu) &= \sum_{i=1}^m \sum_{k \geq 1} p_k \int_T \cdots \int_T \left\{ \int_T \cdots \int_T \right. \\ &\quad \left( 2f \left( x_1, \dots, x_i + \sum_{j=1}^k r_j x_{m+j}, \dots, x_m \right) \right. \\ &\quad \left. \left. + kf(x_1, \dots, 0, \dots, x_m) - (2+k)f(x_1, \dots, x_m) \right) \right. \\ &\quad \left. \nu(dr_1) \cdots \nu(dr_k) \right\} \mu(dx_1) \cdots \mu(dx_{m+k}) \end{aligned} \quad (7.1.8)$$

— see Dawson 2.1.2 [61] for a proof.

To show that this agrees with our previous definition of  $G$ , we will first need the following lemma.

**Lemma 7.1.1.** *Let  $f \in \bar{C}(T^m)$ , let  $\mu$ ,  $\pi_+$ , and  $\pi_-$  be probability measures on  $T$ , and define the signed measure  $\pi = \pi_+ - \pi_-$ . Then for any  $\epsilon \geq 0$ ,*

$$\left| \int_{T^m} f(x)(\mu + \epsilon\pi)^m(dx) - \int_{T^m} f(x)\mu^m(dx) - \epsilon \sum_{i=1}^m \int_{T^m} f(x)\pi(dx_i) \prod_{j \neq i} \pi(dx_j) \right| \leq \epsilon^2 m^2 \|f\|_\infty (1 + \epsilon)^m. \quad (7.1.9)$$

*Proof of Lemma 7.1.1.* First note that

$$\int_{T^m} f(x)(\mu + \epsilon\pi)^m(dx) = \sum_{I \subset [m]} \epsilon^{|I|} \int_{T^m} f(x) \prod_{i \in I} \mu(dx_i) \prod_{j \notin I} \pi(dx_j), \quad (7.1.10)$$

where the sum is over subsets  $I$  of  $[m] = \{1, 2, \dots, m\}$ . Therefore, since  $|\pi(A)| \leq 1$  for every  $A \in \mathcal{B}(T)$ ,

$$\begin{aligned} \left| \int_{T^m} f(x)(\mu + \epsilon\pi)^m(dx) - \int_{T^m} f(x)\mu^m(dx) - \epsilon \sum_{i=1}^m \int_{T^m} f(x)\pi(dx_i) \prod_{j \neq i} \mu(dx_j) \right| &\leq \sum_{k=2}^m \binom{m}{k} \epsilon^k \|f\|_\infty \\ &= ((1 + \epsilon)^m - 1 - m\epsilon) \|f\|_\infty \\ &\leq \epsilon^2 m(m-1)(1 + \epsilon)^m \|f\|_\infty. \end{aligned} \quad (7.1.11)$$

□

Now we show that our two definitions of  $G$  on  $\mathcal{C}^m$  coincide.

**Lemma 7.1.2.** *Fix  $C > 1$ . Then for every  $f \in \bar{C}(T^m)$  and every  $N > C(2 + L_{\max})$ ,*

$$\sup_{\mu \in M_1(T)} |G^N F_f(\mu) - G F_f(\mu)| \leq \frac{m^2(1 + 1/C)^m \mathbb{E}[(2 + L)^2] \|f\|_\infty}{N}, \quad (7.1.12)$$

where by  $GF$  we mean the expression in (7.1.8).

*Proof of Lemma 7.1.2.* Fix  $\mu \in M_1(T)$ . Conditioned on the value of  $L$  used in the definition of  $S_N(\mu)$ , let

$$\pi = \frac{1}{2 + L} S_N(\mu),$$

in other words,  $\pi$  is  $S_N(\mu)$  normalized to to be the signed difference of two probability measures.

Then using Lemma 7.1.1 and conditioning on  $L$ , we get that if  $N > C(2 + L_{\max})$ ,

$$\begin{aligned} |G^N F(\mu, f) - GF(\mu, f)| &= \left| N\mathbb{E} \left[ \int_{T^m} f(x) \left( \mu + \frac{2+L}{N} \pi \right)^m(dx) - \int_{T^m} f(x) \mu^m(dx) \right. \right. \\ &\quad \left. \left. - \frac{2+L}{N} \sum_{i=1}^m \int_{T^m} f(x) \pi(dx_i) \prod_{j \neq i} \mu(dx_j) \right] \right| \\ &\leq N\mathbb{E} \left[ \frac{1}{N^2} m^2 (2+L)^2 \left( 1 + \frac{2+L}{N} \right)^m \right] \\ &\leq m^2 \mathbb{E}[(2+L)^2] (1 + 1/C)^m \end{aligned}$$

Note that all quantities inside the expectations are bounded, so we are justified in exchanging the order of integration. □

**Theorem 7.1.1.** *If  $F \in \mathcal{C}^m$  then (7.1.4) and (7.1.8) agree.*

Now we turn to the generator of the forwards-time process  $D^N$ . As above, fix a measure  $\mu \in E_N$ , let  $R_1, R_2, \dots$  be a sequence of independent copies of  $R$ , and given  $L$  let  $(\xi_1, \xi_2, \chi_1, \dots, \chi_L)$  be a sample without replacement from  $\mu$ . Then define the random measure  $B_N(\mu)$  by

$$B_N(\mu) = L\delta_{(\xi_1 + \xi_2)/2} - \sum_{i=1}^L \delta_{\chi_i}.$$

**Definition 7.1.2.** For each  $F \in \bar{C}(M_1(T))$  such that the following is finite for all  $\mu \in E_N$ , define

$$H^N F(\mu) = N\mathbb{E} [F(\mu + B_N(\mu)/N) - F(\mu)]. \quad (7.1.13)$$

Again, it is easy to see that  $H^N$  is a generator of  $D^N$ , with domain

$$\{F \in \bar{C}(M_1(T)) : H^N F(\mu) < \infty \forall \mu \in E_N\}.$$

in the same sense as in Definition 7.1.1.

We also define for  $F \in \bar{C}(T)$

$$HF(\mu) = \lim_{N \rightarrow \infty} H^N F(\mu_N), \quad (7.1.14)$$

as long as the limit exists independently of the approximating sequence  $\mu_N \in E_N$  with  $\mu_N \rightarrow \mu$ .

Now let  $\xi$ ,  $\xi_1$ , and  $\xi_2$  be iid with distribution  $\mu$ . Then, again as long as several exchanges of limits and integrals are justified, we would have that

$$HF(\mu) = \mathbb{E}[L] \mathbb{E}[\delta F(\mu; R\xi_1 + (1-R)\xi_2) - \delta F(\mu; \xi)]. \quad (7.1.15)$$

For  $F \in \mathcal{C}^m$ , this has the following form, where recall that  $\nu$  is the distribution of  $R$ , and  $p_k = \mathbb{P}\{L = k\}$ :

$$\begin{aligned} HF(\mu, f) &= \mathbb{E}[L] \sum_{i=1}^m \int_T \cdots \int_T \left\{ \int_T (f(x_1, \dots, rx_i + (1-r)x_{m+1}), \dots, x_m) \right. \\ &\quad \left. - f(x_1, \dots, x_m) \right\} \nu(dr) \mu(dx_1) \cdots \mu(dx_{m+1}). \end{aligned} \quad (7.1.16)$$

The following lemma shows that our two definitions of  $H$  on  $\mathcal{C}^m$  coincide. It is proved similarly to Lemma 7.1.2.

**Lemma 7.1.3.** *Fix  $C > 1$ . Then for every  $f \in \bar{C}(T^m)$  and every  $N > C(2 + L_{\max})$ ,*

$$\sup_{\mu \in M_1(T)} |H^N F(\mu, f) - HF(\mu, f)| \leq \frac{m^2(1 + 1/C)^m \mathbb{E}[L^2] \|f\|_\infty}{N}, \quad (7.1.17)$$

where by  $HF$  we mean the expression in (7.1.16).

**Theorem 7.1.2.** *If  $F \in \mathcal{C}^m$  then (7.1.14) and (7.1.16) agree.*

## 7.2 Finite-time convergence

In this section we prove Theorem 6.3.1, the convergence of  $A^N$  to  $A$  and of  $D^N$  to  $D$  as càdlàg processes in  $M_1(T)$ . This involves no rescaling in time — thus, “finite-time convergence”. To do this, we first prove that the limiting generator restricted to the monomial functions uniquely determines the limiting process, and then use our bounds on the differences between the generators proved above to obtain convergence in distribution of the processes. For both steps, we rely on machinery from Ethier & Kurtz [91].

The proofs that  $A^N \xrightarrow{d} A$  and that  $D^N \xrightarrow{d} D$  are very similar, so we give a full proof that  $A^N \xrightarrow{d} A$ , and indicate how the proof is modified to obtain  $D^N \xrightarrow{d} D$ .

Recall that in Section 7.1 we defined  $\mathcal{C}^m$  to be the set of bounded continuous functions on  $M_1(T)$  of the form  $F_f(\mu)$  for some  $f \in \bar{C}(T^m)$ , and that for  $F$  of this form we found that the operator  $G$  was given by equation (7.1.8).

**Lemma 7.2.1.** *Let  $V$  be a càdlàg  $M_1(T)$ -valued Markov process such that*

$$F(V_t, f) - F(V_0, f) - \int_0^t GF(V_s, f) ds \quad (7.2.1)$$

*is a martingale for all  $f \in \bar{C}(T^m)$ . Then  $V_t$  is the ancestral flow as defined in Definition 6.3.2.*

*Proof of Lemma 7.2.1.* Let  $z = (z_1, \dots, z_m)$  be a collection of finite sequences from  $T$ , so  $z_i = (z_{i,1}, \dots, z_{i,n(i)})$  for each  $1 \leq i \leq m$ . Let  $n = \sum_i n(i)$ . For  $f \in \bar{C}(T^m)$  define  $z \circ f \in \bar{C}(T^{n(1)} \times \dots \times T^{n(m)})$  by

$$z \circ f(x_1, \dots, x_n) = f \left( \sum_{j=1}^{n(1)} z_{1,j} x_j, \sum_{j=1}^{n(2)} z_{2,j} x_{n(1)+j}, \dots, \sum_{j=1}^{n(m)} z_{m,j} x_{n-n(m)+j} \right), \quad (7.2.2)$$

where we define an empty sum to have the value 0. Let  $f \in \bar{C}(T^m)$ , and for the remainder of this proof write  $F_f(\mu) = F(\mu, f)$ . If  $Z^m$  is a sequence of  $m$  independent copies of the descendant branching process  $Z$  defined in Definition 6.2.2, an easy computation shows that for each  $\mu \in M_1(T)$ ,

$$F(\mu, Z^m(t) \circ f) - F(\mu, Z^m(0) \circ f) - \int_0^t GF(\mu, Z^m(s) \circ f) ds \quad (7.2.3)$$

is a martingale.

This implies that  $V$  and  $Z^m$  are Feynman-Kac *dual processes* with respect to  $\{F(\cdot, \cdot \circ f) : f \in \bar{C}(T^m)\}$ , and since the functions  $F$  are bounded, namely

$$F(\mu, z \circ f) \leq \|f\|_\infty, \quad (7.2.4)$$

for all  $\mu \in M_1(T)$  and  $z \in (\bigsqcup_{n \geq 0} T^n)^m$ , by Corollary 4.4.14 in [91], we have that for each  $f \in \bar{C}(T^m)$ ,  $V_0 \in M_1(T)$ ,  $Z_0^m \in \left(\bigsqcup_{n \geq 0} T^n\right)^m$ , and  $t \geq 0$ ,

$$\mathbb{E}[F(V_t, Z_0^m \circ f)] = \mathbb{E}[F(V_0, Z_t^m \circ f)]. \quad (7.2.5)$$

Now take  $f(x_1, \dots, x_m) = \prod_{i=1}^m \phi_i(x_i)$ , with  $\phi_i \in \bar{C}(T)$  for each  $1 \leq i \leq m$ . If in (7.2.5) we let each coordinate of  $Z_0^m$  begin with a single particle of type 1, then since the coordinates of  $Z^m$  are independent,

$$\mathbb{E} \left[ \int_T \cdots \int_T \prod_{i=1}^m \phi_i(x_i) V_t(dx_i) \right] = \prod_{i=1}^m \mathbb{E} \left[ \int_T \cdots \int_T \phi_i \left( \sum_{j=1}^{M(t)} Z_j(t)x_j \right) \prod_{j=1}^{M(t)} V_0(dx_i) \right], \quad (7.2.6)$$

where the expectation on the right is with respect to a single copy of  $Z$  begun with a single particle of type 1. Therefore,

$$\begin{aligned} \text{var} \left[ \int_T \phi(x) V_t(dx) \right] &= \mathbb{E} \left[ \left( \int_T \phi(x) V_t(dx) \right)^2 \right] - \mathbb{E} \left[ \int_T \phi(x) V_t(dx) \right]^2 \\ &= \mathbb{E} \left[ \int_T \int_T \phi(x) \phi(y) V_t(dx) V_t(dy) \right] - \mathbb{E} \left[ \int_T \phi(x) V_t(dx) \right]^2 \\ &= 0. \end{aligned} \quad (7.2.7)$$

Since this is true for each  $\phi \in \bar{C}(T)$ ,  $V_t$  is deterministic. The specific form of the flow given in (6.3.5) follows from (7.2.5). □

Now we prove that  $A^N$  converges to  $A$ , and that  $D^N$  converges to  $D$ , if the initial conditions converge. We state the proof in terms of  $A^N$  and  $A$ , but the proof for  $D^N$  and  $D$  is almost identical.

*Proof of Theorem 6.3.1.* We would like to use our Lemma 7.1.2 and Corollary 4.8.7 of Ethier & Kurtz [91], but this requires  $T$  to be compact, so first we must extend our definitions to a compactification of  $T^*$ .

Note that  $S = \{f \in T : 0 \leq \|f\|_\infty \leq 1\}$  is compact by the Banach-Alaolgu theorem, and that  $T = \bigcup_{n=1}^\infty nS$ . Therefore, if we let  $T^* = T \sqcup \{\infty_T\}$ , and say that  $x_n \rightarrow \infty_T$  in  $T^*$  iff  $\|x_n\|_\infty \rightarrow \infty$ , then  $T^*$  is the one-point compactification of  $T$ .

Returning to Definitions 6.1.2, 6.3.2, 7.1.1, and 7.1.8, we see that there is no difficulty in extending  $A^N$  and  $A$  to be processes on  $M_1(T^*)$ , and extending  $G^N$  and  $G$  to  $\mathcal{C}^* = \bigsqcup_{m \geq 0} \{F(f, \cdot) :$

$f \in \bar{C}((T^*)^m)\}$ , where  $F(f, \mu)$  is defined in the same way as before for  $f \in \bar{C}((T^*)^m)$  and  $\mu \in M_1(T^*)$ .

To use 4.8.7 in [91], we need to know that  $\mathcal{C}^*$  is convergence determining, that  $G$  as we have defined is a core for the full generator of  $A$ , that  $A^N$  does not escape to  $\infty_T$  with positive probability, and that  $G^N$  converges to  $G$  on  $\mathcal{C}^*$  in a certain sense.

We know that  $\mathcal{C}$  separates points in  $M_1(T)$ , and if  $\mu|_T = \nu|_T$ , then also  $\mu(\infty_T) = 1 - \mu(T) = \nu(\infty_T)$ , so  $\mathcal{C}^*$  separates points in  $M_1(T^*)$ . Since  $\mathcal{C}^*$  is an algebra that contains the constants and  $M_1(T^*)$  is compact, the Stone-Weierstrass theorem implies that  $\mathcal{C}^*$  is dense in  $C(M_1(T^*))$ , and is hence convergence determining in  $M_1(M_1(T^*))$ .

From Lemma 7.2.1 we know that  $G$  is the generator for the ancestral flow  $A$  restricted to  $\mathcal{C}^*$ . If  $\tilde{G}$  is the full generator of  $A$ , since  $\mathcal{C}^*$  is dense in  $C(M_1(T^*))$ , we know that  $\mathcal{C}$  is a core for  $\tilde{G}$ , and hence  $\tilde{G}$  is the closure of  $G$ .

Our bound in Lemma 7.1.2 implies that for any  $f \in \bar{C}((T^*)^m)$ ,

$$\lim_{N \rightarrow \infty} \sup_{\mu \in E_N} |G^N F_f(\mu) - G F_f(\mu)| = 0.$$

It is also clear that if  $A_0^N(\infty_T) = 0$ , then  $\mathbb{P}\{A_t^N(\infty_T) > 0 \text{ for some } 0 \leq t \leq T\} = 0$ , so by Corollary 4.8.7 in [91], if  $A_0^N \xrightarrow{d} A_0$  with  $A_0(\infty_T) = 0$ , then  $A^N \xrightarrow{d} A$  as  $N \rightarrow \infty$ .

□